UNIVERSITY OF WEST LONDON
The Career University

UWL REPOSITORY

repository.uwl.ac.uk

Automated recognition of Pre-Eclamptic women in early pregnancy

Puvanendran, Suhirthakumar (2025) Automated recognition of Pre-Eclamptic women in early pregnancy. Doctoral thesis, University of West London.

https:doi.org/10.36828/thesis/14921

This is the Submitted Version of the final output.

UWL repository link: https://repository.uwl.ac.uk/id/eprint/14921/

Alternative formats: If you require this document in an alternative format, please contact: open.research@uwl.ac.uk

Copyright:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy: If you believe that this document breaches copyright, please contact us at open.research@uwl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Rights Retention Statement:



AUTOMATED RECOGNITION OF PRE-ECLAMPTIC WOMEN IN EARLY PREGNANCY

by

Suhirthakumar Puvanendran

A thesis

submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

in

Bioinformatics

University of West London

2025

Author's Declaration

I hereby declare that I am the sole author of this thesis. I grant the University of West London the right to make this thesis available to other institutions and individuals for scholarly research.

I affirm that this thesis, entitled "Automatic Recognition of Pre-eclamptic Women in Early Pregnancy," represents my original work, and all sources of information, data, and citations have been appropriately acknowledged. I confirm that I have not committed plagiarism or engaged in any unauthorised use of intellectual property. Furthermore, this thesis has not been submitted, in whole or in part, for any academic award or qualification before this submission.

I have conducted this research in full compliance with ethical guidelines and academic integrity. The findings presented in this thesis contribute to the field of Bioinformatics, explicitly advancing the understanding of the complex relationship between PE in pregnancy and genomic biomarkers.

Submitted on 17 April 2025

Suhirthakumar Puvanendran
Student ID: 21322991
School of Biomedical Sciences
University of West London
United Kingdom

I further authorise the University of West London to reproduce this thesis through photocopying or other means, in whole or in part, at the request of other institutions or individuals for scholarly research.

Suhirthakumar Puvanendran

Acknowledgements

I express my deepest gratitude to my esteemed supervisors, Professor Maddie Ohl, Professor Massoud Zolgharni, and Dr Nasim Dadashi Serej, for their unwavering support, expert guidance, and invaluable insights throughout this research journey. Your mentorship has been pivotal in shaping my academic endeavours and helping me navigate the challenges of this study.

I sincerely thank the University of West London for providing a stimulating academic environment, comprehensive resources, and opportunities that have fostered my intellectual development. Your dedication to excellence has been an inspiration and a cornerstone in completing this thesis.

I owe my dearest family my profound appreciation for your unconditional love, encouragement, and sacrifices. Your unwavering belief in me has been my most significant source of strength and motivation. You have been my constant anchors, and your presence has made this journey achievable and profoundly meaningful.

I extend my heartfelt thanks to my friends and colleagues for their encouragement, thought–provoking discussions, and companionship during this academic endeavour. Your support and collaboration have made this journey both enriching and enjoyable.

Lastly, I wish to acknowledge all those whose names may not appear here but who, in various ways, have contributed to completing this thesis. Your kindness, encouragement, and support, no matter how small, have impacted this work.

This thesis represents an academic achievement, and a journey made possible by many individuals' collective efforts, support, and inspiration. I am profoundly grateful to each of you.

Abstract

Pre-eclampsia (PE) is a severe pregnancy disorder characterised by hypertension and multi–organ dysfunction, affecting 2–8% of pregnancies worldwide. It remains a leading cause of maternal and neonatal morbidity and mortality, contributing to complications such as renal failure, cerebral oedema, intrauterine growth restriction (IUGR), and preterm birth. Despite its clinical significance, early detection of PE remains challenging due to its complex and multifactorial aetiology. Current diagnostic methods rely on maternal symptoms that emerge in the second or third trimester, limiting the potential for early interventions. This study develops an AI-assisted predictive model for early PE detection to address this gap by integrating genomic and transcriptomic biomarkers.

A bioinformatics–driven approach was employed to analyse cell-free RNA (cfRNA) sequencing data from gestational weeks 8 to 16, allowing for a longitudinal assessment of molecular changes preceding PE onset. Identified unique differentially expressed genes (UDEGs) and unique consistently expressed genes (UCEGs), highlighting transcriptomic changes linked to placental dysfunction, maternal immune dysregulation, oxidative stress, and endothelial dysfunction. Additionally, unique variant signatures (UVS) associated with PE were detected through variant calling analysis, further elucidating the genetic landscape of PE.

A multi-step machine learning (ML) pipeline was applied to refine key biomarkers and optimise predictive performance. Several ML models –including Decision Trees (DT), Random Forests (RF), Logistic Regression (LR), Gradient Boosting Machines (GBM), Support Vector Machines (SVM), and Linear Discriminant Analysis (LDA –were trained and evaluated. SVM emerged as the best–performing model, achieving the highest accuracy (≥92%), sensitivity, and specificity in distinguishing PE cases from normal pregnancies. The model was further validated using cross–validation, which demonstrated robust generalisability.

The results reveal distinct transcriptomic and genetic markers uniquely associated with PE, particularly in pathways related to angiogenesis, immune regulation, and oxidative stress response. The AI-assisted model significantly outperforms conventional risk assessment methods, offering a non–invasive, automated, and interpretable approach for PE risk stratification.

This study advances precision medicine in maternal health by integrating bioinformatics, machine learning, and maternal–foetal genomics into a clinically relevant diagnostic framework. The findings underscore the potential of genomic–based AI models in early PE detection, enabling personalised risk assessment, improved clinical decision–making, and timely interventions. Future research should validate the model in diverse populations, incorporate omics data (transcriptomics, proteomics, metabolomics, and epigenomics), and refine AI algorithms to enhance predictive accuracy for real–world clinical applications.

Keywords: Pre-eclampsia, AI-assisted prediction, bioinformatics, cell-free RNA, gene expression, variant analysis, machine learning, maternal health, precision medicine

Table of Contents

List of Abbreviations xiv
List of Figures xviii
List of Tables xxiii
Chapter 1 General Introduction 1
1.1 Overview 1
1.2 General Background 2
1.3 Aim and Objectives. 4
1.4 Methodology 5
1.5 Thesis Structure 6
1.5.1 Chapter 1: General Introduction 6
1.5.2 Chapter 2: Literature Review 6
1.5.3 Chapter 3: General Methods 6
1.5.4 Chapter 4: Identification of Pre-eclampsia Associated Unique Differentially Expressed Gene Signatures at Each Gestational Week in Early Pregnancy 6
1.5.5 Chapter 5: Identification of Pre-eclamptic Unique Consistent Gene Expression Signatures at Each Gestational Week in Early Pregnancy 7
1.5.6 Chapter 6: Identification of Pre-eclamptic Unique Variant Signatures at Each Gestational Week in Early Pregnancy 7
1.5.7 Chapter 7: Developing an Automated AI-aided Disease Prediction Model for Detecting Women at Risk of Pre-eclampsia in Early Pregnancy. 7
1.5.8 Chapter 8: Overall Discussion and Conclusion 7
Chapter 2 Literature Review 8
2.1 Introduction 8
2.1.1 Research Hypothesis 9
2.1.2 Research Question 9

2.1.3 Research Aim and Objectives 9
2.1.4 Definition of Pre-eclampsia 10
2.1.5 Aetiology of Pre-eclampsia 14
2.1.6 Risk Factors of Pre-eclampsia 26
2.1.7 Pathogenesis of Pre-eclampsia 37
2.1.8 Biomarkers Associated with Pre-eclampsia. 46
2.1.9 The Role of Gene Expression Changes in cfRNA in Maternal Pre-eclampsia 63
2.1.10 Biomarkers in Pre-eclampsia Prediction 64
2.1.11 Early Prediction of Pre-eclampsia 70
2.2 Conclusion 85
2.3 Summary of the Chapter 86
Chapter 3 General Methods 89
3.1 Summary 89
3.2 The Population of the Study 89
3.3 Sample Size Calculation. 89
3.4 Ethical Approval 91
3.5 General Workflow 92
3.6 Data Collection 93
3.6.1 Data Collection of cfRNA Sequencing Data 93
3.7 Data Transformation 102
3.8 Rationale for Selecting Bioinformatics Tools for Pre–processing of Genomic Data 103
3.9 Data Quality Control and Pre–processing 104
3.9.1 Quality Control and Pre-processing of cfRNA Data 105
3.10 Alignment 107
3.10.1 Indexing the Reference Genome for Alignment 108
3.10.2 Mapping the Reads with the Reference Genome 109
3.11 Post–Alignment Quality Control 110

Chapter 4 Identification of Pre-eclampsia Associated Unique Differentially Expressed Gene Signatures at Each Gestational Week in Early Pregnancy.... 112
4.1 Introduction.... 112
4.1.1 Understanding Gene Expression 113
4.1.2 Changes in cfRNA Gene Expression Across Gestational Weeks in Pre-eclampsia.... 116
4.1.3 Differential Gene Expression Analysis and Pre-eclampsia Prediction 117
4.1.4 Importance of DEGs and UDEGs in Pre-eclampsia Prediction 119
4.1.5 Research Hypothesis 120
4.1.6 Research Question 120
4.1.7 Research Aim and Objectives 120
4.2 Method 120
4.2.1 Quantification 121
4.2.2 Pre–processing of Gene Expression Count Data.... 122
4.2.3 Differential Gene Expression Analysis 126
4.2.4 Overlap Analysis for Identifying Unique Gene Expression Signatures. 126
4.2.5 Downstream Analysis.... 127
4.3 Results 129
4.3.1 Identification of DEGs 131
4.3.2 Identification of UDEGs.... 135
4.3.3 Downstream Analysis.... 136
4.4 Discussion.... 162
4.4.1 Identification of Differentially Expressed Genes (DEGs) 162
4.4.2 Unique Differentially Expressed Genes (UDEGs) 162
4.4.3 Biological Significance of DEGs and UDEGs.... 163
4.4.4 Contextualisation of Overlapping Key Genes.... 164
4.4.5 Gene Ontology Analysis.... 164

4.4.6 Pathway and Network Analysis Insights 165
4.5 Conclusion 166
4.6 Clinical Implications and AI-Assisted Prediction Models 167
4.7 Limitations and Future Directions 167
4.8 Summary of the Chapter 169
Chapter 5 Identification of Pre-eclamptic Unique Consistent Gene Expression Signatures at Each Gestational Week in Early Pregnancy 171
5.1 Introduction 171
5.1.1 Consistently Expressed Gene Signatures 172
5.1.2 Housekeeping Genes and Consistently Expressed Gene Signatures 173
5.1.3 Ubiquitously Expressed Genes and Consistently Expressed Gene Signatures 173
5.1.4 Research Hypothesis 175
5.1.5 Research Question 175
5.1.6 Research Aim and Objectives 175
5.2 Method 176
5.2.1 Datasets 176
5.2.2 Consistent Gene Expression Analysis (CGEA) 176
5.2.3 Downstream analysis 188
5.3 Results and Discussion 189
5.3.1 Identification of SCGs 189
5.3.2 Identification of BRCGs 190
5.3.3 Downstream Analysis for Unique Consistent Genes 191
5.3.4 Thematic Interpretation of Key Consistent Genes 225
5.3.5 Molecular Synthesis and Alignment with PE Pathogenesis 226
5.4 Conclusion 229
5.5 Future Directions 230
5.6 Summary of the Chapter 231

Chapter 6 Identification of Pre-eclamptic Unique Variant Signatures at Each Gestational Week in Early Pregnancy 234
6.1 Introduction 234
6.1.1 Overview of cfRNA and Its Relevance in Gestational Monitoring 235
6.1.2 Unique Variant Signatures and PE-associated Genetic Variants 236
6.1.3 Significance of Analysing Variant Signatures in cfRNA 237
6.1.4 The Importance of Gestational Age–Specific Biomarkers 238
6.1.5 Current Challenges in Identifying Temporal Genetic Variants 238
6.1.6 Research Hypothesis 240
6.1.7 Research Question 240
6.1.8 Research Aim and Objectives 241
6.2 Method 241
6.2.1 Data Pre–processing for Variant Call Analysis 241
6.2.2 Variant Calling 244
6.3 Results and Discussion 249
6.3.1 Identifying Unique Variant Signatures 249
6.3.2 Identifying PE-associated Significant Variants 251
6.3.3 Detailed Analysis of Key Variant Signatures 254
6.3.4 Distribution of UVS Region Types Across Gestational Weeks 256
6.3.5 Downstream analysis 263
6.4 Conclusion 319
6.5 Future Direction 320
6.6 Summary of the Chapter 321
Chapter 7 Developing an Automated AI-aided Disease Prediction Model for Detecting PE–Risk Women in Early Pregnancy 323
7.1 Introduction 323
7.1.1 Challenges in Developing AI–Driven Predictive Models 324
7.1.2 Significance of a Multi–Dimensional Approach 325
7.1.3 Research Hypothesis 326

7.1.4 Research Question 326
7.1.5 Research Aim and Objectives 326
7.2 Method 327
7.2.1 Data Pre–processing 327
7.2.2 Machine Learning Model Development 333
7.3 Results 340
7.3.1 Data Pre–processing 340
7.3.2 ML–Model Development 350
7.4 Discussion 353
7.4.1 Significance of Feature Selection in Different Gestational Weeks 354
7.4.2 Dimensionality Reduction and Model Optimisation 355
7.4.3 Machine Learning Model Performance 356
7.4.4 Implications for Early Pre-eclampsia Detection 356
7.4.5 Future Directions 357
7.5 Conclusion 358
7.6 Summary of the Chapter 359
Chapter 8 Overall Discussion 362
8.1 Introduction 362
8.2 Overall Discussion 362
8.2.1 Alignment with Existing Literature 364
8.2.2 Chapter 4: Temporal Dynamics of UDEGs 365
8.2.3 Chapter 5: Stability of UCEGs 368
8.2.4 Chapter 6: Genomic Insights from UVS 370
8.2.5 Chapter 7: AI–Driven Predictive Modelling 371
8.2.6 Synthesis and Clinical Implications 372
8.2.7 Interpretation and Implications 373
8.2.8 Summary of the Chapter 375
Chapter 9 Conclusion 377
9.1 Strengths 379

9.2 Limitations 382
9.3 Recommendations for Future Directions 385
9.4 Novel Contributions to Knowledge 390
9.5 Recommendations 391
Appendices 393
9.6 Appendix A 393
9.6.1 Step 1: Load the libraries and per–gestational week data 393
9.6.2 Step 2: Week 8 393
9.6.3 Step 3: Week 9 396
9.6.4 Step 4: Week 10 398
9.6.5 Step 5: Week 11 400
9.6.6 Step 6: Week 12 402
9.6.7 Step 7: Week 13 406
9.6.8 Step 8: Week 14 410
9.6.9 Step 9: Week 15 413
9.6.10 Step 10: Week 16 415
9.6.11 Step 11: Merging volcano plots 418
9.7 Appendix B 419
9.7.1 Step 1: Set the working directory 419
9.7.2 Step 2: Load the required Python libraries 419
9.7.3 Step 3: Load the data and data pre–processing for consistent genes analysis process 419
9.7.4 Step 4: Calculate rank product 420
9.7.5 Step 8: RP sorted table with Mean Expression and Standard Deviation 422
9.7.6 Step 9: Permutation for ‘p’ value calculation for RP table in Ascending order 423
9.8 Appendix C 430
9.8.1 Load the data 430

9.8.2 Read the selected feature data 430
9.8.3 Dimensionality Reduction – PCA 431
9.8.4 Modelling 435
9.8.5 Overlap – UDEGS with Known PE Genes 449
References 451

276 # List of Abbreviations
277

ACOG	American College of Obstetrics and Gynaecology
ADAM	Adaptive Moment Estimation
ADMA	Asymmetrical dimethylarginine
AGEs	Advanced Glycation End products
AI	Artificial Intelligence
ALT	Alanine aminotransferase
AOPP	Advanced Oxidation Protein Products
APS	Antiphospholipid Syndrome
AS	Alternative Splicing
AST	Aspartate aminotransferase
AUC	Area Under the ROC curve
AUC	The area under the curve
BMI	Body Mass Index
BP	Backpropagation
BRCGs	Biologically Relevant Consistent Genes
CAT	Catalase
CC	Cellular Components
CEGS	Consistently Expressed Gene Signatures
cfDNA	cell-free DNA
cfNA	cell-free Nucleic Acids
cfRNA	cell-free RNA
CI	Confidence Interval
CKD	Chronic Kidney Disease
CPM	Counts Per Million
CSDNN	Cost–sensitive Deep Neural Network
CSDNN–FL	Cost–sensitive Deep Neural Network – Focal Loss
CSDNN–WCE	Cost–sensitive Deep Neural Network – Weighted Cross–Entropy
CV	Cross Validation
dbGaP	database of Genotypes and Phenotypes
DEGs	Differentially Expressed Genes
DGEA	Differential Gene Expression Analysis
DL	Deep Learning
DSA	Differential Splicing Analysis
DT	Decision Tree

EBI	European Bioinformatics Institute
EGA	European Genome–Phenome Archive
eIFs	eukaryotic initiation Factors
ENA	European Nucleotide Archive
EOPE	Early–Onset Pre-eclampsia
eQTL	expression Quantitative Trait Loci
FDR	False Discovery Rate
FPR	False Positive Rate
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
GATK	Genome Analysis Toolkit
GEL	Genomics England Limited
GEO	Gene Expression Omnibus
GO	Gene Ontology
GPx	Glutathione Peroxidase
GH	Gestational Hypertension
GW	Gestational Week
GWAS	Genome–Wide Association Studies
hCG	Human chorionic gonadotropin
HELLP	Haemolysis, Elevated Liver enzymes and Low Platelets
HKGs	Housekeeping Genes
ICM	Inner Cell Mass
INDELs	INsertions/DELetions
IPI	Interpregnancy Interval
ISSHP	International Society for the Study of Hypertension in Pregnancy
IUGR	Intrauterine Growth Restriction
IVF	In Vitro Fertilisation
KEGG	Kyoto Encyclopedia of Genes and Genomes
KNN	k–Nearest Neighbors
LOOCV	Leave–One–Out cross–validation
LR	Logistic Regression
LSTM	Long Short–Term Memory
MCODE	Molecular Complex Detection
MDA	Malondialdehyde
MDGI	Mean Decrease Gini Index
MF	Molecular Functions

miRNAs	microRNAs
ML	Machine Learning
MOFA	Multi–Omics Factor Analysis
MOMI	Magee Obstetric Medical and Infant
NB	Naïve Bayes
NCBI	National Centre for Biotechnology Information
NGS	Next-generation sequencing
NIPT	Non-Invasive Prenatal Testing
NN	Neural Network
NO	Nitric Oxide
NP	Normal Pregnancy
NPV	Negative Predictive Value
OS	Operative System
PAPP-A	Pregnancy-Associated Plasma Protein A
PE	Pre-eclampsia
PE-AGV	PE-Associated Genetic Variants
PI	Pulsatility Index
PlGF	Placental Growth Factor
PP13	Placental Protein 13
PPI	Protein–Protein Interaction
PPV	Positive Predictive Value
PSO	Particle Swarm Optimisation
PTMs	Post–translational modifications
PUDF	Public Use Data Files
QC	Quality Control
RAGE	Receptor for AGE
RF	Random Forest
RFE	Recursive Feature Elimination
RLE	Relative Log Expression
RNN	Recurrent neural networks
ROS	Reactive Oxygen Species
RPKM	Reads Per Kilobase per Million mapped reads
SCGs	Significant Consistent Genes
SDEGs	Specific differential gene expression signatures
sFlt1	Soluble fms-like Tyrosine Kinase–1
SGB	Stochastic Gradient Boosting
SLE	Systemic Lupus Erythematosus
SNPs	Single Nucleotide Polymorphisms

SOD	Superoxide Dismutase
SPE	Severe Pre-eclampsia
SRA	Sequence Read Archive
STAR	Spliced Transcripts Alignment to a Reference
STB	Syncytiotrophoblast
SVM	Support Vector Machine
SVM–Lin	Support Vector Machine with a Linear Kernel
SVM–RBF	Support Vector Machine with a Radial Basis Function
TE	Trophectoderm
TGF–β	Transforming Growth Factor–beta
TGVs	Temporal Genetic Variants
TMM	Trimmed Mean of M values
TPM	Transcripts Per Million
TPR	True Positive Rate
UA	Uterine Artery Doppler
UDEGs	Unique Differentially Expressed Gene Signatures
UEGs	Ubiquitously Expressed Genes
UPCR	Urine Protein Creatinine Ratio
UVS	Unique Variant Signatures
VCF	Variant Call Format
VEGF	Vascular Endothelial Growth Factor
vWF	Von Willebrand Factor
XAI	Explainable artificial intelligence

List of Figures

Figure 1: Schematic representation of PE two-stage model update over time. .. 25
Figure 2: Schematic presentation of blastocyst implantation. 40
Figure 3: General workflow of genomic analysis in the study. 93
Figure 4: Workflow for differential gene expression analysis for each gestational week. 121
Figure 5: Volcano plots of DEGs in each GW. 133
Figure 6: Heatmap of DEGs in GW 13..... 134
Figure 7: UpsetR plot for visualising the intersections between DEGs in GWs 136
Figure 8: Results of Gene ontology analysis for DEGs in GW 8, 10, and 13. ... 142
Figure 9: Results of Gene Ontology analysis for DEGs in GW 15 and 16. 143
Figure 10: Results of Gene Ontology analysis for UDEGs in GW 8, 10 and 13. 146
Figure 11: Results of Gene Ontology analysis for UDEGs in GW 15 and 16.... 147
Figure 12: KEGG pathway enrichment analysis by enrichR for DEGs and UDEGs in GW–13, 15, 16..... 153
Figure 13: PPI network for DEGs in GW13 as generated using STRING..... 157
Figure 14: Top two significant clusters in the PPI analysis for GW13 using Cytoscape A, B – Significant cluster 1 and cluster 2 in the PPI network for GW13..... 158
Figure 15: KEGG pathway analysis of top two significant clusters in GW13 PPI analysis using enrichR..... 160
Figure 16: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 8 in NP and PE 194
Figure 17: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 9 in NP and PE 195

Figure 18: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 10 in NP and PE 203
Figure 19: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 11 in NP and PE 204
Figure 20: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 12 in NP and PE 205
Figure 21: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 13 in NP and PE 206
Figure 22: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 14 in NP and PE 207
Figure 23: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 15 in NP and PE 208
Figure 24: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 16 in NP and PE 209
Figure 25: KEGG pathway analysis for unique, consistent genes at GW 8 and 9 in NP and PE 212
Figure 26: KEGG pathway analysis for unique, consistent genes at GW 10 and 11 in NP and PE 213
Figure 27: KEGG pathway analysis for unique, consistent genes at GW 12 and 13 in NP and PE 216
Figure 28: KEGG pathway analysis for unique, consistent genes at GW 14 and 15 in NP and PE 217
Figure 29: KEGG pathway analysis for unique, consistent genes at GW 16 in NP and PE 223
Figure 30: Workflow for identifying unique variant signatures from cfRNA sequencing data. 246
Figure 31: Distribution of Unique Variant Signatures (UVS) across gestational weeks 8-16. 253

Figure 32: Distribution of PE-associated significant variants across gestational weeks. 254
Figure 33: Genomic region distribution of Unique Variant Signatures (UVS) at Gestational Weeks 8 and 9. 259
Figure 34: Genomic region distribution of Unique Variant Signatures (UVS) at Gestational Weeks 10 and 11. 260
Figure 35: Genomic region distribution of Unique Variant Signatures (UVS) at Gestational Weeks 12 and 13. 261
Figure 36:Genomic region distribution of Unique Variant Signatures (UVS) at Gestational Weeks 14 and 15. 262
Figure 37: Genomic region distribution of Unique Variant Signatures (UVS) at Gestational Week 16. 263
Figure 38: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 8. 265
Figure 39: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 8. 266
Figure 40: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 9. 268
Figure 41: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 9. 269
Figure 42: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 10. 271
Figure 43: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 10. 272
Figure 44: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 11. 274
Figure 45: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 11. 275

Figure 46: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 12. 278
Figure 47: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 12. 279
Figure 48: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 13. 282
Figure 49: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 13. 283
Figure 50: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 14. 286
Figure 51: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 14. 287
Figure 52: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 15. 290
Figure 53: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 15. 291
Figure 54: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 16. 294
Figure 55: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 16. 295
Figure 56:Gene Ontology (GO) enrichment analysis for PE-associated significant genetic variants across Gestational Weeks 8 to 11. 299
Figure 57: Gene Ontology (GO) enrichment analysis for PE-associated significant genetic variants across Gestational Weeks 12 to 14. 300
Figure 58: KEGG pathway enrichment analysis for Unique Variant Signatures (UVS) in NP and PE at Gestational Weeks 8 and 9. 308
Figure 59: KEGG pathway enrichment analysis for Unique Variant Signatures (UVS) in NP and PE at Gestational Weeks 10 and 11. 309

Figure 60: KEGG pathway enrichment analysis for Unique Variant Signatures (UVS) in NP and PE at Gestational Weeks 12 and 13. 310
Figure 61: KEGG pathway enrichment analysis for Unique Variant Signatures (UVS) in NP and PE at Gestational Weeks 14 and 15. 311
Figure 62: KEGG pathway enrichment analysis for Unique Variant Signatures (UVS) in NP and PE at Gestational Week 16. 312
Figure 63: KEGG pathway analysis for PE-associated significant genetic variants across Gestational Weeks 8 to 11. 317
Figure 64: KEGG pathway analysis for PE-associated significant genetic variants across Gestational Weeks 12 to 15. The chart displays the enriched pathways. 318
Figure 65: KEGG pathway analysis for PE-associated significant genetic variants at Gestational Week 16. 319
Figure 66: Workflow for feature selection and developing ML model for PE prediction. 330
Figure 67: Venn Diagram showing the overlap between UDEGS and UCEGS genes. 342
Figure 68: Overlap of PE-associated Genes from Databases and UTS 345
Figure 69: Distribution of intersecting UTS between PE–linked genes from databases and UTS identified in UDEGS, UCEGS, and CDEGS. 346
Figure 70: Determination of the optimal number of PCA components across different CV fold values. 349
Figure 71: Model Accuracy Comparison. 351
Figure 72: Receiver Operating Characteristic (ROC) Curves for Different ML Models. 353
Figure 73: Overview of the general methodology. 364
Figure 74: Summary of the UDEGS identification method. 367
Figure 75: Summary of the methodology for identifying UCEGS. 369
Figure 76: Overview of the AI-Driven model development. 372

List of Tables

Table 1 The diagnostic criteria for PE are from professional societies. 12
Table 2: Summary of Key Risk Determinants for Pre-eclampsia. 26
Table 3 Summary of Key Biomarkers from Literature Review. 63
Table 4: Summary of previously published studies and the variable types used in the PE predictive model. 66
Table 5: The candidate biomarkers and factors utilised in the PE predictive models from various studies. 67
Table 6: Accuracy and Area Under the Curve (AUC) Values of Various PE Predictive Models. 70
Table 7: Number of raw data for NP and PE in the Study. 94
Table 8: Demographic and Recruitment Characteristics of Pooled GEO Datasets for cfRNA Analysis 96
Table 9: Source Details of cfRNA Sequencing Data by Gestational Age (GW 8–16). 97
Table 10: Breakdown of Data Per Gestational Week from GSE192902 and GSE154377 102
Table 11: Number of Genes Retained for Differential Gene Expression Analysis (DGEA) After Gene Filtering Across Gestational Weeks 8 to 16 131
Table 12: Differential Gene Expression Analysis Results in Early Pregnancy. . 131
Table 13: Significantly enriched pathways for DEGs in GW 13 using enrichR . 151
Table 14: Summary statistics of two significant clusters 156
Table 15: KEGG pathway analysis results for nodes in Cluster 1 in the PPI network for GW13. 159
Table 16: KEGG pathway analysis results for nodes in Cluster 2 in the PPI network for GW13. 161

Table 17: Consistent genes across gestational weeks from 8 to 16 in NP and PE 191
Table 18: Synthesis of Unique Consistent Gene Signatures (UCEGS) and Alignment with Preeclampsia Pathogenesis (PE vs. NP) 227
Table 19: Gestational–Week specific Distribution of Unique SNPs and INDELs in NP and PE 253
Table 20. Summary of gene variants identified in genomic windows associated with pre-eclampsia. 256
Table 21: Number of unique transcriptomic and variant signatures identified per gestational week. 341
Table 22: Summary of the ML algorithms and CV approaches used for feature selection in GW–specific datasets. 344
Table 23: Feature Selection Results Using Recursive RFE–CV Across Different ML Algorithms 345
Table 24: Key statistical metrics of 20 intersecting genes between PE–linked genes from databases and UTS, including gene symbols, expression categories, base mean expression, log2 fold change, standard error (lfcSE), test statistic (stat), p–value, and adjusted p–value (padj). 347
Table 25: Key statistical metrics of 47 intersecting genes between PE–linked genes from databases and UTS in UCEGS, including gene symbols, expression categories, base mean expression, log2 fold change, standard error (lfcSE), test statistic (stat), p–value, and adjusted p–value (padj). 348
Table 26: Performance evaluation of various ML models for disease prediction. 350

Chapter 1
General Introduction

The first chapter provides an outline of the thesis and the background rationale for the research concept. It also describes the thesis's aims and objectives and summarises the methods used. Finally, the thesis structure outlines how the thesis will be broken down into sub-projects to build an AI-assisted automatic Pre-eclampsia
prediction model.

1.1 Overview

Pre-eclampsia (PE) is a pregnancy–related disease caused by the placenta that is linked to significant maternal and neonatal morbidity and mortality (Huppertz, 2008; Redman and Sargent, 2010; Burton et al., 2019; Yagel and Verlohren, 2020; Cao, Saxena and Gray, 2024). Moreover, it is a complex hypertensive disorder that can be diagnosed with routine antenatal clinical investigations. However, this type of diagnosis aids in the identification of PE–risk women after the onset of PE symptoms. Therefore, late detection of this condition leads to severe complications in the mother and foetus. Maternal complications include renal failure, liver failure, cerebral oedema with seizures and HELLP syndrome (Haemolysis, Elevated Liver enzymes and Low Platelets). Similarly, foetal complications consist of Intra Uterine Growth Restriction (IUGR), oligohydramnios, stillbirth, iatrogenic prematurity, and increased risk of perinatal death (Nirupama et al., 2021). Consequently, determining which pregnant women are most likely to have PE is crucial for personalising preventive treatment, better resource allocation, and increased observation and interventions (Leslie, Thilaganathan and Papageorghiou, 2011a; Zhao et al., 2025). Early prediction of PE is challenging owing to poorly understood causes.

There is no effective preventive mechanism for preventing PE complications, irrespective of the significant clinical manifestation. The primary clinical management technique involves close monitoring and early detection, which facilitates prompt management (Jhee et al., 2019). Relying solely on traditional diagnostic methods such as blood pressure measurement and proteinuria testing is inefficient for early diagnosis of PE, as these approaches detect the condition only after clinical symptoms appear, thus reducing the potential for timely preventive interventions. To address this problem, cutting-edge technology such as Artificial Intelligence (AI) can be used to develop prediction models for recognising PE risk in women using PE-associated biomarkers and other clinical data.

The development of PE risk prediction models has drawn much attention over the past two decades (De Kat et al., 2019). On the other hand, there are no automated AI-assisted prediction models for diagnosing PE risk in early pregnancy in current clinical practice. Therefore, this study used the available maternal characteristics, clinical data, and significant genomic signatures in the AI platform to develop an automated PE risk prediction model. PE–linked genomic biomarkers must be discovered before the prediction model can be created. Genomic and transcriptomic data, including maternal, foetal, placental, and cell-free nucleic acid data, are used to identify PE-linked biomarkers and significant gene expression signatures associated with pre-eclampsia. This study used a variety of bioinformatics analyses to identify significant gene expressions and variant signatures from those data.

1.2 General Background

PE is a dangerous maternal health condition that affects 2 to 8% of all pregnancies worldwide (Tannetta and Sargent, 2013; Richards et al., 2021; Taher, Ali and Otman, 2022). It causes significant maternal and neonatal morbidity and mortality, as well as a substantial increase in foetal preterm birth. PE is also a multi–systemic

illness with genetic and environmental factors involved in its development and pathophysiology, and the only recognised treatment is the delivery of the foetus and placenta. Because the maternal symptoms of hypertension, proteinuria, renal and liver failure, and oedema diminish when placental tissue is removed during delivery, it has been suggested that the placenta, under pathological conditions, may be the leading cause of PE pathogenesis (Yun-Qing et al., 2021; Egbor et al., 2006). Even though the cause of PE is unknown, the pathogenesis of PE is thought to be dominated by five traditional hypotheses such as placental or trophoblast ischemia and hypoxia, vascular endothelial injury, oxidative stress, genetics, and immunological dysregulation (Liu, N. et al., 2021; Jena et al., 2020). PE has two stages in its pathogenesis: aberrant placentation and the emergence of the maternal syndrome (Phipps et al., 2019). Furthermore, endothelial cell dysfunction and endovascular cytotrophoblast shallow penetration in the spiral arteries are two essential aspects in the pathogenesis of PE in abnormal placentation (Dekker and Sibai, 1998). Poor trophoblastic perfusion causes oxidative stress, which leads to the release of trophoblast factors into the maternal circulation, triggering a secondary inflammatory response and endothelial dysfunction, resulting in aberrant placentation (Valenzuela et al., 2011; Borzychowski, Sargent and Redman, 2006). As a result, systemic endothelial dysfunction develops in the later stages of the disease, leading to the typical clinical symptoms of PE, such as hypertension, proteinuria, and oedema (Valenzuela et al., 2011).

PE has a complicated aetiology since it goes from an asymptomatic state in the first trimester to a symptomatic stage later in pregnancy (Huang et al., 2021). Therefore, PE is typically diagnosed after a woman develops clinical symptoms such as a new-onset headache that is resistant to treatment, seizure, dyspnoea, severe persistent right upper quadrant or epigastric pain, and significant placental abruption, or during a routine antenatal consultation if she is asymptomatic but hypertensive (Chappell et al., 2021; Karrar, Martingano and Hong, 2025). Furthermore, despite much research, there is currently no cure for PE other than

delivery (Nirupama et al., 2021). Accordingly, early detection of women at high risk for PE could aid prevention and treatment. There were no acceptable diagnostics to diagnose PE before clinical symptoms appeared until recently, and the diagnosis was based on repeated evaluations of women with established risk factors (Schlembach et al., 2019). As a result of the wide range of PE symptoms, potential biomarkers are needed for early identification (Shafabakhsh et al., 2019).

When the placenta is oxidatively stressed, it produces placental cellular debris, as well as anti-angiogenic substances such as soluble fms-like tyrosine kinase-1 (sFlt1) and soluble Endoglin (sEng), cytokines, and oxidants, into the maternal circulation (Aouache et al., 2018). Furthermore, increased placental apoptotic debris in maternal blood circulation could play a role in the deterioration of inflammatory and immunological responses (Shafabakhsh *et al.,* 2019). In individual cases, there is no single marker that may be used to predict PE (Smets et al., 2006; Tjoa, M. L. et al., 2003; Tjoa, May Lee et al., 2001; Levine et al., 2004; Levine et al., 2005; Salomon et al., 2003; Bosio et al., 2001; Ong et al., 2004; Thadhani et al., 2004; Qiu et al., 2004). A significant number of biochemical biomarkers detected in maternal circulation have been investigated, but no optimal one has yet been identified (Hassan *et al.,* 2019). The most common first-trimester biochemical biomarkers include Pregnancy-Associated Plasma Protein A (PAPP-A), Placental Growth Factor (PlGF), soluble fms-like tyrosine kinase 1 (sFlt1), placental protein 13 (PP13), and triglycerides (Kornacki and Wender-Ożegowska, 2020).

1.3 Aim and Objectives.

This study hypothesised that using gene expression signatures, which include differential gene expression signatures, consistently expressed gene signatures, and variant signatures in the PE prediction model with other clinical and maternal characteristics data will increase the accuracy of PE prediction in early pregnancy. Therefore, this research project aimed to identify genomic signatures and develop

an automated AI-aided prediction model to detect PE risk in women in early pregnancy. The following objectives were created to accomplish the aim of this research study.

1. Identify the gene expression signatures that are differentially expressed and uniquely expressed in cfRNA sequencing data for each gestational week for developing the AI-assisted PE risk prediction model (Chapter 4).
2. Identify the consistent gene expression signatures and unique, consistent gene expression signatures for each gestational week from cfRNA sequencing data for the development of the AI-assisted PE risk prediction model (Chapter 5).
3. Identify the variant signatures per gestational week cfRNA sequencing data for developing a disease prediction model (Chapter 6).
4. Develop an automated AI model incorporating identified genomic biomarkers and assess the effectiveness of the AI (ML) predictive model in detecting the onset of PE in early pregnancy (Chapter 7).

1.4 Methodology

This thesis includes in silico research methodologies based on various bioinformatics and AI-based analyses for developing PE prediction models. Many databases were searched to get the study data for discovering PE-associated biomarkers, including the Sequence Read Archive (SRA), the database of Genotypes and Phenotypes (dbGaP), the Gene Expression Omnibus (GEO) from the National Centre for Biotechnology Information (NCBI), ArrayExpress, European Nucleotide Archive (ENA), European Genome–Phenome archive (EGA) and Genomics England Limited (GEL). Data procured from these platforms were subsequently streamlined into specialised datasets tailored for bioinformatics evaluation and AI-centric predictive modelling. The exploration for genomic biomarkers was facilitated using an array of public bioinformatics tools operating on the R, Python, and Linux platforms.

1.5 Thesis Structure

The structure of this thesis consists of a general introduction (chapter 1), seven chapters, and a conclusion chapter. Except for the first two and last two chapters, each chapter contains independent work contributing to the development of an automated PE prediction model.

1.5.1 Chapter 1: General Introduction

This chapter explains the general information of the thesis, such as the background of the research topic, the aim and objectives, and the thesis structure.

1.5.2 Chapter 2: Literature Review

This chapter explains the existing research and debates relevant to PE and its prediction with AI, as well as an overview of what is known and what is still unknown about PE and its early prediction.

1.5.3 Chapter 3: General Methods

This chapter explains the common methods used in multiple projects, repeated in each project with different data types.

1.5.4 Chapter 4: Identification of Pre-eclampsia Associated Unique Differentially Expressed Gene Signatures at Each Gestational Week in Early Pregnancy

This chapter elucidates the differential gene expression signatures derived from cell-free RNA sequencing data during each week of early pregnancy. Subsequently, with these identified gene expression patterns, an artificial intelligence-assisted predictive model is developed to identify women at risk of PE during early pregnancy.

1.5.5 Chapter 5: Identification of Pre-eclamptic Unique Consistent Gene Expression Signatures at Each Gestational Week in Early Pregnancy

This chapter introduces the novel method for identifying consistently expressed genes per gestational age from cfRNA sequencing data and how to use artificial intelligence to develop a disease prediction model for identifying women at risk for PE. Moreover, this chapter explores further downstream analysis, such as pathway analysis.

1.5.6 Chapter 6: Identification of Pre-eclamptic Unique Variant Signatures at Each Gestational Week in Early Pregnancy

This chapter of the thesis explores identifying the variants in the cfRNA sequencing data to validate the consistent genes identified in the previous chapter and to develop an AI-assisted prediction model.

1.5.7 Chapter 7: Developing an Automated AI-aided Disease Prediction Model for Detecting Women at Risk of Pre-eclampsia in Early Pregnancy.

This chapter explores the development of an AI-aided disease prediction model using genomic biomarkers from gene expression (differential and consistent), variant calling, and differential methylation analyses.

1.5.8 Chapter 8: Overall Discussion and Conclusion

This final chapter presents a comprehensive discussion of the project focused on developing an AI-aided PE prediction model. It concludes with a summary of the key research findings, addresses the study's limitations, and highlights the significance of these findings for future research on early PE prediction using AI.

Chapter 2
Literature Review

2.1 Introduction

Pre-eclampsia (PE) remains a leading contributor to maternal and perinatal morbidity and mortality, affecting approximately 2% to 8% of pregnancies globally (Karrar, Martingano and Hong, 2025; Sun *et al.*, 2025). This condition accounts for an estimated 46,000 maternal deaths annually, with the burden disproportionately falling on low- and middle-income countries, particularly in sub-Saharan Africa and South Asia (Melesse et al., 2025). The prevalence of PE varies geographically, with higher rates reported in Africa (5.6%) and Latin America (5.3%) compared to high-income countries (3.7%) (Abalos et al., 2013; Wang et al., 2021). Sociodemographic factors such as advanced maternal age (>35 years), ethnicity, and socioeconomic status significantly influence PE risk even within resource-rich settings (Ives et al., 2020. PE contributes to 11%–14% of global maternal mortality, highlighting the urgent need for enhanced diagnostic, preventative, and management strategies worldwide (Pfeiffer et al., 2023; Melesse et al., 2025).

No definitive pharmacological therapy for PE will enhance maternal and foetal outcomes (Zhaoer *et al.,* 2022). Furthermore, aberrant placental development is the root cause of several severe pregnancy problems, including PE, foetal growth restriction, recurrent miscarriage, and stillbirth (Turco and Moffett, 2019; Brosens *et al.,* 2011). Therefore, delivering the placenta remains the known treatment option to avert additional maternal and foetal complications, suggesting that the placenta plays a significant role in the pathogenesis of PE (Young, Levine and Karumanchi, 2010). Even though the pathophysiology of heterogeneous PE is still poorly understood, it is widely believed that abnormal placenta development beginning in the first trimester is a major contributing factor (Petca *et al.,* 2022). Moreover, PE also increases both the mother's and the child's long–term risk of

developing cardiovascular disease, in addition to the substantial health risks it poses to the mother and her baby during pregnancy and labour (Staff, 2019). Therefore, it is essential to understand the aetiology of PE and, more specifically, how the placenta develops in healthy and dysfunctional pregnancies to comprehend the pathogenesis of PE for early detection, prevention, and management.

2.1.1 Research Hypothesis

PE arises due to abnormal placentation, immune dysregulation, oxidative stress, and endothelial dysfunction, which collectively disrupt maternal vascular function and lead to hypertensive disorders in pregnancy.

2.1.2 Research Question

What are the key molecular and immunological mechanisms contributing to the development of PE, and how can early detection strategies be improved?

2.1.3 Research Aim and Objectives

Research Aim

The research aimed to identify the biomarkers linked with PE and investigate the pathophysiological mechanisms underlying PE. It focuses on placental dysfunction, immune dysregulation, oxidative stress, and endothelial dysfunction while identifying early risk assessment and intervention biomarkers.

Research Objectives

1. To examine the role of abnormal placentation and shallow trophoblast invasion in the onset of PE.
2. To analyse the impact of immune system dysregulation, particularly regulatory T cells and natural killer cells, on PE progression.
3. To assess the contributions of oxidative stress and endothelial dysfunction to PE pathophysiology.

4. To identify key molecular factors involved in PE development, such as sFlt–1, sEng, VEGF, and PlGF.
5. To explore potential early biomarkers and predictive models for PE diagnosis.
6. To evaluate maternal risk factors, including chronic hypertension, obesity, diabetes, and autoimmune diseases, in relation to PE susceptibility.
7. To review the literature on emerging technologies for PE prediction, including the role of cell-free RNA (cfRNA) as a non-invasive diagnostic source and the application of Artificial Intelligence (AI) models.

2.1.4 Definition of Pre-eclampsia

PE is a pregnancy disease that begins in the placenta and results in a variety of maternal and foetal complications (Redman, Christopher W. and Sargent, 2005). PE is a multisystemic condition with a poorly characterised aetiology, pathogenesis, and pathophysiology. Variations in the disease's description have influenced the definition of PE over time, and a greater understanding of its pathophysiology has also influenced the definitions.

PE is traditionally defined as a condition characterised by the development of a new onset of hypertension and a new onset of proteinuria after 20 weeks of gestation (Lai *et al.,* 2021; Reddy *et al.,* 2021; Brown, M. A. *et al.,* 2001). Furthermore, the traditional diagnosis of PE is usually based on the fulfilment of the following criteria: pregnancy of more than 20 weeks, proteinuria (2+ on dipstick or more than 300 mg/24 h), and arterial hypertension ≥ 140/90 mmHg (Sisti and Colombi, 2019). Even though the classic definition is determined by the new onset of hypertension in conjunction with proteinuria, some pregnant women present with multisystemic symptoms even when proteinuria is absent. As a result, classical definitions of PE must be broadened with additional multisystemic signs. The expanded definition of PE, as opposed to the traditional definition, better identifies women who develop PE (Magee *et al.,* 2021).

Recently, many professional societies for the study of hypertension have broadened the PE definitions to increase the diagnostic prevalence of PE (Table 1). In 2013, the American College of Obstetricians and Gynaecologists (ACOG) Committee revised the definition of PE to include hypertension with either proteinuria or at least one severe feature (Woelkers *et al.,* 2015). Similarly, in 2014, the International Society for the Study of Hypertension in Pregnancy (ISSHP) revised its definition of PE to include haematological complications, renal insufficiency, impaired liver function, neurological symptoms, or uteroplacental dysfunction in addition to new-onset hypertension in the absence of proteinuria (Kallela *et al.,* 2016). Also, the ISSHP updated its PE definitions in 2018; however, the diagnostic criteria remained the same as in 2014 (Bouter and Duvekot, 2020; Brown, Mark A. *et al.,* 2018). The ACOG and ISSHP have ruled out the possibility of PE diagnosis being based on proteinuria in the updated definitions (Wójtowicz *et al.,* 2019). Moreover, in 2019, the National Institute for Health and Care Excellence (NICE) outlined the definition of PE as the new onset of SBP >140 mm Hg or DBP >90 mm Hg after 20 weeks of pregnancy and one or both of the following medical conditions have coexisted:– proteinuria (urine protein: creatinine ratio ≥30 mg/mmol or albumin: creatinine ratio ≥8 mg/mmol, or ≥1 g/L [2+] on dipstick testing) and other maternal organ dysfunction, including features such as renal or liver involvement, neurological or haematological complications, or uteroplacental dysfunction (such as foetal growth restriction, abnormal umbilical artery Doppler waveform analysis, or stillbirth) (Webster *et al.,* 2019). The updated ACOG 2013 and ISSHP 2014 classifications allow for a more sensitive diagnosis of PE in women with new-onset signs and symptoms of PE when proteinuria is absent (Kallela *et al.,* 2016). According to Woelkers et al., 2015 study, redefining PE as per the updated 2013 ACOG criteria will increase the diagnostic prevalence of PE by 20%. Among these PE definitions, the one proposed by the ISSHP is the most widely accepted (Poon, Liona C. *et al.,* 2019).

Table 1 The diagnostic criteria for PE are from professional societies.

Name of professional society	Characteristics	Diagnosing criteria
International Society for the Study of Hypertension in Pregnancy (ISSHP 2014)	Blood pressure	Systolic blood pressure at ≥140 mmHg and/or diastolic blood pressure at ≥90 mmHg on at least two occasions measured four hours apart in previously normotensive women and is accompanied by ≥1 of the following new-onset conditions at or after 20 weeks gestation.
	Proteinuria	≥30 mg/mol protein: creatinine ratio; ≥300 mg/24hr; or ≥2+ dipstick
	Evidence of Maternal Organ Dysfunction	• Acute kidney injury (creatinine ≥90umol/L; 1 mg/dL), • Liver involvement (elevated transaminases, e.g., alanine aminotransferase or aspartate aminotransferase >40 IU/L with or without right upper quadrant or epigastric abdominal pain) • Neurological complications (e.g., eclampsia, altered mental status, blindness, stroke, clonus, severe headaches, and persistent visual scotomata) • Haematological complications (thrombocytopenia–platelet count <150 000/μL • Disseminated intravascular coagulation. • Haemolysis
	Uteroplacental dysfunction	• Foetal growth restriction • Abnormal umbilical artery Doppler waveform analysis or stillbirth
International Society for the Study of Hypertension in Pregnancy (ISSHP 2018)	Blood pressure	New onset of hypertension (blood pressure of ≥140 mmHg systolic and/or ≥90 mmHg diastolic) after 20 weeks of gestation accompanied by one or more of the following new-onset conditions at or after 20 weeks gestation.
	Proteinuria	≥30 mg/mol protein: creatinine ratio; ≥300 mg/24hr; or ≥2+ dipstick
	Evidence of Maternal Organ Dysfunction	• Renal insufficiency (creatinine > 90 μmol/L; 1 mg/dL) • Liver involvement (elevated transaminases with or without right upper quadrant or epigastric abdominal pain) • Neurological complications (examples include eclampsia, altered mental status, blindness, stroke, hyperreflexia with clonus, severe headaches with hyperreflexia, and persistent visual scotomata) • Haematological complications (thrombocytopenia with platelet count below 150,000/dL) • Disseminated intravascular coagulation.

Name of professional society	Characteristics	Diagnosing criteria
		• Haemolysis
	Uteroplacental dysfunction	• Foetal growth restriction • Abnormal umbilical artery Doppler waveform analysis or stillbirth
American College of Obstetricians and Gynaecologists (ACOG 2013)	Blood pressure	Systolic blood pressure of 140 mm Hg or more or diastolic blood pressure of 90 mm Hg or more on two occasions at least 4 hours apart after 20 weeks of gestation in a woman with a previously normal blood pressure. Systolic blood pressure of 160 mm Hg or more or diastolic blood pressure of 110 mm Hg or more. (Severe hypertension can be confirmed within a short interval (minutes) to facilitate timely antihypertensive therapy).
	Proteinuria	Proteinuria 300 mg or more per 24–hour urine collection (or this amount extrapolated from a timed collection) or Protein/creatinine ratio of 0.3 mg/dL or more or Dipstick reading of 2+ (used only if other quantitative methods not available)
	Evidence of Maternal Organ Dysfunction	Or, in the absence of proteinuria, new-onset hypertension with the new onset of any of the following: 1. Thrombocytopenia: Platelet count less than 100,000 3 109/L 2. Renal insufficiency: Serum creatinine concentrations greater than 1.1 mg/dL or a doubling of the serum creatinine concentration in the absence of other renal disease 3. Impaired liver function: Elevated blood concentrations of liver transaminases to twice normal concentration 4. Pulmonary oedema 5. New-onset headache unresponsive to medication and not accounted for by alternative diagnoses or visual symptoms.
American College of Obstetricians and Gynaecologists (ACOG 2019)	Blood pressure	Namely, new-onset elevated blood pressure (BP) ≥ 140/90 mmHg on two occasions at least 4 hours apart or persistent high BP ≥160/110 mmHg combined (with significant proteinuria, defined as 24–hour urine protein ≥300 mg or urine protein to creatinine ratio ≥ 0.3 or dipstick 2+).
	Proteinuria	Significant proteinuria is defined as 24–hour urine protein ≥300 mg or urine protein to creatinine ratio ≥ 0.3 or dipstick 2+.
	Evidence of Maternal Organ Dysfunction	In cases without proteinuria, PE was diagnosed if the patient had hypertension with any of the following: • Thrombocytopenia (platelet count <100 000 × 109/L)

Name of professional society	Characteristics	Diagnosing criteria
		• Renal insufficiency (serum creatinine >1.1 mg/dL) • Impaired liver function (serum aspartate aminotransferase or serum alanine aminotransferase ≥70 IU/L) • Pulmonary oedema • Headache or visual disturbances.
National Institute for Health and Care Excellence (NICE 2019)	Blood pressure	Concerns could include any of the following: • Sustained systolic BP ≥160 mmHg.
	Evidence of Maternal Organ Dysfunction	• Any maternal biochemical or haematological investigations that cause concern, for example, a new and persistent: 1. rise in creatinine (≥90 micromol/l or more, ≥1 mg/100 ml), or 2. rise in alanine transaminase (>70 IU/l, or twice the upper limit of normal range), or 3. fall in platelet count (<150,000/microlitre). • Signs of impending eclampsia, pulmonary oedema, or other signs of severe PE • Suspected foetal compromise. • Any other clinical signs that cause concern.

2.1.5 Aetiology of Pre-eclampsia

PE is a complicated pregnancy-specific syndrome thought to be brought on by a confluence of genetic, environmental, and immune variables (Williams and Pipkin, 2011). Although the exact cause of PE is not fully understood, numerous theories about how the condition develops, and advances have been proposed, each with a different mechanism or pathway. Several theories have been proposed to explain the phenomenon, including the two-stage model theory, placental dysfunction theory, immunological dysregulation theory, oxidative stress and endothelial dysfunction theory, genetic predisposition theory, and cardiovascular dysfunction theory. The two-stage model theory and the placental dysfunction theory were widely accepted among researchers and healthcare professionals studying PE. These theories have gained significant support due to their comprehensive approach and ability to explain various aspects of the condition.

2.1.5.1 Two-stage Model

One of the most widely accepted theories is the "two-stage" model, which suggests that PE is caused by abnormal placental implantation followed by a maternal systemic inflammatory response (Staff, 2019; Geldenhuys *et al.,* 2018; Gathiram and Moodley, 2016). According to this theory, insufficient trophoblast invasion and spiral artery remodelling in early pregnancy lead to placental ischemia and hypoxia, releasing pro-inflammatory cytokines and oxidative stress. These factors trigger an inflammatory response in the maternal endothelium, leading to hypertension, proteinuria, and other clinical manifestations of PE. PE is a trophoblastic disease rather than a haemodynamic, renal, or even endothelial system, according to the two-stage model theory, first presented by Redman in 1991 from the expansion of an earlier research study published in 1990. It was formerly exclusively recognised as a "hypertensive disease" of pregnancy, but the two-stage approach has fundamentally altered how scientists view PE (Staff, 2019). Redman added that PE is caused by abnormal placental trophoblasts, rather than the foetus or the uterus. The placenta, a component of the foetal tissue, is thought to play a significant role in the development of PE, even though the foetus does not directly cause the disorder (Roberts and Escudero, 2012).

The first two-stage PE model was introduced by Redman in 1991 and featured two linked stages, such as pre-clinical and clinical stages (Figure 1). The pre-clinical stage (placental stage) was brought on by shallow spiral artery remodelling during the first half of pregnancy, which resulted in structural defects in the spiral arteries supplying the intervillous space and placental ischemia (Staff *et al.,* 2022; Gathiram and Moodley, 2016). Additionally, it was suggested that poor placentation-related uteroplacental malperfusion might enhance the discharge of placental products or activities into the mother's circulation. The clinically evident maternal disease in the second stage was thought to be a result of placental

ischemia brought on by spiral artery insufficiency in Stage 1 (Albrecht and Pepe, 2020).

The two-stage model of PE development and its progression is central to understanding the pathogenesis of this pregnancy-related disorder. The model was not static and underwent significant refinements with advancing research. A seminal refinement came from Christopher Redman and colleagues in 1999, which built upon and significantly enhanced understanding of the disease. They proposed a clear distinction between the two stages: stage 1 was defined by the abnormal development of the placenta. Redman emphasised that the problem began at the level of trophoblast invasion and inadequate remodelling of maternal spiral arteries. The result was a poorly perfused placenta. This stage can occur without overt clinical symptoms, making it challenging to detect (Redman, C. and Sargent, 2000). The next stage, 2, was the disease's clinical manifestation, where the ischemic placenta's factors entered the maternal circulation and led to widespread maternal endothelial dysfunction. Here, Redman emphasised the release of factors due to ischemia and highlighted the role of oxidative stress and inflammation in amplifying the disease's symptoms (Redman, Christopher WG, Sacks and Sargent, 1999).

Redman's 1999 model of PE introduced pivotal changes to the understanding of the disorder. This model shifted the focus towards early placental events, particularly highlighting the trophoblast's role in arterial remodelling, distinguishing itself from previous models, which mainly acknowledged placental ischemia. Furthermore, the model emphasised the significant contributions of inflammation and oxidative stress, proposing that in conjunction with placental ischemia, these processes underlie the development of the maternal syndrome of PE. Notably, Redman's model brought together intricate interactions among immunological,

vascular, and inflammatory factors, offering a comprehensive framework to comprehend the multifaceted aetiology of PE.

The extension of the two-stage model in 2014 by Staff et al. provided a comprehensive reframing of the earlier understanding of PE. This revised approach offered a more intricate and nuanced depiction of the disorder's progression, emphasising the multifaceted interplay of factors contributing to its onset and development. The genesis of this evolved understanding finds its roots in the work of Staff et al. (2014), who built upon the foundational tenets of the two-stage model, proposing a nuanced multi-step progression of PE. At the heart of this development was the emphasis on the mother's immune response to the growing foetus. The first stage delineated the dysregulation of maternal immune tolerance towards allogeneic trophoblasts. This phase was no longer mentioned but was demarcated into two distinct stages in this new understanding, bringing forth the intricate processes leading to poor placentation. Following this, the third stage encompassed the consequences of such immune intolerance. Poor immune response was posited to affect placental development directly, culminating in dysfunctional uteroplacental perfusion. The ramifications of this were two–fold: a failure in vascular adaptation and a surge in inflammation.

Then, in the fourth stage, the placenta – under stress and duress – began releasing several inflammatory factors into the maternal bloodstream. This influx of inflammatory agents instigated widespread maternal vascular stress, ultimately manifesting as the clinical symptoms of PE. Simultaneously, Staff et al. introduced complementary pathways that could work alongside or independently of the primary stages. There was a recognition that certain maternal conditions, such as Pre-existing vascular inflammation from ailments like chronic hypertension, autoimmune disorders, or obesity, could both disrupt proper placentation and

amplify the maternal clinical manifestations. This understanding was pivotal as it widened the scope of factors influencing the onset and progression of PE.

Although foundational, the two-stage model was expanded into a comprehensive multi-step model highlighting the complex interactions between the maternal immune response, placental development, and subsequent clinical manifestations. Staff et al.'s work enhanced scientific understanding of this condition and opened new avenues for targeted research and therapeutic interventions.

The two-stage model of PE underwent further refinement in 2019 through the work of Staff et al. This updated model presents a comprehensive insight into the intricate mechanisms underlying the progression of PE. The cornerstone of the model lies in its elucidation of two primary pathways culminating in Stage 1 of PE, characterised by syncytiotrophoblast (STB) stress. The first pathway, termed the "poor placentation" pathway, delineates a sequence wherein suboptimal placental development leads to STB stress. The second, the "congested placenta" pathway, highlights how placental congestion can similarly result in this stress endpoint. Notably, the model acknowledges the potential contribution of unidentified factors to the development of Stage 1.

A notable candidate in this regard is excessive trophoblast senescence, a phenomenon discussed by Cox and Redman in 2017. This pathway involves the accelerated ageing of placental cells, potentially augmenting the overall cellular stress within the placenta. A pivotal departure from the original two-stage model lies in considering maternal risk factors. Unlike the initial model, the 2019 update incorporates and underscores the role of maternal influences. The earlier concept of "maternal type" PE, introduced by Ness and Roberts in 1996, posited instances of hypertension and proteinuria in expectant mothers that weren't intrinsically tied to placental dysfunction or altered uteroplacental circulation.

In contrast, the 2019 model underscores the impact of maternal factors on the pathogenesis of PE. Conditions such as chronic arterial disease, obesity, and specific autoimmune diseases are posited to affect various facets of placentation, from its dimensions to its function. These maternal factors converge on the dysfunction of Stage 1. Furthermore, they may heighten maternal cardiovascular sensitivity to placental-released substances, ultimately giving rise to the clinical signs observed in Stage 2 of PE.

A notable projection of this updated model is its assertion that "maternal" PE is contingent upon the presence of a placenta. This firmly underscores the indispensable role of placental dysfunction in the condition's aetiology. Through its incorporation of maternal risk factors and the multifaceted pathways contributing to PE, the 2019 updated two-stage model offers a more holistic and nuanced understanding of this complex disorder.

2.1.5.2 Placental Dysfunction Theory

The placental dysfunction theory has emerged as a prominent explanation for the pathogenesis of PE. This theory posits that inadequate placental development and function play a central role in the development of PE. The placenta, a remarkable organ that forms during pregnancy, is the interface between the maternal and foetal circulatory systems, facilitating nutrient and oxygen exchange. It also plays a critical role in modulating maternal immune responses to ensure the foetus is tolerated as a semi–allograft. In cases of PE, however, the placenta's normal development is disrupted, leading to a cascade of events contributing to the disorder. One key aspect of the placental dysfunction theory is the concept of shallow trophoblast invasion and impaired spiral artery remodelling (Vogtmann et al., 2021). During healthy pregnancies, specialised trophoblast cells invade the uterine lining and spiral arteries, transforming them into low-resistance vessels that can accommodate the increased blood flow required to nourish the growing foetus.

In PE, this invasion is insufficient, leading to inadequate remodelling of the arteries and decreased blood flow to the placenta. This phenomenon results in placental ischemia, which is characterised by reduced oxygen and nutrient supply to the developing foetus (Redman and Sargent, 2009).

Placental ischemia triggers a series of responses aimed at restoring oxygen supply, but these adaptations inadvertently contribute to the pathogenesis of PE. The stressed placenta releases various factors, including anti-angiogenic factors like soluble fms-like tyrosine kinase-1 (sFlt-1) and soluble endoglin (sEng). These factors disrupt the balance between pro–angiogenic and anti-angiogenic signals, causing endothelial dysfunction and impaired maternal vascular health (Torres-Vergara, Rivera and Escudero, 2022; Ramma and Ahmed, 2011; Steinberg, Khankin and Karumanchi, 2009).

Endothelial dysfunction is a hallmark of PE, leading to increased peripheral resistance and elevated blood pressure. Moreover, the release of sFlt-1, which acts as a decoy receptor for vascular endothelial growth factor (VEGF), further exacerbates the endothelial dysfunction and contributes to the clinical manifestations of PE, including proteinuria and oedema (Levine et al., 2004). The immunological dimension of the placental dysfunction theory is also crucial. The placenta is a unique immunological interface where maternal immune tolerance to the foetus is maintained. In PE, an altered immunological response to the placenta can trigger systemic inflammation and immune dysregulation. Aberrant immune responses may contribute to the endothelial dysfunction observed in PE (Wallace, Fraser and Cartwright, 2012).

Research continues to unravel the intricate mechanisms underlying the placental dysfunction theory. Genetic factors, oxidative stress, and the role of other angiogenic and anti-angiogenic factors are areas of ongoing investigation. However, the placental dysfunction theory provides a comprehensive framework that explains multiple aspects of PE, from its initiation to the maternal and foetal

consequences (Schneider, 2017). Understanding the placental dysfunction theory has significant clinical implications. Early detection and intervention strategies targeting placental dysfunction, endothelial dysfunction, and immune dysregulation may offer avenues for mitigating the severity of PE and improving outcomes for both mother and baby (McElwain et al., 2020). By addressing the root causes of the disorder, researchers and healthcare professionals are taking critical steps towards reducing the burden of PE on maternal and neonatal health.

The placental dysfunction theory has gained prominence in explaining the complex pathogenesis of PE. Shallow trophoblast invasion, impaired spiral artery remodelling, and subsequent placental ischemia lead to the release of factors that disrupt the angiogenic balance, cause endothelial dysfunction, and trigger immune responses. Understanding these mechanisms offers hope for developing targeted interventions and personalised approaches to manage and prevent PE, ultimately enhancing the well-being of pregnant women and their unborn babies.

2.1.5.3 Immunological Dysregulation Theory

While its precise aetiology of PE remains elusive, there has been a significant shift towards understanding the immune system's role in its onset and progression, particularly the concept of immunological dysregulation. The Immunological Dysregulation Theory postulates that an altered maternal immune response, especially involving specific immune cells and regulatory molecules, is pivotal to the pathogenesis of PE. Under normal circumstances, the maternal immune system adapts to tolerate the semi-allogeneic foetus, ensuring the pregnancy progresses without complications. This adaptation involves the maternal immune system recognising and accepting the presence of paternal antigens while maintaining the ability to ward off infections (Burton *et al.,* 2019). A fundamental aspect of this tolerance is the interaction of maternal immune cells with the invading trophoblast cells, which anchor the placenta to the uterine wall and facilitate nutrient exchange. In a typical pregnancy, trophoblasts modulate their

expression of immune–related molecules, effectively promoting an immunosuppressive environment (Fu *et al.,* 2017). However, the landscape shifts dramatically in the context of PE. Trophoblasts in affected pregnancies may incite excessive maternal immune responses or inadequately elicit the necessary immunosuppression, leading to aberrant immune reactions. This malfunction is believed to be partly due to shallow trophoblast invasion and impaired uterine spiral artery remodelling (Redman, C. W., Sargent and Staff, 2014). Diving deeper, the role of particular immune cells, namely regulatory T cells (Tregs) and natural killer (NK) cells, is now being keenly studied. Tregs, characterised by their immunosuppressive function, are crucial for maintaining tolerance towards the foetus (Sasaki *et al.,* 2004). Research has illuminated that in PE, there is a substantial decrease in the population and function of Tregs, thereby failing to suppress deleterious immune responses (Prins *et al.,* 2009). Similarly, NK cells, particularly uterine NK cells, have a unique phenotype during pregnancy and are involved in modulating placental development. Aberrant activation of NK cells, observed in PE, can lead to trophoblast apoptosis, subsequently contributing to placental dysfunction (Kemp *et al.,* 2016).

A consistent theme in the Immunological Dysregulation Theory is the upsurge of inflammation. The altered immune landscape in PE skews towards an inflammatory phenotype. There is an increased production of pro-inflammatory cytokines and chemokines, along with evidence of oxidative stress. This inflammatory milieu, notably characterised by elevated TNF–α, IL–6, and other inflammatory markers, exacerbates endothelial dysfunction, a key characteristic of PE (Duley, 2009). These findings underscore the importance of the Immunological Dysregulation Theory in deciphering the aetiology of PE. Nevertheless, it is vital to acknowledge the intricate, multifaceted nature of the immune response during pregnancy. While the Immunological Dysregulation Theory provides a foundation, the nuances of the interplay between the immune system, placenta, and overall maternal physiology are vast and necessitate further exploration. A comprehensive

understanding of these complex interactions is key to unveiling PE's mysteries. Moreover, such insights are quintessential for devising targeted therapeutic interventions that could revolutionise the management of this condition, ultimately safeguarding maternal and foetal health.

2.1.5.4 Oxidative Stress and Endothelial Dysfunction Theory

Despite being the subject of extensive research, the precise aetiology of PE remains elusive. However, the intertwined roles of oxidative stress and endothelial dysfunction stand out as potential culprits in the pathogenesis of this condition. Oxidative stress arises from an imbalance between Reactive Oxygen Species (ROS) production and the body's antioxidative defences. In PE, this imbalance tilts towards the overproduction of ROS, leading to increased evidence of lipid peroxidation, indicative of oxidative damage to cell membranes (Redman, Christopher W. and Sargent, 2005). A major underlying factor for this oxidative surge is believed to be placental ischaemia, a result of impaired spiral artery remodelling. The placenta of affected pregnancies often displays diminished antioxidants like superoxide dismutase while exhibiting increased lipid peroxidation products. This heightened oxidative stress is implicated in releasing anti-angiogenic factors such as sFlt-1 and soluble endoglin, which further sabotage placental and endothelial function (Maynard *et al.,* 2003).

Simultaneously, endothelial dysfunction, defined by a diminished vasodilation capacity, a pro-inflammatory state, and increased prothrombotic properties, emerges as a prominent feature in PE. Factors released by an ischemic placenta, including ROS and anti-angiogenic proteins, are postulated to propagate systemic endothelial dysfunction (Levine *et al.,* 2004). A direct outcome of this dysfunction is the reduced bioavailability of Nitric Oxide (NO), a pivotal vasodilator and platelet aggregation inhibitor. Increased oxidative stress quenches NO, intensifying the hypertensive state seen in PE (Eddy, Bidwell and George, 2018). Furthermore, endothelial cells in preeclamptic conditions manifest elevated expression of

adhesion molecules, promoting leukocyte adhesion and inflammation, further complicating the picture (Goulopoulou and Davidge, 2015).

Oxidative stress and endothelial dysfunction in PE are not isolated phenomena; they are intricately linked and reciprocally reinforce each other. ROS, a byproduct of oxidative stress, can directly undermine endothelial function by suppressing the production and activity of NO. Conversely, a compromised endothelium can amplify oxidative stress by fostering inflammation and thrombosis, which, in turn, spur more ROS production. This bidirectional relationship between oxidative stress and endothelial dysfunction sets off a pernicious cycle, further worsening the clinical manifestations of PE. Moreover, factors excreted by the dysfunctional endothelium can influence the placenta, further exacerbating its oxidative stress. This is evident in the observation that anti-angiogenic factors, whose release is stimulated by placental oxidative stress, can induce systemic endothelial dysfunction, further accentuating the disease state (Chaiworapongsa *et al.,* 2014).

Understanding the relationship between oxidative stress and endothelial dysfunction offers a glimpse into the complex pathogenesis of PE. Their roles, fuelled by poor placentation and subsequent placental ischaemia, generate a vicious feedback loop. Deciphering these mechanisms could pave the way for innovative therapeutic interventions and strategies for managing this challenging obstetric condition.

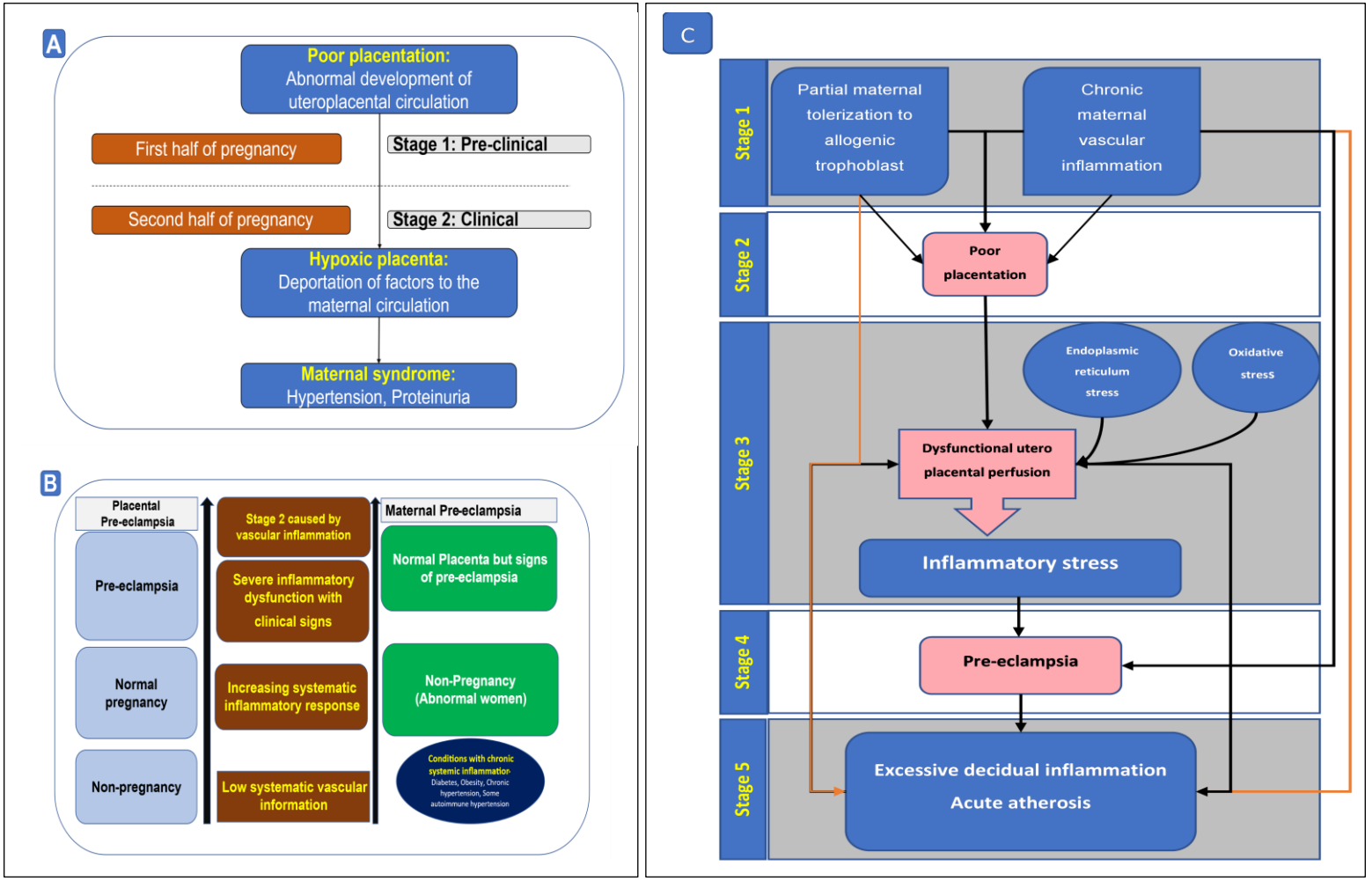


Figure 1: Schematic representation of PE two-stage model update over time.

A - Illustration of the initial two-stage model for PE, introduced in 1991. **B** - Depiction of the extended two-stage model for PE, from 1999. **C** - Update of the 1999 two-stage model for PE, as revised in 2014

2.1.6 Risk Factors of Pre-eclampsia

Early identification and treatment of PE can significantly improve the outcomes for both mothers and newborns. The start and severity of the PE illness are surprising, even in cases where expectant mothers have access to advanced healthcare facilities. Many risk factors are linked to PE, even if its exact genesis is unknown. Although the precise cause of PE is still unknown, several risk factors have been found. Comprehending these risk variables is crucial for suitable screening, prompt identification, and action. Table 1 displays the high and moderate risk factors for PE.

Table 2: Summary of Key Risk Determinants for Pre-eclampsia.

Principal Risk Factors of Pre-eclampsia
1. Prior history of preeclampsia
2. Chronic hypertension
3. Renal disease
4. First Pregnancy
5. Multiple Pregnancy
6. Maternal Age
7. Obesity
8. Autoimmune Disorders
9. Diabetes: Pre-existing type I or type II diabetes mellitus
10. Family History
11. IVF Pregnancies
12. Interval Between Pregnancies
13. Ethnicity

2.1.6.1 Prior History of Pre-eclampsia

A history of PE in prior pregnancies is one of the main risk factors for the illness. Pregnant women who experienced PE in a previous pregnancy are more likely to experience it again (Havers-Borgersen et al., 2023; De Marco et al., 2025). One recent retrospective analysis showed that among women with a history of PE, about 42.9% experienced recurrence, with a higher risk associated with prior early-onset preeclampsia and chronic hypertension (Nie et al., 2024). For instance, women who had early–onset PE (before 34 weeks of gestation) or severe features during their initial episode are more likely to encounter the disorder in subsequent pregnancies (Dong et al., 2021; Wainstock and Sheiner, 2022). Several theories explain this recurrence risk. Firstly, individual predispositions like genetic factors, immune responses, or underlying endothelial dysfunction can persist across pregnancies. Secondly, women with a history of PE may have subtle chronic vascular or kidney diseases that become exacerbated by the physiological demands of pregnancy (De Marco et al., 2025). It's essential to note that while a prior history of PE increases the risk, it does not guarantee its occurrence in subsequent pregnancies. Conversely, the absence of PE in previous pregnancies doesn't necessarily protect against its onset in future ones, though the risk is substantially lower. For women with a history of PE, comprehensive prenatal care, including early and regular antenatal checks, close blood pressure monitoring, and potential prophylactic interventions like low-dose aspirin, is pivotal in managing the heightened risk in subsequent pregnancies.

2.1.6.2 Chronic Hypertension

Chronic hypertension is a well–established high–risk factor for the development of PE. By definition, chronic hypertension indicates a Pre-existing vascular tone and responsiveness alteration. It underscores an underlying endothelial dysfunction,

which, combined with the placental and vascular adaptations required in pregnancy, can compound the risk for superimposed PE (Lei *et al.*, 2022). Furthermore, the systemic inflammation often accompanying chronic hypertension can exacerbate placental oxidative stress, which has been pinpointed as a significant driver of PE (Martini *et al.*, 2025). Moreover, vascular remodelling, crucial for adequate placental perfusion and prevention of placental ischemia, is often compromised in women with chronic hypertension (Shaw *et al.*, 2024). Inadequate remodelling can precipitate ischemic placental disease, of which PE is a key manifestation. Given the Pre-existing vascular compromise in women with chronic hypertension, the physiological demand for increased cardiac output and expanded blood volume during pregnancy may further strain the vascular system, heightening the risk of PE. Hence, regular antenatal monitoring of such high–risk women is essential to detect early markers of superimposed PE and to ensure early intervention.

2.1.6.3 Renal Disease

Renal disease, with its associated alterations in kidney function and systemic haemodynamics, poses a significant risk for the development of PE during pregnancy. The kidney plays a pivotal role in regulating blood pressure, fluid balance, and waste excretion, which can be compromised in patients with underlying renal disorders. Research has consistently indicated that women with chronic kidney disease (CKD) have a heightened risk of developing PE (Nevis *et al.,* 2011; Zhang, Jing-Jing *et al.,* 2015; Wiles *et al.,* 2019). This susceptibility can be attributed to several factors. Firstly, damaged renal glomeruli, often seen in CKD, can be more susceptible to the systemic endothelial dysfunction characteristic of PE (Hui, D. and Hladunewich, 2019). The impaired glomerular function can further reduce the kidney's ability to handle the increased renal plasma flow during pregnancy, resulting in proteinuria, a hallmark of PE. Furthermore, women with renal disease often have Pre-existing hypertension, a

recognised independent risk factor for PE (Haudiquet *et al.*, 2025). This Pre-existing hypertension can exacerbate the vasoconstrictive and inflammatory responses seen in PE. Additionally, the renin–angiotensin system, pivotal in regulating blood pressure and fluid balance, may be dysregulated in those with renal diseases. An altered renin–angiotensin response can enhance susceptibility to the imbalances in angiogenic and anti-angiogenic factors observed in PE (Flint *et al.,* 2019; Szczepanski *et al.,* 2020). Renal disease stands as a significant predisposing factor for PE, primarily due to the kidney's central role in maintaining cardiovascular homeostasis and its interaction with systemic factors during pregnancy.

2.1.6.4 First Pregnancy

Nulliparity, or being pregnant for the first time, has long been recognised as a significant risk factor for PE. Although the exact mechanisms remain unclear, one hypothesis suggests that the maternal immune system’s response to foreign paternal antigens plays a role, with the risk diminishing in subsequent pregnancies as the body becomes more accustomed to these antigens (Miranda, Edwards and Myers, 2011). During the initial stages of pregnancy, nulliparous women encounter foetal antigens for the first time, which may lead to an exaggerated immune response and endothelial dysfunction, a hallmark of PE (Kornacki *et al.*, 2021). Additionally, the remodelling of spiral arteries supplying the placenta is crucial for a healthy pregnancy, and in nulliparous women, this process may be less effective, contributing to poor placental perfusion and increased susceptibility to PE (Gathiram and Moodley, 2016; Staff, 2019). Epidemiological studies consistently show that nulliparous women have a higher incidence of PE than multiparous women, highlighting the role of physiological adaptation in reducing PE risk in subsequent pregnancies (Harris *et al.*, 2024). The vascular changes that occur during a woman’s first pregnancy, including the remodelling of maternal blood vessels to accommodate increased blood flow, may be more pronounced and

prone to abnormalities, potentially leading to poor placental development and PE (Hong *et al.*, 2021).

2.1.6.5 Multiple Pregnancy

Multiple pregnancies, characterised by the concurrent gestation of two or more foetuses, are a recognised high–risk factor for PE. The association between various pregnancies and an elevated risk of PE is believed to arise from the increased placental mass present in such pregnancies. With more placental tissue, antiangiogenic factors are amplified into the maternal circulation, hindering the vascular remodelling necessary for appropriate placental blood flow (Wang, Wu and Shen, 2021). The impaired vascular remodelling and the resultant placental ischemia have been posited as pivotal events in the pathogenesis of PE. Furthermore, twin or higher–order pregnancies are linked with an intensified maternal systemic inflammatory response, which can exacerbate endothelial dysfunction, a primary characteristic of PE (Wang, Ying, Wu and Shen, 2021a). The increased placental mass in multiple pregnancies releases more syncytiotrophoblast debris into the maternal circulation, triggering inflammation and contributing to the endothelial damage seen in PE (Wang, Wu and Shen, 2021). Moreover, studies have shown that the risk of PE is notably higher in twin pregnancies compared to singleton ones, and this risk further escalates with higher-order multiples, such as triplets (Wang, Ying, Wu and Shen, 2021). Given the augmented risk in multiple pregnancies, rigorous prenatal monitoring is paramount for early detection and management of PE in these populations, ensuring optimal maternal and foetal outcomes.

2.1.6.6 Maternal Age

Maternal age is increasingly recognised as a salient risk factor in the development of PE. As women worldwide tend to delay childbirth for various personal and societal reasons, understanding the ramifications of maternal age on PE risk is

becoming paramount (Zakour Khadari *et al.*, 2025). Younger mothers, particularly those in their early to mid–teens, were historically considered at a heightened risk for PE. It has been postulated that their physiological immaturity, coupled with the demands of pregnancy, could contribute to adverse placentation, a precursor to preeclamptic manifestations (van der Krogt *et al.*, 2025). However, emerging research has spotlighted the other end of the age spectrum, revealing that advanced maternal age, usually defined as age 35 or older at the time of delivery, is associated with a notably increased incidence of PE. The age–related decline in cardiovascular adaptability, coupled with a higher prevalence of comorbid conditions like hypertension and diabetes in older women, might make them more susceptible to the endothelial dysfunction that underpins PE (Q. Xu *et al.*, 2025). Furthermore, it's worth noting that women of advanced maternal age are more inclined to seek assisted reproductive technologies. Apart from age, such interventions have been independently linked to elevated PE risks (Deng *et al.*, 2025).

2.1.6.7 Obesity

Obesity, defined by a Body Mass Index (BMI) of 30 or greater, is a rapidly growing public health concern that has repercussions beyond metabolic and cardiovascular disorders. In the realm of maternal health, obesity has been spotlighted as a significant predisposing factor for the development of PE, a hypertensive condition specific to pregnancy. This association is especially pertinent given the escalating prevalence of obesity in women of reproductive age globally (Gong *et al.*, 2022). The pathophysiology linking obesity to PE is multifaceted. Adipose tissue is not just an energy storage depot; it actively secretes adipokines, cytokines, and other bioactive substances. In obese pregnant women, a pro-inflammatory state is fostered by an elevated secretion of inflammatory markers from excess adipose tissue. This inflammation may contribute to the endothelial dysfunction that is emblematic of PE (Herrock, Deer and LaMarca, 2023). Additionally, obesity is

linked to insulin resistance, which can result in an overproduction of insulin–like growth factors. These factors may interfere with the vasodilatory effects necessary for healthy placental blood flow, further contributing to the pathogenesis of PE (Hu and Zhang, 2021). Moreover, obesity is often intertwined with other comorbidities like chronic hypertension and diabetes, both of which compound the risk of developing PE. Studies have shown that as maternal BMI increases, so does the risk for PE, making it a dose–dependent relationship (Xiong *et al.*, 2025). Given these findings, clinicians must recognise obesity not just as a standalone risk but as part of an interconnected web of factors that can exacerbate the threat of PE. Proactive measures, such as weight management counselling before conception and vigilant monitoring during pregnancy, can be instrumental in mitigating risks and ensuring optimal maternal and foetal outcomes.

2.1.6.8 Autoimmune Disorders

Autoimmune conditions like Systemic Lupus Erythematosus (SLE) and antiphospholipid syndrome are also associated with PE. One prominent hypothesis is that women with autoimmune disorders possess a heightened systemic inflammatory response, which predisposes them to the development of PE (Dahlstrom, Subramaniam and Faye-Petersen, 2022; Buyon *et al.,* 2015). For instance, SLE is associated with elevated levels of pro-inflammatory cytokines like TNF–α and IL–6, also observed in PE (Castellanos Gutierrez *et al.*, 2022). Antiphospholipid syndrome (APS) is another autoimmune disorder strongly linked to PE (Skoura *et al.*, 2022). Patients with APS produce antiphospholipid antibodies, which can promote clot formation and may interfere with trophoblast differentiation and invasion. Impaired trophoblast invasion can result in incomplete spiral artery remodelling, a defining feature of PE (Wang *et al.*, 2025). The antibodies further instigate a pro-inflammatory state, promoting endothelial dysfunction, a shared feature between both conditions (Feng *et al.*, 2024). Consequently, APS pregnancies are at an elevated risk for PE, necessitating

vigilant monitoring and interventions like aspirin or heparin administration to mitigate risks (Mayer-Pickel *et al.*, 2020). Understanding the link between APS and PE can offer therapeutic and preventive strategies in high–risk populations.

2.1.6.9 Diabetes: Pre-existing Type I or Type II Diabetes Mellitus

Pre-existing type I and type II diabetes mellitus are increasingly recognised as notable risk factors for the development of PE. The interplay between diabetes and PE can be viewed through a lens of metabolic and vascular challenges intrinsic to both disorders, making their confluence especially concerning for pregnant women (Yang and Wu, 2022). Diabetes, irrespective of its type, is associated with endothelial dysfunction. The chronic hyperglycaemia inherent in diabetes mellitus often produces Advanced Glycation End products (AGEs) and activates the receptor for the AGE (RAGE) pathway. This activation is implicated in oxidative stress and inflammation, leading to vascular damage that closely mirrors the endothelial dysfunction observed in PE (Herrock, Deer and LaMarca, 2023). Additionally, insulin resistance, particularly pronounced in type II diabetes, results in compensatory hyperinsulinemia. Elevated insulin levels can exert vasoconstrictive effects and impede the necessary vasodilation in placental blood vessels, further aggravating the risk for PE (McElwain *et al.*, 2020). Furthermore, the presence of nephropathy in women with longstanding type I diabetes can compound the risk. The renal dysfunction and proteinuria, commonly seen in diabetic nephropathy, are hallmarks of severe PE, making early differentiation between the two conditions a clinical challenge (Auger *et al.*, 2025). Given these intertwined pathophysiological pathways, it becomes paramount for clinicians to employ vigilant prenatal monitoring in pregnant women with Pre-existing diabetes. Early detection, rigorous glycaemic control, and appropriate interventions can play a pivotal role in mitigating the compounded risks of diabetes and PE, ensuring more favourable maternal and neonatal outcomes.

2.1.6.10 Family History

A family history of PE has long been identified as a crucial determinant in a woman's susceptibility to this hypertensive disorder of pregnancy. Delving into genetic predispositions, the inheritance patterns associated with PE suggest the convergence of multifactorial influences that span both genetic and environmental domains (Slobodchikova *et al.*, 2025). Twin studies provide compelling evidence of the hereditary nature of PE. For instance, monozygotic twins exhibit higher concordance rates for PE compared to dizygotic twins, underscoring the genetic underpinnings of this disorder (Jasper *et al.*, 2024). Women whose mothers or sisters had PE are notably at an increased risk, suggesting a direct line of inheritance. Specifically, a woman with a positive family history can have her risk of developing the condition augmented two to five-fold (Serrano *et al.*, 2020). It is worth noting that while the genetic markers and pathways that predispose individuals to PE are not yet fully elucidated, ongoing research suggests links to genes related to angiogenesis, immunity, and cardiovascular function. Recent advances in genome–wide association studies (GWAS) continue to shed light on potential genetic loci associated with the disorder (Tyrmi *et al.*, 2023). However, genetics alone does not tell the entire story. The interplay between genetic susceptibilities and environmental factors, such as diet, lifestyle, and overall health, further compounds the risk (Jasper *et al.*, 2024). Thus, while a family history provides a significant clue to potential predispositions, a holistic assessment that encompasses genetic and environmental risk factors becomes essential.

2.1.6.11 IVF Pregnancies

In vitro fertilisation (IVF) has revolutionised reproductive medicine, offering countless couples an opportunity to conceive. However, pregnancies resulting from assisted reproductive technologies (ART), such as IVF, have been consistently associated with a heightened risk of maternal complications, most notably PE (Donno *et al.*, 2025). While the exact aetiology of PE remains

multifaceted and not entirely understood, recent literature suggests that pregnancies conceived via IVF might have a different placentation process than those conceived naturally. This divergence can lead to suboptimal trophoblast invasion and impaired remodelling of spiral arteries, factors known to precipitate PE (Nayak *et al.*, 2024). It's worth noting that IVF pregnancies often involve multiple embryos being implanted, leading to a higher prevalence of twin or higher–order gestations. Multiple pregnancies, in turn, are established risk factors for PE due to the increased placental mass and associated vascular demands. However, even singleton pregnancies resulting from IVF have been shown to exhibit an elevated risk of PE compared to naturally conceived singletons, indicating intrinsic risk factors associated with the IVF procedure itself (Rahu *et al.*, 2019; von Versen-Höynck *et al.*, 2019). The underlying mechanisms linking IVF and PE risk might also be tied to the ovarian stimulation protocols utilised in IVF. Exogenous hormone treatments can influence endometrial receptivity, potentially affecting placentation and subsequent pregnancy outcomes (Teh, McBain and Rogers, 2016; Parisi *et al.*, 2023).

2.1.6.12 Interval Between Pregnancies

The interval between pregnancies, often referred to as the Interpregnancy Interval (IPI), has garnered attention for its potential implications on various maternal and foetal outcomes, including PE (Njoroge *et al.*, 2021). Short IPIs, often defined as intervals less than 18 months, have been associated with an increased risk of PE in subsequent pregnancies. One proposed mechanism underlying this increased risk is the insufficient time for maternal physiological recovery following a previous pregnancy. The rapid succession of pregnancies might prevent the complete restoration of nutritional reserves, particularly folate, which plays a crucial role in placental development and function. Furthermore, the lingering inflammation from a recent pregnancy can predispose the maternal system to the endothelial dysfunction characteristic of PE (Ni *et al.*, 2023). Conversely, long IPIs, typically

intervals exceeding 59 months, have also been linked to heightened PE risk. It's hypothesised that the protective immunological adaptations garnered from an initial pregnancy may wane over extended periods. The loss of such immunotolerance could render the maternal system more susceptible to the placental and immunological triggers of PE during subsequent pregnancies (Kornacki *et al.*, 2024). Fundamentally, the interplay of IPIs with PE risk appears to follow a 'Goldilocks' principle: both extremely short and notably extended IPIs have been linked to heightened chances of developing PE. While other confounding factors may contribute to these observations, IPI remains essential for maternal health counselling.

2.1.6.13 Ethnicity

Ethnicity is increasingly being recognised as an influential determinant in the susceptibility to PE, a pregnancy–associated hypertensive disorder that poses significant risks to both mother and foetus. The ethnic disparities observed in PE incidence highlight the intricate interplay between genetic, environmental, and socio–cultural factors in its pathogenesis (Conti-Ramsden, de Marvao and Chappell, 2025). Women of African descent consistently exhibit higher PE prevalence and more severe disease forms compared to Caucasian women, likely influenced by genetic predispositions affecting hypertension, endothelial function, and inflammatory responses (Nakimuli *et al.*, 2014; Fasanya *et al.*, 2021). Conversely, East Asian women, particularly from China and Japan, tend to have lower PE incidence, potentially due to protective genetic factors along with dietary and lifestyle differences (Shi *et al.*, 2021; Conti-Ramsden, de Marvao and Chappell, 2025). The risk profile among Hispanic women in the United States often deviates from their regions of origin, showing elevated PE risk attributed to socio-economic stressors, healthcare access challenges, and lifestyle modifications associated with acculturation (Boakye *et al.*, 2021; Johnson and Louis, 2022). While ethnicity undeniably influences PE predisposition, it must be examined

alongside broader genetic, socio-cultural, and environmental contexts. Such a comprehensive approach allows for tailored prenatal care and interventions to enhance maternal-foetal outcomes (Fasanya *et al.*, 2021; Conti-Ramsden, de Marvao and Chappell, 2025).

2.1.7 Pathogenesis of Pre-eclampsia

The exact cause of PE remains elusive; however, it is widely believed that its origins are rooted in the placenta. Early in NP, invasive extravillous trophoblasts transform the maternal spiral arteries into large-diameter, low-resistance vessels that ensure adequate blood flow to the developing foetus. In PE, this remodelling process is incomplete, leading to reduced placental perfusion and a cascade of subsequent events (Redman, Christopher W. and Sargent, 2005). This maladapted placental perfusion is thought to result in placental ischemia and hypoxia. The placenta releases many factors into the maternal circulation as a compensatory response. These include pro-inflammatory cytokines, antiangiogenic factors like soluble fms-like tyrosine kinase-1 (sFlt-1) and soluble endoglin (sEng), and other agents that contribute to endothelial dysfunction (Levine *et al.,* 2004). Elevated levels of sFlt-1, for example, have been observed to sequester vascular endothelial growth factor (VEGF) and placental growth factor (PlGF), leading to endothelial dysfunction, one of the hallmarks of PE (Maynard *et al.,* 2003).

Endothelial dysfunction is pivotal in the manifestation of clinical symptoms observed in preeclamptic patients, prominently featuring hypertension, proteinuria, and organ damage. This dysfunction is further intensified by a systemic inflammatory response, which triggers coagulation pathways, induces vasoconstriction, and escalates oxidative stress (Redman, Christopher W. and Sargent, 2005). At the heart of PE's pathogenesis lies a complex web of events: it initiates from suboptimal placental perfusion, progressing to the secretion of angiogenic and inflammatory factors, culminating in systemic endothelial

dysfunction. Hence, a comprehensive understanding of normal and abnormal placentation processes is indispensable for grasping the intricate pathogenesis of PE.

2.1.7.1 Process of Normal Placentation

The placenta is a specialised organ that forms during pregnancy and plays a crucial role in exchanging nutrients, gases, and waste products between the mother and the developing foetus. It is a complex structure formed from maternal and foetal tissues and is essential for the foetus's growth and survival.

In more detail, the haploid sperm and haploid egg join during fertilisation to form the one–cell diploid fertilised egg (zygote), which begins the placenta and foetal development (Li, L., Lu and Dean, 2013). The zygote passes through a series of mitotic divisions as it travels down the fallopian tube to the uterus, moving through 2–cell, 4–cell, 8–cell, and 16–cell stages. This process is known as cleavage, and the cells that result from it are known as blastomeres. Furthermore, in a zygote, segmentation is the process of cleavage. The zygote is partitioned into a cluster of solid spheres of cells within the zona pellucida when the cells number around sixteen. This is known as a morula, and it begins a compaction process to become the blastocyst, the following primary stage in embryogenesis.

At the 8-cell stage, a process known as compaction begins, in which the spherical and loosely linked blastomeres take on a flattened polarised cell shape. The cohesiveness of cells within tissues can change dramatically during embryonic development. Tissue compaction is the process of increasing cell cohesiveness. The process of enhancing cell cohesion is known as tissue compaction. When cells come into closer contact with their neighbours, tissue compaction develops, which is linked to cell spreading (Turlier and Maître, 2015). During compaction, it will eventually cause cells to form the maternal component of the placenta, known as

the decidua, which is responsible for supplying nutrition and oxygen to the embryo through the blood vessels formed. When cells are compacted, they create an inner and outer layer. The inner layer will eventually give rise to the inner cell mass, whilst the outer layer will give rise to the polarised trophoblast. Pools of clear fluid accumulate between some of the internal cells by the time 30 blastomeres have been created, and these gaps soon consolidate into a single subcentral cavity. The resulting hollow cellular ball is a blastula of a form found in humans called a blastocyst, with a blastocoel as its cavity. The human blastocyst comprises the trophectoderm (TE), Inner Cell Mass (ICM), and blastocoel, all covered by zona pellucida.

2.1.7.2 Implantation and Early Placentation

The active blastocyst and the receptive uterus interact in a complicated way during the implantation process, which is a highly coordinated process (Jansen *et al.,* 2020; Kim, Su-Mi and Kim, 2017; Paria, Huet-Hudson and Dey, 1993; Zhang, Shuang *et al.,* 2013). Once the blastocyst has hatched from the zona pellucida, the embryo implants into the uterine wall one week after conception (Figure 1) (Ortega *et al.,* 2018). The blastocyst comprises an outside single-layered epithelial cover, known as the trophoblast, and a totipotent inner cell mass, known as the embryoblast, at this stage (Selwood and Johnson, 2006). The embryoblast is the direct precursor of the epithelial sections of the placental foetomaternal barrier. In contrast, the trophoblast is the direct precursor of the placental mesenchyme and the foetal vascular system.

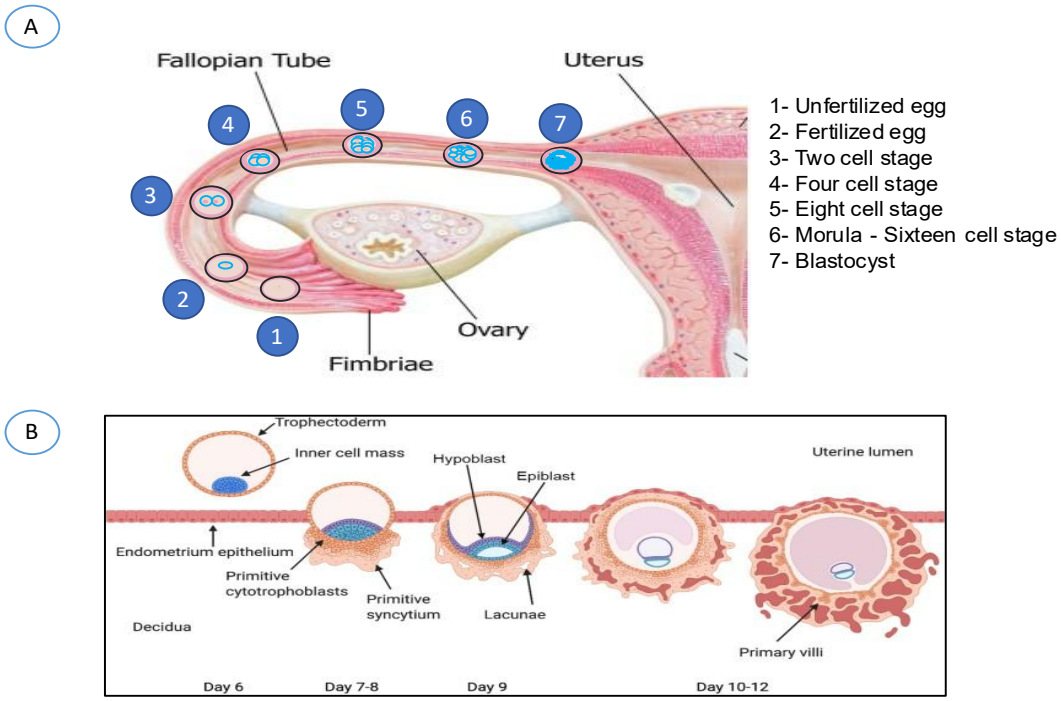


Figure 2: Schematic presentation of blastocyst implantation.

A shows the Blastocyst's development from the unfertilised egg, and B shows the implantation process into the endometrium epithelium.

The trophoblast cells give rise to the foetal chorionic villi (Arutyunyan *et al.,* 2023; Huppertz, B., 2006). The chorionic villi development begins around the end of the first week of pregnancy when the trophoblast cells differentiate into two distinct layers: the cytotrophoblast and syncytiotrophoblast (Haram *et al.,* 2020). The syncytiotrophoblast layer produces and secures Human chorionic gonadotropin (hCG) into the maternal bloodstream. The levels of hCG in the maternal bloodstream start to rise rapidly after implantation and continue to increase throughout the first trimester of pregnancy. The production of hCG by the syncytiotrophoblast is critical for the maintenance of pregnancy, as hCG helps to support the developing embryo and foetus by stimulating the production of oestrogen and progesterone by the ovaries. hCG is also used as a marker for

pregnancy testing and monitoring, as it can be detected in urine and blood samples.

The cytotrophoblast cells proliferate and extend into the syncytiotrophoblast, forming finger-like projections called primary villi. From the second week of pregnancy, mesodermal cells from the embryonic disc migrate into the chorionic villi, creating a network of blood vessels. These vessels are then covered by an inner layer of cytotrophoblast cells and an outer layer of syncytiotrophoblast cells, forming secondary villi. By the third week of pregnancy, the villi continue to grow and branch, forming tertiary villi. The tertiary villi are highly branched and contain numerous capillaries, which allow for the exchange of nutrients, oxygen, and waste products between the foetal and maternal bloodstreams (Huppertz, Berthold, 2023).

Apposition between the foetal chorionic villi and the maternal decidua basalis occurs during the early stages of human placenta development. Specifically, apposition occurs during the implantation process, which begins around 5–6 days after fertilisation and is complete by day 14. During implantation, the blastocyst, a hollow ball of cells that will eventually become the embryo, attaches to the uterine wall and burrows into the endometrial lining (Hamada and Saadeldin, 2022). As the blastocyst penetrates the endometrium, it comes into close contact with the decidua basalis, part of the endometrial lining that will ultimately form the maternal side of the placenta. Over the next several weeks, the chorionic villi will continue to grow and branch, creating a complex network of finger-like projections that extend into the maternal blood spaces. As the chorionic villi extend further into the uterine lining, they encounter the mother's blood vessels. Blood vessels from the embryo begin to grow into the chorionic villi, forming a network of blood vessels. The apposition between the chorionic villi and the decidua basalis allows for the exchange of nutrients, oxygen, and waste products between the foetal and

maternal circulations, essential for foetal growth and development. As the chorionic villi grow and develop, they form the placenta, a bridge between the mother and foetus (Kliman, 1999). The placenta acts as a filter, preventing harmful substances from reaching the foetus while allowing necessary nutrients and oxygen to pass through (Burton and Fowden, 2015). The placenta continues to mature throughout the pregnancy, with the chorionic villi multiplying and branching to increase the surface area available for exchange. As the pregnancy progresses, the placenta becomes more efficient at transporting nutrients, oxygen, and waste products between the mother and foetus.

2.1.7.3 Placenta Development per Gestational Week

The human placenta develops throughout pregnancy and undergoes several biological changes in structure and function per gestational week. Understanding the biological changes that occur during each week of placental development is essential for identifying biomarkers in PE.

In weeks 4–5, the trophoblast cells differentiate into two layers – the cytotrophoblast and syncytiotrophoblast. The cytotrophoblast cells proliferate and invade the maternal decidua, while the syncytiotrophoblast layer forms from the fusion of the outer cytotrophoblast cells (Horii, Boyd and Parast,2018). The villi also begin to form from the trophoblast and project into the decidua. In weeks 6–8, the villi continue to grow and branch, increasing the surface area for exchanging nutrients and waste products between the maternal and foetal bloodstreams. The placenta also begins to produce hormones such as hCG and progesterone, which help maintain the pregnancy. The syncytiotrophoblast layer further differentiates into villous and extravillous syncytiotrophoblast, which have different functions in placental development.

In weeks 9–12, the placenta becomes more organised, with distinct layers of cells that specialise in different functions. The villi are covered by a layer of syncytiotrophoblast cells, which fuse to form a syncytium, facilitating nutrient and gas exchange. The cytotrophoblast cells form columns that anchor the villi to the decidua. The foetal blood vessels within the villi are surrounded by a layer of cytotrophoblast cells, which help regulate blood flow. Placental cells called Hofbauer cells, which are immune cells, also appear (Ingman *et al.,* 2010).

In weeks 13–16, the placenta continues to grow and develop, accompanied by increased villi and blood vessels. During this period, the syncytiotrophoblast layer starts to thin while the cytotrophoblast layer maintains its proliferation. Furthermore, the placenta generates additional hormones, such as oestrogen and relaxin, contributing to maternal body preparation for childbirth.

Advancing to weeks 17–20, the placenta attains its maximum thickness and weight. The villi and blood vessels persist in both number and size enlargement. The syncytiotrophoblast layer experiences a further reduction in thickness while the cytotrophoblast layer's proliferation continues. Notably, the placenta transfers maternal antibodies into the foetal bloodstream, imparting passive immunity to the developing foetus.

Transitioning to weeks 21–30, the growth rate of villi and blood vessels decelerates compared to preceding weeks. Concurrently, the syncytiotrophoblast layer undergoes additional thinning, and the cytotrophoblast layer continues its proliferation. Enhanced nutrient and gas exchange efficiency characterises the placenta at this stage, including the ability to transfer larger molecules, such as proteins, to the foetus.

Progressing into weeks 31–40, the placenta begins ageing and degeneration. This phase involves a gradual reduction in both the quantity and functionality of villi and blood vessels (Salafia and Popek, 2008; Kojima *et al.*, 2022). The syncytiotrophoblast layer diminishes in thickness, increasing prominence in the cytotrophoblast layer. Despite these changes, the placenta maintains its essential role in offering support and nourishment to the foetus until the moment of birth.

2.1.7.4 Process of Abnormal Placentation

Abnormal placentation is a key factor in the development of PE, a serious complication of pregnancy that can lead to maternal and foetal morbidity and mortality (James *et al.,* 2022). Several types of abnormal placentation can contribute to the pathophysiology of PE, including impaired trophoblast invasion, inadequate spiral artery remodelling, and placental ischaemia (Staff *et al.,* 2022).

The unique and highly invasive nature of human placentation is thought to be an underlying reason for the existence of PE. Humans possess a haemochorial placenta, where foetal trophoblast cells directly invade the maternal decidua and remodel the spiral arteries, establishing direct contact with maternal blood (Phipps *et al.*, 2019). This highly invasive strategy is shared by a small number of species, including some higher apes. In contrast, most other mammalian species possess less invasive placentas (e.g., epitheliochorial or endotheliochorial), where there are more layers of maternal tissue separating the foetal and maternal circulations. This evolutionary trade-off for highly efficient nutrient transfer in humans creates a delicate immunological and vascular balance (Barbeito and Miglino, 2023). The failure of this deep trophoblast invasion and subsequent remodelling—the core pathology of PE-is therefore a direct consequence of the unique physiological demands placed on the maternal system by the invasive human placenta. This comparative perspective underscores why PE is predominantly a human (and

certain primate) disease, providing a deeper mechanistic justification for studying the failure of early placentation.

Trophoblast invasion is the process by which placental cells, known as trophoblasts, invade the uterine wall and remodel the maternal blood vessels to establish a functional placenta. Inadequate trophoblast invasion is a hallmark of PE, and it has been linked to abnormalities in the expression and function of several molecules involved in trophoblast migration and invasion. For example, studies (Lim *et al.,* 1997; Gathiram and Moodley, 2016) have shown that pre-eclamptic placentas exhibit decreased expression of integrins, which are cell surface proteins that mediate trophoblast adhesion and migration. Other studies (Dimitriadis *et al.,* 2023; Hong, Park and H Cha, 2021; Varughese *et al.,* 2010; Verlohren, Stepan and Dechend, 2012) have shown that Pre-eclamptic placentas have increased expression of soluble fms-like tyrosine kinase-1 (sFlt-1), a protein that binds to and neutralises vascular endothelial growth factor (VEGF) and placental growth factor (PlGF), two key molecules involved in trophoblast invasion and angiogenesis.

Inadequate spiral artery remodelling is another type of abnormal placentation that can contribute to PE. During a typical pregnancy, the spiral arteries in the maternal uterine wall undergo extensive remodelling to accommodate the increased blood flow required by the developing foetus (Osol and Mandala, 2009). This remodelling process involves the infiltration of trophoblasts into the spiral arteries, resulting in their dilation and reduced resistance to blood flow. However, in cases of PE, the spiral arteries do not remodel properly, leading to diminished blood flow to the placenta. Consequently, placental factors are released, contributing to the maternal inflammatory response. This inflammatory response can cause endothelial dysfunction, subsequently triggering hypertension, proteinuria, and other characteristic features of PE.

Finally, placental ischemia, or reduced blood flow to the placenta, is a third type of abnormal placentation that can contribute to the pathophysiology of PE (Chaiworapongsa *et al.,* 2014). Placental ischemia can result from impaired trophoblast invasion or inadequate spiral artery remodelling, and it can cause the placenta to release factors that contribute to maternal endothelial dysfunction and inflammation. Overall, abnormal placentation in PE is a complex and multifactorial process that involves dysfunction of the placenta, maternal immune system, and maternal vascular system (McMaster, Zhou and Fisher, 2004; Fisher, 2015). A better understanding of the underlying mechanisms of abnormal placentation may lead to improved strategies for preventing and managing PE.

2.1.8 Biomarkers Associated with Pre-eclampsia.

Building on the established clinical definitions and pathogenic mechanisms, research has focused on identifying biomarkers for early PE detection and prognosis (Ng et al., 2024). As the classic symptoms of hypertension and proteinuria can be late manifestations, molecular signatures offer a promising avenue for earlier risk assessment (MacDonald *et al.*, 2022). Scientists have delved into identifying such signatures, which predominantly originate from the key pathogenic sources discussed earlier: the placenta, endothelium, and immune system, as well as from indicators of oxidative stress, renal dysfunction, and various circulating factors (Kaitu'u-Lino *et al.*, 2025).

2.1.8.1 Placenta–Derived Biomarkers

The placenta is at the forefront of PE's aetiology, an organ quintessential for foetal nourishment and growth. Aberrations in placental development and function give rise to distinct biomarkers (Roberts and Hubel, 2009). For instance, the anti-angiogenic factor sFlt-1 (Soluble fms like tyrosine kinase-1) is elevated in PE. By counteracting the effects of vascular endothelial growth factor (VEGF) and

placental growth factor (PlGF), sFlt-1 is implicated in the endothelial dysfunction characteristic of the disorder (Maynard *et al.,* 2003). Additionally, reduced levels of PlGF, vital for placental vascular development, have been noted in women with PE (Levine *et al.,* 2004).

Another noteworthy biomarker is soluble endoglin (sEng), a coreceptor for transforming growth factor-β (TGF-β). sEng is an anti-angiogenic molecule whose levels are found to be elevated in women with PE. Elevated sFlt-1 levels can enhance vascular permeability and contribute to the endothelial dysfunction characteristic of PE (Venkatesha *et al.,* 2006; Ives *et al.,* 2020). Pregnancy-associated plasma protein-A (PAPP-A) is yet another molecule studied concerning PE. PAPP-A is involved in the cleavage and, thus, the activation of insulin–like growth factor (IGF). Reduced serum levels of PAPP-A during the first trimester have been associated with an increased risk of PE, especially in its early–onset form (Smith, G. C. S. *et al.,* 2002; Jayamol, Dey and Sahoo, 2019; Baron and Baron, 2023). Although these biomarkers indicate the underlying pathophysiological processes in PE, they have limitations. The specificity and sensitivity of these markers vary, and no single marker can predict PE with absolute certainty. Moreover, while the aforementioned markers might indicate an increased risk of developing PE, they don not precisely predict its severity or the timeframe of its onset. Understanding placental biomarkers is crucial to creating more efficient screening tools and potential therapeutic strategies for PE. By detecting the disease early on, managing the condition and reducing its associated risks might be possible. Nevertheless, as understanding of these markers deepens, it is important to remain cognisant of the challenges in their clinical application and to seek more comprehensive, multi-marker models that improve prediction accuracy.

2.1.8.2 Endothelial-Derived Biomarkers

The endothelium, a single layer of cells lining blood vessels, is pivotal in maintaining vascular homeostasis. It orchestrates vital processes such as vascular tone regulation, cell adhesion, inflammation, and thrombosis (Davignon and Ganz, 2004; Krüger-Genge *et al.,* 2019). Underlying the clinical manifestations of PE is endothelial dysfunction, a state in which the endothelium loses its homeostatic properties, leading to vasoconstriction, inflammation, and hypercoagulability (Redman, Christopher W. and Sargent, 2005). Given the pivotal role of the endothelium in PE's pathogenesis, endothelial-derived biomarkers have garnered significant attention for their potential in early detection, diagnosis, and prognosis of the disorder. Notably, sFlt-1, an anti-angiogenic factor, inhibits vascular endothelial growth factor (VEGF) and placental growth factor (PlGF). Its elevation in maternal circulation is strongly associated with PE and may even predict its onset before clinical manifestations appear (Maynard *et al.,* 2003; Dimitriadis *et al.,* 2023). Similarly, sEng hinders the binding of transforming growth factor–beta (TGF-β) to its receptor, and elevated levels, especially when paired with sFlt-1, can hint at severe PE forms, emphasising its role in intensifying endothelial dysfunction (Venkatesha *et al.,* 2006). Furthermore, increased concentrations of the Von Willebrand Factor (vWF) glycoprotein in preeclamptic women underscore endothelial damage, with its larger, more prothrombotic forms being particularly notable (Bergmann *et al.,* 1991; Schapkaitz *et al.,* 2021).

Endothelin-1 (ET-1), a potent vasoconstrictor, has its escalation linked with hypertension, a hallmark of PE. Its heightened levels in affected patients point to its pivotal role in the ensuing endothelial dysfunction and elevated blood pressure (Taylor *et al.,* 2003; George and Granger, 2011; LaMarca, 2012). On the other end of the spectrum, reduced PlGF levels in PE, coupled with increased sFlt-1, tilt the angiogenic balance, fuelling endothelial dysfunction and disease manifestation (Chaiworapongsa *et al.,* 2014). Beyond their academic intrigue, these biomarkers have profound clinical ramifications. Their early detection can shape management

strategies, staving off grave complications and illuminating PE's molecular underpinnings, which might unlock novel therapeutic avenues.

2.1.8.3 Oxidative Stress Biomarkers

Oxidative stress, characterised by an imbalance between the production of free radicals, specifically reactive oxygen species (ROS), and the body's capacity to detoxify these radicals or repair the damage they cause, is increasingly recognised for its role in PE (Agarwal *et al.,* 2012). Malondialdehyde (MDA), a product of lipid peroxidation, stands out as a key biomarker, with research indicating notably higher levels in preeclamptic women than in normotensive pregnant counterparts (Kumar and Jain, 2024). This elevation suggests significant ROS-induced cellular damage, particularly to cell membranes. Additionally, nitric oxide (NO), a gaseous signalling molecule essential for vasodilation, experiences disrupted production or bioavailability, leading to hypertension. Elevated levels of asymmetrical dimethylarginine (ADMA), which inhibits nitric oxide synthase (NOS), have been linked with PE (Petkova-Parlapanska *et al.*, 2025). This elevation potentially results in decreased NO production and subsequent vasodilation impairment. Intriguingly, circulating antibodies in preeclamptic patients have also been tied to NOS dysfunction and diminished NO availability (Harmon *et al.,* 2016; Motta-Mejia *et al.,* 2017). Delving into the realm of placental antioxidant enzymes, it's evident that superoxide dismutase (SOD), catalase (CAT), and glutathione peroxidase (GPx) play critical roles in neutralising ROS effects. Preeclamptic pregnancies, however, have exhibited decreased activities of placental SOD and CAT, suggesting an impaired defence against oxidative stress in the placental environment (Myatt and Webster, 2009; Marín *et al.,* 2020). Further markers spotlighting oxidative stress in PE include Advanced Oxidation Protein Products (AOPP) and 8-isoprostane. Preeclamptic women displayed heightened levels of AOPP, indicating oxidative protein modification, and 8-isoprostane, another lipid peroxidation product, underscored the escalated oxidative stress state (Walsh, S. W. *et al.,* 2000; Drejza

et al., 2022). To encapsulate, the multi–factorial pathogenesis of PE undeniably intertwines with oxidative stress. A profound understanding of these biomarkers can illuminate PE's pathophysiology and pave the way for innovative diagnostic methods and treatment strategies.

Oxidative stress biomarkers such as malondialdehyde (MDA), asymmetrical dimethylarginine (ADMA), and advanced oxidation protein products (AOPP) have been widely studied in preeclampsia. Importantly, the timing of their detection during pregnancy varies across studies, influencing their interpretability and potential clinical utility. ADMA has been investigated for early prediction of preeclampsia, with several studies measuring elevated levels as early as the first or early second trimester, suggesting it may serve as an early biomarker for risk stratification (e.g., Petkova-Parlapanska et al., 2025). In contrast, foundational research on markers like MDA, AOPP, and 8-isoprostane primarily reports elevated levels closer to clinical diagnosis, typically in mid-to-late pregnancy, or in placental tissues obtained post-delivery, reflecting established disease and pathological oxidative damage (Ibrahim et al., 2024; Arogbokun et al., 2021). This temporal dynamics underscores the importance of selecting appropriate biomarker windows in both research and clinical settings to optimise prediction, diagnosis, and understanding of preeclampsia pathophysiology.

Oxidative stress, characterised by an imbalance between the production of free radicals (ROS) and the body's capacity to detoxify them, is increasingly recognised for its role in PE (Assani *et al.*, 2025). As this thesis relates to different gestational time points, it is important to note that the timing of biomarker detection varies. Some markers are investigated for early prediction, while others reflect established disease.

Malondialdehyde (MDA), a product of lipid peroxidation, stands out as a key biomarker, with research indicating notably higher levels in preeclamptic women (Junaid, 2025). Additionally, nitric oxide (NO), a gaseous signalling molecule

essential for vasodilation, experiences disrupted production (Socha *et al.*, 2022). Elevated levels of asymmetrical dimethylarginine (ADMA), which inhibits nitric oxide synthase (NOS), have been linked with PE. Notably, ADMA has been investigated for early prediction, with some studies measuring elevated levels as early as the first or early second trimester, suggesting it may serve as a risk stratification biomarker. Intriguingly, circulating autoantibodies, particularly those targeting the angiotensin II type 1 receptor (AT1-AA), in preeclamptic patients have also been tied to NOS dysfunction.

Delving into the realm of placental antioxidant enzymes, it's evident that superoxide dismutase (SOD), catalase (CAT), and glutathione peroxidase (GPx) play critical roles in neutralising ROS effects, with their activity often decreased in preeclamptic placental tissue. Further markers spotlighting oxidative stress in PE include Advanced Oxidation Protein Products (AOPP) and 8-isoprostane. In contrast to early prediction markers like ADMA, foundational research on markers like MDA, AOPP, and 8-isoprostane primarily reports elevated levels closer to the time of clinical diagnosis (mid-to-late pregnancy) or in placental tissues obtained post-delivery, reflecting established pathological oxidative damage. To encapsulate, the multi–factorial pathogenesis of PE undeniably intertwines with oxidative stress. A profound understanding of these biomarkers across different gestational windows can illuminate PE's pathophysiology and pave the way for innovative diagnostic methods.

2.1.8.4 Markers of Organ Dysfunction

PE is a complex multisystem disorder characterised by vascular dysfunction, leading to maternal organ dysfunction. Understanding the various markers that indicate organ dysfunction is pivotal for the timely identification and management of women at risk. Notably, many markers of organ dysfunction associated with PE provide insights into the underlying pathophysiology of this condition and can offer guidance for clinical decisions. One of the primary organs affected by PE is the

kidney (Lopes van Balen *et al.,* 2017; Kristensen *et al.,* 2019). Proteinuria, or an increased amount of protein in the urine, has historically been a defining feature of the disorder (von Dadelszen *et al.,* 2011). While a 24–hour urine collection was the gold standard, recent advancements in diagnostics utilise the protein–to–creatinine ratio in random urine samples, allowing for a faster diagnosis. (Podymow and August, 2010). Beyond proteinuria, other renal markers, like elevated serum creatinine levels, signify a reduction in the glomerular filtration rate, which indicates kidney dysfunction.

PE has a multisystem impact and can notably affect the liver (Duley, 2009; Alese, Moodley and Naicker, 2021). The liver plays an essential role in several physiological processes, including the synthesis of vital proteins, detoxification, and metabolism. Therefore, any injury to the liver can have a systemic impact. In PE, liver involvement is primarily due to endothelial dysfunction, leading to liver ischemia and, consequently, liver cell necrosis (Hammoud and Ibdah, 2014a; Aouache *et al.,* 2018). This cell death results in the release of liver enzymes into the bloodstream, which can be detected through laboratory tests. Among the liver enzymes, aspartate aminotransferase (AST) and alanine aminotransferase (ALT) are the most checked when assessing liver function (Dacaj *et al.,* 2016; Lala, Zubair and Minter,2023). Elevated levels of AST and ALT in the bloodstream often indicate liver cell damage. In patients with PE, significantly increased levels of these enzymes are a cause for concern and can point to a severe form of the condition known as HELLP syndrome (Rao *et al.,* 2016; Hammoud and Ibdah, 2014). HELLP stands for Haemolysis, Elevated Liver enzymes, and Low Platelet count. It's a life–threatening pregnancy complication that typically occurs in the later stages of pregnancy or soon after childbirth. Importantly, hepatic complications can manifest clinically in several ways. Some women might report right upper quadrant or epigastric pain, while others could have nausea and vomiting, both of which can be symptoms of liver involvement in PE (Barton and Sibai, 2008). When these clinical signs are coupled with elevated liver enzyme

levels, it underscores the importance of prompt diagnosis and management to prevent further complications (Wagner, 2004). Regular monitoring of liver enzyme levels in pregnant women diagnosed with or suspected of having PE is vital. Swift therapeutic interventions can mitigate the progression of liver injury and related complications. Given that PE is a leading cause of maternal morbidity and mortality worldwide, understanding its systemic impacts, including on the liver, is crucial for safeguarding maternal health.

The cardiovascular system is equally implicated. Elevated blood pressure is a primary hallmark of PE. Besides, abnormalities in echocardiographic findings, such as left ventricular hypertrophy and diastolic dysfunction, provide evidence of cardiac strain and potential long–term cardiovascular complications (Melchiorre *et al.,* 2011). On a vascular level, studies have indicated that an increased concentration of soluble fms-like tyrosine kinase 1 (sFlt–1) and reduced levels of placental growth factor (PlGF) in the blood, can serve as markers for endothelial dysfunction, a characteristic feature of PE (Maynard et al., 2003; Kim, Shin-Young et al., 2007; Gurnadi et al., 2015). Moreover, neurologic symptoms such as persistent headaches, visual disturbances, and hyperreflexia could suggest cerebral oedema or heightened cerebral perfusion pressure, both of which underscore the neurological involvement in PE (Wagner, 2004).

This is a critical point because the multisystem nature of PE results in heterogeneous clinical presentation. Although the extent of involvement of individual organs, such as the kidney or liver, can vary between patients, the markers themselves are consistently recognised as defining features of the syndrome. These markers include new-onset hypertension, proteinuria, elevated liver enzymes (AST and ALT), and neurological symptoms (Ng *et al.*, 2024; Tomkiewicz and Darmochwał-Kolarz, 2024). Their consistency arises from a shared underlying pathophysiology of systemic endothelial dysfunction, characterised by anti-angiogenic imbalance and inflammatory activation, which is

a primary driver that subsequently manifests across multiple organ systems (LaMarca, 2012; McElwain *et al.*, 2020). Therefore, although the pattern of organ dysfunction varies, the emergence of these markers remains a consistent hallmark of disease progression.

The multifaceted nature of PE affects various organ systems, each with its set of markers indicative of dysfunction. Recognising and monitoring these markers is instrumental in diagnosing, managing, and preventing potential complications arising from PE.

2.1.8.5 Immune Response Biomarkers

Research efforts have been directed towards identifying reliable biomarkers that can aid in early PE prediction, diagnosis, and monitoring, thus improving maternal and foetal outcomes. Several immune response biomarkers have emerged as potential candidates for understanding and managing PE. One of the notable biomarkers is soluble fms-like tyrosine kinase-1 (sFlt-1), which is a soluble receptor for vascular endothelial growth factor (VEGF) and placental growth factor (PlGF). Increased levels of sFlt-1 have been observed in preeclamptic women, contributing to the disruption of angiogenesis and endothelial dysfunction, hallmark features of the disorder (Maynard *et al.,* 2003; Kim, Shin-Young *et al.,* 2007; Powe, Levine and Karumanchi, 2011; Palmer *et al.,* 2015). Another biomarker, placental growth factor (PlGF), has garnered attention due to its role in angiogenesis and vascular homeostasis. Reduced levels of PlGF, often accompanied by elevated sFlt-1 levels, have been associated with PE and may serve as a predictive marker (Levine *et al.,* 2004; Herraiz *et al.,* 2015). The immune system's involvement in PE has led to the investigation of various immune-related biomarkers. Tumour necrosis factor-alpha (TNF-α), a pro-inflammatory cytokine, has been linked to the pathogenesis of PE by promoting endothelial dysfunction and impairing trophoblast invasion (Goswami *et al.,* 2006; Gathiram and Moodley, 2016; Fakhr *et al.,* 2022). Interleukin-6 (IL-6), another pro-inflammatory cytokine, is elevated in

preeclamptic women, potentially contributing to the systemic inflammatory response and endothelial dysfunction seen in the disorder (Shahshahan, Hashemi and Rasouli, 2014; Amirian, Mahani and Abdi, 2020; Hezareh, Moghaddam-Banaem and Shahali, 2021). Emerging evidence suggests a role for immune regulatory molecules like regulatory T cells (Tregs) and their associated cytokines in PE. Reduced numbers and dysfunction of Tregs have been reported in preeclamptic pregnancies, possibly contributing to immune maladaptation and excessive inflammation (Sasaki *et al.,* 2007). Furthermore, the imbalance between pro-inflammatory cytokines like IL-17 and anti-inflammatory cytokines like IL-10 has been implicated in PE, underscoring the intricate immune dysregulation in the disorder (Aggarwal *et al.,* 2019; Eghbal-Fard *et al.,* 2019; Nath *et al.,* 2020).

In recent years, interest has grown in the potential of microRNAs (miRNAs) as diagnostic and prognostic biomarkers for PE. MiRNAs are small, non-coding RNAs that regulate gene expression and have been detected in various biological fluids. MiR-210, for instance, is elevated in preeclamptic women and is associated with hypoxia and endothelial dysfunction (Ura *et al.,* 2014). PE remains a significant challenge in maternal–foetal medicine, necessitating the identification of reliable biomarkers for timely diagnosis and effective management. Immune response biomarkers, including sFlt-1, PlGF, TNF-α, IL-6, Tregs, and miRNAs, have shown promise in shedding light on the complex pathophysiology of PE. miRNAs are key post-transcriptional regulators of gene expression, and their roles extend fundamentally into the immune system (Artimovič *et al.*, 2024). They fine-tune the differentiation, proliferation, and function of almost all immune cells, including T cells, B cells, macrophages, and natural killer (NK) cells. By targeting the mRNA of key immune genes, miRNAs help maintain immune homeostasis and are critical in initiating, dampening, and resolving inflammatory responses. Dysregulation of specific miRNAs is strongly implicated in autoimmune disorders and pathological inflammatory states, making their profile a vital marker in conditions like PE, where immune maladaptation is central to the pathogenesis.

2.1.8.6 Biomarkers in Cell-free Nucleic Acids

Maternal and foetal health dynamics during pregnancy are intricate and influenced by many genetic and environmental factors. Recent advancements in molecular diagnostics have unveiled the potential of circulating nucleic acids, particularly cell-free RNA (cfRNA) and DNA (cfDNA), as valuable biomarkers for predicting and understanding complex pregnancy–associated conditions, prominently PE (Wong, B. C. and Lo, 2003; Hahn *et al.,* 2011; Rahat *et al.,* 2020; Palei, 2021). Their origin, dynamics in maternal circulation, and reflection of underlying physiological and pathological processes underscore this potential.

The clinical application of cfDNA and cfRNA extends beyond PE to routine prenatal screening, particularly for foetal aneuploidies (e.g., Trisomy 21) via Non-Invasive Prenatal Testing (NIPT) (Moufarrej *et al.*, 2023). Furthermore, altered cfRNA profiles have been successfully linked to gestational age prediction and risk of preterm delivery, reflecting dynamic changes in the maternal-foetal unit (Gray, Hemberg and Karumanchi, 2022). The utility of cfNA is also established in conditions that share a common pathogenic basis with PE, such as Intrauterine Growth Restriction (IUGR) and Gestational Diabetes Mellitus (GDM) (Morano *et al.*, 2018). For instance, GDM, a major risk factor for PE, is associated with specific cfDNA methylation and transcriptomic signatures, suggesting overlapping molecular pathways (e.g., vascular stress and inflammation) that compromise placental function across multiple adverse pregnancy outcomes (Zhou *et al.*, 2025).

In the context of PE, the potential benefits of employing cfNA are multifaceted. Firstly, the non–invasive nature of obtaining cfNA from maternal blood eliminates the risks associated with invasive procedures such as amniocentesis or chorionic villus sampling, thus ensuring safety for both the mother and the foetus (Takabayashi *et al.,* 1995; Lo, Y. M. *et al.,* 1997; Wright, C. F. and Burton, 2009; Romão *et al.,* 2012). Early identification of PE, potentially even before clinical

symptoms, allows for timely interventions, close monitoring, and tailored care plans, optimising maternal and foetal outcomes (Chaiworapongsa *et al.,* 2014). Elevated levels of cfDNA in maternal plasma have been associated with the development of PE, reflecting the heightened placental stress or increased apoptosis of trophoblasts in affected pregnancies (Levine *et al.,* 2004). Moreover, differential methylation patterns of cfDNA might offer insights into placental epigenetic changes pivotal to PE's pathogenesis (Hogg *et al.,* 2013).

Recent advancements in sequencing and bioinformatics now enable the analysis of both cfDNA and cfRNA methylation patterns, which are critical in diagnosing complex pregnancy disorders(Meng *et al.*, 2025; Vliet *et al.*, 2025). Contemporary studies leverage cfDNA methylation signatures to predict pregnancies at risk of adverse outcomes, including PE, demonstrating a non-invasive tool for tracking placental epigenetic status throughout gestation (Chotaliya, Ladumor and Dhorajiya, 2025). The integration of these epigenomic profiles with transcriptomic data using advanced AI models is now considered a key future direction for holistic molecular profiling in precision maternal health.

Additionally, certain placental-derived microRNAs, a subset of cfRNA, whose expression profiles are altered in preeclamptic pregnancies, may play roles in the disease's pathophysiology, influencing processes like angiogenesis and immune modulation (Ura *et al.,* 2014). The benefits of cell-free Nucleic Acids (cfNA) as biomarkers in early pregnancy extend beyond mere early detection, providing profound insights into the molecular intricacies underlying PE, and hold promise in refining prenatal care strategies. Therefore, this research study employed cfRNA sequencing data obtained during early pregnancy to identify a diverse range of genomic biomarkers. These included differentially uniquely expressed gene signatures, consistent gene signatures, and variant signatures. The overarching objective was to leverage these genomic biomarkers in combination with artificial intelligence to develop a robust disease prediction model.

2.1.8.6.1 Production and Origin of cfRNA

The discovery of circulating cfNAs in maternal plasma has revolutionised prenatal diagnostics, offering a non–invasive approach to assess the genetic and physiological status of the foetus. These cfNAs, which include both cfDNA and cfRNA, serve as valuable biomarkers for various diagnostic applications. The primary source of cfNAs in pregnant women is the placenta, particularly the apoptotic and necrotic trophoblasts. As the placenta develops, trophoblasts in direct contact with maternal blood undergo continuous cellular turnover, releasing cellular fragments, including nucleic acids, into the maternal circulation (Lo, YM Dennis *et al.,* 1997; Manokhina, Wilson and Robinson, 2015; Knöfler *et al.,* 2019). This transfer of nucleic acids into the maternal bloodstream is a form of microchimerism, which explains the placental origin of a substantial portion of cfDNA. While the placenta is the dominant source, other foetal cells, such as those shed from developing foetal organs, also contribute to the pool of circulating nucleic acids (Rijnders *et al.,* 2001; Roberts and Escudero, 2012; Taglauer, Wilkins-Haug and Bianchi, 2014). Furthermore, as the placenta grows and the pregnancy progresses, the absolute quantity of cfNA in maternal circulation tends to increase, peaking towards the end of the pregnancy (Fan *et al.,* 2010; Chiu, Rossa W. K. and Lo, 2021).

2.1.8.6.1.1 Foetal microchimerism

Cell-free nucleic acids (cfNAs) of foetal origin circulate in maternal blood and have been proposed as potential biomarkers for various pregnancy-related complications, including PE (Chiu, Rossa WK *et al.,* 2011). Elevated concentrations of cfNA have been consistently observed in women with PE compared to normotensive pregnant women. It is postulated that placental stress or damage associated with PE leads to increased shedding of cfNA into the maternal circulation (Tsui *et al.,* 2011). Microchimerism refers to the presence of a small number of cells or genetic material from one individual within the tissues

or circulation of another genetically distinct individual. During pregnancy, foetal cells can persist in maternal circulation and tissues for years post-delivery. These persisting foetal cells have been implicated in the development of maternal autoimmune conditions. In the context of PE, studies have explored the possibility that an imbalance in maternal-foetal microchimerism could contribute to its pathogenesis (Gammill *et al.,* 2010). Both cfNA and microchimerism highlight the intricate interplay between maternal and foetal systems. While cfNA reflects placental health and function, microchimerism points to longer-term cellular interactions that could influence maternal health. Understanding the dynamics of cfNA and microchimerism offers promising insights into the complex aetiology of PE and may improve its early prediction. Continued research in these areas is essential for establishing their clinical relevance.

2.1.8.6.1.2 Characteristics of cfNA

The cfNA has emerged as a ground–breaking entity in the field of prenatal medicine and diagnostics (Xu *et al.*, 2022). Found in the maternal bloodstream, these fragments of DNA and RNA are derived from the foetus, providing a non–invasive window into foetal health and development. The ability to isolate and analyse these nucleic acids has offered novel avenues for detecting genetic and chromosomal abnormalities, assessing foetal health, and monitoring pregnancy progress (Wynn *et al.*, 2025). As the potential of cfNA becomes more pronounced, understanding their distinct characteristics becomes vital (Meng *et al.*, 2025). These attributes not only shed light on the biological nature of cfNA but also influence their clinical applications, diagnostic accuracy, and the overall interpretation of test results.

2.1.8.6.1.2.1 Distinct Size Profile

The unique characteristics of cffNA have reshaped non-invasive prenatal testing methodologies and have broad implications for future clinical applications. A

distinguishing hallmark of these foetal-derived molecules is their distinct size profile, setting them apart from maternal nucleic acids in circulation. cfDNA displays a characteristic pattern where fragments predominantly measure 160–180 base pairs in length. This size suggests DNA wrapped around a nucleosome, indicating a connection to the specific chromatin structures and nucleosome packaging found within the placental origin from which most cfDNA is derived (Lo et al., 2010). This size specificity is a mere biological observation and holds significant practical implications. In Non-Invasive Prenatal Testing (NIPT), this unique size profile can be exploited to enrich foetal DNA from maternal plasma. By selectively targeting these shorter fragments, increasing the relative proportion of cfDNA is possible, enhancing the sensitivity and accuracy of downstream genetic analyses (Jiang *et al.,* 2015). On the other hand, cfRNA, while less extensively studied than cfDNA, has also displayed distinct size patterns. cfRNA often appears as shorter transcripts, a phenomenon believed to be attributed to the inherent instability of RNA molecules, as well as the selective release or preservation of specific RNA species from the placental cells (Aucamp *et al.,* 2018; Moufarrej *et al.,* 2020). These shorter RNA fragments represent specific transcripts or degradation products that can provide insights into foetal gene expression patterns, physiology, or placental health. Furthermore, the size profiles of cfDNA and cfRNA can be crucial quality controls in diagnostic assays. By ensuring that the isolated fragments align with the expected sizes for foetal nucleic acids, researchers and clinicians can better ascertain the authenticity of the cfNA and differentiate it from potential contaminants or maternal nucleic acid artefacts (Chiu, Rossa WK *et al.,* 2011). The unique size profiles of cfDNA and cfRNA stand out as defining characteristics that not only shed light on their biological origins but also present valuable opportunities for optimising non-invasive prenatal diagnostics. As research unravels the intricacies of these foetal–derived molecules, their size profiles will undoubtedly remain central to their utility and application in clinical settings.

2.1.8.6.1.2.2 Source and Origin.

Another unique characteristic of cfNA is its distinct source and origin, which has heralded a transformative era in prenatal diagnostics. Central to understanding cfNA's diagnostic potential is elucidating their source and origin. Foremost, the placenta, specifically the trophoblast cells forming the blastocyst's outer layer, emerges as the primary contributor. As these cells experience apoptosis, a process of programmed cell death, they liberate nucleic acid fragments into the maternal bloodstream, with syncytiotrophoblast being the major contributors (Lo, Y. M. *et al.,* 1997; Elmore, 2007; Sharp *et al.,* 2010). Additionally, beyond the placenta, foetal tissues and organs undergoing development might also release nucleic acids, albeit their contribution is comparatively minor (Bischoff *et al.,* 2002; Bianchi, 2004). Notably, the characteristics of cfNA distinguish them from maternal nucleic acids. For instance, foetal DNA fragments tend to be shorter, reflecting differences in chromatin structures and nucleosome packaging (Chan *et al.,* 2004). The dynamics of cfNA become even more intricate in twin or multiple pregnancies. In such scenarios, each foetus and its corresponding placenta contribute to the cfNA pool, and differential growth rates or placental mass variances can influence the cfNA proportions in maternal circulation (Suzumori *et al.,* 2018). Furthermore, pathological conditions like placental insufficiency might amplify the release of cfNA, shedding light on compromised placental barriers or heightened cellular turnover in response to stressors (Lun *et al.,* 2008). The multifaceted origins and sources of cfNA, underscored by the placenta's pivotal role, have profound implications for non-invasive prenatal diagnostics, providing an invaluable window into foetal health and development.

2.1.8.6.1.2.3 Dynamics in Maternal Circulation

The dynamic behaviour of cfNA in maternal circulation provides crucial insight into prenatal diagnostics and maternal-foetal interactions. Predominantly originating from the placenta, these foetal fragments find their way into the maternal

bloodstream as the trophoblasts, which form the vital link between the foetus and mother, undergo regular cellular turnover processes such as apoptosis and necrosis (Lo, Y. M. *et al.,* 1997; Sharp *et al.,* 2010). Interestingly, as the pregnancy progresses, there's a notable rise in cfNA concentration in maternal plasma. This increase is closely tied to the enlarging placenta, heightened cellular activity, and the maturity of the foetus (Li, Y. *et al.,* 2004). Post–childbirth, the maternal circulation exhibits a rapid clearance of cfNA, with studies showing an impressively short half-life for cfDNA of around 16 minutes, ensuring its almost complete disappearance hours after delivery (Lo, YM Dennis *et al.,* 1999). This swift clearance has critical clinical implications, especially when monitoring post–partum complications or placental remnants. Various factors can influence cfNA levels, including maternal weight, gestational age, and even sampling times, emphasising the need for standardisation in sample collection and interpretation (Cogswell and Yip, 1995; Wataganara *et al.,* 2003; Deng and Liu, 2022). Moreover, any sudden changes in cfNA concentrations can serve as red flags, with drops potentially indicating foetal complications and spikes suggesting conditions like placental lesions or multiple pregnancies (Chen, S. *et al.,* 2017). The intricate dynamics of cfNA within maternal circulation are not mere biological phenomena but are closely entwined with the broader tapestry of pregnancy, offering invaluable insights for refining non-invasive prenatal diagnostics.

To provide a clearer quantitative and temporal overview, the following table synthesizes the key biomarkers discussed in the preceding sections. This summary consolidates information regarding their typical detection levels, gestational timing, and the sample source from which they are measured. This table serves as a comparative baseline from the existing literature, setting the stage for evaluating the novel, gestational-week-specific cfRNA signatures identified in the results chapters of this thesis.

Table 3 Summary of Key Biomarkers from Literature Review.
The table categorises established biomarkers for PE discussed in the literature, detailing their biological source, typical gestational timing of detection, and observed levels in PE compared to NP. This serves as a baseline for comparison with novel biomarkers identified in this study.

Biomarker Category	Biomarker	Sample Source	Timing (Gestational Age)	Level in PE (vs. NP)
Placental / Angiogenic	sFlt-1 (Soluble fms-like tyrosine kinase-1)	Maternal Circulation / Plasma	Predictive before clinical onset	Elevated
Placental / Angiogenic	PlGF (Placental Growth Factor)	Maternal Circulation / Plasma / Serum	First Trimester / Predictive	Reduced
Placental / Angiogenic	sEng (Soluble Endoglin)	Maternal Circulation	Not specified - hints at severe PE	Elevated
Placental / Biochemical	PAPP-A (Pregnancy-associated plasma protein-A)	Maternal Serum	First Trimester	Reduced
Endothelial	vWF (Von Willebrand Factor)	Maternal Circulation (implied)	Not specified	Increased concentrations
Endothelial	ET-1 (Endothelin-1)	Not specified	Not specified	Escalated / Heightened levels
Oxidative Stress	MDA (Malondialdehyde)	Not specified	Not specified	Higher levels
Oxidative Stress	ADMA (Asymmetrical dimethylarginine)	Not specified	Not specified	Elevated levels
Oxidative Stress	SOD (Superoxide dismutase) & CAT (Catalase)	Placenta	Not specified	Decreased activities
Oxidative Stress	AOPP & 8-isoprostane	Not specified	Not specified	Heightened levels
Organ Dysfunction	Proteinuria	Urine	Diagnosis (post-20 weeks)	Increased
Organ Dysfunction	Serum Creatinine	Serum	Diagnosis (post-20 weeks)	Elevated
Organ Dysfunction	AST & ALT (Liver Enzymes)	Bloodstream	Diagnosis (esp. HELLP)	Elevated
Immune Response	TNF-α & IL-6 (Cytokines)	Maternal Circulation (implied)	Not specified	Elevated
Immune Response	Regulatory T cells (Tregs)	Blood / Decidua (implied)	Not specified	Reduced numbers / Dysfunction
cfNA	miR-210	Biological Fluids	Not specified	Elevated

2.1.9 The Role of Gene Expression Changes in cfRNA in Maternal Pre-eclampsia

The cfRNA sequencing has revealed significant gene expression changes associated with PE, providing valuable insights into the condition's pathophysiology and potential for early prediction. These changes are detectable early in gestation, well before the onset of clinical symptoms. Several genes have been consistently identified as differentially expressed in PE cases, including CLDN7, PAPPA2, SNORD14A, PLEKHH1, MAGEA10, TLE6, and FABP1, among others. These genes are involved in critical processes related to PE, such as placental development, cell junction formation, embryonic development, and fatty acid metabolism. The altered expression of these genes forms the basis for

potential early prediction of PE, with studies showing that a panel of 18 genes measured between 5 and 16 weeks of gestation can identify mothers at risk before clinical symptoms appear (Moufarrej et al., 2022). Furthermore, a classifier based on 13 cfRNA signatures and two clinical features has demonstrated high predictive performance for preterm and early–onset PE. The cfRNA changes reflect key aspects of preeclampsia physiology, including enrichment for genes specific to neuromuscular, endothelial, and immune cell types and tissues, as well as the downregulation of microRNAs that may play vital roles in the condition (Moufarrej et al., 2022). The finding that altered gene expression patterns are independent of clinical factors such as maternal age, body mass index (BMI), and race is a significant observation reported in high-resolution cfRNA studies. While many traditional predictive models rely on integrating these maternal characteristics with biomarkers for risk stratification, the demonstrated independence of certain cfRNA signatures highlights that the molecular changes being captured are likely direct consequences of placental dysfunction, rather than simply correlating with maternal risk factors (Castillo-Marco *et al.*, 2025). This distinction is crucial for developing novel, pathology-specific molecular diagnostic tools. These findings demonstrate that cfRNA sequencing can reveal complex transcriptomic changes associated with PE, offering new insights into the condition's pathophysiology and potential for early, non–invasive prediction, which could significantly improve maternal and foetal outcomes.

2.1.10 Biomarkers in Pre-eclampsia Prediction

The imperative to develop precise predictive models has prompted researchers to investigate various biomarkers and factors over the years. Many studies have woven a tapestry of indicators, ranging from maternal and clinical characteristics to intricate genomic signatures (Boutin *et al.,* 2018; Tarca *et al.,* 2022). Table 4 presents a compilation of biomarkers and factors utilised in constructing predictive models for PE. Notably, few investigations have incorporated genomic biomarkers

into their models. Furthermore, incorporating diverse genomic signatures, including but not limited to differential gene expression signatures, variant signatures, methylation signatures, and eQTL signatures, has been inadequately explored in existing studies. This suggests a potential avenue for future research to explore a broader range of genomic information and signatures when developing predictive models for PE.

The quest for a definitive predictive model continues; integrating emerging technologies and more expansive longitudinal studies may hold the key to bridging this gap. A summary of the performance metrics, including accuracy and Area Under the Curve (AUC), achieved by various Traditional, Machine Learning, and Deep Learning predictive models for PE in recent literature is presented in Table 6.

Table 4: Summary of previously published studies and the variable types used in the PE predictive model.

No	Studies	Maternal characteristics	Clinical characteristics	UtA Doppler measure	Medical history	Laboratory data	Socioeconomic factors	Genetic variants
1.	Jhee et al. 2019	Y	Y	N	Y	Y	N	N
2.	Marić et al. 2020	Y	Y	Y	Y	Y	N	N
3.	Martinez–Velasco, Martinez–Villasenor and Miralles–Pechuan. 2018	Y	N	N	Y	N	Y	N
4.	Marin et al. 2019	Y	Y	Y	Y	Y	N	N
5.	Neocleous et al. 2009	Y	Y	Y	Y	Y	N	Y
6.	Tahir, Muhlis, Badriyah and Syarif. 2018a	Y	Y	Y	Y	Y	N	N
7.	Sufriyana, Wu and Su. 2020a	Y	N	Y	Y	N	Y	N
8.	Sufriyana, Wu and Su . 2020b	Y	N	Y	Y	N	Y	N
9.	Yu et al, 2005	Y	N	Y	Y	N	N	N
10.	North et al, 2011	Y	N	N	N	N	N	N
11.	Odibo et al, 2011	Y	N	Y	Y	N	N	Y
12.	Caradeux et al, 2013	Y	N	Y	Y	N	N	N
13.	Parra–Cordero et al, 2013	Y	N	Y	Y	N	N	Y
14.	Scazzocchio et al, 2013	Y	N	Y	Y	N	N	Y
15.	Wright et al, 2019	Y	N	Y	Y	N	N	Y
16.	Bennett et al., 2022	Y	N	N	Y	N	N	N
17.	Manoochehri et al., 2021	Y	N	N	Y	N	N	N
18.	Tahir, Badriyah and Syarif, 2018b	Y	Y	N	Y	N	N	N
19.	Villalaín et al., 2022	Y	Y	N	Y	Y	N	N
20.	Lin et al., 2022	Y	Y	N	Y	Y	N	N
21.	Zhang et al., 2022	Y	Y	N	Y	Y	N	N
22.	Schmidt et al., 2022	Y	Y	N	Y	Y	N	N
23.	Carreno and Qiu, 2020	Y	Y	N	Y	Y	N	N
24.	Liu et al., 2022	Y	Y	N	Y	Y	N	N

Y = Used; N = Not used

While numerous biomarkers and factors have been incorporated into PE predictive models, particular ones exhibit significant influence in accurately predicting the risk of PE in women. Table 5 presents the distinct biomarkers and factors highlighted in the published articles, demonstrating their crucial role in augmenting the models' capacity to evaluate the probability of developing PE accurately. Their pronounced impact underscores their potential as valuable tools for identifying and managing PE risk during pregnancy.

Table 5: The candidate biomarkers and factors utilised in the PE predictive models from various studies.

No	Study	Candidate biomarkers and factors
1	Marin et al., 2019	1. Blood pressure 2. Age 3. Weight 4. The growth factor of the placenta
2	Marić et al., 2020	1. Diagnosis of chronic hypertension 2. Number of babies 3. History of PE 4. Diabetes mellitus type 1 5. Mean systolic and maximum diastolic blood pressure 6. Laboratory results: –High level of glucose $–4^{+}$ –Protein$–3^{+}$
3	Jhee et al.,2019	1. Systolic BP 2. Serum Bun and Creatinine level 3. Platelet count 4. White blood cell counts 5. Serum calcium level 6. Serum Magnesium level 7. Urine protein to creatinine ratio (UPCR) 8. Urine albumin to creatinine ratio (UACT)

No	Study	Candidate biomarkers and factors
4	Sufriyana, Wu and Su, 2020a	1. Maternal weight 2. BMI Values 3. BMI categories (<25 kg/m2 vs ≥25 kg/m2) 4. RI–tA 5. PI–UtA 6. Presence or Absence of a bilateral notch 7. sFlt-1 8. PIGF 9. sFlt–1/PlGF ratio 10. Angiogenic Factors: – Soluble fms-like tyrosine kinase 1 (sFlt–1) – Placental growth factor (PlGF)
5	Sandström et al., 2019	1. Increasing BMI 2. Maternal Age 3. Assisted reproduction 4. Country of birth 5. Chronic hypertension 6. Pre-existing diabetes 7. Chronic kidney disease 8. Family history of PE 9. MAP
6	Moreira, M. W. L., Rodrigues, J. J. P. C., Oliveira, A. M. B., et al. (2018)	1. Proteinuria 2. Hypertension
7	Sovio, U. et al. (2020)	1. 4–hydroxyglutamate 2. C–glycosyltryptophan 3. 1–palmitoyl–2–linoleoyl–GPE 4. sFlt–1: PlGF ratio 5. Pregnancy–associated plasma protein–A (PAPP–A) 6. placenta growth factor (PlGF)

No	Study	Candidate biomarkers and factors
8	Chaemsaithong, P. et al. (2019)	1. placental growth factor 2. Mean arterial pressure 3. Measurement of UtA–PI (uterine artery Pulsatility index) 4. placental growth factor
9	Martinez–Velasco, A., Martinez–Villasenor, L. and Miralles–Pechuan, L. (2018)	1. Duration of a completed pregnancy in weeks 2. Poverty 3. Water retention/Oedema in pregnancy 4. Toxaemia 5. Education (Completed years of schooling) 6. Highest completed year of school or degree 7. Pregnancy outcome 8. Labour force status
10	Tahir, M., Badriyah, T. and Syarif, I. (2018a)	1.Maternal_Age 2.MAP 3.Classfication_BMI 4.Previous_PE_Case 5.Personal_History_of_Hypertension 6.Personal_Pregnancy_History_of_Hypertension 7.Diabetes_Pragestasioal 8.Proteinuria_Reagen_Strip 9.Childbirth_Process gram/L

Despite the abundance of research and the multifaceted nature of PE predictive models, a universal challenge remains that achieving an optimal level of accuracy in the clinical setting, particularly during early pregnancy, has not yet been realised. Many of these models (Table 6Table 5), while statistically significant in research environments, lack the predictive power or specificity needed for real–world clinical application, leaving a gap in early, reliable detection of preeclamptic women (M. Liu *et al.*, 2022; Melinte-Popescu *et al.*, 2023). As the quest for a definitive

predictive model continues, integrating emerging technologies and more expansive longitudinal studies may hold the key to bridging this gap.

Table 6: Accuracy and Area Under the Curve (AUC) Values of Various PE Predictive Models.

No	Studies	Types of	AUC	Accuracy
1	Rokotyanskaya *et al.*, 2020	Traditional	0.733	–
2	Serra *et al.*, 2020	Traditional	0.96	–
3	Modak *et al.*, 2020	Traditional	0.949	–
5	Jhee *et al.*, 2019	ML model	–	97.3%
6	Marić *et al.*, 2020	ML model	0.79	–
7	Marin *et al.*, 2019	ML model	–	80%
8	Liu et al., 2022b	ML model	0.86	74%
9	Li et al., 2021	ML model	0.955	92%
10	Carreno and Qiu, 2020	ML model	0.93	–
11	Martínez–Velasco, Martínez–Villaseñor	ML model	–	85.3%
12	Bosschieter et al., 2022	ML model	0.770 ±	–
13	Schmidt et al., 2022	ML model	–	89% ± 3%
14	Sufriyana, Wu and Su, 2020b	ML model	–	90.6%
15	Sufriyana, Wu and Su, 2020a	ML model	0.95	–
16	Zhang, X. et al., 2022	ML model	0.8974	–
17	Lin et al., 2022	ML model	0.84 ±	–
18	Tahir, Badriyah and Syarif, 2018b	DL model	–	96.66%
19	Sakinah et al., 2019	DL model	–	90.22%
20	Tahir, Badriyah and Syarif, 2018a	DL model	–	95.68%
21	Manoochehri et al., 2021	DL model	–	79.1%
22	Han et al., 2020	DL model		79.8%
23	Bennett et al., 2022	DL model	0.724	72.2%

Key: ML = Machine Learning, DL = Deep Learning.

2.1.11 Early Prediction of Pre-eclampsia

Numerous predictive models have been developed to identify women at risk of PE, utilising various biomarkers. Prediction models have emerged as critical tools in modern healthcare, aiding clinicians in evaluating the likelihood of various medical conditions. In the context of maternal-foetal medicine, prediction models have

garnered substantial attention due to their potential to identify pregnancies at risk of PE. Accurate prediction of PE holds the potential to revolutionise prenatal care and improve maternal and foetal outcomes (Simas *et al.,* 2007; Kenny *et al.,* 2020). To achieve this transformative impact, these prediction models must embody essential characteristics. They should be simple and efficient, rapidly identifying the risk factors associated with PE. Furthermore, the methods employed for PE prediction should be non–invasive, affordable, and applicable early in gestation, ensuring accessibility for a broad population. Safety considerations are paramount, necessitating minimal risks to both the pregnant individual and the foetus. Widespread accessibility, validity, reliability, and reproducibility are equally vital. Additionally, these models should offer a high likelihood ratio for positive results (>10), signifying a strong association with PE development, and a low likelihood ratio for negative results (<0.1), indicating a reliable means of ruling out PE when it is unlikely to occur. In adhering to these criteria, PE prediction models prove not only effective but also practical, making them accessible tools for healthcare providers and contributing significantly to early intervention and improved outcomes in prenatal care (Conde-Agudelo, Villar and Lindheimer, 2004; Leslie, Thilaganathan and Papageorghiou, 2011). In recent years, significant progress has been made in developing prediction models for PE, integrating a wide range of maternal biophysical, biochemical, and imaging markers. The journey to accurately predict PE has evolved from relying on traditional methods to including advanced AI techniques.

2.1.11.1 Traditional Prediction Models for Pre-eclampsia

Historically, traditional statistical predictive models have been deeply rooted in years of clinical research and observations, providing a foundation for PE risk assessment. These models are crucial in identifying high–risk pregnancies, enabling timely interventions, and improving maternal and foetal outcomes. One widely used approach incorporates factors such as maternal age, body mass index (BMI), blood pressure, and the presence of pre-existing conditions like diabetes or

hypertension into logistic regression models (North *et al.*, 2011). These models assign a probability score to each patient, indicating their risk of developing PE during pregnancy. Combining these predictors allows healthcare providers to stratify patients into different risk categories, facilitating personalised care and monitoring. Additionally, some models incorporate biomarkers such as maternal serum levels of placental growth factor (PlGF) and soluble fms-like tyrosine kinase-1 (sFlt-1) (Zeisler *et al.*, 2016). These markers help improve predictive accuracy by providing biological insights into the pathophysiology of PE. While traditional statistical models have made significant strides in predicting PE, ongoing research continues to refine them, incorporating newer variables and ML techniques to enhance their accuracy and clinical utility.

2.1.11.1.1 Related Studies of Traditional Pre-eclampsia Prediction Model

A study by Modak et al., 2020 and his research team was conducted in India and concentrated on a cohort of 116 pregnant women. The selection of participants for the study was meticulous, involving stringent inclusion and exclusion criteria to ensure the dataset's integrity and exclude any outliers. The primary objective of this study was to assess the effectiveness of utilising UA Doppler and Urine Protein Creatinine Ratio (UPCR) measurements taken at the onset of pregnancy to predict the development of PE. Additionally, the study aimed to compare the precision of these two techniques in PE prediction. Upon enrolment, the patients underwent screening involving spot UA Doppler and UPCR assessments. Throughout their pregnancy, they were continuously monitored through blood pressure measurements and clinical examinations to identify the occurrence of PE. Midstream urine samples were collected and analysed using the immunoturbidimetric microalbumin method to quantify protein levels, while creatinine measurements were performed using a modified Jaffe's method. Individuals with a UPCR ratio equal to or exceeding 35.5 mg/mmol were considered positive for PE. Statistical analysis was conducted on all collected data, and the screening test's effectiveness was evaluated using parameters such as

the ROC curve, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Results showed that women with PE had a significantly higher median UPCR of 44.8 mg/mmol than unaffected women, who's median UPCR was 26.6 mg/mmol. An optimal spot UPCR cut–off value of 35.5 mg/mmol was determined for PE detection, with the test demonstrating a specificity of 94.06%, sensitivity of 80%, NPV of 96.94%, and PPV of 66.76%. The area under the ROC curve (AUC) for spot UPCR was 0.949 (95% CI: 0.891–1.000). When identifying PE using a mean UA resistance index exceeding 0.7, the specificity was 97.03%, sensitivity was 60%, NPV was 94.23%, and PPV was 75%, with an AUC of 0.856 (95% CI 0.742–0.971). Combining spot UPCR and UA Doppler measurements enabled the prediction of PE with an accuracy of 92.24% (Modak *et al.*, 2020; Aljameel *et al.*, 2023).

In 2020, Serra and colleagues conducted a study in Spain to develop a multivariate Gaussian distribution model for early–onset PE (EOPE) screening during the first trimester, particularly in a routine care setting with low–risk expectant mothers. Their dataset consisted of 13 distinct features derived from 6,893 singleton deliveries within the general population. The construction of the PE screening model involved three key steps: calculating the multiple of the median, defining prior risk, and defining posterior risk. The researchers discovered that combining biophysical parameters, maternal characteristics, and placental growth factor (PlGF) achieved the most effective detection rates. This combination yielded a remarkable 94% detection rate for a 10% false–positive rate (FPR) and a 59% detection rate for a 5% FPR. The AUC for this combined model was 0.96, with a 95% confidence interval (CI) ranging from 0.94 to 0.98. Notably, the inclusion of PlGF among the biophysical indicators significantly increased the detection rate, enhancing the accuracy of early–onset PE screening in low–risk pregnancies (Serra *et al.*, 2020; Aljameel *et al.*, 2023).

Another study in 2020 conducted by Rokotyanskaya and colleagues in Russia aimed to develop predictive tools for assessing the risk of PE in pregnant women. The study involved 457 pregnancies at gestational stages between 22 and 36

weeks and incorporated various molecular–genetic and biomedical indicators. Using logistic regression as their primary modelling method and utilising the Open Epi system to determine risk characteristics, the researchers conducted a retrospective analysis of gestational progress and labour outcomes. Ultimately, their efforts yielded a predictive system with an area under the receiver operating characteristic curve (AUC) of 0.733, indicating a moderate ability to distinguish between pregnant individuals at higher and lower risk of developing PE. This study contributes to the ongoing endeavours to improve the early prediction and management of PE, a critical complication during pregnancy (Rokotyanskaya *et al.*, 2020; Aljameel *et al.*, 2023).

In a prospective observational study conducted in 2020, Soongsatitanon and colleagues investigated the predictive potential of uterine artery (UA) Doppler, Pulsatility index (PI), and serum placental protein 13 (PP13) levels during the first trimester for the development of PE. The study gathered data from 353 pregnant women in Thailand. Statistical analysis using SPSS software was applied to the UA Doppler and PP13 test results to assess their predictive capacities. The study's findings revealed that the combined assessment of UA, PI, and serum PP13 levels yielded the highest accuracy in predicting PE. This combined approach resulted in a negative predictive value (NPV) of 94.4%, indicating its effectiveness in ruling out PE when negative, a specificity of 62.9%, sensitivity of 58.6% for detecting PE when present, and a positive predictive value (PPV) of 12.4%, signifying its ability to predict the condition when positive. These results highlight the potential of these first-trimester biomarkers in identifying individuals at risk of developing PE, offering valuable insights for early intervention and management (Soongsatitanon and Phupong, 2022; Aljameel *et al.*, 2023).

Traditional statistical models used for predicting PE often fall short due to their limited capacity to handle the intricate and nonlinear nature of the data. PE, a multifactorial condition, is influenced by numerous genetic, clinical, and environmental factors, making it challenging for traditional statistical approaches (Gray, Hemberg and Karumanchi, 2022; Chang, Seow and Chen, 2023). These

models typically rely on predetermined equations and may not adequately capture complex relationships within the data. In contrast, AI-based prediction models, such as deep learning and ensemble methods, excel in handling these challenges. They can automatically identify nonlinear patterns, adapt to evolving datasets, and effectively deal with missing data. Additionally, AI models can leverage large datasets, enabling more accurate predictions and early detection of PE by uncovering subtle biomarkers and patterns (Hedley *et al.*, 2023). While AI models may have been criticised for their interpretability, their superior predictive performance makes them a promising tool for improving PE prediction and ultimately enhancing maternal and foetal health outcomes. However, rigorous validation and integration into clinical practice are essential to fully realising their potential benefits.

2.1.11.2 AI Prediction Models for Pre-eclampsia

AI represents a dynamic field that delves into many principles aimed at the ambitious goal of fabricating machines possessing a remarkable semblance to human cognitive faculties. This intricate pursuit encompasses a multifaceted spectrum of technologies, methodologies, and paradigms that collectively strive to imbue machines with the ability to emulate and simulate various facets of human–like thinking, reasoning, and intentionality (Santoro *et al.*, 2021; Li, 2023). Central to AI's essence is its pursuit of harmonising with established human reactions to a diverse array of stimuli, fostering a seamless interaction between machines and their surroundings. At its core, AI involves a delicate interplay of cutting-edge software architectures and robust hardware infrastructure (Gill *et al.*, 2022). These elements work in unison, synergising knowledge–based processing models, vast repositories of data, and sophisticated algorithms. Such intricate orchestration aims to replicate, and sometimes even enhance, the intricate decision–making processes humans execute naturally. AI endeavours to transcend mere automation through the fusion of these components, venturing into autonomous, adaptive, and contextually aware systems.

The applications of AI traverse a tapestry of industries, leaving indelible imprints. In the realm of scientific exploration, AI acts as a tireless data miner, uncovering intricate patterns within vast datasets that human analysts might overlook(Chen, Jhangri and Chandra, 2014). In medicine, AI algorithms crunch complex medical data to predict and diagnose diseases, enhancing patient care and treatment planning (Rajkomar, Dean and Kohane, 2019; Ben-Israel *et al.*, 2020). As AI perpetually evolves, its journey is marked by a symbiotic relationship with cognitive science, data analytics, computer science, and numerous other disciplines. It is a testament to human ingenuity, challenging the boundaries of what machines can achieve and inspiring a redefinition of understanding of intelligence and creativity in the digital age.

2.1.11.2.1 Related Studies of AI Prediction Models for Pre-eclampsia

In recent years, the application of AI in predicting PE has opened new avenues for enhancing prenatal care and alleviating the burden of this condition. One prominent AI application in PE prediction revolves around the comprehensive analysis of various maternal biomarkers. Researchers have developed AI algorithms capable of amalgamating data from blood tests, medical histories, other clinical factors, and genomic biomarkers to formulate precise PE prediction models (Huang *et al.*, 2023). These diverse AI-assisted PE prediction models underscore the versatility of AI techniques, which harness various types of data and methodologies to enhance the early identification and management of this crucial pregnancy complication. The selection of a specific model hinges on factors such as the availability of data sources, the unique clinical context, and the desired balance between interpretability and predictive accuracy. However, it's worth noting that the two primary branches of AI, Machine Learning (ML) and Deep Learning (DL), are the most frequently employed techniques in developing PE prediction models in the latest research studies.

2.1.11.2.1.1 ML–Based Related Studies.

Many ML algorithms have recently emerged, leveraging diverse biomarkers and factors to identify women at risk of PE. Martínez–Velasco and colleagues employed various standard ML algorithms to predict the occurrence of PE (Martínez-Velasco, Martínez-Villaseñor and Miralles-Pechuán, 2018). These algorithms encompassed Random Forest (RF), Support Vector Machines (SVM), Bayesian networks, Neural Networks (NN), Naïve Bayes (NB), C4.5–like decision trees, Logistic Model Trees, C5.0, Boosted Logistic Regression, and Multivariate Adaptive Regression Splines. The dataset utilised in their research consisted of 25 parameters relevant to PE prediction. A strategy was adopted to address missing data, where missing values were imputed with the mean of the available data. Remarkably, within the array of ML algorithms explored, RF, coupled with the Leave–One–Out cross–validation (LOOCV) method, exhibited the most impressive performance with accuracy, specificity, and sensitivity metrics reaching values of 0.8530, 0.8614, and 0.6846, respectively. It's worth noting that this study focused on predictive modelling for PE without incorporating genomic biomarkers into the model development. Furthermore, the researchers aimed to enhance interpretability by utilising RF to extract factor significance through the Mean Decrease Gini Index (MDGI) metric and by visualising the most influential factors through a Decision Tree (DT), shedding light on the key determinants of PE prediction.

In another research study, Jhee and their team harnessed electronic medical records from Yonsei University Hospital, amassing data from a substantial cohort of 11,006 pregnant women to predict late–onset PE (Jhee *et al.*, 2019). This dataset encompassed maternal information from the early second trimester to the 34th week of pregnancy, meticulously extracted from electronic medical records. The development of prediction models was an intricate process involving parameter selection through cluster analysis and pattern recognition. The study explored an array of ML techniques, including Logistic Regression (LR), DT modelling, RF algorithms, Stochastic Gradient Boosting (SGB), Support Vector Machine (SVM) approaches, and the Naïve Bayes (NB) classification method. C–

statistics were employed as evaluation metrics to gauge the performance of these models. The results unveiled the C–statistics for each model, where the RF algorithm achieved a value of 0.894, LR attained 0.806, the DT model reached 0.857, SVM exhibited 0.573, NB classification achieved 0.776, and the SGB method excelled with a C–statistic of 0.924. The most impressive performance was observed in the SGB model, showcasing a False Positive Rate (FPR) of 0.009 and an accuracy of 0.973. This underscores the effectiveness of the SGB method in late–onset PE prediction, offering promising avenues for enhanced prenatal care and risk assessment. Furthermore, it's noteworthy that this study utilised electronic records encompassing data from the early second trimester in the ML algorithms for constructing predictive models, resulting in notably higher accuracy levels. This underscores the potential benefit of identifying genomic biomarkers from early pregnancy cfRNA sequencing data and incorporating them into prediction models, which could significantly enhance accuracy and aid in the early detection of women at risk of PE.

Marin and colleagues embarked on a study aimed at harnessing the potential of ML algorithms to predict the occurrence of PE, leveraging essential maternal information such as blood pressure, age, and weight (Marin *et al.*, 2019). Their approach involved the utilisation of the Viterbi ML algorithm, renowned for its ability to unveil significant connections between nodes within multiple datasets. This task would be considerably challenging through conventional means. With this medical data, the algorithm employed a built–in computer program that estimated the probability of PE development. Study participants were equipped with smart bracelets housing sensors and electronic communication modules to facilitate data collection and monitoring, collectively called the "i–bracelet system." These bracelets allowed for real–time blood pressure readings, which were transmitted to a mobile device via Bluetooth pairing. The Viterbi algorithm was pivotal in determining whether pregnant women wearing these bracelets were at risk of developing PE. It accomplished this by selecting the most probable path, utilising a set of hidden states within a hidden Markov model. One hundred and

five individuals participated in the study, and the results were promising, with an overall accuracy rate of 80%. Additionally, the model exhibited a specificity of 72% and an impressive sensitivity of 92.5%. This innovative approach demonstrated the potential of ML algorithms in PE prediction and showcased the integration of wearable technology for continuous health monitoring, highlighting its relevance in modern healthcare practices.

Marić and colleagues developed an ML model for EOPE prognosis, notable for automatically selecting the most relevant features from many variables (Marić *et al.*, 2020). Their objective was to create a prediction tool based on routine prenatal visit data accessible to all expectant mothers, facilitating the identification of high–risk patients through an initial screening process. The dataset encompassed records of 16,370 deliveries that occurred between April 2014 and January 2018 at Lucile Packard Children's Hospital in California. The predictive model was constructed using two statistical learning algorithms: gradient boosting and elastic net. It considered 67 factors, including medication usage, routine clinical data, medical histories, and prenatal laboratory results. The model's performance was rigorously evaluated using cross–validation, with key metrics such as the AUC, TPR, and FPR being assessed. Notably, the elastic net algorithm emerged as the most effective, yielding an AUC of 0.79 (95% CI 0.75–0.83), an FPR of 8.1%, and a sensitivity of 45.2% for overall PE prediction. Furthermore, the model predicted EOPE, achieving an AUC of 0.89 (95% CI 0.84–0.95), a TPR of 72.3%, and an FPR of 8.8%. This research demonstrates the potential of ML–driven early prognosis tools in enhancing PE risk assessment during routine prenatal care.

2.1.11.2.1.2 DL–Based Related Studies

DL offers advantages over traditional ML for predicting PE because it can automatically extract intricate features from complex, multi–modal data. Deep neural networks excel in capturing hierarchical patterns and temporal dependencies, making them suitable for large–scale datasets and tasks involving medical imaging or patient histories (Zhang *et al.*, 2019). Transfer learning and

ensemble methods further enhance their utility. However, deep learning requires substantial data and computational resources, can be challenging to interpret, and may risk overfitting. The choice between deep learning and ML should align with data characteristics and interpretability needs, emphasising collaboration between domain experts and data scientists for accurate and clinically relevant PE prediction models.

In a comprehensive study by Han and colleagues, medical data from 568 women receiving care at the Obstetrics Department of Fujian Maternal and Child Health Hospital were meticulously analysed (Han, Q. *et al.,* 2020). This diverse sample encompassed pregnancies afflicted with PE, normal–term pregnancies, and pregnancies characterised by gestational hypertension (GH) during the period spanning from September 2014 to September 2018. Among these pregnancies, 216 were diagnosed with PE, 136 with GH, and 216 remained unaffected by either condition. The study's primary objective was to pinpoint the most reliable indicators of PE utilising a sophisticated backpropagation (BP) Neural Network (NN) model. To execute this task, the researchers harnessed TensorFlow software to construct a three–layer BP NN. This neural network incorporated 25 distinct features as input nodes. The output layer of the NN was designed to categorise the data into three sample types, thereby enabling a meticulous examination of the factors influencing PE risk. The weight values (W1) corresponding to the input layer neuron nodes were derived through the training of the NN model, and their significance was established based on the magnitude of these weights. Ultimately, the BP NN analysis identified a set of prominent predictors for PE, with a notable emphasis on albumin, mean platelet volume, blood urea nitrogen, lactate dehydrogenase, and triglyceride as the most influential factors. These findings offer valuable insights into the critical determinants of PE. Moreover, the BP NN demonstrated a notable accuracy rate of 78.8%, underscoring its potential as a useful tool in the early identification and management of PE.

Tahir and colleagues conducted a study from December 2017 to February 2018. They involved the collection of 17 parameters from a substantial dataset comprising 1077 patient records sourced from two hospitals in Makassar and Haji General Hospital in Surabaya (Tahir, Badriyah and Syarif, 2018a). Notably, the experiment employed two distinct types of data: the original dataset containing all 17 features and a dataset that underwent feature selection through the Particle Swarm Optimisation (PSO) algorithm, resulting in a reduced set of seven features. To evaluate the performance of the prediction models, the researchers employed various methods, including rule induction, DL, SVM, DT, NN, NB, and k–nearest Neighbors (KNN). Upon comparing these methodologies, it was observed that DL consistently outperformed the others in terms of accuracy for both data types. Specifically, when utilising the original dataset, DL achieved an accuracy rate of 95.12%. Impressively, when applying DL to the dataset optimised by the PSO algorithm, the accuracy improved to 95.68%. Consequently, it can be concluded that DL exhibited superior predictive accuracy when the feature selection process was integrated, underscoring its potential as a robust tool for PE risk assessment.

Moreover, Tahir and colleagues embarked on another study to harness Neural Networks (NN) as a predictive tool for estimating the likelihood of PE, putting it on par with other established algorithms like Naïve Bayes (NB) and linear regression (Tahir, Badriyah and Syarif, 2018b). Their dataset, sourced from Surabaya Haji General Hospital, encompassed 239 samples and took into consideration a comprehensive set of 17 risk factors relevant to PE prediction. To optimise the predictive power of the NN model, the researchers conducted experiments with a single hidden layer to identify the ideal number of neurons that yielded the lowest error rate, which ultimately stood at 17 neurons. In the final phase, the model underwent rigorous validation and was subjected to comparison using three distinct validation methods. NN emerged as the top performer, boasting an impressive 96.66% accuracy rate, as validated through the LOOCV technique.

This study underscores the potential of NN in enhancing the accuracy of PE prediction, with implications for improving maternal and foetal health outcomes.

Sakinah and colleagues employed a specialised variant of recurrent neural networks (RNN), known as Long Short–Term Memory (LSTM), to develop an accurate prognosis model (Sakinah *et al.*, 2019). Achieving precision in this method necessitated the determination of an optimal combination of several critical parameters. These parameters encompassed factors such as the count of neurons within the hidden layers, maximum epochs, and the selection of training and testing data. The LSTM network incorporated the Adaptive Moment Estimation (ADAM) optimisation technique during the learning process, which facilitated the discovery of optimal weight values, ultimately minimising the system's error rate. The researchers employed LOOCV and tenfold cross–validation methods to validate their approach. The dataset for their study was sourced from Haji General Hospital in Surabaya. The study achieved the highest accuracy rates for training data (96.62%) and testing data (90.22%). These impressive percentages were attained through a combination of factors, including LOO cross–validation, 30 hidden neurons, a maximum of 100 epochs, and a single parameter series pattern. This study underscores the potential of LSTM–based predictive models in PE prognosis, offering a promising avenue for improving risk assessment and maternal care.

In another study, Bennett and colleagues harnessed extensive datasets from various sources, including the Public Use Data Files (PUDF) for Texas, the Magee Obstetric Medical and Infant (MOMI) database, and the Oklahoma PUDF, with the primary objective of achieving early prediction of PE through the application of ML techniques, specifically the cost–sensitive deep neural network (CSDNN) method (Bennett *et al.*, 2022). To optimise the efficiency of their models, the research employed a combination of network architectures and hyperparameter optimisation algorithms, such as Hyperband, Bayesian optimisation, and random

search. The performance of CSDNN was rigorously assessed in comparison to several other predictive models, including Support Vector Machine with a linear kernel (SVM–Lin), Logistic Regression (LR), and Support Vector Machine with a radial basis function (SVM–RBF), along with their respective cost–sensitive variants. Among the models evaluated, CSDNN with focal loss (CSDNN–FL) demonstrated the best performance for the Oklahoma and Texas datasets, achieving Area Under the Curve (AUC) values of 64% and 66%, respectively. In the case of the MOMI dataset, both CSDNN–FL and CSDNN with the loss function weighted cross–entropy (CSDNN–WCE) outperformed other methods, yielding an AUC of 76%. While models like WSVM–RBF, CSDNNs, and WLR achieved the highest G–mean values, the AUC values for Native Americans in Texas reached 57.1%, and for African Americans in Texas, it was 66.7%. In Oklahoma, the best results for Native Americans were attained using DNN and a balanced batch approach, resulting in an AUC of 58%. Meanwhile, the best outcomes for Oklahoma African Americans were obtained with CSDNN–WCE (AUC 62.3%). CSDNN–FL with a balanced batch approach or CSDNN–WCE were the most effective models for the MOMI African American dataset. This comprehensive evaluation of DL methods provides valuable insights into the prediction of PE, highlighting the superiority of CSDNN models in specific datasets and underscoring the importance of tailored approaches for different demographic groups.

In 2021, Manoochehri and colleagues undertook a study to identify patients at risk of PE syndrome and uncover the associated risk factors (Manoochehri *et al.*, 2021). Their approach involved the development of a data mining–based screening model. The dataset utilised for this research comprised medical records from 1452 pregnant women, each characterised by nine distinct features. These records were sourced from Hamadan City, Iran, from April 2005 to March 2015. The investigation applied six data mining techniques: RF, LR, discriminant analysis, KNN, SVM, and C5.0 DT. The study's findings pointed to the most

significant risk factors contributing to the detection of PE, which included variables such as the number of pregnancies, maternal age, and the season during which pregnancy occurred. Among the six data mining methods, the highest prediction accuracy, reaching 0.791, was achieved by deploying the SVM model. These results underscore the potential of data mining techniques in identifying PE risk and provide valuable insights for improving prenatal care and risk assessment.

Traditional statistical predictive models and AI-based approaches –including both ML and DL models –have demonstrated enhanced accuracy in predicting the risk of PE. Nevertheless, these models, despite their advancements, still encounter challenges in accurately identifying women at risk for PE. PE prediction in pregnancies, though vital, presents challenges due to its complex and silent symptom presentation. Selecting the appropriate vital signs for diagnosis is tricky, often requiring medical experts' consultation to ensure accurate feature selection. Another significant challenge lies in safeguarding patients' data privacy. While data sharing is essential for refining AI models, protecting patient confidentiality remains paramount. Solutions like federated learning allow AI models to learn without compromising data security.

Metrics like sensitivity, specificity, and accuracy judge AI models' reliability in clinical settings. High sensitivity is crucial to detect as many positive cases as possible. Explainable artificial intelligence (XAI) can make these models more understandable, boosting trust among health professionals and patients. Dataset quality is equally critical; inadequate size or missing values can hamper the performance of AI models. Imbalanced datasets, arising from low disease prevalence or single–centre data sources, might introduce biases, emphasising the need for larger, multicentre datasets.

Another issue is the black–box nature of ML models. Their inherent complexity makes them hard for humans to decipher, potentially eroding trust in life–critical

decisions. Techniques like XAI can demystify these models, making them more transparent. However, human barriers persist in AI adoption in healthcare, particularly in determining accountability for errors between algorithms and physicians. Improved human–AI interaction can mitigate this with clearer algorithmic interpretability. However, one cannot overlook the potential for model bias. If not carefully calibrated, AI systems could inadvertently perpetuate societal biases. To address this, it's essential to use diverse datasets during training and to scrutinise the model's decisions across different subgroups. However, data from the first trimester was unavailable, as most pregnant women participated only in the antenatal evaluation program, and there were few recorded incidents of PE.

2.2 Conclusion

PE is a serious pregnancy complication that impacts maternal and perinatal health. Early detection can enhance outcomes, but currently, no reliable screening test exists to predict its onset, particularly at the most prevalent term gestations. Numerous predictive models for PE have been developed, some showing potential. However, the variation among these models creates challenges in choosing the most effective one or aggregating their findings. A significant barrier is the high heterogeneity observed between studies.

The reviewed models showed moderate predictive capabilities but were not without flaws. Key limitations include poor calibration, likely resulting from overfitting during their development phase, and considerable heterogeneity. Furthermore, their net benefit across various settings was minimal. This means the empirical support for using these models for clinical decision–making regarding PE remains limited. Before considering their practical application, any models that have not been validated should be rigorously assessed for their predictive accuracy, net benefit, and variability across multiple contexts.

Research has identified potential biomarkers for PE using samples from diagnosed cases. While these studies have proposed promising biomarkers, their primary

focus is understanding the disease rather than predicting it. The ideal approach might be combining multiple sources' biomarkers to achieve the highest predictive accuracy. This literature review highlights a notable gap in automated AI-aided PE prediction models. Specifically, many of these models have overlooked including a diverse range of genomic biomarkers. Such overlooked biomarkers encompass differential gene expression signatures, methylation patterns, variant signatures, and eQTL signatures derived from various sources. Large–scale studies on unselected populations offer the best opportunity for discovering these new biomarkers. However, to evaluate their potential to enhance outcomes for expectant individuals and their offspring, these identified markers, or their combinations, must be externally validated in other cohorts.

2.3 Summary of the Chapter

PE is a significant pregnancy–related hypertensive disorder contributing to maternal and foetal morbidity and mortality worldwide. While extensive research has identified key mechanisms such as abnormal placentation, immune dysregulation, oxidative stress, and endothelial dysfunction, the precise cause remains unclear. Current predictive models for PE have limitations, as they fail to capture the dynamic molecular changes across gestational weeks and do not effectively integrate genomic biomarkers, reducing their accuracy and clinical applicability.

The literature review highlights several critical gaps in existing research. Traditional pathophysiological models, including the two-stage model and placental dysfunction theory, explain PE development but lack integration into ML–based predictive models. Existing ML models primarily rely on clinical parameters rather than genomic data, limiting their ability to identify early molecular warning signs. Additionally, immunological dysregulation and oxidative stress theories underscore the role of altered immune responses and endothelial dysfunction in PE, but these molecular insights are not fully leveraged in predictive approaches.

Another major gap is the lack of explainability and external validation in current ML models, reducing their reliability for real–world clinical use.

This research aimed to bridge these gaps by developing a gestational week–specific, multi–omics–based ML model that integrates genomic biomarkers such as differential gene expression signatures, and variant signatures. By utilizing cell-free RNA sequencing data from early pregnancy and applying advanced AI techniques, this study enhances predictive accuracy, improves early detection, and ensures model interpretability and validation across diverse cohorts. Ultimately, this work contributes to a more reliable, non–invasive, and clinically applicable PE prediction model, facilitating early risk assessment and personalised interventions for better maternal and foetal health outcomes.

Key Findings of the Literature Review:

- The study explores the complex pathogenesis of PE and highlights multiple theories, including the two-stage model, placental dysfunction, immunological dysregulation, oxidative stress, and endothelial dysfunction.
- The two-stage model suggests that abnormal placental implantation leads to maternal systemic inflammatory responses, causing the symptoms of PE.
- Placental dysfunction theory proposes that shallow trophoblast invasion and inadequate spiral artery remodelling result in placental ischemia, triggering endothelial dysfunction and hypertension.
- The immunological dysregulation theory explains PE as a result of altered maternal immune responses, particularly involving regulatory T cells and natural killer cells.
- Oxidative stress and endothelial dysfunction theory suggests that increased reactive oxygen species (ROS) production and reduced antioxidant defences contribute to endothelial dysfunction and hypertension in PE.

- The study discusses the evolution of the two-stage model from 1991 to 2019, incorporating maternal risk factors, immune responses, and oxidative stress into the disease framework.
- Key molecular factors implicated in PE pathogenesis include soluble fms-like tyrosine kinase-1 (sFlt-1), soluble endoglin (sEng), vascular endothelial growth factor (VEGF), and placental growth factor (PlGF).
- The study emphasises the importance of early detection and intervention strategies to mitigate maternal and foetal complications associated with PE.
- Most predictive models do not consider week–by–week gene expression changes, limiting their ability to capture dynamic molecular shifts that precede PE.
- Current models rely on single–layer biomarkers (e.g., clinical or transcriptomic data) instead of multi–omics (genomics, proteomics, metabolomics) and longitudinal datasets, reducing predictive accuracy.
- Many models function as black–box predictors without biological interpretation, and they often lack validation on diverse populations, limiting clinical applicability.

Chapter 3
General Methods

3.1 Summary

This thesis comprises a series of bioinformatics analyses, including differential gene expression analysis, consistent gene expression analysis, and variant calling analysis aimed at constructing an automated recognition model with AI assistance to identify pre-eclamptic risk in pregnant women during early pregnancy. This chapter delineates the typical workflow steps for conducting each experiment in this study. Moreover, the unique methodology of each experiment is explored in detail within its respective chapter.

3.2 The Population of the Study

This study aimed to develop an AI-assisted prediction model for identifying women at risk of PE during early pregnancy. Consequently, the transcriptomic and variant data of pregnant women are considered the population for this study. Data on pregnant women that had previously been gathered were collected for this study. Furthermore, this research project did not explicitly recruit or study human subjects.

3.3 Sample Size Calculation.

Sample size calculation in Bioinformatics analysis differs from traditional processes utilised in experimental or medical investigations. In bioinformatics, the sample size is broadly speaking determined by factors inclusive of statistical power and effect length estimation (Bi and Liu, 2016). Additionally, the specified level of accuracy, data imbalance, the quantity of the effect being evaluated, and the desired statistical power all play a function in organising the proper sample size for gene expression research. The sample size calculation serves to pick out the minimal quantity of samples essential to estimate the parameters with an

appropriate level of precision (Hajian-Tilaki, 2014). In contrast, use of a statistically incorrect sample size may lead to inadequate results in both clinical and laboratory studies as well as resulting in time loss, cost, and ethical problems (Serdar *et al.,* 2021).

Sample size calculation through power analysis is fundamental in the research design phase. This method permits researchers to confidently gauge the number of individuals required to reveal an existing effect in a research study. However, power analysis is significantly influenced by the nature of the NGS data and the goal of the research studies (Li, C. *et al.,* 2018). It is underpinned by four foundational principles: effect size, Alpha (α), power, and the chosen statistical test. These principles collectively shape a research study's reliability and robustness.

The rigorous principles of sample size calculation, including defining Effect Size, Alpha (α), and Power (β), served as the foundational rationale for assessing the data sufficiency in this study. Unlike traditional prospective clinical trials, this research utilised pooled public Next-Generation Sequencing (NGS) data, making retrospective power calculations based on final sample size (N=228) essential. The complexity and heterogeneity inherent in transcriptome data, alongside the primary aim of Differential Gene Expression Analysis (DGEA), necessitated strict criteria for data selection. Specifically, the total sample size across all gestational weeks was deemed appropriate and statistically sufficient because it adhered to the common bioinformatics convention of providing the power required to detect a medium effect size (Cohen's d = 0.5). This application of statistical principles ensures that the large volume of sequencing data collected was robust and appropriately sized for the subsequent complex statistical and machine learning analyses performed in this thesis.

3.4 Ethical Approval

The ethical concerns that emerge when conducting genomic research are numerous and include consent, privacy, collection, storage, and sharing of genomic data (de Vries *et al.,* 2011; Kaye *et al.,* 2010; Caulfield *et al.,* 2008). Although no humans were recruited in this study, it is necessary to obtain ethical approval from the appropriate ethics committee and electronic repositories to conduct in silico experiments because of the use of patients' data. There are two steps in obtaining ethical approval for this study: approval from the Biomedical Sciences Research Ethics Committee of the University of West London and approval to access the data from electronic repositories. The proposal for the proposed study (Application ID: UWL/REC/CNMH–00959) was approved by the Biomedical Sciences Research Ethics Committee of the University of West London.

The most important ethical consideration for this research is handling data from electronic data repositories. However, this study used the publicly available cfRNA sequencing data from the National Centre for Biotechnology Information (NCBI). Handling these data from repositories comes with associated ethical considerations. Although the data are made accessible to the public, it does not necessarily mean they can be freely used for any purpose. A primary ethical concern is the protection of individual privacy. Even if the data are de–identified, there is a risk of re–identification in some contexts.

It is crucial to confirm that the publicly accessible RNA sequencing data have been de–identified and to examine the applicable terms of use before conducting any analysis. This guarantees that the permissions established by the original authors are compatible. Adopting strict precautions is essential to prevent the potential re–identification of specific samples. It is necessary to strictly adhere to the licence requirements under which the data was initially obtained. Such care ensures that

the research stays true to the core goal of the initial data gathering, upholding the highest standards of privacy and ethical considerations for the dataset.

Accessing data for this study from the Gene Expression Omnibus (GEO) hosted by the NCBI database did not require specific permission since it is openly available. Despite its public accessibility, the study was committed to ethical handling, particularly concerning potentially identifiable details or human genetic data. The data acquired was strictly for research and was not repurposed for unrelated activities like marketing or insurance. Furthermore, personal data, such as hospital health records, were managed per the General Data Protection Regulation (GDPR) standards.

3.5 General Workflow

This study uses a variety of genomic analyses for the identification of biomarkers, which include recurring processes (Figure 3) from initial data collection to alignment. Some common tools were also frequently employed in the genomic analyses' data gathering, pre–processing, and alignment. This chapter presents the detailed workflow for each genomic analysis.

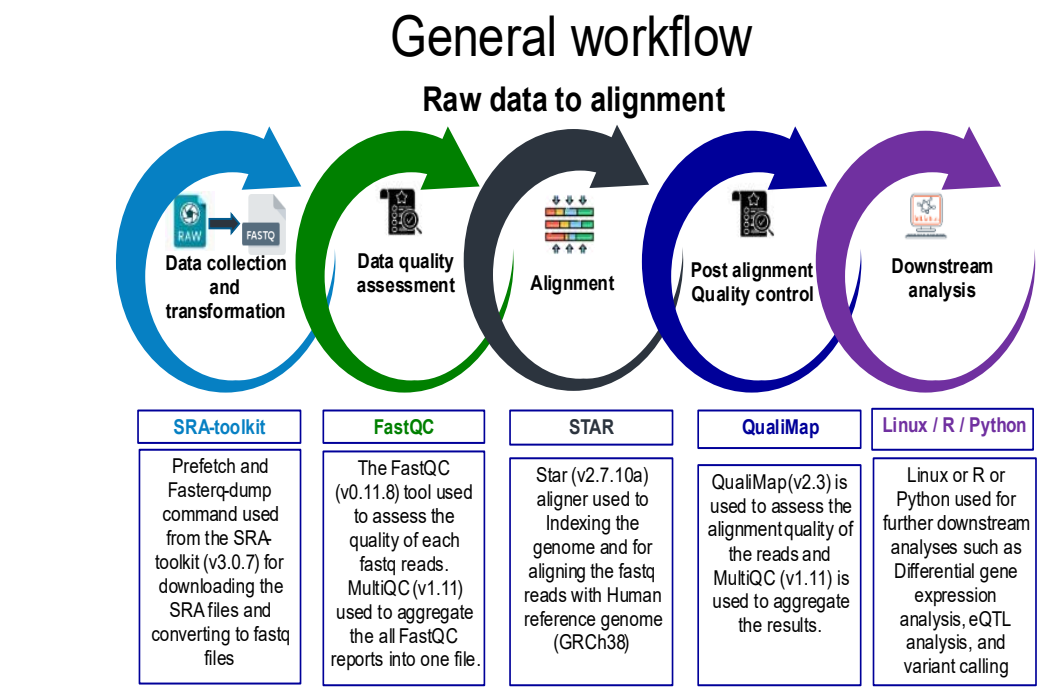


Figure 3: General workflow of genomic analysis in the study.

The workflow shows the typical steps in this study's genomic analyses, from collecting initial data to aligning.

3.6 Data Collection

The development of the PE prediction model in this study involved identifying gene expression signatures, unique, consistent gene expression signatures, and variation signatures using cfRNA sequencing data. Therefore, various databases were searched for relevant data sets for each project.

3.6.1 Data Collection of cfRNA Sequencing Data

The GEO and ArrayExpress from the European Bioinformatics Institute (EBI) are two prominent omics repositories. They compile and offer gene expression data from a vast array of studies (Wang, Zichen *et al.,* 2016; Barrett *et al.,* 2012; Rustici *et al.,* 2012). To gather human cfRNA sequencing data from longitudinal studies, these databases were queried using the keywords "Preeclampsia" and "cell-free DNA and RNA" from January 2021 to January 2023. This study's inclusion criteria comprised human data, cell-free RNA sequencing data from maternal blood, and

high–throughput sequencing data. Any datasets containing non–human data, or those focusing on microarray and methylation sequencing, were excluded. From the search, the following cfRNA sequencing datasets derived from maternal blood were incorporated into this study: GSE154377 and GSE192902. These datasets were respectively submitted by Del Vecchio et al., 2021, and Moufarrej et al., 2021.

3.6.1.1 GSE154377

The GSE154377 dataset contains 134 samples, including 102 maternal, 25 cord plasma, and seven non–pregnant women samples. These data were generated by the platform GPL20301 Illumina HiSeq 4000 (Homo sapiens). A total of 16 single–end early pregnancy (up to gestational week 16) raw sequencing data (NP = 8, PE = 8) were used in this experiment from the GSE154377 data series (Table 7).

Table 7: Number of raw data for NP and PE in the Study.

GEO Accession ID	NP	PE	Platforms	Total
GSE154377	8	8	GPL20301 Illumina HiSeq 4000 (Homo sapiens)	**16**
GSE192902	157	60	GPL24676 Illumina NovaSeq 6000 (Homo sapiens)	**217**
Total	**165**	**68**		**233**

3.6.1.2 GSE192902

The selected dataset, GSE192902, consists of 404 blood plasma samples from 199 pregnant mothers. From this dataset, samples from early pregnancy (up to 16 weeks) were filtered, resulting in 217 samples. Of these, 157 are labelled NP, and 60 are labelled PE. The samples underwent paired–end cfRNA sequencing using the GPL24676 Illumina NovaSeq 6000 platform (Homo sapiens). However, some of the data in GSE192902 were collected from the same patients over various gestational weeks (Moufarrej et al., 2022). It would be preferable to have longitudinal data broken down by gestational week, which could help to identify the potential biomarkers. This kind of longitudinal data aids in accurately determining

the gene expression pattern in early pregnancy and comprehending the biology of the disease. However, insufficient per–gestational cfRNA sequencing data is available from the transcriptome databases. In total, this study employed 233, NP (n=165), and PE (n=68) cfRNA sequencing data from maternal blood (Table 10).

The source datasets (GSE154377 and GSE192902) provided maternal plasma cfRNA sequencing data from early pregnancy. While some individuals in these datasets contributed samples across multiple gestational weeks, the data were not consistently or fully longitudinal across all patients, precluding a comprehensive time-series analysis. Consequently, the primary analytical approach adopted in this thesis was structured cross-sectionally. For each gestational week (GW), all available samples from that specific week were pooled and treated as an independent cohort. Therefore, the sample numbers presented represent the total number of samples analysed per week and do not equate to unique pregnancies, as some individuals contributed multiple samples across different weekly cohorts. This approach was strategically implemented to maximise statistical power for identifying reliable GW-specific signatures, but it introduces a limitation in the study's ability to track individual longitudinal temporal

The demographic composition of the pooled datasets, originating from diverse cohorts in the United States, provides a degree of ethnic representation but is not comprehensive across all global populations (Table 8). This geographic and ethnic profile introduces a critical limitation in the initial scope of the study, as it may affect the generalisability of the identified gene expression and variant signatures to women from distinct socioeconomic or ancestral backgrounds. This critical assessment of the data source underscores the necessity for prospective validation in a wider multi-ethnic cohort, as outlined in the future directions (Chapter 9).

Table 8: Demographic and Recruitment Characteristics of Pooled GEO Datasets for cfRNA Analysis

GEO ID	Publication (Year)	Recruitment Location and Setting	Types of Participants Included (by self-identified race/ethnicity)
GSE192902	Moufarrej et al. (2022)	United States cohort (Stanford-based, multi-site)	Pregnant individuals across a diverse set of racial and ethnic groups (e.g., White, Asian, Black/African American, Hispanic/Latina)
GSE154377	Del Vecchio et al. (2021)	United States cohort (UCLA, Los Angeles)	Pregnant individuals representing a racially and ethnically diverse population (e.g., White, Asian, Hispanic, Black)

Table **9** provides detailed source information for the cfRNA sequencing dataset used in this study, including Sample IDs, GEO accession numbers, SRA run identifiers, gestational age (GW 8–16), and clinical classifications (PE, Severe Pre-eclampsia (SPE), and NP).

Table 9: Source Details of cfRNA Sequencing Data by Gestational Age (GW 8–16).

No	Sample ID	GEO accession	SRA Run	Gestational Age	Condition	No	Sample ID	GEO accession	SRA Run	Gestational Age	Condition
1	P3072PE	GSM5768388	SRR17333072	8	PE	30	P3013PE	GSM5768447	SRR17333013	10	PE
2	P3330PE	GSM5768534	SRR17333330	8	PE	31	P3078PE	GSM5768383	SRR17333078	10	PE
3	P3113SPE	GSM5768347	SRR17333113	8	SPE	32	P3256PE	GSM5768608	SRR17333256	10	PE
4	P3240SPE	GSM5768624	SRR17333240	8	SPE	33	P3302PE	GSM5768561	SRR17333302	10	PE
5	P3052NP	GSM5768408	SRR17333052	8	NP	34	P3012NP	GSM5768448	SRR17333012	10	NP
6	P3090NP	GSM5768370	SRR17333090	8	NP	35	P3059NP	GSM5768401	SRR17333059	10	NP
7	P3095NP	GSM5768364	SRR17333095	8	NP	36	P3098NP	GSM5768363	SRR17333098	10	NP
8	P3107NP	GSM5768353	SRR17333107	8	NP	37	P3165NP	GSM5768699	SRR17333165	10	NP
9	P3202NP	GSM5768662	SRR17333202	8	NP	38	P3171NP	GSM5768693	SRR17333171	10	NP
10	P3230NP	GSM5768633	SRR17333230	8	NP	39	P3222NP	GSM5768642	SRR17333222	10	NP
11	P3275NP	GSM5768589	SRR17333275	8	NP	40	P3260NP	GSM5768605	SRR17333260	10	NP
12	P3354NP	GSM5768509	SRR17333354	8	NP	41	P3267NP	GSM5768597	SRR17333267	10	NP
13	P3367NP	GSM5768498	SRR17333367	8	NP	42	P3268NP	GSM5768596	SRR17333268	10	NP
14	P3372NP	GSM5768492	SRR17333372	8	NP	43	P3269NP	GSM5768595	SRR17333269	10	NP
15	P3376NP	GSM5768488	SRR17333376	8	NP	44	P3299NP	GSM5768565	SRR17333299	10	NP
16	P3387NP	GSM5768477	SRR17333387	8	NP	45	P3336NP	GSM5768528	SRR17333336	10	NP
17	P3003PE	GSM5768457	SRR17333003	9	PE	46	P3343NP	GSM5768520	SRR17333343	10	NP
18	P3191PE	GSM5768673	SRR17333191	9	PE	47	P3386NP	GSM5768479	SRR17333386	10	NP
19	P3074SPE	GSM5768386	SRR17333074	9	SPE	48	P3393NP	GSM5768471	SRR17333393	10	NP
20	P3062NP	GSM5768398	SRR17333062	9	NP	49	P3009PE	GSM5768451	SRR17333009	11	PE
21	P3102NP	GSM5768358	SRR17333102	9	NP	50	P3128PE	GSM5768332	SRR17333128	11	PE
22	P3140NP	GSM5768321	SRR17333140	9	NP	51	P3197PE	GSM5768667	SRR17333197	11	PE
23	P3179NP	GSM5768684	SRR17333179	9	NP	52	P3209PE	GSM5768655	SRR17333209	11	PE
24	P3258NP	GSM5768606	SRR17333258	9	NP	53	P3246PE	GSM5768618	SRR17333246	11	PE
25	P3271NP	GSM5768593	SRR17333271	9	NP	54	P3247PE	GSM5768617	SRR17333247	11	PE

No	Sample ID	GEO accession	SRA Run	Gestational Age	Condition	No	Sample ID	GEO accession	SRA Run	Gestational Age	Condition
26	P3277NP	GSM5768585	SRR17333277	9	NP	55	P3319PE	GSM5768544	SRR17333319	11	PE
27	P3278NP	GSM5768586	SRR17333278	9	NP	56	P3323PE	GSM5768541	SRR17333323	11	PE
28	P3347NP	GSM5768517	SRR17333347	9	NP	57	P3154SPE	GSM5768710	SRR17333154	11	SPE
29	P3350NP	GSM5768514	SRR17333350	9	NP	58	P3229SPE	GSM5768635	SRR17333229	11	SPE
59	P3377SPE	GSM5768487	SRR17333377	11	SPE	88	P3290NP	GSM5768574	SRR17333290	11	NP
60	P3390SPE	GSM5768475	SRR17333390	11	SPE	89	P3293NP	GSM5768571	SRR17333293	11	NP
61	S4611PE	GSM4669997	SRR12214611	11	PE	90	P3308NP	GSM5768556	SRR17333308	11	NP
62	P2994NP	GSM5768466	SRR17332994	11	NP	91	P3311NP	GSM5768553	SRR17333311	11	NP
63	P2995NP	GSM5768464	SRR17332995	11	NP	92	P3364NP	GSM5768500	SRR17333364	11	NP
64	P3031NP	GSM5768429	SRR17333031	11	NP	93	P3370NP	GSM5768494	SRR17333370	11	NP
65	P3041NP	GSM5768419	SRR17333041	11	NP	94	P3380NP	GSM5768484	SRR17333380	11	NP
66	P3046NP	GSM5768414	SRR17333046	11	NP	95	P3388NP	GSM5768476	SRR17333388	11	NP
67	P3056NP	GSM5768404	SRR17333056	11	NP	96	P3148PE	GSM5768716	SRR17333148	12	PE
68	P3082NP	GSM5768378	SRR17333082	11	NP	97	P3006SPE	GSM5768454	SRR17333006	12	SPE
69	P3085NP	GSM5768376	SRR17333085	11	NP	98	P3135SPE	GSM5768324	SRR17333135	12	SPE
70	P3124NP	GSM5768336	SRR17333124	11	NP	99	P3239SPE	GSM5768625	SRR17333239	12	SPE
71	P3142NP	GSM5768319	SRR17333142	11	NP	100	P3244SPE	GSM5768620	SRR17333244	12	SPE
72	P3186NP	GSM5768679	SRR17333186	11	NP	101	P3248SPE	GSM5768616	SRR17333248	12	SPE
73	P3187NP	GSM5768677	SRR17333187	11	NP	102	P3249SPE	GSM5768615	SRR17333249	12	SPE
74	P3194NP	GSM5768670	SRR17333194	11	NP	103	P3250SPE	GSM5768614	SRR17333250	12	SPE
75	P3203NP	GSM5768660	SRR17333203	11	NP	104	S4596PE	GSM4669982	SRR12214596	12	PE
76	P3211NP	GSM5768653	SRR17333211	11	NP	105	S4606PE	GSM4669992	SRR12214606	12	PE
77	P3214NP	GSM5768650	SRR17333214	11	NP	106	S4616PE	GSM4670002	SRR12214616	12	PE
78	P3225NP	GSM5768640	SRR17333225	11	NP	107	S4621PE	GSM4670007	SRR12214621	12	PE
79	P3235NP	GSM5768629	SRR17333235	11	NP	108	P2996NP	GSM5768463	SRR17332996	12	NP
80	P3237NP	GSM5768627	SRR17333237	11	NP	109	P2999NP	GSM5768461	SRR17332999	12	NP
81	P3243NP	GSM5768621	SRR17333243	11	NP	110	P3000NP	GSM5768459	SRR17333000	12	NP

No	Sample ID	GEO accession	SRA Run	Gestational Age	Condition	No	Sample ID	GEO accession	SRA Run	Gestational Age	Condition
82	P3252NP	GSM5768612	SRR17333252	11	NP	111	P3001NP	GSM5768460	SRR17333001	12	NP
83	P3255NP	GSM5768609	SRR17333255	11	NP	112	P3002NP	GSM5768458	SRR17333002	12	NP
84	P3264NP	GSM5768602	SRR17333264	11	NP	113	P3021NP	GSM5768439	SRR17333021	12	NP
85	P3265NP	GSM5768599	SRR17333265	11	NP	114	P3050NP	GSM5768411	SRR17333050	12	NP
86	P3279NP	GSM5768587	SRR17333279	11	NP	115	P3081NP	GSM5768379	SRR17333081	12	NP
87	P3281NP	GSM5768584	SRR17333281	11	NP	116	P3087NP	GSM5768373	SRR17333087	12	NP
117	P3110NP	GSM5768350	SRR17333110	12	NP	146	P3089NP	GSM5768371	SRR17333089	13	NP
118	P3120NP	GSM5768340	SRR17333120	12	NP	147	P3201NP	GSM5768663	SRR17333201	13	NP
119	P3176NP	GSM5768688	SRR17333176	12	NP	148	P3217NP	GSM5768647	SRR17333217	13	NP
120	P3182NP	GSM5768682	SRR17333182	12	NP	149	P3254NP	GSM5768610	SRR17333254	13	NP
121	P3193NP	GSM5768671	SRR17333193	12	NP	150	P3259NP	GSM5768604	SRR17333259	13	NP
122	P3238NP	GSM5768626	SRR17333238	12	NP	151	P3266NP	GSM5768598	SRR17333266	13	NP
123	P3251NP	GSM5768613	SRR17333251	12	NP	152	P3270NP	GSM5768594	SRR17333270	13	NP
124	P3253NP	GSM5768611	SRR17333253	12	NP	153	P3272NP	GSM5768592	SRR17333272	13	NP
125	P3261NP	GSM5768603	SRR17333261	12	NP	154	P3276NP	GSM5768588	SRR17333276	13	NP
126	P3273NP	GSM5768591	SRR17333273	12	NP	155	P3326NP	GSM5768537	SRR17333326	13	NP
127	P3280NP	GSM5768583	SRR17333280	12	NP	156	S4523NP	GSM4669909	SRR12214523	13	NP
128	P3287NP	GSM5768576	SRR17333287	12	NP	157	S4538NP	GSM4669924	SRR12214538	13	NP
129	S4528NP	GSM4669914	SRR12214528	12	NP	158	S4543NP	GSM4669929	SRR12214543	13	NP
130	P3204PE	GSM5768659	SRR17333204	13	PE	159	P3181SPE	GSM5768683	SRR17333181	14	SPE
131	P3208PE	GSM5768657	SRR17333208	13	PE	160	S4591PE	GSM4669977	SRR12214591	14	PE
132	P3284PE	GSM5768580	SRR17333284	13	PE	161	P3005NP	GSM5768455	SRR17333005	14	NP
133	P3152SPE	GSM5768711	SRR17333152	13	SPE	162	P3071NP	GSM5768389	SRR17333071	14	NP
134	P3161SPE	GSM5768705	SRR17333161	13	SPE	163	P3185NP	GSM5768678	SRR17333185	14	NP
135	P3192SPE	GSM5768672	SRR17333192	13	SPE	164	P3190NP	GSM5768674	SRR17333190	14	NP
136	P3205SPE	GSM5768661	SRR17333205	13	SPE	165	P3236NP	GSM5768628	SRR17333236	14	NP
137	P3207SPE	GSM5768656	SRR17333207	13	SPE	166	S4518NP	GSM4669904	SRR12214518	14	NP

No	Sample ID	GEO accession	SRA Run	Gestational Age	Condition	No	Sample ID	GEO accession	SRA Run	Gestational Age	Condition
138	P3242SPE	GSM5768622	SRR17333242	13	SPE	167	P3125PE	GSM5768335	SRR17333125	15	PE
139	P3245SPE	GSM5768619	SRR17333245	13	SPE	168	P3174PE	GSM5768689	SRR17333174	15	PE
140	P3297SPE	GSM5768568	SRR17333297	13	SPE	169	P3200PE	GSM5768664	SRR17333200	15	PE
141	S4601PE	GSM4669987	SRR12214601	13	PE	170	P3234PE	GSM5768631	SRR17333234	15	PE
142	P2998NP	GSM5768462	SRR17332998	13	NP	171	P3322PE	GSM5768542	SRR17333322	15	PE
143	P3007NP	GSM5768453	SRR17333007	13	NP	172	P3360SPE	GSM5768504	SRR17333360	15	SPE
144	P3011NP	GSM5768449	SRR17333011	13	NP	173	S4586PE	GSM4669972	SRR12214586	15	PE
145	P3065NP	GSM5768395	SRR17333065	13	NP	174	P3004NP	GSM5768456	SRR17333004	15	NP
175	P3080NP	GSM5768380	SRR17333080	15	NP	202	P3282PE	GSM5768581	SRR17333282	16	PE
176	P3084NP	GSM5768375	SRR17333084	15	NP	203	P3150SPE	GSM5768715	SRR17333150	16	SPE
177	P3093NP	GSM5768367	SRR17333093	15	NP	204	P3228SPE	GSM5768637	SRR17333228	16	SPE
178	P3105NP	GSM5768355	SRR17333105	15	NP	205	P3233SPE	GSM5768630	SRR17333233	16	SPE
179	P3119NP	GSM5768341	SRR17333119	15	NP	206	P3314SPE	GSM5768550	SRR17333314	16	SPE
180	P3158NP	GSM5768706	SRR17333158	15	NP	207	P3035NP	GSM5768425	SRR17333035	16	NP
181	P3196NP	GSM5768669	SRR17333196	15	NP	208	P3045NP	GSM5768415	SRR17333045	16	NP
182	P3224NP	GSM5768639	SRR17333224	15	NP	209	P3055NP	GSM5768405	SRR17333055	16	NP
183	P3257NP	GSM5768607	SRR17333257	15	NP	210	P3060NP	GSM5768399	SRR17333060	16	NP
184	P3286NP	GSM5768578	SRR17333286	15	NP	211	P3068NP	GSM5768392	SRR17333068	16	NP
185	P3289NP	GSM5768575	SRR17333289	15	NP	212	P3069NP	GSM5768391	SRR17333069	16	NP
186	P3303NP	GSM5768560	SRR17333303	15	NP	213	P3101NP	GSM5768359	SRR17333101	16	NP
187	P3310NP	GSM5768554	SRR17333310	15	NP	214	P3164NP	GSM5768700	SRR17333164	16	NP
188	P3328NP	GSM5768536	SRR17333328	15	NP	215	P3188NP	GSM5768675	SRR17333188	16	NP
189	P3329NP	GSM5768535	SRR17333329	15	NP	216	P3189NP	GSM5768676	SRR17333189	16	NP
190	P3335NP	GSM5768529	SRR17333335	15	NP	217	P3198NP	GSM5768666	SRR17333198	16	NP
191	P3342NP	GSM5768523	SRR17333342	15	NP	218	P3199NP	GSM5768665	SRR17333199	16	NP
192	P3346NP	GSM5768518	SRR17333346	15	NP	219	P3216NP	GSM5768648	SRR17333216	16	NP
193	P3349NP	GSM5768515	SRR17333349	15	NP	220	P3241NP	GSM5768623	SRR17333241	16	NP

No	Sample ID	GEO accession	SRA Run	Gestational Age	Condition	No	Sample ID	GEO accession	SRA Run	Gestational Age	Condition
194	P3366NP	GSM5768497	SRR17333366	15	NP	221	P3292NP	GSM5768572	SRR17333292	16	NP
195	P3369NP	GSM5768496	SRR17333369	15	NP	222	P3307NP	GSM5768557	SRR17333307	16	NP
196	S4513NP	GSM4669899	SRR12214513	15	NP	223	P3325NP	GSM5768539	SRR17333325	16	NP
197	P3077PE	GSM5768382	SRR17333077	16	PE	224	P3353NP	GSM5768511	SRR17333353	16	NP
198	P3127PE	GSM5768333	SRR17333127	16	PE	225	P3373NP	GSM5768491	SRR17333373	16	NP
199	P3180PE	GSM5768685	SRR17333180	16	PE	226	P3379NP	GSM5768485	SRR17333379	16	NP
200	P3183PE	GSM5768680	SRR17333183	16	PE	227	S4533NP	GSM4669919	SRR12214533	16	NP
201	P3262PE	GSM5768601	SRR17333262	16	PE	228	S4548NP	GSM4669934	SRR12214548	16	NP

Of the 233 available sequencing datasets, 228 were chosen for comprehensive genomic analyses, differential gene expression analyses, consistent gene expression analyses, and variant calling analyses (Table 10). These analyses were carried out for each gestational week, from week 8 to week 16.

Table 10: Breakdown of Data Per Gestational Week from GSE192902 and GSE154377

Gestational ages	GSE192902		GSE154377		Total
	NP	*PE*	*NP*	*PE*	
8 Weeks	12	4	0	0	**16**
9 Weeks	10	3	0	0	**13**
10 Weeks	15	4	0	0	**19**
11 Weeks	34	12	0	1	**47**
12 Weeks	21	8	1	4	**34**
13 Weeks	14	11	3	1	**29**
14 Weeks	5	1	1	1	**8**
15 Weeks	22	6	1	1	**30**
16 Weeks	20	10	2	0	**32**
Total	**153**	**59**	**8**	**8**	**228**

3.7 Data Transformation

This study used the SRAtoolkit (v3.0.7) to acquire the Sequence Read Archive (SRA) data. This toolkit, designed for the Linux platform, is a suite of programs and libraries optimised for handling large–scale genomic datasets. Once the required SRA datasets were successfully retrieved, the next task was to make them compatible with most conventional bioinformatics tools. Typically, these tools operate on data in the fastq format, a de facto standard for raw sequence data combined with quality scores. Although the SRA stores data in its proprietary format, the SRA toolkit allows utilities to convert it into the more universally

accepted fastq format. Leveraging the capabilities of the SRAtoolkit, the downloaded SRA data were converted into fastq files. This conversion facilitated further downstream analyses and ensured that data was ready for the subsequent quality assessment step in the research pipeline. Following data transformation, various tools were utilised at each stage of the bioinformatics pipelines. The landscape of bioinformatics is vast, with numerous algorithms and tools available for every step of the analysis, from quality control to final interpretation. Nonetheless, the choice of tools can significantly influence the outcome. Selecting the right tools is not just a matter of convenience or speed but scientific integrity. The wrong choice can lead to misinterpretations, inaccurate results, and false positives. Therefore, careful consideration and validation of each selected tool is crucial to enhance the accuracy of the results and minimise potential errors. As the field evolves, staying updated with the latest tools and best practices becomes even more essential to ensure rigorous and reliable analyses.

3.8 Rationale for Selecting Bioinformatics Tools for Pre–processing of Genomic Data

This study utilised cfRNA sequencing data sourced from two separate studies. Given the limited availability of cfRNA sequencing data in early pregnancy within transcriptomic databases, combining the gene expression count data from both sets became essential. By consolidating these datasets, the study aimed to provide a more comprehensive insight into early pregnancy dynamics. However, combining gene expression count data from two research studies can be complex due to the variations in library preparation methods, sequencing kits, and other algorithms and their respective versions in the RNA sequencing pipeline.

The method of sequencing used significantly impacts the outcome of RNA–Seq experiments. Depending on the sequencing platforms employed, both the quality of the data and the depth of biological comprehension could significantly vary (Kukurba and Montgomery, 2015). The cfRNA sequencing data chosen for this study were derived from Illumina sequencing technology. However, the specific

platforms differ: GSE192902 utilises the GPL20301 Illumina HiSeq 4000, while GSE154377 employs the GPL24676 Illumina NovaSeq 6000 (Table 8). While both platforms hail from Illumina and utilise comparable sequencing chemistries, subtle technical discrepancies might exist between them. Such nuances can introduce batch effects in the data.

Different alignment and quantification tools or even versions of the same tool can yield divergent results (Srivastava *et al.,* 2020). For instance, alignment tools like STAR, HISAT2, or TopHat introduce unique biases or systematic disparities in the data. This is because algorithms can be updated, optimised, or even fundamentally changed between versions, leading to potential discrepancies when comparing or merging data analysed using different iterations. Furthermore, inconsistencies can arise when different studies use varying gene annotations, reference genomes, or counting tools like HTSeq and featureCounts. The choice of normalisation method, whether Reads Per Kilobase of transcript per Million mapped read (RPKM), Fragments Per Kilobase of transcript per Million mapped read (FPKM), Transcripts Per Million (TPM), or Trimmed Mean of M-value (TMM), further complicates data integration as datasets normalised differently can misalign when combined. Moreover, differential gene expression analysis tools, such as DESeq2, edgeR, and limma–voom, employ distinct statistical models that can produce varied outcomes. This study ensured consistency using the same type and version of algorithms at every step, from the quality assessment of the fastq files to quantification. Additionally, batch effects in the combined gene expression count data were eliminated, followed by a thorough normalisation process.

3.9 Data Quality Control and Pre–processing

Next–generation sequencing (NGS) generates many short reads containing a large quantity of genomic data. It has a significant impact on biological and medical research as well as diagnostics (Fabbro *et al.,* 2013). However, this advanced NGS technology, on the other hand, can yield unanticipated artefacts (Alkhateeb and Rueda, 2017; Waszak *et al.,* 2014; Lavezzo *et al.,* 2016). Therefore, data Quality

Control (QC) and pre-processing of raw sequences are mandatory for further downstream analysis.

The first and most crucial phases in handling NGS data are data quality control and pre-processing, which provide the foundation for further downstream analysis (He *et al.,* 2020). In bioinformatics, Data QC refers to the procedures used to detect and rectify data errors and inconsistencies, ensuring optimal quality. This QC process encompasses several steps. These include raw data quality assessment, post-alignment quality evaluation, adapter trimming, filtering of low–quality reads, and duplicate removal. Each step is crucial to ensuring the reliability and accuracy of subsequent downstream analyses. Once data QC is performed, preprocessing steps transform the raw data into a more suitable form for downstream analyses. This step is more time–consuming from a computational standpoint than the other stages in the genomic analysis pipeline (Shahryary, Hazarika and Johannes, 2020).

3.9.1 Quality Control and Pre-processing of cfRNA Data

Modern high-throughput sequencing technologies can yield many sequence reads in a single experiment. However, not every sequencing technology is perfect, and each instrument will output a unique set of defects, such as wrong nucleotide bases being called. These erroneously called bases are due to the technical constraints of each sequencing technology. Therefore, it is necessary to understand, identify and exclude error types that may impact the interpretation of the downstream analysis. Sequence quality control is an essential first step in RNA–sequencing analysis. Early detection of errors saves time and aids in identifying the correct biomarkers for detecting disease conditions. Consequently, the quality of fastq files must be evaluated to ensure high–quality reads before further downstream analysis (Kukurba and Montgomery, 2015).

Presently, there are several tools available for assessing the quality of high throughput data, such as FastQC, FASTX–Toolkit, PRINSEQ, QC–Chain, and NGS QC Toolkit (Zhou, Q. *et al.,* 2018; Zhou, Q. *et al.,* 2013; Patel, R. K. and Jain, 2012; Fabbro *et al.,* 2013). Despite the availability of various quality control tools, FastQC (v0.11.8) was utilised in the Linux Operative System (OS) to assess the quality of each RNA-sequencing data before proceeding with downstream analysis. It intends to make quality control checks on raw fastq sequence data from high-throughput sequencing workflows as simple as possible. FastQC is a widely used open–source tool for quality control of raw NGS data, and it is a fast and efficient way to assess the quality of the raw data generated by the sequencing platform. It can be used to identify potential problems that can impact the downstream analysis. Moreover, it offers a modular set of studies that can quickly determine whether the data contains any concerns that need to be addressed before further investigation.

FastQC produces individual HTML reports for each FASTQ file, a feature that can be time–consuming when manually reviewing results from numerous files. This manual perusal can be labour–intensive and less efficient, particularly in large–scale studies. As a result, there is a pressing need for a tool that can aggregate these individual reports into a single, comprehensive file. Such aggregation would streamline the quality control process and offer a holistic view of the data, enabling rapid identification of trends and anomalies across samples. In this context, a tool that consolidates these outputs would save considerable time, ensuring efficient data processing and more immediate insights into potential sequencing issues.

MultiQC has become a key tool in the world of bioinformatics, expertly combining the outcomes of numerous bioinformatics investigations into a single, comprehensive report and facilitating the visualisation and interpretation of data across multiple samples (Ewels *et al.,* 2016). MultiQC (v1.11) was used in Linux OS to aggregate the FastQC results from numerous samples into a single

report. This consolidation by MultiQC expedites the quality control process and prominently highlights any inconsistencies or anomalies that might otherwise be overlooked. One of the notable features of MultiQC is its interactive interface, allowing users to delve into intricate metrics and gain a nuanced understanding of their sequencing quality and overall analytical outcomes. Adding to its versatility, the tool can be expanded with custom modules, ensuring compatibility with a broader range of bioinformatics tools, even those not supported initially. Given the ever–evolving landscape of high–throughput sequencing, MultiQC has become an indispensable asset, adeptly bridging individual tool outputs with holistic data interpretation. After thoroughly examining the FastQC and MultiQC reports, the raw sequencing reads underwent adapter removal using Fastp (v0.23.4). Subsequently, the alignment phase of the RNA sequencing pipeline commenced utilising the adapter–removed reads exhibiting the highest quality.

3.10 Alignment

The next crucial step in the RNA sequencing workflow is the alignment, which begins once high-quality data is obtained from the pre-processing step. The short–read sequences originating in the reference genome must be identified to determine which genes they belong to. The mapping of short FASTQ reads to a larger reference genome known as alignment. It is challenging and complicated for RNA sequencing due to a variety of reasons, including variants in the sequences, RNA splicing, sequencing errors, high mapping error rates, the length of short reads, larger reference genome, and the vast volume of short reads that need to be mapped (Trapnell and Salzberg, 2009). Moreover, it consumes more time to process and necessitates the creation of fast and precise alignment algorithms (Hatem *et al.,* 2013). However, numerous fast and splice–aware aligner algorithms are available for accurately and effectively aligning the reads (Grant *et al.,* 2011). On the other hand, traditional mapping algorithms, like Bowtie and BWA, are not advised for aligning RNA–Sequence reads to the reference genome due to their limitations in processing spliced transcripts (Kukurba and Montgomery, 2015). As

a result, high–quality sequencing reads were aligned to the human reference genome (hg38) using the STAR alignment software (v2.7.10a) on the Linux OS.

Spliced Transcripts Alignment to a Reference (STAR), an ultrafast universal RNA–sequencing aligner (v2.7.10a), was used to align the cfRNA short reads with the human reference genome (hg38). It is demonstrated to have better accuracy and performs more quickly when compared to other aligners by a factor of more than 50, and STAR was designed to align the non–contiguous sequences directly to the reference genome (Dobin *et al.,* 2013). Aligning reads using STAR is a two–step process – generating a reference genome index and mapping the reads to the reference genome.

3.10.1 Indexing the Reference Genome for Alignment

Numerous sequence reads are produced in high–throughput sequencing, although it is initially unclear where in the genome they originated specifically. The reference genome acts as an established benchmark in genomic studies. It represents a fully sequenced and annotated version of a genome, against which sequence reads from experimental samples can be compared and aligned. Researchers can accurately reassemble the many reads produced during high–throughput sequencing using this standard reference genome. This process helps determine the position and context of each read within the broader genomic landscape, facilitating a deeper understanding of the genetic information and its potential implications.

An aligner must fragment the reference genome into small chunks before aligning the short–read sequences onto the reference genome. Reference genome FASTA file is used as an input for generating genome indices (Dobin and Gingeras, 2015). This process must be carried out only once for a genome assembly. The retrieved genome indices can be saved in a folder and used repeatedly with the same reference genome and different samples.

3.10.2 Mapping the Reads with the Reference Genome

Mapping sequencing reads to a reference genome is integral to the foundation of genomic studies (Alser *et al.,* 2021). This alignment step acts as the anchor of most genomic analysis workflows. By aligning raw reads to a known template or reference, researchers can accurately determine the position, structure, and context of these reads within the genome. This information is invaluable for deciphering genetic variations, understanding gene expression patterns, or simply piecing together the genomic landscape of an organism. Without this foundational mapping process, it would be challenging to derive meaningful insights or accurately interpret the raw sequencing data. Therefore, the alignment ensures precision and sets the stage for subsequent in–depth analyses and discoveries in genomics.

This mapping process has two–pass stages, which are the detection of the novel junctions and insertion into the genome indices, and re–map all reads using an annotated GTF file and detected novel junctions (Dobin and Gingeras, 2015). In the first pass, the primary goal is the detection of novel junctions. This stage meticulously scans the sequencing reads to identify potential splicing events that have not been previously annotated in the reference genome. Once these novel junctions are detected, they are integrated into the genome indices, essentially updating or augmenting the reference with these newfound points of interest.

The second pass then utilises this enriched genome index for a more refined mapping. During this stage, not only are the original genome annotations (typically provided through a GTF file) considered, but the newly detected novel junctions from the first pass are also used. The alignment process becomes more thorough and accurate by re–mapping all the reads with the annotated GTF file and the detected novel junctions. This ensures that known and novel splicing events are captured, providing a more complete picture of the transcriptomic landscape. This two-stage approach thereby enhances the reliability and depth of the mapping results, paving the way for more insightful downstream analyses. After the

alignment process, determining the alignment's quality becomes crucial for improving the precision of subsequent genomic analysis.

3.11 Post–Alignment Quality Control

Post–alignment quality control becomes essential once this alignment is done to ensure the derived information's accuracy and reliability. It offers insights into the quality of sequenced samples. While raw data quality control checks provide an initial assessment, the post–alignment quality control often reveals nuances about potential issues with the samples. It's crucial for identifying samples that might have passed initial raw data checks but still contain inaccuracies or anomalies (Guo, Y. *et al.,* 2014). The significance of this step cannot be overemphasised, especially for genomic variant identification, gene expression analysis, and other genomic inquiries.

A crucial first step after alignment is to check the alignment rate. A high alignment rate indicates that the majority of reads originate from the reference genome or transcriptome, which ensures the reliability of the data (Conesa *et al.,* 2016). Reads that do not map or map multiple locations can be discarded or further analysed to understand their origins. The distribution of reads across genomic features, such as exons, introns, and intergenic regions, is another valuable metric. An excess of intronic or intergenic alignments can hint at DNA contamination or issues with the RNA quality (Martin, 2011). Additionally, examining the distribution of reads along transcripts provides insights into potential 3' bias, often stemming from the library preparation method (Oshlack and Wakefield, 2009). Uniform coverage is usually ideal, although specific library prep methods can inherently introduce biases. Strand–specificity is another factor. For strand–specific RNA–seq protocols, an unexpected off–strand mapping indicates potential issues in the library preparation (Levin *et al.,* 2010).

Several tools, including QualiMap, RseQC, RNA–seQC, SAMtools, and BAMtools, have been developed to help researchers evaluate the quality of alignment data. These tools validate the alignment process and ensure that downstream analyses

are based on reliable, high–quality data. Tools like SAMtools (Li, H. *et al.,* 2009a) and Picard provide insights into the number of mapped reads, mapping quality, and coverage depth. Anomalies in these metrics may indicate problems with the alignment process or the quality of the input data. Another critical consideration is detecting and evaluating potential biases in the data. Batch effects, sequence–specific biases, and GC content bias can impact the accuracy of expression quantification. Tools like RSeQC (Wang, Liguo, Wang and Li, 2012) and RNA–SeQC (DeLuca *et al.,* 2012) enable the identification of these biases, allowing researchers to apply appropriate corrections and improve the reliability of downstream analyses. Moreover, post–alignment quality control encompasses assessing transcriptome coverage and identifying outliers. This involves examining the uniformity of read distribution across genes and evaluating the representation of known transcripts. Tools like Qualimap (García-Alcalde *et al.,* 2012) provide detailed reports on coverage uniformity and aid in identifying potential outliers that may require further investigation.

Among these tools, Samtools (v1.6) is used to perform post–alignment quality control for the cfRNA sequencing data. It is widely used and the most appropriate for post–alignment quality assessment as it provides a range of functionalities, such as viewing, sorting, filtering, merging, and indexing (Bao *et al.,* 2014; Li, H. *et al.,* 2009b). Additionally, the flagstat function in Samtools provides a basic summary of the alignment, including the percentage of mapped reads.

After the post–alignment quality assessment of the BAM files produced from the alignment process, they are utilised in the gene expression analysis and variant calling pipelines. In the gene expression analysis pipeline, BAM files are employed in the quantification process to generate gene expression count files. Likewise, BAM files are used in the variant calling pipeline to generate Variant Call Format (VCF) files to identify variant signatures.

Chapter 4

Identification of Pre-eclampsia Associated Unique Differentially Expressed Gene Signatures at Each Gestational Week in Early Pregnancy

This chapter delves into identifying the Pre-eclamptic Unique Differentially Expressed Gene Signatures (UDEGs) at each gestational week of early pregnancy, utilising cfRNA sequencing data. The exploration extends to performing pathway and network analyses to enhance understanding of the pathophysiology of PE. To culminate, the chapter focuses on developing AI-assisted prediction models to identify women at risk of PE using gestational week–specific differential gene expression signatures (SDEGs) derived from comprehensive differential gene expression analyses. This approach integrates molecular insights, computational methodologies, and predictive modelling to contribute to the understanding and management of PE.

4.1 Introduction

Gene expression signatures, which are distinct patterns of gene activity, are crucial in predicting and determining the prognosis of diseases. These signatures can give insight into the unique biological responses associated with diseases, thereby improving diagnostic accuracy and allowing for personalised treatment plans (Lamb *et al.,* 2006; Gliddon *et al.,* 2018; Wang, Susan Siyu, Lewis and Pitzalis, 2023). For instance, in cancer research, gene expression profiles have proven helpful for predicting disease progression, response to therapy, and overall survival rate (Sotiriou and Pusztai, 2009). Similarly, in cardiovascular disease, specific gene signatures can help anticipate disease progression and guide preventative strategies (Wingrove *et al.,* 2008). Thus, these distinctive molecular 'fingerprints' open new avenues for disease management and personalised medicine, showing promising potential to revolutionise healthcare. In the era of precision medicine, there is a growing interest in identifying molecular signatures

that can predict disease susceptibility, severity, and response to treatment. Gene expression profiling has been employed in various diseases to identify such biomarkers. Differential gene expression analysis provides insight into the pathogenesis of diseases, potentially revealing novel therapeutic targets and enabling the development of predictive models (Hoque *et al.,* 2022).

The application of advanced genomic techniques to the field of obstetrics has significantly expanded understanding of the molecular mechanisms underlying various pregnancy complications (Reilly *et al.,* 2023; Guo, Z. *et al.,* 2020). One promising technique is differential gene expression analysis, which enables the identification of unique gene expression profiles associated with certain conditions. This enables earlier detection and more accurate prediction.

4.1.1 Understanding Gene Expression

Gene expression, the process of transforming the information within a gene into a functional product such as a protein or RNA molecule, is essential for all living organisms and involves complex multi-step biological mechanisms utilising genetical information (Armingol *et al.,* 2021; Volgin, 2014). Gene expression mainly contains transcription, translation, and Post–Translational Modification (PTM) steps in which DNA is converted to RNA and then to functional protein, respectively.

Transcription is the first step in gene expression, and the genetic information in a gene in DNA is transcribed to produce the complementary RNA molecule. This RNA molecule, known as mRNA, carries the genetic codes for protein synthesis from the DNA to the ribosomes in the cytoplasm. The transcription process can be divided into three stages: initiation, elongation, and termination (Porrua and Libri, 2015; Uptain, Kane and Chamberlin, 1997). Transcriptional initiation is initiated by the RNA polymerase enzyme binding to the promoter region of DNA. After that, it starts synthesising a new RNA strand from a DNA template. Before being ready for translation, it undergoes additional processing steps like capping, splicing, and

polyadenylation. Transcription is tightly controlled and governed by transcription factors, enhancers, and repressors. Moreover, it is an essential process for the stringent multi–level control of gene expression in different cell types and under varying conditions (Guo, J., 2014).

The encoded genetic information in the mRNA molecule used to produce the protein is called translation and occurs in the ribosomes. The translation process starts with the binding of the small ribosomal subunit to the mRNA molecule for forming an initiation complex, and it is facilitated by initiation factors, such as eukaryotic initiation factors (eIFs) (Brito Querido *et al.,* 2024; Hinnebusch and Lorsch, 2012). Then, the small ribosomal subunit scans the mRNA molecule to locate the start codon (AUG) and signal the beginning of the translation process. Next to the initiation process, the large ribosomal subunit combines with the initiation complex to form the functional ribosome. Then, transfer RNA (tRNA) molecules carrying amino acids are transferred and bind to the mRNA codon based on complementary base pairs. During the process of base pairing on the binding sites, Aminoacyl (A), Peptidyl (P), and Exit (E) sites play a crucial role. These sites, A, P and E, represent the ribosome's first, second and third tRNA binding locations (Blanchard *et al.,* 2004). At the A site, peptide bonds develop between the amino acids the closest tRNAs carry when the proper tRNA binds with mRNA molecules. The peptidyl transferase enzyme, a ribozyme, a type of RNA molecule with catalytic activity, facilitates this process. Then, the ribosome moves along the mRNA, shifting one codon forward with each step. During this movement, it exchanges tRNA molecules to ensure the accurate addition of amino acids to the developing polypeptide. As the process unfolds, vacant tRNA molecules are cleared from the E site. This sequential cycle repeats, facilitated by elongation factors until a stop codon is detected on the mRNA. Termination factors recognise this stop signal, leading to the ribosome's release of the finalised polypeptide chain and subsequent disassembly of the ribosomal complex (Hellen, 2018).

TRNA molecules carrying specific amino acids attach to the mRNA codons by matching bases, helping build acids into a polypeptide chain during translation. This process consists of three stages: initiation, elongation, and termination. In the initiation phase, the small ribosomal subunit binds to the mRNA molecule, and the initiator tRNA carrying methionine binds to the start codon (AUG) (Hershey, Sonenberg and Mathews, 2012). Then, the large ribosomal subunit joins the complex, forming the functional ribosome. Elongation involves the addition of amino acids one after another to the growing polypeptide chain, guided by the codons on the mRNA and complementary anticodons on the tRNA molecules (Spencer and Groudine, 1990; Reines, Conaway and Conaway, 1999). Finally, termination occurs when a stop codon (UAA, UAG, or UGA) is encountered, leading to the release of the completed polypeptide chain from the ribosome (Proudfoot, 2016).

Post–translational modifications (PTMs) are chemical alterations that impact the properties of proteins, whether through proteolytic cleavage or the attachment of modifying groups like methyl, glycosyl, phosphoryl, or acetyl, to one or more amino acids (Ramazi and Zahiri, 2021; Ramazi, Allahverdi and Zahiri, 2020). These alterations often involve phosphorylation, glycosylation, acetylation, methylation, and ubiquitination (Sharma *et al.,* 2019). These modifications wield significant influence over various aspects of protein structure and function. Firstly, they can directly impact protein functionality by activating or deactivating enzymatic activity, often achieved through conformational changes or altered interactions with other molecules (Kokkinidis, Glykos and Fadouloglou, 2020). Additionally, PTMs are key in regulating Protein–Protein Interactions (PPI), influencing essential cellular processes like signal transduction and protein complex assembly. Furthermore, they can dictate the subcellular localisation of proteins, guiding them to specific compartments within the cell or earmarking them for secretion. PTMs also control protein stability, shielding proteins from degradation or earmarking them for breakdown via cellular degradation pathways. Moreover, these modifications can

induce structural alterations in proteins, affecting their folding, stability, and overall conformation (Zhong *et al.,* 2023; Tikhonov *et al.,* 2021). Lastly, PTMs often serve as critical signalling events, orchestrating cellular responses to internal and external stimuli. PTMs are indispensable regulators of protein function, cellular signalling, and the myriad physiological processes underpinning cellular function.

4.1.2 Changes in cfRNA Gene Expression Across Gestational Weeks in Pre-eclampsia

Research findings indicate a strong correlation between alterations in PE-associated gene expression and gestational age. In particular, gene expression levels have been shown to change significantly during gestation, influencing gene expression patterns in both healthy and preeclamptic pregnancies. Several studies have explored the complex relationship between changes in gene expression in cfRNA and the occurrence of PE (Gong et al., 2021; Zhou, S. et al., 2023; Tsang et al., 2017; Jang et al., 2017; Rasmussen et al., 2022; Moufarrej et al., 2022). However, there remains a notable gap in research focused on a per–gestational–week approach to identifying gene expression signatures. This approach, which involves examining gene expression patterns in cfRNA across different gestational weeks, offers a more nuanced understanding of the temporal dynamics of gene regulation during early pregnancy and its relation to PE onset and progression. Furthermore, comprehending the evolution of gene expression changes across gestational weeks may provide valuable insights into the underlying pathophysiology of PE. This understanding can inform the development of early intervention strategies and more effective risk management approaches. Through the analysis of cfRNA sequencing data obtained at different stages of the early prenatal period, scientists may follow the progression of changes in gene expression and possibly pinpoint markers specific to a particular gestational age linked to PE. These markers have the potential to be utilised in non–invasive diagnostic tools for identifying women at risk of PE during early pregnancy (Zhou, S. *et al.,* 2023).

Gene expression changes in PE may indeed occur progressively throughout pregnancy. However, identifying gene expression changes in early pregnancy holds significant potential for detecting the disease condition before complications arise. Moufarrej et al. (2022) identified that gene expression changes in cfRNA seq data of PE show a peak in early gestation before 20 weeks; however, they did not specifically report which week shows the most significant changes in gene expression. Another study by Jang et al (2017) also discovered that changes in global gene expression in the amniotic fluid cell-free transcriptome begin at 19 gestational weeks. Moreover, Poon, Leona C. and Nicolaides, 2014 suggest that the most excellent time to reliably predict PE is between weeks 11 and 13 of gestation (Poon, Leona C. and Nicolaides, 2014). As a preventive strategy, low-dose aspirin could be started before 16 weeks of gestation if this early prediction window leads to a decrease in the prevalence of PE (Roberge, Giguère *et al.,* 2012; Roberge, Villa *et al.,* 2012; Bujold *et al.,* 2010). Therefore, identifying the specific gestational week that exhibits significant gene expression changes in early pregnancy cfRNA sequencing data could facilitate the early management of PE.

4.1.3 Differential Gene Expression Analysis and Pre-eclampsia Prediction

High-throughput RNA sequencing technologies significantly impact understanding of human diseases by identifying genomic signatures that expand knowledge of molecular processes in biological and biomedical sciences (Li, D. *et al.,* 2022). Differential Gene Expression Analysis (DGEA), a valuable technique that has become essential to many biological study domains, including developmental biology, neurobiology, cancer research, and, more recently, personalised medicine, is executed using this RNA sequencing data (Han, Y. *et al.,* 2015). This analytical approach is built on the idea that various cells, tissues or situations might exhibit gene activity patterns, offering insights into their underlying biology. Essentially, gene expression is the mechanism through which genetic information is utilised to generate an output, a protein. Nevertheless, not every gene in a cell is actively expressed simultaneously. Depending on the cell type, status or

exposure to factors, distinct groups of genes will be activated or deactivated. This specificity serves as the foundation of gene expression.

DGEA involves comparing gene expression levels between two or more biological samples under different conditions. The main aim is to identify Differentially expressed genes (DEGs) that show significant changes in expression levels between these conditions (Costa-Silva, Domingues and Lopes, 2017). The techniques utilised for DGEA have indeed evolved dramatically alongside technological advancements. Moreover, in the era of AI, DGEA is increasingly performed using advanced ML techniques to identify candidate features for developing disease prediction models.

Identifying gene expression patterns linked to PE marks substantial advancement, yet effectively applying these discoveries in clinical settings necessitates the creation of dependable predictive disease models. Multiple research studies have integrated data on differential gene expression with ML techniques to construct disease prediction models for identifying PE–risk women (Huang et al., 2022; Li, X. et al., 2022; Wang, Hao et al., 2023; Bai et al., 2023; Moufarrej et al., 2022). The resultant model, trained on gene expression profiles, exhibited considerable accuracy in forecasting PE, highlighting the promise of amalgamating genomic data and ML for disease prediction. In an independent investigation, Moufarrej et al (2022) and their team employed ML algorithms to develop a predictive model using cfRNA sequencing data. This model achieved a notable level of accuracy in predicting PE, further emphasising the potential of cfRNA as a non–invasive biomarker for early disease detection. Moreover, Wang, Hao et al. (2023) employed transcriptomic data in conjunction with integrated ML frameworks, demonstrating enhanced accuracy in predicting PE. These findings also underscore the potential of integrating genomic data with ML to improve the accuracy and usefulness of disease prediction models.

4.1.4 Importance of DEGs and UDEGs in Pre-eclampsia Prediction

The DEGs identified during gestational weeks less than 20 suggest associations with the onset of PE during early pregnancy stages (Moufarrej et al., 2022; Ogoyama et al., 2022). These signatures encompass genes involved in various biological processes, including angiogenesis, inflammation, oxidative stress, and placental development (Leavey *et al.,* 2016). For instance, Nishizawa et al.'s (2007) study demonstrates that DEGs are associated with some essential pathways that could be involved in PE, such as the TGF–β and CEBPA–related pathways. In another investigation, Sitras et al. (2009) showed that many DEGs were associated with Alzheimer's disease, angiogenesis, and the Notch, TGFβ, and VEGF signalling pathways. A recent study by Huang et al. (2022) further demonstrated that DEGs are connected with gonadotropin secretion, immune cell infiltration, and the SMAD and MAPK pathways associated with PE. Hence, identifying DEGs during each gestational week could provide valuable insights into the progression of PE development in early pregnancy. Furthermore, leveraging these DEGs in ML algorithms to construct PE prediction models could significantly improve accuracy. Moreover, such an approach would facilitate the identification of the DEGs, which is crucial for developing diagnostic tools to detect women at risk of developing PE.

Comprehending the fluctuation in gene expression throughout GW and its correlation with PE is paramount for uncovering the intricate pathophysiological mechanisms of this disorder. Recent studies indicate that distinctive gene expression patterns at various gestational weeks could offer valuable insights into the initial molecular processes that predispose individuals to PE (Moufarrej et al., 2022; Huang et al., 2022; Li, X. et al., 2022; Lian et al., 2013). Consequently, identifying UDEGs at each gestational week is crucial for broadening understanding of the progression of PE during gestation. UDEGs refer to those genes that exhibit differential expression, specifically in one GW but not in any other GWs. These genes are uniquely differentially expressed in that particular GW and do not overlap with the set of DEGs from other GWs. Furthermore, it could

aid in pinpointing the specific gestational week for non–invasive specimen collection to detect women at risk of developing PE.

4.1.5 Research Hypothesis

The expression of UDEGs at each GW provides valuable biomarkers for the early prediction of PE, enabling improved risk assessment and preventive interventions.

4.1.6 Research Question

Can GW-specific unique differential gene expression signatures (UDEGs) in cfRNA be identified and characterised as candidate biomarkers for PE prediction?

4.1.7 Research Aim and Objectives

This study aimed to identify and characterise gestational week–specific DEGs and UDEGs in early pregnancy and integrate them into a predictive model for PE. To achieve this aim, the following objectives were established:

1. Identify DEGs and UDEGs for each GW from 8 to 16 using cfRNA sequencing data.
2. To understand biological processes and signalling pathways linked to PE, perform pathway enrichment analysis.
3. Conduct protein–protein interaction (PPI) analysis to identify key regulatory networks.
4. Compare identified gene signatures with previously known PE–linked genes to assess their novelty and relevance.

4.2 Method

Chapter 3, titled 'General Methods', provides a detailed outline of the processes for data collection, quality assessment, and pre–processing. This section further elaborates on the methodologies for quantifying aligned gene expression data, its pre–processing, differential gene expression analysis, and subsequent downstream analyses (Figure 4).

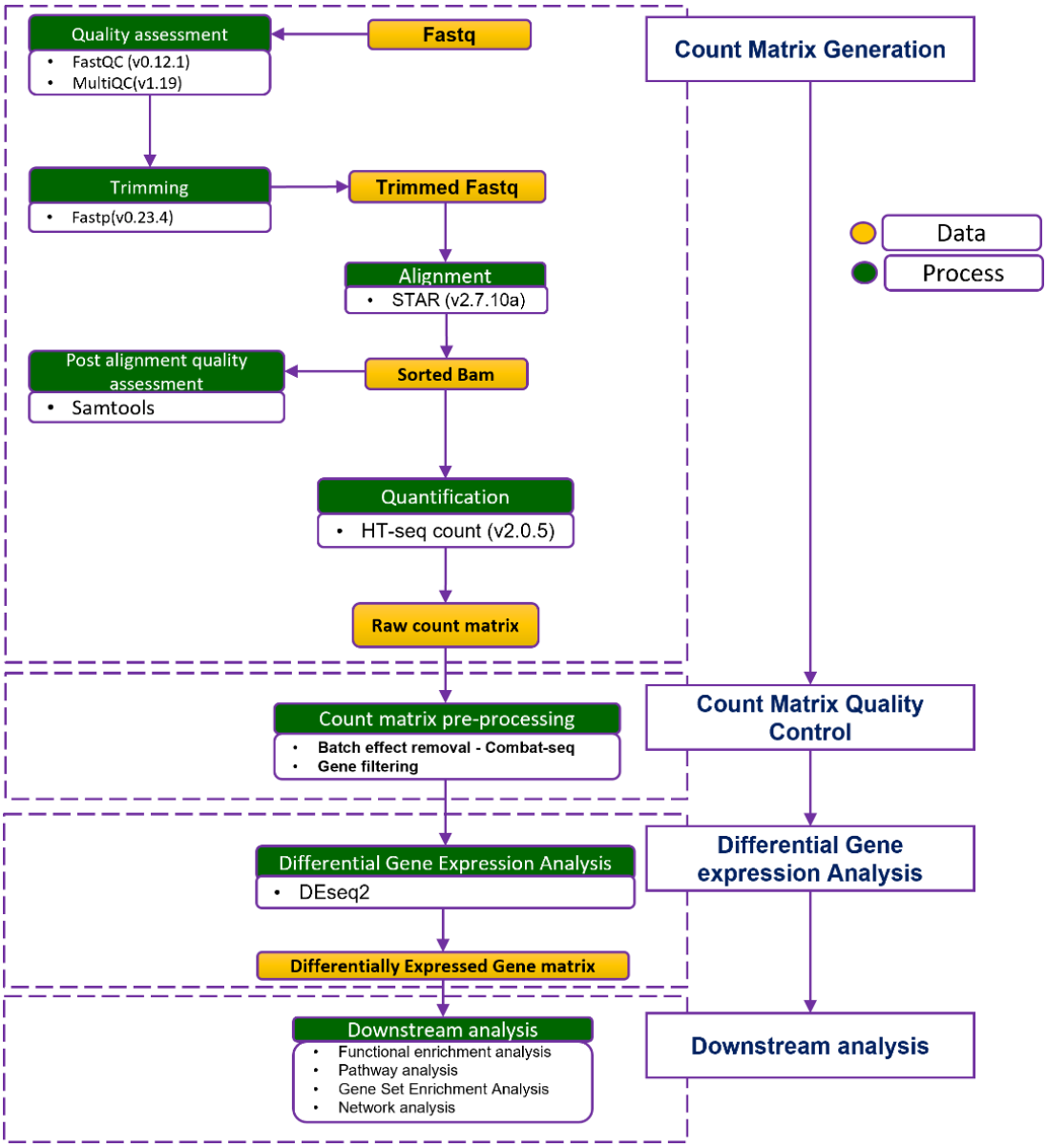


Figure 4: Workflow for differential gene expression analysis for each gestational week.

4.2.1 Quantification

Quantification is quantifying the gene expression by RNA–seq by counting the number of reads that align to each gene using programs such as FeatureCounts, HTseq–count, Salmon, and Kallisto. Among these, HT–seq count (v2.05), a Python–based tool, was used for counting mapped reads and producing the gene expression matrix, representing the gene expression level for multiple genes

across multiple samples. It is a well–established and widely used tool, making it compatible with many RNA–seq analysis workflows (Anders, Pyl and Huber, 2015). Additionally, it provides a robust and accurate method for counting reads that align with annotated genomic features. It relies on genomic annotations (like GTF files) to assign reads to features. This annotation–based approach enhances the specificity and accuracy of gene expression quantification.

4.2.2 Pre–processing of Gene Expression Count Data

The gene expression count data generated for this study requires preprocessing before progressing to downstream analyses, aiming to identify transcriptomic signatures. This step is critical because it integrates data from two distinct studies. The pre–processing procedure involves several essential steps, each tailored to enhance the quality and reliability of the gene expression data for each gestational week. These steps include removing batch effects, gene filtering, and normalisation applied to the respective gene expression matrices for each gestational week. This meticulous pre-processing aimed to eliminate unwanted sources of variation introduced during experimental procedures, refine the dataset by focusing on informative genes, and ensure that gene expression values are comparable across samples and conditions. The sequential application of these steps contributes to the robustness and accuracy of downstream analyses, providing a solid foundation for uncovering meaningful biological insights specific to each gestational week.

4.2.2.1 Batch Effect Removal.

In RNA sequencing, the term "batch" pertains to a grouping of samples that undergo processing or sequencing collectively within a singular experimental run. Batches become part of the experimental workflow at multiple stages, encompassing sample preparation, library construction, and sequencing. Including batches in RNA–seq datasets can introduce technical variations unrelated to biological distinctions. This technical variability, often known as batch effects, is a

systematic and non–biological source of variation introduced during the experimental process that can impact the measurements of high–throughput sequencing data (Yi *et al.,* 2018; Espín-Pérez *et al.,* 2018). Batch effects emerge from dissimilarities among samples not inherent to the planned experimental design. These variations can arise from multiple sources, including differences introduced by various handlers, variations in experiment locations, utilisation of distinct batches of reagents, and even influences from biological factors like growth locations. If left unaddressed, these non–biological sources of variation can introduce systematic biases that may obscure genuine biological signals within the dataset.

Researchers typically employ normalisation methods, statistical techniques, and careful experimental design to mitigate the impact of batch effects and enhance the accuracy and interpretability of their findings. Nonetheless, normalisation methods and procedures often fail to reduce batch effects ultimately. Normalisation techniques, such as transforming raw counts into logarithmic values of Counts Per Million (CPM), Transcripts Per Million (TPM), or Reads Per Kilobase per Million mapped reads (RPKM/FPKM), as well as methods like the Trimmed Mean of M values (TMM) or Relative Log Expression (RLE), can effectively address differences in the overall expression distribution of each sample across batches, they may not entirely correct for batch effects in composition (Zhang et al., 2020).

Several methods have been developed to address this issue in response to technical bias in gene expression data. ComBat stands out as a frequently employed method, especially in scenarios where the number of batches is explicitly known (Zhang, Yuqing, Parmigiani and Johnson, 2020). Methods like SVASeq and RUVSeq are commonly employed for unknown batches. However, traditional differential expression programs necessitate raw, untransformed counts and recommend the inclusion of batch variables as covariates. This limitation impedes conventional methods like edgeR and DESeq2, as these tools require untransformed data. A breakthrough in 2020 led to the development of ComBat–seq, a tool derived from the original ComBat method but explicitly tailored for RNA–

Seq data. ComBat–seq outputs adjusted counts, enabling their use in downstream analyses, including differential expression analysis. This innovation provides a solution for incorporating batch information into analyses without sacrificing the utility of widely used differential expression programs. This research study applied the ComBat–seq function in the SVA R package to mitigate batch effects in the combined gene expression data (see *Appendix A: Differential Gene Expression Analysis* for details on Batch effect removal).

4.2.2.2 Gene Filtering

Gene filtering during differential gene expression analysis involves selecting a subset of genes from the complete set measured in high–throughput genomic experiments like microarrays or RNA–seq data. The primary objective of filtering lowly expressed genes is to concentrate on a more manageable and biologically pertinent set of genes, enhancing the precision and interpretability of the outcomes.

In this research study, gene filtering is performed using the DESeq2 R package based on a standardised criterion that targets lowly expressed genes across all samples. DESeq2 applies an independent filtering process to remove genes with low mean normalised counts, ensuring that only genes with sufficient expression levels are retained for downstream analysis. This approach helps to refine the dataset by eliminating genes with high variability and low statistical power, thereby enhancing the robustness and reliability of the differential expression analysis. This study's filtering strategy aligns with best RNA-seq analysis practices, contributing to improved detection of biologically meaningful differentially expressed genes.

4.2.2.3 Normalisation

Normalisation in the context of data analysis refers to the process of adjusting or scaling data to account for various factors that could introduce biases or variations. Normalisation aims to make data comparable across different samples or conditions, especially when systematic differences arise due to technical artefacts

or experimental procedures . This is a crucial step in various fields, including genomics, transcriptomics, and other high–throughput technologies. Many types of normalisation methods, such as RPKM, TPM, TMM, RLE, Quantile normalisation, Cyclic Loess normalisation, and Total Count Normalisation, are available for RNA sequencing, each designed to address specific challenges and biases inherent in high–throughput sequencing technologies. (Cole *et al.,* 2019; Feng, Y. and Li, 2021). Selecting a normalisation method is paramount in guaranteeing the precision and dependability of subsequent analyses among the array of available normalisation methods. Researchers often navigate through the myriad of normalisation options, considering the characteristics of their data, experimental design, and specific analysis goals to judiciously select the most appropriate normalisation method for their RNA–seq dataset. By carefully choosing and implementing these normalisation techniques, researchers enhance the precision and reliability of their gene expression analyses, uncovering meaningful biological insights.

The TMM normalisation method is highly beneficial for RNA–seq data analysis because it addresses several key challenges inherent to this type of high–throughput sequencing. One of its primary advantages is its adept handling of differences in library sizes, a common occurrence in RNA–seq experiments, where variations in sequencing depth can lead to disparities in the number of reads obtained for each sample (Wang, Beibei and Luan, 2023; Evans, Hardin and Stoebel, 2018). TMM normalisation mitigates this effect, ensuring that gene expression values are comparable across samples. Moreover, TMM is robust to outliers, employing a trimmed mean approach that disregards extreme log–fold changes, enhancing its resilience to highly variable expression levels in specific genes. Importantly, TMM normalisation tackles composition bias, a phenomenon often observed in RNA–seq data, where the relative proportions of different transcripts can vary between samples. By accounting for these factors, TMM preserves the biological variability in gene expression, which is crucial for accurate downstream analyses such as differential gene expression. Widely implemented

in popular R packages like DESeq2 and edgeR, TMM normalisation has become a standard and trusted approach, acknowledging the unique characteristics of RNA–seq data and contributing to the reliability of results in transcriptomic studies.

4.2.3 Differential Gene Expression Analysis

DGEA is a crucial step in deciphering the molecular mechanisms underlying biological processes, and a multitude of bioinformatics tools and packages have been developed to facilitate this analysis. These packages are designed to handle high–throughput sequencing data, particularly from technologies like RNA sequencing (RNA–seq), providing researchers with diverse options to suit their experimental designs and preferences. One widely used package is DESeq2, a R/Bioconductor package that employs a negative binomial distribution model for differential expression analysis. DESeq2 is known for its robustness in handling various experimental designs, including complex designs with multiple conditions and replicates, and it incorporates size factor normalisation and advanced statistical methods for accurate results (Stupnikov *et al.,* 2021; Love, Huber and Anders, 2014). Another prominent R package is edgeR, tailored explicitly for RNA–seq data analysis. edgeR implements empirical Bayes methods and utilises the negative binomial distribution for modelling count data, offering statistical power and flexibility in the presence of biological variability and low replicate numbers (Robinson, McCarthy and Smyth, 2010).

The DESeq2 package was employed to conduct differential gene expression analysis on gene expression matrices corresponding to each gestational week (see Appendix A: Differential Gene Expression Analysis). After that, the identified set of significant DEGs was utilised in downstream analyses to unravel their functional implications and biological significance.

4.2.4 Overlap Analysis for Identifying Unique Gene Expression Signatures.

Differentially expressed genes during each gestational week were employed in the overlap analysis to identify genes unique to each week of gestation. Exploring the

identification of unique genes associated with each gestational week in the context of PE represents a promising strategy for developing a predictive model for this complex pregnancy–related disorder. PE exhibits dynamic changes throughout pregnancy, and by focusing on gene expression patterns specific to individual gestational weeks, this study aimed to uncover temporal molecular signatures indicative of distinct stages in the development of the disorder. This approach facilitates the discovery of potential biomarkers and provides valuable insights into the underlying pathophysiological mechanisms, allowing for a more nuanced understanding. The ultimate goal is to contribute to developing a predictive model aligned with the principles of precision medicine, tailoring risk assessments and interventions to the specific characteristics of individual pregnancies. Integrating molecular information with relevant clinical data enhances the model's robustness, offering the potential for early identification and targeted interventions, thereby advancing the ability to mitigate the severity or prevent the onset of PE.

4.2.5 Downstream Analysis

Downstream analyses encompassed a variety of bioinformatics and statistical approaches, such as gene ontology annotation, pathway enrichment analysis, and network–based analyses. The objective was to comprehensively understand the molecular processes and pathways in which the identified DEGs are involved. Additionally, the DEGs set served as a foundation for constructing interaction networks and deciphering potential regulatory mechanisms. Integrative analyses allowed for extracting meaningful biological insights, aiding in interpreting the observed gene expression changes and contributing to a more thorough characterisation of the underlying biological phenomena associated with the experimental conditions under investigation.

4.2.5.1 Gene Ontology Analysis

Gene Ontology (GO) is a crucial resource in bioinformatics, offering a standardised vocabulary for describing the functions of genes across diverse organisms. The

ClusterProfiler package was utilised to conduct GO analysis for the DEGs identified in each gestational week. This package in R is designed explicitly for comparing biological themes among gene clusters (Yu *et al.,* 2012). This approach allows for the identification of Biological Processes (BP), Molecular Functions (MF), and Cellular Components (CC) associated with the genes that are differentially expressed during specific stages of gestation. Additionally, GO analysis aids in comprehending the functional significance of genes that exhibit differential expression across various gestational weeks. By identifying enriched GO terms, researchers can glean insights into the biological mechanisms driving the developmental processes at each stage of pregnancy. This information holds significance for understanding typical foetal development and recognising any anomalies or deviations that might manifest in different pregnancy–related conditions.

DEGs and UDEGs were identified within each gestational week, and all DEGs and UDEGs in early pregnancy were analysed using GO analysis. Significantly enriched GO terms, determined based on a p–value threshold of <0.05, were subsequently analysed to gain insights into the functional characteristics associated with these gene sets. This approach allows for exploring the biological processes, molecular functions, and cellular components overrepresented within the differentially expressed gene sets, providing valuable information about the underlying biological mechanisms driving the observed changes. By examining the enriched GO terms, this study elucidated the functional roles of the genes involved and gained a deeper understanding of the biological pathways and processes relevant to the studied condition.

4.2.5.2 Pathway Analysis

The pathway analysis was conducted to elucidate the molecular pathways associated with the DEGs identified in GW 8, 10, 13, 15 and 16. The enrichR package (v3.2) is used in R for the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. Enriched terms and pathways were identified based on

statistical significance, with a p–value threshold of <0.05 using Benjamini–Hochberg correction for multiple tests. Subsequently, pathway analysis was conducted to assess the involvement of DEGs and UDEGs in specific biological pathways and signalling cascades.

4.2.5.3 Protein–Protein Interactions

DEGs and UDEGs were further analysed at the protein level using Protein–Protein Interaction (PPI) network functional enrichment analysis through the Search Tool for the Retrieval of Interacting Genes (STRING (v12.0)) database (https://string-db.org/) (Szklarczyk et al., 2023). It is a database for searching interactions of known and predicted proteins, which was used to study the molecular mechanism of disease. The interaction between proteins with a confident score cutoff >0.4 was considered statistically significant. Results from STRING were further analysed using Cytoscape (v3.10.1) (https://cytoscape.org/) to visualise molecular interaction networks and integrate gene expression profiles. This can be useful in identifying clusters of protein interaction that are highly related to PE. The Molecular Complex Detection (MCODE; version 1.4.2) plugin in Cytoscape was then utilised to screen the top three modules in the PPI network (used for clustering a given network based on topology to find densely connected regions). Parameters were set as follows: Degree cutoff, 2; node score cutoff, 0.2; k–core, 2; and maximum depth, 100. The list of genes generated in the cluster was then analysed again in enrichR for significantly enriched ontology terms. This comprehensive approach helps elucidate the molecular pathways and biological processes associated with PE.

4.3 Results

The gene expression count data generated from HTSeq count includes 60,705 coding and non–coding genes. Before conducting downstream analysis, it is essential to correct for batch effects, which may arise due to technical variations across experimental batches. To mitigate these variations, the Combat–seq

function from the sva R package was applied. This function, which implements the ComBat algorithm, effectively adjusts for batch effects, improving the accuracy and reliability of the dataset. After batch correction, the normalised and gene–filtered expression count data were prepared for further analysis.

This study performed differential gene expression analysis (DGEA) separately for each gestational week (GW) from 8 to 16, treating each dataset as an independent entity. This approach aimed to identify unique differentially expressed genes (UDEGs) specific to each GW, allowing for a more precise understanding of gene regulation during early pregnancy. Both coding and non–coding genes (n = 62,700) were included in the DGEA for each GW. The DESeq2 R package was used to filter out lowly expressed genes based on mean normalised counts across all samples.

Initially, raw count data were normalised to account for sequencing depth differences. The mean expression level for each gene was then calculated, and genes with very low mean normalised counts were filtered out. These genes tend to exhibit high variability and lack sufficient statistical power to detect differential expression reliably. This filtering process was conducted independently of experimental conditions to ensure unbiased results. The DESeq2 package automatically determined an optimal filtering threshold to maximise the number of significantly differentially expressed genes while controlling the false discovery rate (FDR). By reducing the multiple testing burden, this filtering step enhances the detection of meaningful DEGs. The resulting dataset (**Error! Reference source not found.**) provides a more robust foundation for understanding gene expression dynamics across gestational weeks, ensuring biologically relevant findings.

Table 11: Number of Genes Retained for Differential Gene Expression Analysis (DGEA) After Gene Filtering Across Gestational Weeks 8 to 16

Gestational Week	Total Features in Raw Data (Approx.)	Number of Genes Retained (Post-Filtering)	Percentage Retained
8	60,705	19,332	31.8%
9	60,705	15,249	25.1%
10	60,705	17,007	28.0%
11	60,705	23,915	39.4%
12	60,705	25,794	42.5%
13	60,705	26,061	42.9%
14	60,705	14,578	24.0%
15	60,705	23,978	39.5%
16	60,705	23,506	38.7%

4.3.1 Identification of DEGs

In total, 2078 DEGs (Upregulated genes –2035 and downregulated genes –43) were identified in the cfRNA sequencing data from early pregnancy. GW13 exhibited significant gene expression changes, with 1886 upregulated and 37 downregulated genes, representing a notable deviation from other GWs (Table 12). The typical gene expression changes observed at GW13 are visualised in the volcano plot presented in Figure 5.

Table 12: Differential Gene Expression Analysis Results in Early Pregnancy.

Gestational week	Upregulated DEG	Downregulated DEG	Upregulated UDEG	Downregulated UDEG
8	4	0	3	0
9	0	0	0	0
10	1	0	1	0
11	40	3	30	3
12	87	1	49	1
13	1886	37	1836	37
14	0	0	0	0
15	5	1	4	1
16	12	1	7	0
Total	**2035**	**43**	**1930**	**42**

No DEGs were detected in GW9 and GW14. This absence of DEGs could be attributed to several factors. Firstly, it's possible that during these particular gestational weeks, the gene expression profiles were relatively stable, with minimal changes occurring compared to other stages of gestation. This stability could indicate a period of homeostasis or regulatory equilibrium in the developmental process. Additionally, the absence of DEGs could result from biological factors such as the timing of key developmental events. GW9 and GW14 may represent stages where critical molecular processes are already established or have not yet initiated significant changes, leading to a lack of detectable differential expression. Furthermore, technical factors such as sample size, variability, and the sensitivity of the analysis method could influence the detection of DEGs. GW14, characterised by the lowest sample size compared to other GWs, may present challenges in detecting subtle changes in gene expression. This reduced sample size could decrease statistical power and sensitivity in DEGs.

● Significantly differentially expressed genes (DEGs)

● Non-significant genes

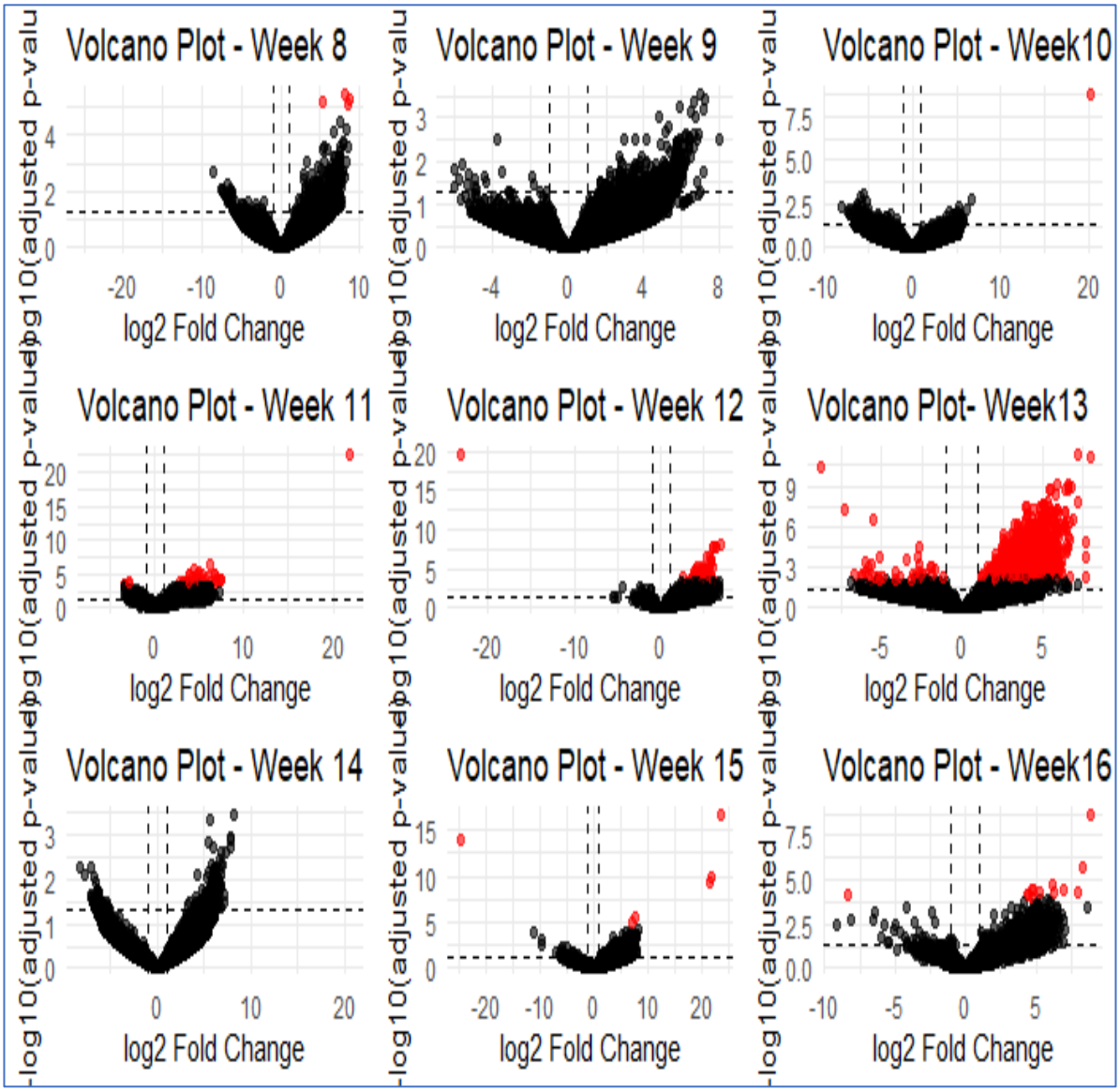


Figure 5: Volcano plots of DEGs in each GW.

Volcano plots demonstrating the Differential Gene Expression Analysis (DGEA) results across gestational weeks (GW 8–16). Each plot displays the magnitude of gene expression change (log_2 fold change) on the x-axis and statistical significance ($-log_{10}$ adjusted p-value) on the y-axis.

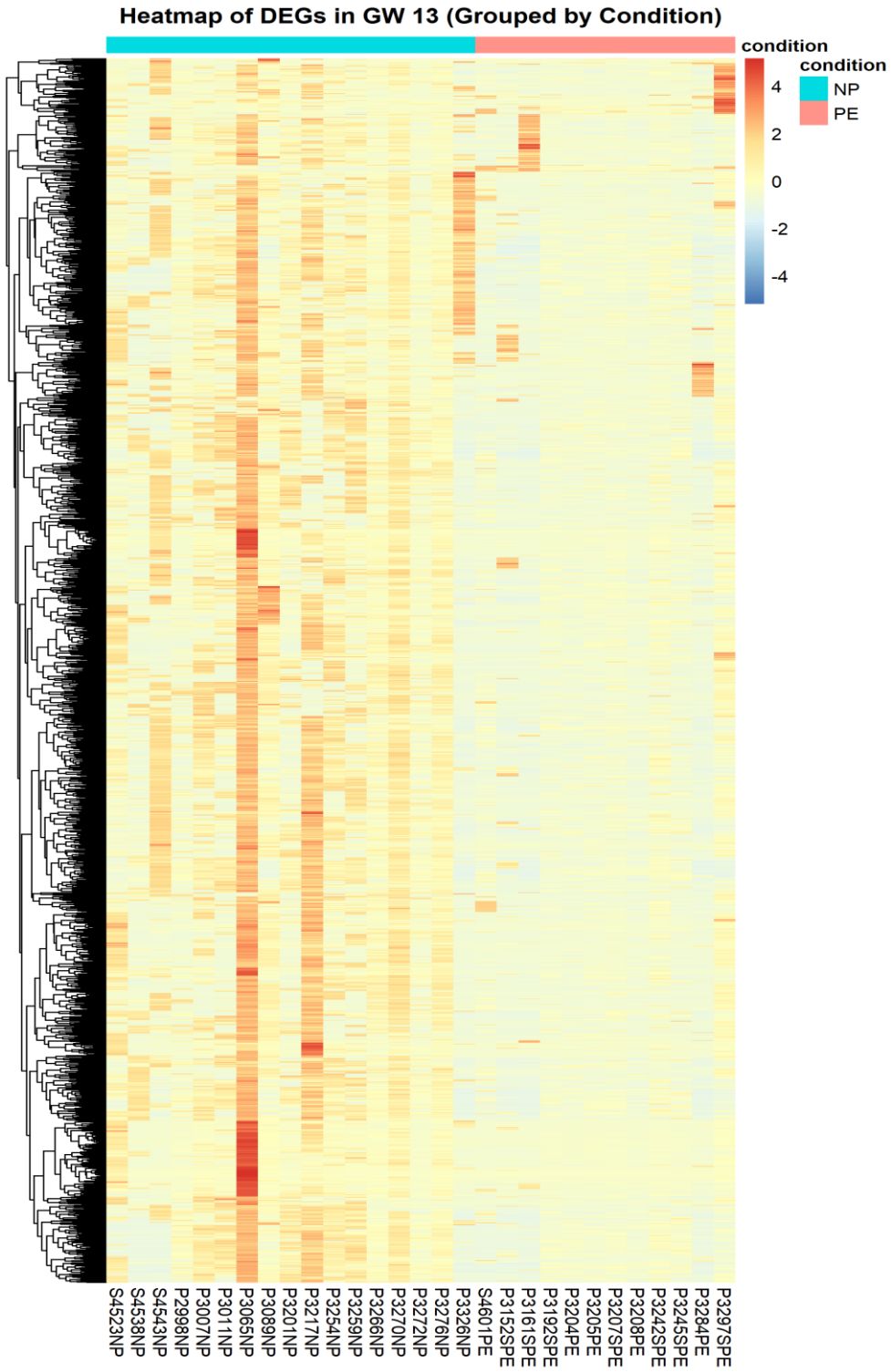


Figure 6: Heatmap of DEGs in GW 13

The heatmap presented in Figure 6 Visualises the differential expression of genes in GW 13, comparing two conditions: NP and PE. The columns represent individual samples labelled according to their condition, while the rows correspond to the expression profiles of different genes. The colour gradient, ranging from dark blue to red, reflects the varying levels of gene expression, with blue indicating low expression and red indicating high expression. Hierarchical clustering of both genes and samples reveals distinct patterns, with samples from NP and PE groups forming separate clusters, suggesting a significant difference in gene expression between the two conditions at GW 13. Notably, specific genes show pronounced upregulation or downregulation in PE compared to NP, with the intensity of the red colour indicating differential expression. These variations highlight potential biomarkers and pathways relevant to PE. While the heatmap offers an insightful visual representation of gene expression differences, statistical tests such as p–values from differential expression analysis (e.g., DESeq2 or edgeR) would further validate these findings.

4.3.2 Identification of UDEGs

DEGs in each gestational week were used in the overlap analysis to identify the UDEGs in Python. This analysis determines which DEGs are differentially expressed uniquely in each GW compared to other GWs. It is observed that the number of UDEGs is significantly higher, reaching 1873 in GW13. This indicates that a substantial proportion of genes exhibit differential expression, specifically at this gestational stage, as shown in Table 12. This information is valuable for understanding the unique molecular characteristics and biological processes associated with GW13 compared to other time points during pregnancy. A total of 38 DEGs were found to intersect between GW12 and GW13. Similarly, 7 DEGs from GW11 and 5 DEGs from GW16 were intersected with GW13. These intersections were identified through overlap analysis and visualised using an UpSetR plot in Figure 7. UpSetR plots help visualise intersections between data sets, such as DEGs from different GWs.

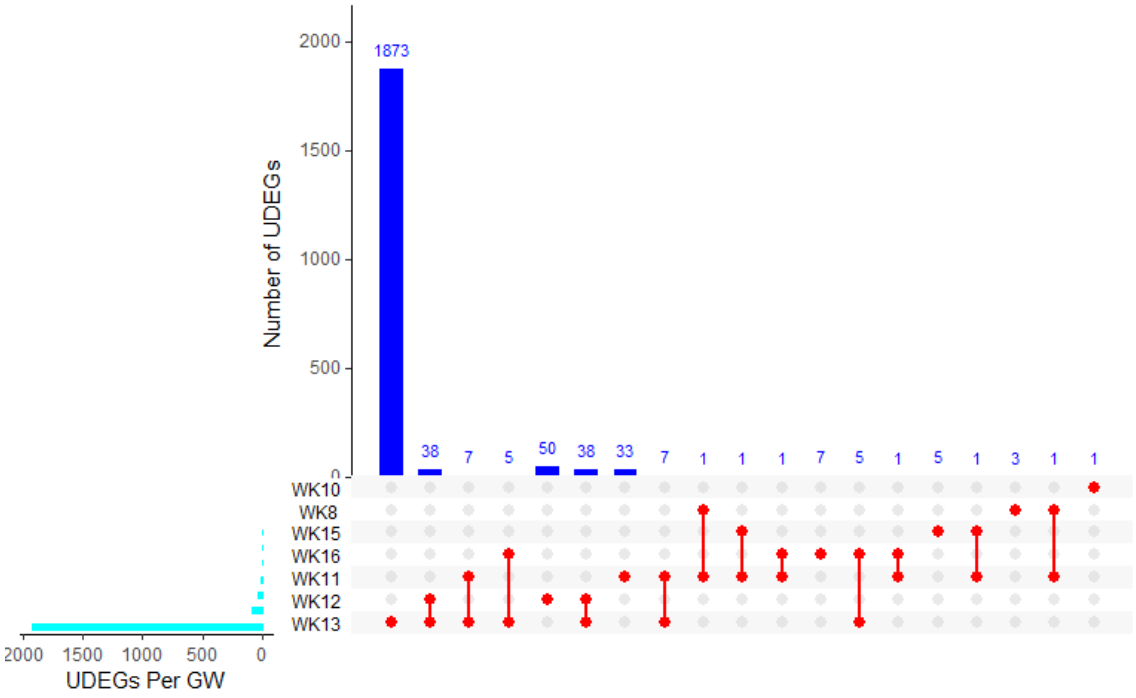


Figure 7: UpsetR plot for visualising the intersections between DEGs in GWs

4.3.3 Downstream Analysis

4.3.3.1 Gene Ontology Analysis for DEGs in Each Gestational Week

Initially, DEGs in each GW were individually subjected to GO analysis using the ClusterProfiler R package. In this analysis, the top 15 GO–enriched term categories were considered in each GW. Interpreting these terms is crucial as they represent the most statistically significant and biologically relevant enrichments, aiding in hypothesis generation and resource optimisation. Focusing on these key terms facilitates clearer interpretation, efficient prioritisation of findings, and enhanced communication of research outcomes to the scientific community.

In GW8, the identified DEGs were enriched in several biological processes. This included liver regeneration, animal organ regeneration, response to cAMP, response to organophosphorus, response to purine–containing compound, lipoprotein metabolic process, liver development, hepaticobiliary system development, regeneration, gland development, and lipid transport. This

enrichment provides valuable insights into the biological activities and processes occurring during this stage of pregnancy, particularly in relation to liver and organ development, metabolic responses, and tissue regeneration. However, it is noteworthy that no statistically significant results were identified regarding MF and CC in this analysis. This suggests that the differential gene expression patterns observed in GW8 primarily relate to biological processes rather than specific molecular functions or cellular components (Figure 8). In contrast to GW8, there were no DEGs identified in GW9, and only one DEGs was detected in GW10. DEGs in GW10 showed the enrichment in MF and CC only. Specifically, the GO terms microtubule binding and tubulin binding were enriched in the MF category. These terms indicate the involvement of the identified genes in binding to microtubules and tubulin proteins, suggesting roles in microtubule dynamics and organisation. In the CC category, several terms related to the mitotic spindle were enriched, including mitotic spindle pole, spindle microtubule, spindle pole, mitotic spindle, and spindle. These terms highlight the localisation of the identified genes to key structures involved in mitosis, such as the spindle apparatus and spindle poles, indicating their roles in regulating mitotic processes and chromosome segregation.

GO analysis conducted for GW11 and GW12 did not yield any enriched terms in BP, MF, or CC. This suggests that the differential gene expression patterns observed in these weeks may not be significantly associated with specific functional categories or cellular components based on the GO annotations. However, in GW13, the identified DEGs showed significant enrichment terms in BP, MF, and CC. In biological process, these terms include response to light stimulus, canonical NF–kappaB signal transduction, T cell differentiation, response to radiation, mitophagy, response to UV, circadian rhythm, lymphocyte differentiation, B cell activation, regulation of small GTPase mediated signal transduction, regulation of canonical NF–kappaB signal transduction, double–strand break repair, regulation of GTPase activity, cellular response to interleukin–4, ARF protein signal transduction, regulation of ARF protein signal transduction,

response to interleukin–4, regulation of B cell activation, stress–activated protein kinase signalling cascade, and stress–activated MAPK cascade. These enriched terms provide insights into the biological processes associated with gene expression changes during this stage of pregnancy. In MF, the enriched terms include guanyl–nucleotide exchange factor activity, GTPase activator activity, small GTPase binding, transcription coregulator binding, GTPase binding, and protein serine kinase activity. These findings shed light on the specific molecular functions and interactions associated with the differentially expressed genes during gestational GW13. The enriched CC terms, which include essential structures such as the extrinsic component of the plasma membrane, the RNA polymerase II transcription regulator complex, and the extrinsic component of the membrane, are pivotal in elucidating the complex pathogenesis of PE.

Similar to GW13, the DEGs identified in GW15 also demonstrated enrichment in BP, MF, and CC terms. This indicates that significant molecular alterations and biological processes are ongoing during this GW. However, in contrast to GW13 and GW15, the DEGs identified in GW16 showed enrichment only in MF and CC terms, omitting BP terms. This suggests a potential shift in molecular activity or focus during this specific stage of gestation, highlighting the importance of temporal analysis in understanding the dynamic changes occurring throughout pregnancy. Further exploration of the enriched terms in MF and CC may provide insights into the specific molecular functions and cellular components that play crucial roles in foetal development and maternal adaptation during GW16.

In GW15, the enriched BP include retina development in camera–type eye, camera–type eye development, eye development, visual system development, and sensory system development. These findings suggest a focus on the development and organisation of ocular structures and sensory systems during this stage of gestation, highlighting the importance of visual and sensory functions in foetal development at this time (Figure 9). The top 15 statistically significant MF in GW15 reveal a diverse array of transporter activities and ion channel functions that play crucial roles in foetal development and cellular homeostasis. These

include phosphate ion transmembrane transporter activity, sodium: phosphate symporter activity, L–glutamate transmembrane transporter activity, and potassium: proton antiporter activity, indicating active processes related to ion transport across cell membranes. Additionally, the presence of acidic amino acid transmembrane transporter activity, solute: potassium antiporter activity, and neurotransmitter transmembrane transporter activity suggests dynamic neurotransmitter signalling and cellular communication within the developing foetal nervous system. Furthermore, metal cation: proton antiporter activity and metal cation: monoatomic cation antiporter activity highlights the importance of metal ion homeostasis during this critical stage of gestation. These findings underscore the intricate molecular machinery at work during GW15, contributing to the overall understanding of foetal development and the physiological processes occurring within the developing embryo. Only one GO term is significantly enriched in CF, which is "Clathrin–sculpted vesicle." This enrichment indicates the presence and potential importance of vesicular structures sculpted by the protein Clathrin within the cellular environment during this stage of gestation. Clathrin–coated vesicles play crucial roles in various cellular processes, including endocytosis, intracellular trafficking, and protein sorting. The enrichment of this term suggests active vesicular trafficking and membrane dynamics within cells, possibly involved in important developmental processes or cellular signalling pathways during GW15. Understanding the presence and function of Clathrin–sculpted vesicles provide insights into the intricate cellular mechanisms underlying foetal development and physiological adaptations during this critical stage of gestation.

In GW16, the significantly enriched GO terms are observed exclusively in MF and CC, with no enriched terms found in BP. This suggests a potential shift in molecular activity or focus during this specific stage of gestation, where the biological processes underlying gene expression changes may not be as prominently reflected. However, enriched terms in MF and CC indicate that molecular functions and cellular components remain actively involved in the developmental processes occurring during GW16. Analysis of these enriched terms in MF and CC can

provide valuable insights into the specific molecular functions and cellular structures that play crucial roles in foetal development and maternal physiological adaptations during this stage of pregnancy.

MF during GW16 comprises a broad spectrum of enzymatic activities and molecular interactions that play a key role in cellular processes and signalling pathways (Figure 6). Various phosphatase activities, consisting of inositol tetrakisphosphate phosphatase and phosphatidylinositol phosphate 4–phosphatase, are significant, underscoring their crucial roles in regulating lipid signalling pathways and preserving cellular homeostasis. Additionally, chromatin–protein adaptor activity suggests involvement in chromatin organisation and potential gene regulation mechanisms. Enzymatic functions related to inositol phosphate metabolism and phosphate group hydrolysis highlight their significance in cellular metabolism and signalling cascades. Enriching lipid–related functions such as sphingolipid binding and lipid kinase activity underscores their roles in lipid metabolism and signalling pathways crucial for cellular membrane dynamics. Moreover, adenylyltransferase activity and double–stranded RNA binding suggests contributions to RNA processing and nucleotide signalling pathways. These findings collectively depict the complexity and diversity of MF operatives during GW16, providing insights into the intricate mechanisms governing foetal development and maternal physiology during this critical stage of gestation.

The photoreceptor outer segment, photoreceptor cell cilium, 9+0 non–motile cilium cilium cellular component GO terms were significantly enriched in GW16. The prevalence of these terms implies a particular emphasis on the development and arrangement of sensory structures, particularly within the visual system, at GW16. Photoreceptor outer segments constitute vital elements of photoreceptor cells responsible for light signal transduction, while photoreceptor cell cilia play essential roles in maintaining cellular integrity and facilitating signal transmission. The presence of 9+0 non–motile cilia and non–motile cilia further emphasize the significance of ciliary structures in cellular functionality and signalling pathways during this pivotal gestational phase. Comprehending the

enrichment of these cellular component terms offers valuable insights into the underlying cellular mechanisms governing visual system development and sensory perception at gestational week 16, thus enhancing understanding of foetal development and physiological adaptations during pregnancy.

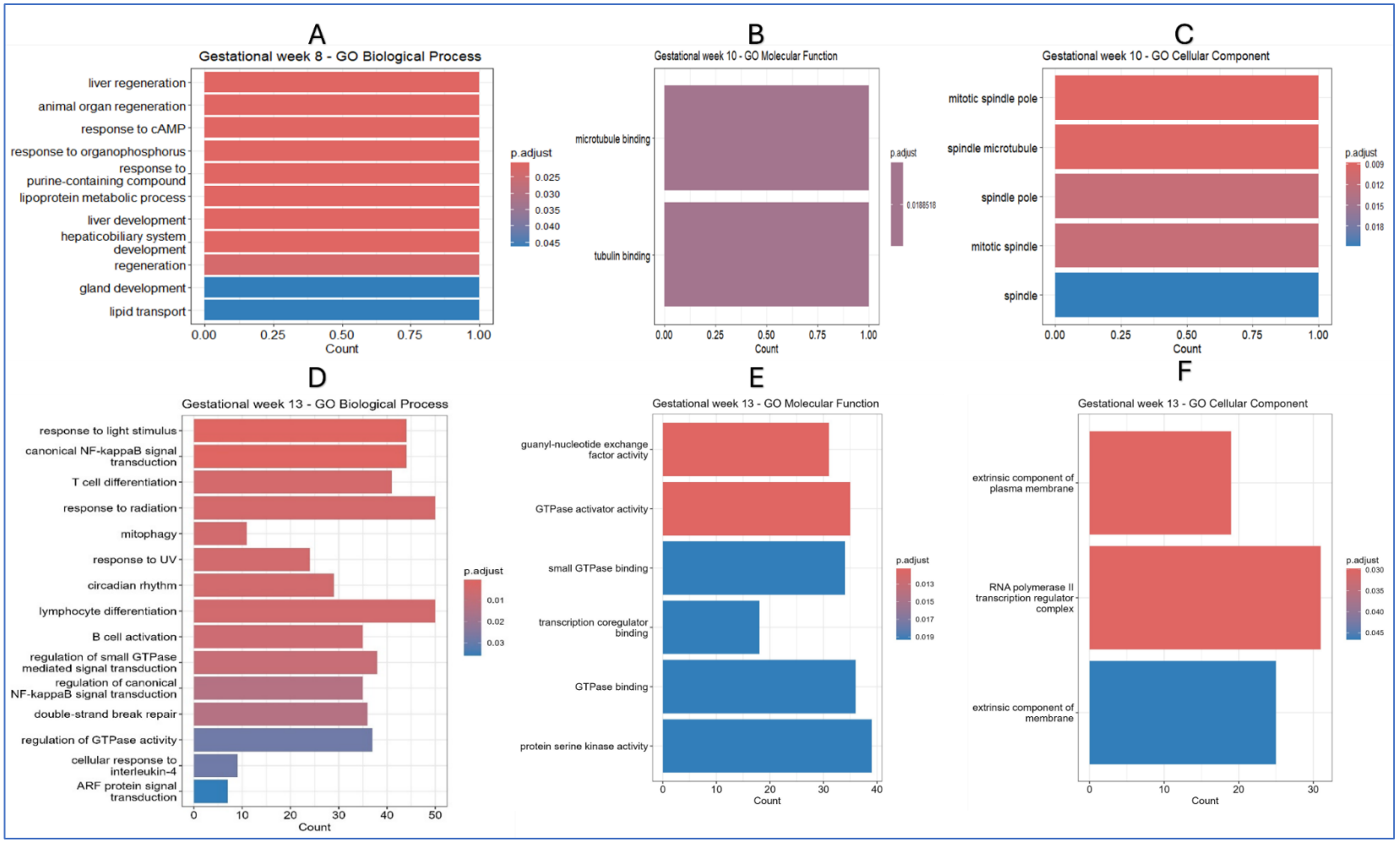


Figure 8: Results of Gene ontology analysis for DEGs in GW 8, 10, and 13.

A –Enriched GO terms in biological process in GW 8, B and C – Enriched GO terms in molecular function and cellular component in GW 10, D, E, and F – Enriched GO terms in biological process, molecular function, and cellular component in GW 13, respectively.

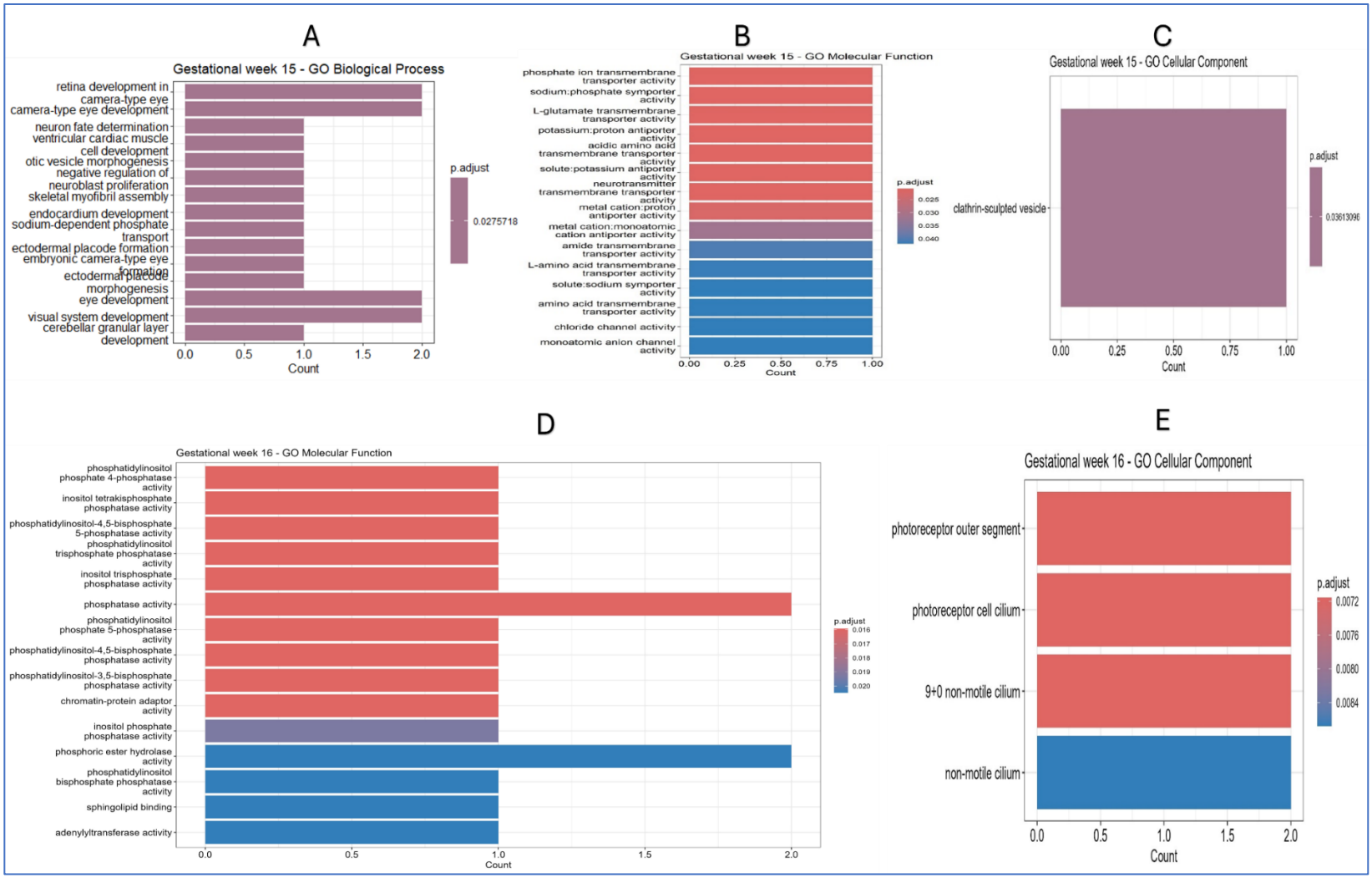


Figure 9: Results of Gene Ontology analysis for DEGs in GW 15 and 16.

A, B, and C –Enriched GO terms in biological process, molecular function, and cellular component respectively in GW 15, D and E – Enriched GO terms in molecular function, and cellular component respectively.

4.3.3.2 Gene Ontology Analysis for UDEGs in Each Gestational Week

GO analysis for UDEGs also produces the same results as the DEGS in GW 8, 10, and 15. The consistency in GO analysis results between UDEGs and DEGs across different gestational weeks indicates the robustness and significance of the identified biological processes, highlighting the dynamic yet coordinated regulation of gene expression throughout prenatal development. However, slight differences were noted in GW 13 and 16. In GW 13, compared to the GO analysis with the DEGs, the UDEGs show differences in two biological processes: regulation of B cell activation and stress–activated protein kinase signalling kinase. Similarly, the biological processes of regulation of GTPase activity and ARF protein signal transduction differ from those of the UDEGs in GW (Figure 10). In molecular function, UDEGs show two additional significant functions compared to the DEGs in GW13, which are phosphatidylinositol phosphate binding and calcium channel activity. The cellular component analysis for both DEGs and UDEGs reveals the same significant components in GW13.

In GW16, the molecular function of UDEGs exhibits double–stranded RNA binding, protein serine/threonine phosphatase activity, protein phosphatase regular activity, phosphatase regular activity, and nucleotidyltransferase activity, as compared to DEGs (Figure 11). In contrast to DEGs, the cellular component in GW16 for UDEGs only has one significant component: the Schaffer collateral–CA1 synapse. The biological process for UDEGs in GW16 displays several vital processes, including chemokine (C–C motif) ligand five production, regulation of chemokine (C–C motif) ligand five production, positive regulation of long–term synaptic depression, regulation of neurotransmitter receptor localisation to postsynaptic specialisation membrane, MDA–5 signalling pathway, regulation of ribonuclease activity, positive regulation of monocyte chemotactic protein–1 production, ARF protein signal transduction, regulation of ARF protein signal transduction, negative regulation of type 1 interferon–mediated signalling pathway, regulation of receptor localisation to synapse, regulation of nuclease activity, protein localisation to

postsynaptic specialisation membrane, neurotransmitter receptor localisation to postsynaptic specialisation membrane, and monocyte chemotactic protein–1 production. However, there were no significant GO terms enriched for DEGs in GW16 in biological processes.

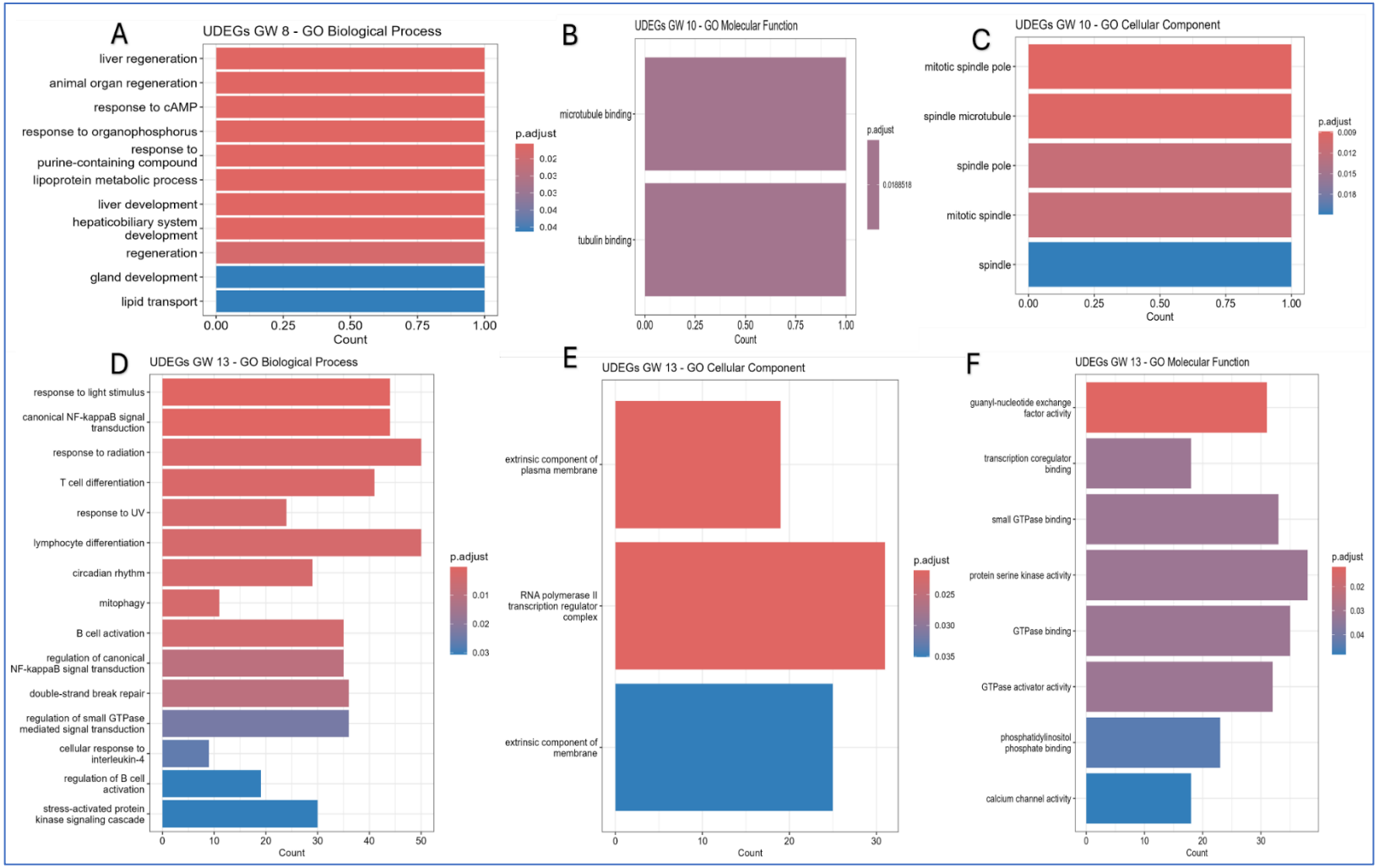


Figure 10: Results of Gene Ontology analysis for UDEGs in GW 8, 10 and 13.

A –Enriched GO terms in biological process in GW 8, B and C – Enriched GO terms in molecular function and cellular component in GW 10, D, E, and F – Enriched GO terms in biological process, molecular function, and cellular component in GW 13, respectively.

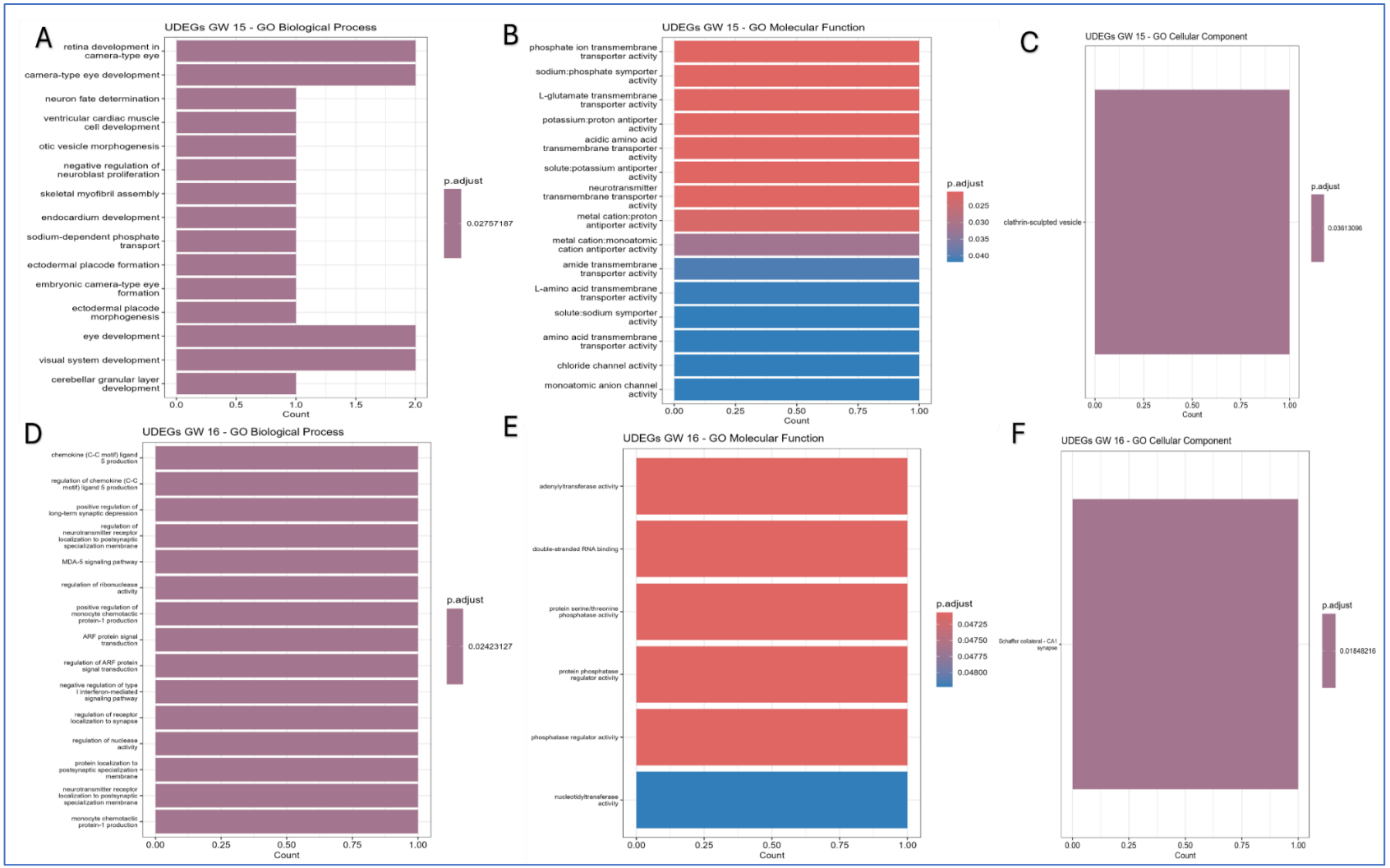


Figure 11: Results of Gene Ontology analysis for UDEGs in GW 15 and 16.

A, B, and C –Enriched GO terms in biological process, molecular function, and cellular component, respectively, in GW 15, D, E, and F – Enriched GO terms in biological process, molecular function, and cellular component, respectively.

4.3.3.3 Pathway Analysis

The functional and signalling pathway enrichment for DEGs and UDEGs were conducted using enrichR package in R with integrated KEGG pathway. This analysis showcased significant enrichment of pathways ($p < 0.05$) during GW 13, 15, and 16. These enriched pathways offer valuable insights into the molecular mechanisms underlying PE development and progression during these critical gestational periods. Intriguingly, this pathway analysis also unveiled a lack of significant pathway enrichment during GW 8, 9, 10, 11, 12 and 14, suggesting potential biological quiescence or a different regulatory landscape during these time points. These findings shed light on the dynamic nature of gene expression changes associated with PE across various stages of gestation, emphasising the importance of considering temporal aspects in understanding the pathophysiology of this complex disorder.

In GW13, the DEGs were significantly enriched in PD–L1 expression and PD–1 checkpoint pathway in cancer, Dopaminergic synapse, C–type lectin receptor signalling pathway, Phosphatidylinositol signalling system, Th17 cell differentiation, GnRH secretion, Inositol phosphate metabolism, Th1 and Th2 cell differentiation, Endometrial cancer, Acute myeloid leukaemia, Colorectal cancer, Circadian entrainment, Osteoclast differentiation, Insulin resistance, Chemokine signalling pathway, Neurotrophin signalling pathway, AMPK signalling pathway, Apoptosis, GnRH signalling pathway, AND Type II diabetes mellitus (Figure 12, Table 13). These significant enrichment pathways provide essential insights into the molecular basis of PE. Moreover, it encompasses diverse biological processes, including immune dysregulation, inflammation, vascular dysfunction, and metabolic abnormalities, all of which are implicated in the pathogenesis of PE. Particularly, pathways connected to immunological checkpoints, such as PD–L1 expression and the PD–1 checkpoint pathway in cancer, as well as those involving cytokine signalling (for example, Th17 cell differentiation and the chemokine

signalling pathway), indicate that immune system malfunction performs a vital role in PE. Moreover, pathways associated with vascular function (for example, phosphatidylinositol signalling system, osteoclast differentiation) and metabolic regulation (insulin resistance, AMPK signalling pathway) underscore the complex interplay of vascular and metabolic factors in the disorder. Furthermore, the enrichment of cancer–related pathways (for example, endometrial and colorectal cancer) suggests that cancer and PE may share molecular processes. Pathway analysis of DEGs in GW13 highlights PE's multifactorial nature and underscores diverse molecular pathways' involvement in its pathogenesis.

The DEGs in GW15 were identified to be significantly enriched in pathways such as Nicotine addiction, Synaptic vesicle cycle, Glutamatergic synapse, and Retrograde endocannabinoid signalling. It offers intriguing insights into potential mechanisms underlying PE (Figure 12). Nicotine, a potent vasoconstrictor found in cigarette smoke, has been linked to causing issues with placental dysfunction and oxidative stress, which are key characteristics of PE. Its vasoconstrictive effects can decrease uteroplacental blood flow, contributing to placental insufficiency and the subsequent development of oxidative stress in the placental environment. Additionally, disruptions in neurotransmitter systems, particularly those involving glutamate, which impacts communication and neuroinflammation, could contribute to the symptoms of PE, like seizures and cerebral oedema. Moreover, retrograde endocannabinoid signalling suggests a possible link between endocannabinoid dysregulation and vascular dysfunction, inflammation, and placental insufficiency observed in PE. These findings highlight the interplay between influences, neurotransmitter imbalances, and vascular issues in the development of PE, indicating directions for further research and therapeutic strategies.

During GW16, the DEGs were significantly enriched in the Inositol phosphate metabolism pathway, which suggests that dysregulation in this pathway may play

a role in the development of PE (Figure 12). Inositol phosphates are involved in various cellular processes, including signal transduction, membrane trafficking, and cytoskeleton assembly, all crucial for normal placental and vascular function during pregnancy. Altered levels of inositol phosphates or disruptions in the enzymes responsible for their metabolism could contribute to the pathophysiology of PE. However, further studies are needed to elucidate the specific mechanisms involved and to explore the potential therapeutic implications of targeting this pathway for the management of PE.

The pathway analysis with UDEGs did not reveal significant differences compared to those enriched with DEGs.

Table 13: Significantly enriched pathways for DEGs in GW 13 using enrichR

KEGG 2021 Human Term	KEGG 2021 Human Overlap	KEGG 2021 Human P value	KEGG 2021 Human Adjusted P value	KEGG 2021 Human Odds Ratio	KEGG 2021 Human Combined Score	KEGG 2021 Human Genes
PD–L1 expression and PD–1 checkpoint pathway in cancer	15/89	0.001569	0.204738	2.641563	17.05759	MAP3K3;MAP2K1;TICAM2;PIK3CD;FOS;NFKB1;IKBKB; EML4;ZAP70; CD4;RPS6KB2;AKT1; RAF1;NFKBIE;TLR2
Dopaminergic synapse	19/132	0.002761	0.204738	2.192733	12.92024	CAMK2B;CACNA1A;ITPR3;CALML4;CACNA1C;ARRB2;F OS;COMT;MAPK8;GNAL;GNGT2;GNG2; PPP2R1B;PPP2R1A;PPP2R5E;GNB2; AKT1;CLOCK;PLCB2
C–type lectin receptor signalling pathway	16/104	0.00294	0.204738	2.369278	13.81113	KSR1;SYK;STAT2;PIK3CD;CBLB;ITPR3;CALML4;LSP1; NFKB1; NFKB2;IKBKB;MAPK8;BCL3; AKT1;RAF1;CLEC1B
Phosphatidylinositol signalling system	15/97	0.003722	0.204738	2.382819	13.32847	IPPK;MTMR14;PIK3CD;OCRL;ITPR3;CALML4;MTMR7;I NPP5F;DGKQ;PLCE1;PIP4K2C;CDIPT; PLCB2;PLCD4;PLCD1
Th17 cell differentiation	16/107	0.003927	0.204738	2.290798	12.69055	SMAD3;FOS;NFKB1;IL27RA;IKBKB;ZAP70;MAPK8;CD4;I RF4; RARA;IL21R;HLA–DRA;STAT6; NFKBIE;JAK3;HLA–DQA1
GnRH secretion	11/64	0.005438	0.204738	2.700155	14.07976	MAP2K1;TRPC5;PIK3CD;AKT1;GNRH1;ITPR3;ARRB2; CACNA1C;RAF1; PLCB2;ESR2
Inositol phosphate metabolism	12/73	0.005472	0.204738	2.560006	13.33286	IPPK;INPP5F;MTMR14;PIK3CD;PLCE1;OCRL;PIP4K2C; CDIPT;PLCB2;PLCD4;PLCD1;MTMR7
Th1 and Th2 cell differentiation	14/92	0.005753	0.204738	2.336873	12.05356	JAG1;FOS;RUNX3;NFKB1;IKBKB;ZAP70;MAPK8;CD4; MAF;HLA–DRA;STAT6;NFKBIE; JAK3;HLA–DQA1
Endometrial cancer	10/58	0.007686	0.204738	2.709211	13.18955	APC2;MAP2K1;GADD45B;TCF7;CTNNA1;ILK; PIK3CD;AKT1;RAF1;DDB2
Acute myeloid leukaemia	11/67	0.007726	0.204738	2.55509	12.42581	IKBKB;CSF1R;MAP2K1;TCF7;RPS6KB2;RARA; PIK3CD;AKT1;RAF1;NFKB1;PML
Colorectal cancer	13/86	0.008066	0.204738	2.317575	11.17095	APC2;MAP2K1;SMAD3;GADD45B;TCF7;PIK3CD; FOS;DDB2;MAPK8;MSH3; RPS6KB2;AKT1;RAF1

KEGG 2021 Human Term	KEGG 2021 Human Overlap	KEGG 2021 Human P value	KEGG 2021 Human Adjusted P value	KEGG 2021 Human Odds Ratio	KEGG 2021 Human Combined Score	KEGG 2021 Human Genes
Circadian entrainment	14/97	0.009166	0.204738	2.195503	10.30191	RYR1;CAMK2B;ITPR3;CALML4;CACNA1C;FOS;GRIN2C;RYR3;GRIN2D;PER1;GNGT2;GNG2;GNB2;PLCB2
Osteoclast differentiation	17/127	0.009272	0.204738	2.012908	9.421906	IFNAR2;CSF1R;MAP2K1;JUND;SYK;STAT2;NCF4;PIK3CD;FOS;LILRB2;NFKB1;NFKB2;IKBKB;MAPK8;AKT1;MAP2K7;JUNB
Insulin resistance	15/108	0.010115	0.204738	2.099729	9.645643	SREBF1;INSR;PIK3CD;NFKB1;AGT;IKBKB;MLXIPL;GYS1;MAPK8; G6PC3;RPS6KB2;AKT1;OGT;SLC27A5;PCK2
Chemokine signalling pathway	23/192	0.010686	0.204738	1.774439	8.053822	CX3CR1;MAP2K1;SHC1;STAT2;WAS;PIK3CD;ARRB2;CXCR6;NFKB1;IKBKB;FGR;PREX1;TIAM1;GRK2;GNGT2;GNG2;GNB2;CXCR2;AKT1;RAF1;DOCK2;JAK3;PLCB2
Neurotrophin signalling pathway	16/119	0.010919	0.204738	2.022594	9.136499	CAMK2B;MAP3K3;MAP2K1;SHC1;PIK3CD;CALML4;NFKB1;IKBKB; MAPK8;MAPK7;ARHGDIA;AKT1;RAF1;MAP2K7;NFKBIE;SH2B1
AMPK signalling pathway	16/120	0.011787	0.208	2.003038	8.895065	SREBF1;PFKFB4;STRADA;INSR;PIK3CD;HMGCR;SIRT1;LIPE;GYS1;G6PC3;PPP2R1B;PPP2R1A;PPP2R5E; RPS6KB2;AKT1;PCK2
Apoptosis	18/142	0.012999	0.216652	1.890582	8.210561	MAP2K1;GADD45B;PARP1;HTRA2;PIK3CD;ITPR3;FOS;NFKB1; CTSS;IKBKB;PIDD1;MAPK8;TUBA1A;AIFM1;AKT1;CTSH;RAF1;CTSB
GnRH signalling pathway	13/93	0.015176	0.232796	2.113987	8.853445	CAMK2B;MAP3K3;MAP2K1;ITPR3;CALML4;CACNA1C; MAPK8;MAPK7;GNRHR;GNRH1;RAF1;MAP2K7;PLCB2
Type II diabetes mellitus	8/46	0.01552	0.232796	2.735366	11.39456	IKBKB;MAPK8;INSR;PDX1;CACNA1A;PIK3CD;CACNA1C;HK2

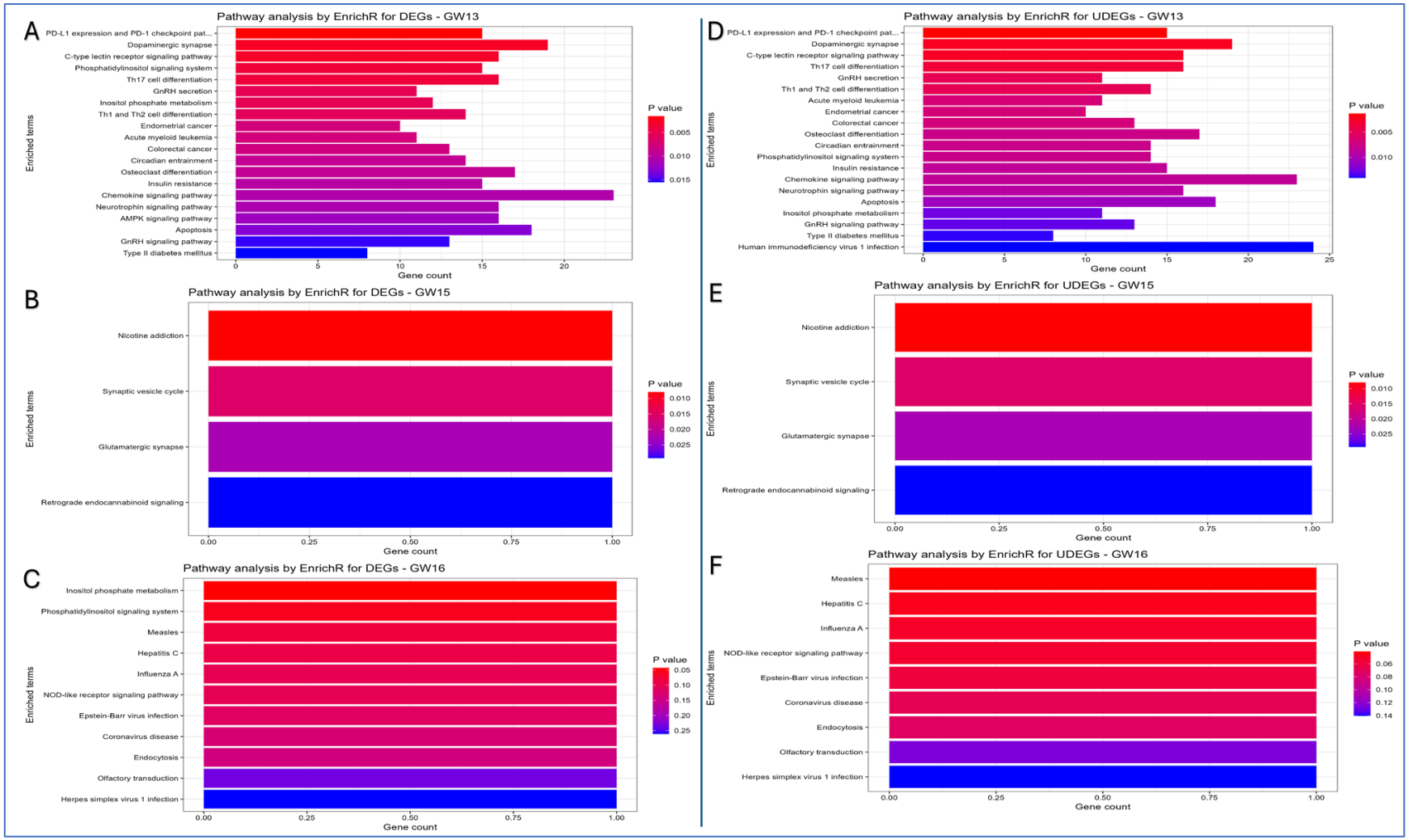


Figure 12: KEGG pathway enrichment analysis by enrichR for DEGs and UDEGs in GW–13, 15, 16

A, B, and C –Significant pathways for DEGs in GW13, GW15, and GW16 and D, E, and F –Significant pathways for UDEGs in GW13, GW15, and GW16 from the KEGG pathways database, respectively.

4.3.3.4 Protein–Protein Interactions

PPI analysis for DEGs identified no significant clusters, except in GW13, across all gestational weeks ranging from 8 to 16. In GW13, 1923, DEGs were queried in the STRING online database, resulting in the construction of a PPI network. The network comprises 1221 nodes representing genes and 5182 edges representing interactions between these genes. However, 702 DEGs were not incorporated into the PPI networks. Additionally, 4650 edges were assessed as the expected number of edges for this PPI network based on random chance. The network's observed edge count, however, was 5182, showing a much higher number of interactions than would be predicted by chance (Figure 13). This implies a higher degree of connectivity within the network than would be predicted by chance interactions alone, which may point to well–coordinated biological processes or functional connections between the relevant genes.

According to their degree of importance, the results from STRING were transferred to Cytoscape. Through Cytoscape's MCODE, this study selected the top two significant clusters for further pathway enrichment analysis using enrichR. Cluster 1, comprising 26 nodes and 161 edges, exhibited substantial enrichment in the following KEGG pathways: ribosome biogenesis in eukaryotes, basal transcription factors, RNA polymerase, nucleotide excision repair, and the cytosolic DNA–sensing pathway (Table 14). In PE, disruptions in ribosome biogenesis might hinder placental development and function, while aberrant transcriptional processes could lead to altered gene expression patterns related to angiogenesis and inflammation. Impaired nucleotide excision repair may contribute to DNA damage accumulation, exacerbating oxidative stress and inflammation. At the same time, dysregulation of the cytosolic DNA–sensing pathway might intensify the inflammatory response observed in the condition. These findings highlight potential molecular mechanisms involved in the development of PE, suggesting avenues for further research and therapeutic targeting.

Similarly, cluster 2 has 15 nodes and 51 edges exhibited significant enrichment in the following KEGG pathways: Inositol phosphate metabolism, Phosphatidylinositol signalling system, Shigellosis, AGE–RAGE signalling pathway in diabetic complications, Thyroid hormone signalling pathway, Calcium signalling pathway, Pathogenic Escherichia coli infection, Phospholipase D signalling pathway, Chemokine signalling pathway, Rap1 signalling pathway, Salmonella infection, African trypanosomiasis, Carbohydrate digestion and absorption, Endocrine and other factor–regulated calcium reabsorption, Long–term depression, GnRH secretion, Cortisol synthesis and secretion, and Long–term potentiation (Table 16). Among these pathways, inositol phosphate metabolism and the phosphatidylinositol signalling system are particularly noteworthy. Dysregulation in these pathways could impact cellular signalling processes crucial for vascular function and placental development, potentially contributing to the pathogenesis of PE. Additionally, the involvement of pathways such as calcium signalling, chemokine signalling, and AGE–RAGE signalling in diabetic complications suggests potential intersections between PE and metabolic disorders. Furthermore, pathways related to infectious diseases, such as shigellosis, Salmonella infection, and pathogenic Escherichia coli infection, hint at the role of microbial factors or immune responses in PE. Overall, the enrichment of these diverse pathways in Cluster 2 highlights the multifactorial nature of PE and underscores the need for comprehensive investigations into its aetiology and potential therapeutic targets.

Table 14: Summary statistics of two significant clusters as determined by Cytoscape

Summary Statistics	Cluster 1	Cluster 2
Number of nodes	26	15
Number of edges	161	51
Avg. number of neighbours	12.385	6.8
Network diameter	3	3
Network radius	2	2
Characteristic path length	1.671	1.848
Clustering coefficient	0.84	0.942
Network density	0.495	0.486
Network heterogeneity	0.198	0.11
Network centralisation	0.157	0.181
Connected components	1	1

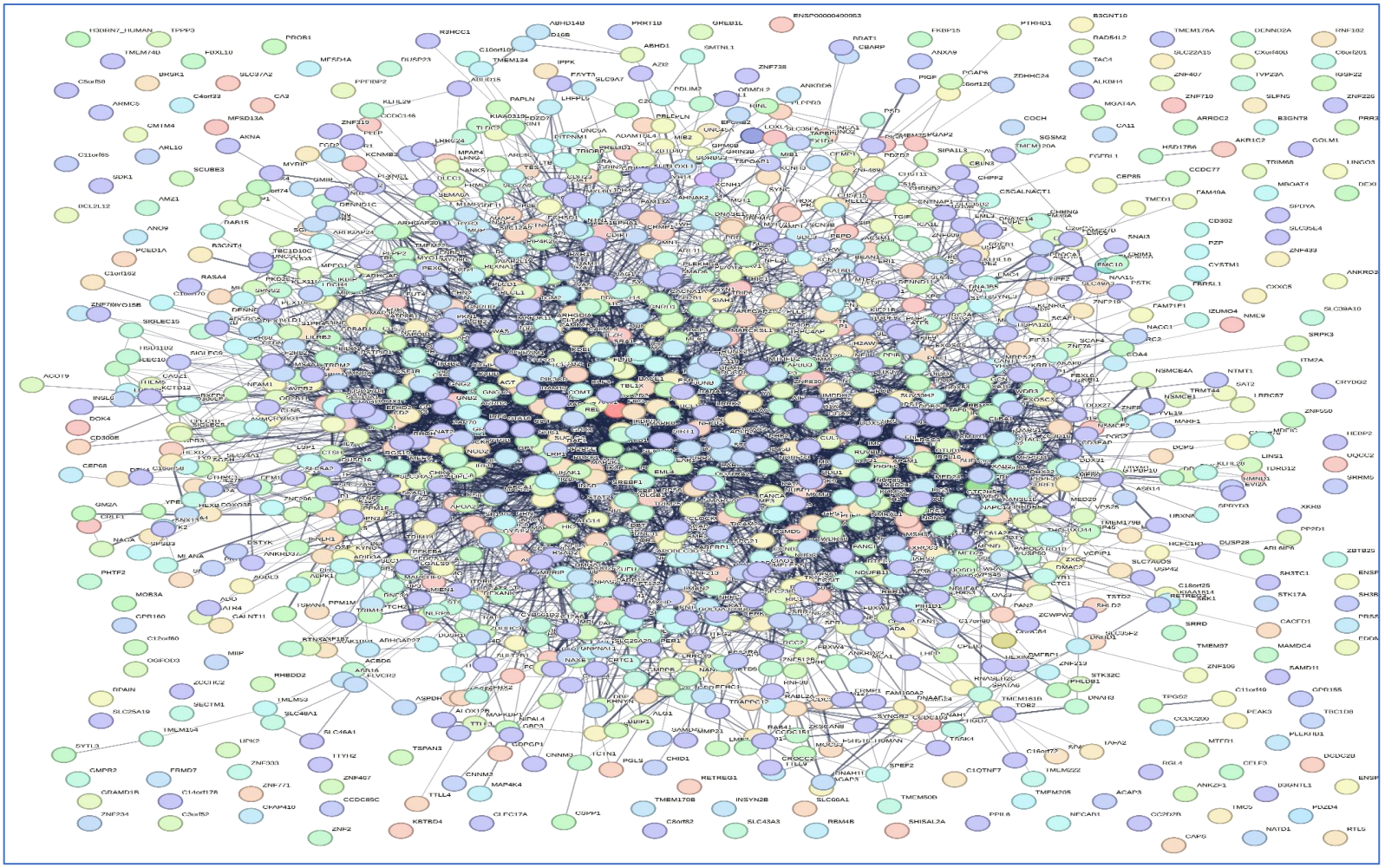


Figure 13: PPI network for DEGs in GW13 as generated using STRING

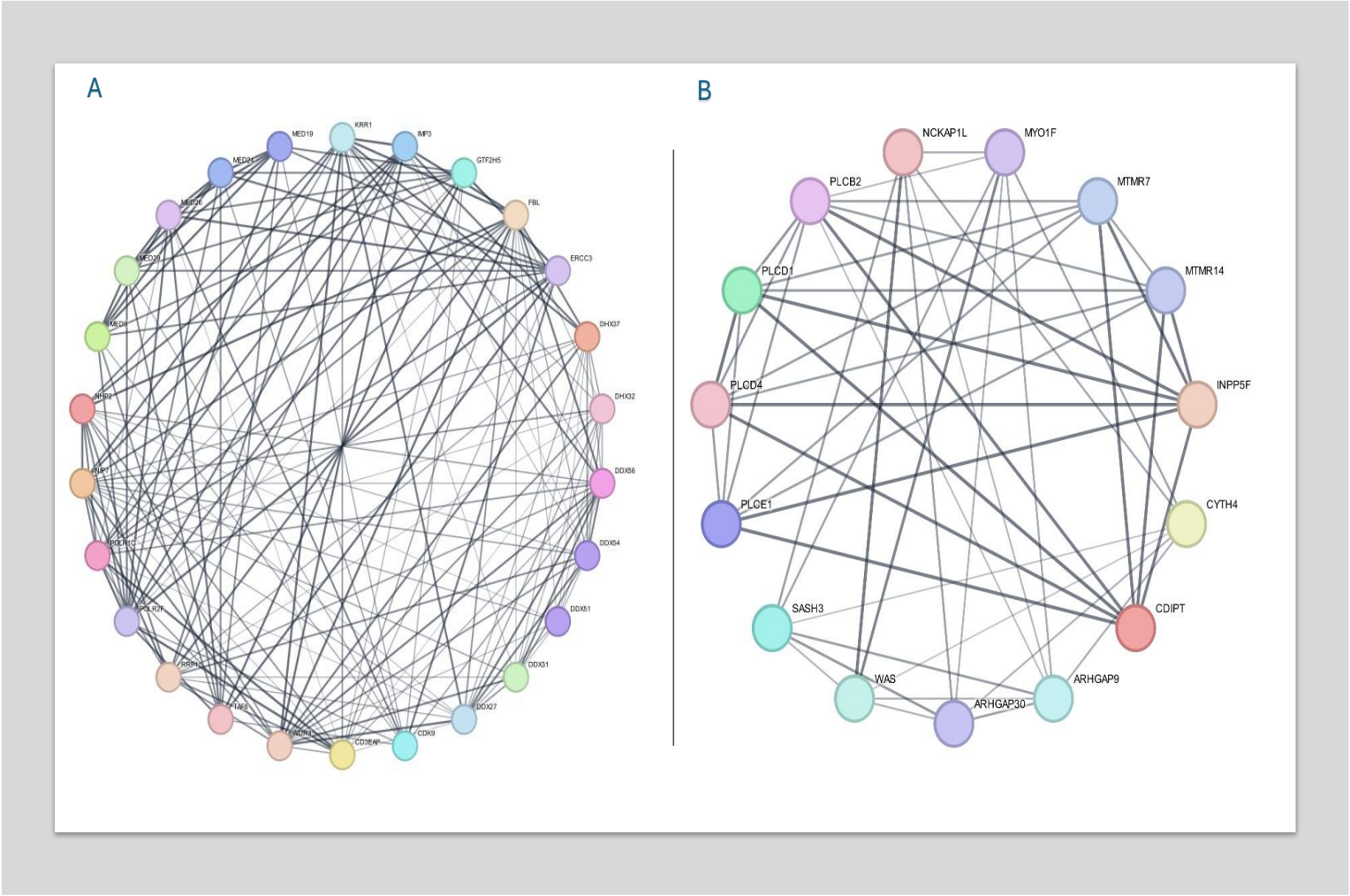


Figure 14: Top two significant clusters in the PPI analysis for GW13 using Cytoscape

A, B – Significant cluster 1 and cluster 2 in the PPI network for GW13

Table 15: KEGG pathway analysis results for nodes in Cluster 1 in the PPI network for GW13.

KEGG 2021 Human Term	KEGG 2021 Human Overlap	KEGG 2021 Human P value	KEGG 2021 Human Adjusted P value	KEGG 2021 Human Odds Ratio	KEGG 2021 Human Combined Score	KEGG 2021 Human Genes
Ribosome biogenesis in	4/108	1.10E–05	8.77E–05	34.73776	396.7222	FBL;IMP3;WDR3;NHP2
Basal transcription factors	3/45	2.67E–05	0.000107	61.90062	651.8928	ERCC3;TAF6;GTF2H5
RNA polymerase	2/31	0.000738	0.001969	57.31322	413.2932	POLR1C;POLR2F
Nucleotide excision repair	2/47	0.001695	0.003389	36.90556	235.4685	ERCC3;GTF2H5
Cytosolic DNA–sensing pathway	2/63	0.003023	0.004836	27.20355	157.8236	POLR1C;POLR2F

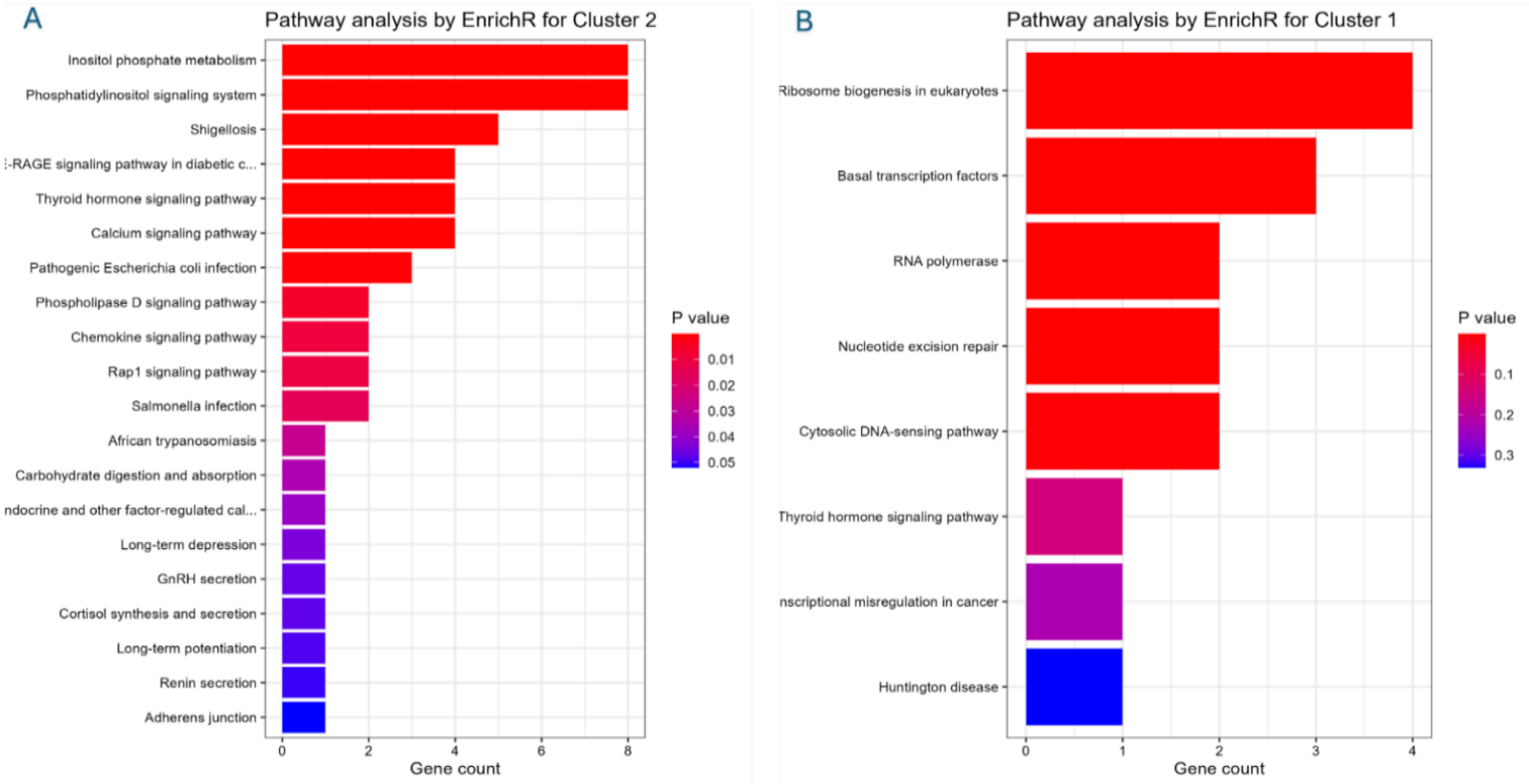


Figure 15: KEGG pathway analysis of top two significant clusters in GW13 PPI analysis using enrichR.

Table 16: KEGG pathway analysis results for nodes in Cluster 2 in the PPI network for GW13.

KEGG 2021 Human Term	KEGG 2021 Human Overlap	KEGG 2021 Human P value	KEGG 2021 Human Adjusted P value	KEGG 2021 Human Odds Ratio	KEGG 2021 Human Combined Score	KEGG 2021 Human Genes
Inositol phosphate metabolism	8/73	1.34E–16	9.89E–15	350.2418	12801.91	INPP5F;MTMR14;PLCE1;CDIPT;PLCD4;PLCB2;PLCD1;MTMR7
Phosphatidylinositol signalling system	8/97	1.43E–15	5.28E–14	255.4864	8733.55	INPP5F;MTMR14;PLCE1;CDIPT;PLCD4;PLCB2;PLCD1;MTMR7
Shigellosis	5/246	7.34E–07	1.42E–05	40.96266	578.5871	CYTH4;PLCE1;PLCD4;PLCB2;PLCD1
AGE–RAGE signalling pathway in	4/100	7.70E–07	1.42E–05	75.33712	1060.534	PLCE1;PLCD4;PLCB2;PLCD1
Thyroid hormone signalling pathway	4/121	1.65E–06	2.45E–05	61.74981	822.0941	PLCE1;PLCD4;PLCB2;PLCD1
Calcium signalling pathway	4/240	2.49E–05	0.000307	30.42989	322.6034	PLCE1;PLCD4;PLCB2;PLCD1
Pathogenic Escherichia coli infection	3/197	0.000392	0.004149	25.50387	200.0304	CYTH4;NCKAP1L;MYO1F
Phospholipase D signalling pathway	2/148	0.005361	0.049592	20.90516	109.3035	CYTH4;PLCB2
Chemokine signalling pathway	2/192	0.008867	0.072906	16.02834	75.74082	WAS;PLCB2
Rap1 signalling pathway	2/210	0.01053	0.077922	14.62796	66.60889	PLCE1;PLCB2
Salmonella infection	2/249	0.014568	0.098003	12.29399	51.99039	CYTH4;NCKAP1L
African trypanosomiasis	1/37	0.027403	0.13946	39.58135	142.3785	PLCB2
Carbohydrate digestion and absorption	1/47	0.034688	0.13946	30.96118	104.0719	PLCB2
Endocrine and other factor–regulated	1/53	0.039034	0.13946	27.38049	88.8036	PLCB2
Long–term depression	1/60	0.044082	0.13946	24.12349	75.30635	PLCB2
GnRH secretion	1/64	0.046955	0.13946	22.5873	69.08455	PLCB2
Cortisol synthesis and secretion	1/65	0.047672	0.13946	22.23326	67.66472	PLCB2
Long–term potentiation	1/67	0.049105	0.13946	21.55736	64.96939	PLCB2

4.4 Discussion

The results of this study highlight significant differential gene expression patterns in early pregnancy, with a notable peak in GW 13. Identifying 2,078 DEGs and 1,930 UDEGs underscores the dynamic molecular landscape associated with PE onset. Existing studies (Moufarrej et al., 2022) have demonstrated early gestational gene expression changes in PE, but this study uniquely identifies the week-specific progression of these changes.

4.4.1 Identification of Differentially Expressed Genes (DEGs)

This study's analysis began with the identification of DEGs across different GWs. Using HTSeq count data and the DESeq2 R package, this research identified a total of 2078 DEGs, with 2035 upregulated and 43 downregulated genes. Notably, GW13 exhibited the most significant gene expression changes, with 1886 upregulated and 37 downregulated genes. This finding underscores the critical period of GW13 in the molecular changes associated with pregnancy and potentially PE.

The absence of DEGs in GW9 and GW14 could be attributed to the stability of gene expression profiles during these weeks, indicating a period of regulatory equilibrium or minimal developmental changes. Additionally, technical factors such as sample size variability may influence the detection of DEGs, as seen in GW14, which has the lowest sample size. Understanding these dynamics is crucial as it highlights the importance of both biological and technical considerations in gene expression studies.

4.4.2 Unique Differentially Expressed Genes (UDEGs)

The overlap analysis revealed a significant number of UDEGs in GW13, highlighting the unique molecular characteristics of this gestational stage. This uniqueness is further emphasised by the identification of intersecting DEGs between GW12 and GW13 and GW11 and GW16, indicating shared molecular

processes during these periods. The UDEGs provide valuable insights into the specific biological activities occurring at different stages of pregnancy. Identifying these unique gene sets is essential as they could reveal specific pathways and regulatory networks active during critical windows of foetal development and maternal adaptation.

The significant number of UDEGs in GW13 suggests that this period may be crucial for the manifestation of PE–related changes. The biological processes active during this week, as indicated by GO analysis, include immune response, cell differentiation, and signal transduction, all of which are relevant to PE pathophysiology. The transition from a normal to a preeclamptic state could be driven by disruptions in these processes, making GW13 a focal point for further investigations.

The observed changes in total gene expression (UDEGs) must be interpreted within the limitations of traditional bulk RNA-seq DGEA. Although changes in overall transcript abundance may be highly significant, such analyses do not account for alternative splicing (AS). AS is a process whereby a single gene can generate multiple distinct mRNA isoforms, potentially leading to substantial functional differences (e.g., alterations in protein activity, localisation, or stability), even when the total gene read count remains unchanged (Liu, Fang and Wu, 2022). This complexity implies that functional dysregulation in PE could occur solely through shifts in isoform usage, without any corresponding change in the metrics used in this study's DGEA.

4.4.3 Biological Significance of DEGs and UDEGs

The study found that GW13 exhibited the most profound gene expression changes, aligning with previous research indicating the significance of this period for placental development and immune system modulation. DEGs in GW13 were enriched in pathways related to immune response (PD–L1 expression, NF–kB signalling), metabolic regulation (AMPK signalling, insulin resistance), and vascular function (phosphatidylinositol signalling). The presence of immune–

related pathways suggests that PE may be influenced by dysregulated maternal immune adaptation, a hypothesis supported by previous studies (Lokki, Heikkinen-Eloranta and Laivuori, 2018; Nieves *et al.,* 2024).

4.4.4 Contextualisation of Overlapping Key Genes

The biological significance of the six genes identified in the overlap analysis (USB1, U2SURP, RBCK1, RPPH1, KLC1, and GIMAP4) is highly relevant to established PE pathology. For instance, the identification of RBCK1 (E3 ubiquitin-protein ligase) is notable, as proteins involved in ubiquitin-mediated degradation are central to regulating cellular stress and inflammatory signalling pathways common in PE (Pei *et al.*, 2025). Similarly, genes involved in foundational RNA processing, such as U2SURP (splicing regulator) and RPPH1 (RNA component), are highly significant (Myhrer *et al.*, 2024). Their dysregulation indicates a broad, systemic failure in fundamental placental cell maintenance mechanisms induced by chronic stress. Furthermore, genes related to transport (KLC1) and immune function (GIMAP4) highlight the multi-systemic cascade of vascular and immunological collapse that characterises PE (Zhou *et al.*, 2024).

4.4.5 Gene Ontology Analysis

GO analysis of DEGs and UDEGs revealed significant enrichment in various biological processes, molecular functions, and cellular components. In GW8, enrichment was observed in liver regeneration, organ development, and metabolic processes, emphasising the developmental activities during early pregnancy. These findings align with the known biological functions during early gestation, where rapid cell growth and organogenesis are predominant.

In some weeks, the lack of significant results in MF and CC categories suggests that the differential gene expression patterns may be more relevant to biological processes during these stages. For example, the molecular mechanisms may be well–established and consistent across samples, leading to less variability and

fewer significant differences. Alternatively, the sample size and variability could play a role in the detection power of these categories.

Significant enrichment was found in biological processes such as T cell differentiation, B cell activation, and the stress–activated MAPK cascade in GW13. The molecular functions included GTPase activator activity and transcription coregulator binding, indicating complex regulatory mechanisms at play. The cellular components enriched in GW13, such as the RNA polymerase II transcription regulator complex, highlight the importance of transcriptional regulation during this stage. These findings suggest that immune modulation and transcriptional control are critical during this period, potentially influencing the stability and functionality of the maternal–foetal interface.

4.4.6 Pathway and Network Analysis Insights

Pathway enrichment analysis using the enrichR package and KEGG pathways revealed significant pathways during GW13, GW15, and GW16. In GW13, pathways such as PD–L1 expression, Th17 cell differentiation, and insulin resistance were significantly enriched. These pathways are crucial for understanding the immunological and metabolic dysregulation associated with PE. Moreover, these pathways are consistent with the underlying pathophysiology of PE, characterised by abnormal placental development, immune dysfunction, and endothelial injury. The enrichment of cancer–related pathways also suggest potential common molecular mechanisms between cancer and PE. This insight is particularly intriguing as it opens new avenues for exploring shared therapeutic targets and biomarkers between these conditions.

The PD–L1 pathway, typically associated with immune evasion in cancer, might indicate immune tolerance mechanisms in pregnancy. Dysregulation in this pathway could lead to an inappropriate immune response against the placenta, contributing to PE. Similarly, Th17 cell differentiation and its role in autoimmune responses might indicate an imbalance in immune cell populations that could underlie the inflammatory component of PE.

In GW15, pathways related to neurotransmitter systems, such as nicotine addiction and glutamatergic synapses, were enriched. These findings suggest a link between neurotransmitter imbalances and the neurological symptoms of PE. The involvement of retrograde endocannabinoid signalling highlights the potential role of endocannabinoid dysregulation in vascular and inflammatory responses. This connection between neurotransmitter pathways and PE symptoms such as headaches, visual disturbances, and seizures could lead to new diagnostic markers or therapeutic targets.

During GW16, the inositol phosphate metabolism pathway was significantly enriched, indicating its role in cellular signalling and homeostasis. The enriched terms in molecular function and cellular component categories during this week further emphasise the dynamic molecular changes occurring at this stage of pregnancy. This pathway's involvement in signal transduction and cellular response mechanisms could be pivotal in the final stages of placental development and function, influencing the overall pregnancy outcome.

Moreover, protein–protein interaction (PPI) analysis identified key gene clusters involved in ribosome biogenesis, calcium signalling, and inositol phosphate metabolism, which have been previously linked to vascular and placental dysfunction in PE.

4.5 Conclusion

This study systematically identified PE-associated differentially expressed gene signatures across gestational weeks in early pregnancy. Notably, GW13 exhibited the highest number of DEGs and UDEGs, emphasising its importance in the pathogenesis of PE. Pathway enrichment analysis revealed significant involvement of immune, metabolic, and vascular pathways, aligning with known mechanisms underlying PE. The findings underscore the value of integrating GW–specific gene expression signatures into AI-assisted predictive models, facilitating early diagnosis and targeted interventions. While the study has limitations, such as batch effects and the need for experimental validation, it lays a strong

foundation for future research in precision obstetric medicine. Integrating multi–omics approaches and validating findings in larger cohorts will be crucial for enhancing the predictive power of PE risk models and improving maternal–foetal health outcomes.

4.6 Clinical Implications and AI-Assisted Prediction Models

Identifying Unique Differentially Expressed Gene Signatures (UDEGs) in early pregnancy provides a novel and temporally resolved biomarker for AI-assisted PE prediction models. Integrating these gestational week (GW)-specific gene expression signatures into Machine Learning (ML) algorithms enhances the accuracy of early risk stratification for PE. This approach directly supports the goals of precision medicine by emphasising the potential for transcriptomic data to serve as a non-invasive diagnostic tool. The observed temporal resolution of UDEGs, particularly the intense molecular activity at GW 13, reinforces the biological urgency of this window for intervention. Identifying these molecular signatures before the clinical manifestation of PE is critical, as the strong predictive potential of biomarkers identified before 16 weeks of gestation is necessary for initiating low-dose aspirin prophylaxis, which is the universally recognized therapeutic window to effectively reduce the risk of preterm PE. This findings validate that molecular risk stratification can be achieved well within this crucial therapeutic period, contributing to the refinement of existing predictive frameworks.

4.7 Limitations and Future Directions

While this study provides significant insights into the molecular mechanisms underlying PE, several limitations and future directions warrant consideration. First, the reliance on publicly available datasets, while invaluable for hypothesis generation, may introduce batch effects and technical biases that could influence the results despite rigorous normalisation and preprocessing efforts. Second, while the DGEA approach successfully identified changes in total gene transcript abundance, it does not account for changes in gene function driven purely by

alternative splicing (AS). AS can produce structurally or functionally distinct protein isoforms from the same gene, a key mechanism in tissue-specific processes like placental development. Future efforts should therefore incorporate Differential Splicing Analysis (DSA) to fully resolve the molecular complexity of PE at the transcript level. Third, the findings from pathway and network analyses, while indicative of key biological processes, require further experimental validation in laboratory settings to establish causative relationships and functional relevance.

Future research should focus on integrating multi–omics data, including transcriptomics, proteomics, epigenomics, and metabolomics, to construct a more comprehensive and holistic molecular profile of PE. Such integrative approaches could uncover novel biomarkers and therapeutic targets by identifying cross–omics correlations and pathway interactions. Moreover, validation studies using more extensive, independent, and ethnically diverse cohorts are essential to ensure the reproducibility and generalisability of these findings. Longitudinal studies tracking changes in gene expression across multiple gestational weeks would also provide critical insights into the dynamic nature of PE progression and its temporal molecular signatures.

Additionally, advances in single–cell technologies could help delineate cell type–specific gene expression patterns, enabling a deeper understanding of the cellular contributors to PE pathophysiology. ML and AI could also play a pivotal role in integrating large–scale data, refining predictive models, and identifying subtle patterns that may not be apparent through traditional statistical methods. Finally, translating these discoveries into clinical practice will require collaboration with clinicians to develop accessible diagnostic tools and targeted therapies that can improve maternal and foetal outcomes. Addressing these limitations and pursuing these future directions will be critical for advancing understanding of PE and its clinical management.

4.8 Summary of the Chapter

PE remains a leading cause of maternal and foetal complications, necessitating early detection strategies. This study employs cfRNA sequencing data to identify GW–specific differentially expressed genes (DEGs) in early pregnancy, enabling the development of a predictive model for PE. By conducting overlap analysis across GWs 8–16, the study uncovers UDEGs per GW, highlighting the temporal dynamics of gene regulation.

A total of 2078 DEGs were identified, with GW13 showing the highest number of significant changes (1886 upregulated, 37 downregulated). The study also compares its findings with previously reported PE–linked genes, identifying six overlapping genes (USB1, U2SURP, RBCK1, RPPH1, KLC1, and GIMAP4). Pathway enrichment analysis highlights the involvement of immune dysregulation, inflammation, metabolic alterations, and angiogenesis, which are crucial in PE pathogenesis.

Protein–protein interaction (PPI) analysis at GW13 reveals a dense network with 5182 interactions, emphasising the coordinated regulatory mechanisms underlying PE. Key pathways, including the PD–L1 checkpoint, phosphatidylinositol signalling, Th17 differentiation, and AMPK signalling, are significantly enriched, offering mechanistic insights. Building on these findings, the next chapter uses this dataset to identify genes that are consistently expressed in normal and pre-eclamptic conditions.

Key Findings of the Study:

- The study identified Unique Differentially Expressed Gene Signatures (UDEGs) for each GW from 8 to 16 using cfRNA sequencing data.
- A total of 2078 differentially expressed genes (DEGs) were identified, with 2035 upregulated and 43 downregulated.
- Gestational week 13 had the highest number of UDEGs, with 1836 upregulated and 37 downregulated.

- The most significant shifts in gene expression were observed in gestational week 13.
- Several pathways were identified, including the PD–L1 checkpoint, phosphatidylinositol signalling, immune response, and metabolic pathways.
- Differentially expressed genes from gestational week 13 formed a dense network with 5182 interactions across 1221 nodes, indicating coordinated biological functions.
- Differentially expressed genes were associated with immune response, inflammation, angiogenesis, and oxidative stress, which are essential for understanding preeclampsia pathophysiology.

Chapter 5
Identification of Pre-eclamptic Unique Consistent Gene Expression Signatures at Each Gestational Week in Early Pregnancy

This chapter explores the identification of pre-eclamptic consistent gene expression signatures and unique, consistent genes in each GW using the cfRNA sequencing data. It investigates further by performing downstream analysis to understand the pathophysiology of PE.

5.1 Introduction

PE usually manifests clinically in the mother after 20 weeks of pregnancy; thus, early diagnosis is necessary to prevent complications. It is evidenced that significant transcriptomic changes occur in the placenta before 20 weeks of gestation, and identifying these alterations could be important for understanding and managing placental disorders like PE (Gupta *et al.,* 2022; Bogias *et al.,* 2022). These changes are seen in the mother's uterus' spiral arteries, which significantly remodel during pregnancy to ensure enough blood supply to the placenta and foetus. The spiral artery remodelling process begins with the invasion of trophoblast cells into the decidual segments around 12–13 GW. It is completed by mid–pregnancy after the second stage of remodelling the myometrial segments from 15 weeks onwards. This extensive vascular remodelling is critical for establishing adequate uteroplacental circulation to support the developing foetus.

Many transcriptomics research studies have revealed variations in gene expression that contribute to the pathophysiology of the disease, offering new insights into the molecular causes of PE . Furthermore, several recent studies have discovered panels of DEGs that may be useful as biomarkers or offer insights into the molecular pathophysiology of PE. Yadama et al. (2020) identified a core set of 117 differentially expressed genes in early pregnancy (10–18 weeks) associated with PE, and they found an overlapping whole–blood gene signature enriched for immune processes across these conditions. Recently, another study by Moufarrej

et al. (2021) identified 544 DEGs between PE cases and controls before 20 weeks gestation. Another study by Huang et al. (2022) identified 441 DEGs linked with PE through bioinformatics analyses, weighted gene co–expression network analysis (WGCNA), and enrichment analysis of multiple datasets. These research studies focused on identifying the DEGs between normal and disease states. By comparing gene expression profiles, researchers aimed to pinpoint specific genes that exhibit altered expression levels in PE compared to healthy pregnancies. This approach helps to uncover potential molecular mechanisms underlying the disease, providing insights that could lead to the development of diagnostic markers and therapeutic targets for PE. Through detailed bioinformatics analyses and techniques, these studies have significantly contributed to understanding the genetic basis of PE. However, some gene pools exhibit consistent expression levels, either higher or lower, in both normal and disease conditions. Differential gene expression analysis did not consider these gene pools, as this approach focuses on comparing expression levels between NP and PE states. Therefore, this study hypothesises that the consistently expressed gene signatures in NP and PE could be linked to PE disease progression. Identifying these gene signatures could reveal new biomarkers for the early detection of PE during early pregnancy.

5.1.1 Consistently Expressed Gene Signatures

The expression of genes differs in diseased and healthy conditions. Nonetheless, independent of status, some genes exhibit consistent expression levels that are higher or lower in all samples. These types of genes are called Consistently Expressed Gene Signatures (CEGS). CEGS are differentiated from the Housekeeping Genes (HKGs), Ubiquitously Expressed Genes (UEGs), and Differentially Expressed Genes (DEGs). Unlike HKGs, which are expressed across all tissues and conditions to support basic cellular functions, CEGS exhibit stability specific to a particular state, reflecting condition–specific resilience rather than ubiquitous necessity (Iyer *et al.,* 2017). UEGs are generally expressed throughout most tissues and conditions, such as NP and PE, with little variability. In contrast,

CEGS are rigidly controlled within specific states, demonstrating their specialised functions in maintaining context–dependent physiological balance (Gu, J. *et al.,* 2023). Conversely, DEGs exhibit notable variations in expression across various conditions, emphasising dynamic reactions to disease or normal conditions, while CEGS stay constant, offering a steady baseline .

5.1.2 Housekeeping Genes and Consistently Expressed Gene Signatures

Even though many genes exhibit different expression levels on various conditions, however some genes, such as HKGs, show the same expression levels in normal and disease states for regulating normal cellular activities (Eisenberg and Levanon, 2013). They are well known for their role in maintaining fundamental biological processes and are extensively expressed in various tissues and cell types. As these genes were thought to express at consistent levels in various biological situations, they were initially used as "control genes" for standardising quantitative expression research. However, additional studies have shown that HKG is tightly controlled and that different cell types express it differently, even though it is broadly expressed. HKG is broadly expressed in a variety of tissues and cell types to maintain a baseline transcriptome, which essentially distinguishes them (Shaw *et al.,* 2011; Chang *et al.,* 2011).

CEGS display fluctuating expression levels based on health or disease status, whereas HKG typically maintains consistent expression regardless of the conditions. These gene expression changes make CEGS valuable biomarkers in disease prediction models. Nonetheless, because of their stable expression under various circumstances, HKGs are usually utilised as controls.

5.1.3 Ubiquitously Expressed Genes and Consistently Expressed Gene Signatures

Ubiquitously expressed genes (UEGs) are consistently transcribed across various tissues and cell types. These genes perform fundamental cellular processes required for the maintenance and survival of nearly all cell types. In a comprehensive study on mammalian tissue transcriptomes, Ramsköld et al. (2009)

identified approximately 8,000 UEGs that account for roughly 75% of total mRNA by message copy number in most tissues (Ramsköld *et al.,* 2009). UEGs predominantly encode intracellular proteins involved in core biological activities, including metabolism, transcription, RNA processing, and translation. These functions are vital for the daily operations of cells, irrespective of their specific roles within tissues. UEGs exhibit stable expression levels across tissues, with slight variation, positioning them as housekeeping genes essential for general cellular function (Joshi *et al.,* 2022).

While not as ubiquitous as UEGs, CEGs maintain stable expression in certain tissues or under specific conditions. These genes often reflect the specialised roles required by a subset of tissues. CEGs are typically involved in more context–dependent processes, such as tissue repair, immune responses, or specialised cell signalling pathways. For example, genes encoding membrane–bound proteins or receptors are consistently expressed where specific signalling functions are necessary. Unlike UEGs, CEGs are more likely to be involved in extracellular or membrane functions, contributing to tissue–specific roles and signalling.

The primary distinction between UEGs and CEGs is their contribution to tissue–specific identity. UEGs are critical for maintaining the fundamental operations of cells across different tissue types, providing the core machinery necessary for intracellular function. CEGs, in contrast, are integral to the specialised tasks of particular tissues, playing roles in tissue–specific processes that may not be essential in all cell types (Ramsköld *et al.,* 2009). Tissues such as muscle and liver, which exhibit simpler transcriptomes dominated by UEGs, contrast sharply with more complex tissues like the brain, kidney, and testis, where CEGs play a more prominent role. This complexity in transcriptomes reflects the diversity of functions carried out in these tissues, indicating that the expression patterns of UEGs and CEGs determine the commonalities and the specialisations of tissue–specific gene regulation.

The pregnancy week approach used to identify DEGs revealed that gestational week 13 exhibited more DEGs than other gestational weeks. This finding underscores the importance of focusing on early pregnancy for biomarker identification, particularly when developing machine–learning models to detect women at risk of PE. Identifying biomarkers during these early gestational weeks offers several advantages, including earlier detection and intervention, which can significantly improve outcomes for both mother and foetus. Additionally, it provides a longer window for monitoring and potential treatment while offering insights into the underlying mechanisms of PE development.

5.1.4 Research Hypothesis

Unique Consistent gene expression signatures (UCEGS) identified at each GW can serve as biomarkers for early detection and understanding of PE.

5.1.5 Research Question

Can GW–specific consistent gene expression signatures (CEGS) differentiate between normotensive and pre-eclamptic pregnancies, and how can they enhance early prediction models for Pre-eclampsia?

5.1.6 Research Aim and Objectives

This chapter aimed to identify GW–specific unique, consistent gene expression signatures (UCEGS) in NP and Pre-eclamptic pregnancies and evaluate their potential as biomarkers for early PE detection. To achieve this aim, the following objectives were established:

1. To identify significant, consistent genes (SCGs) and biologically relevant consistent genes (BRCGs) in NP and PE pregnancies from gestational weeks 8 to 16.
2. To analyse and compare enriched biological processes, molecular functions, and cellular components associated with CEGS in NP and PE.

3. To identify key gestational weeks showing distinct gene expression patterns in PE compared to NP.
4. To propose potential biomarkers and pathways for early PE prediction.

5.2 Method

Chapter 3 (General method) comprehensively explains the cfRNA–seq data collection to read quantification steps in the RNA sequencing analysis workflow. This current method section builds upon that foundation and goes into greater detail about consistent gene expression analysis to identify CEGS.

5.2.1 Datasets

The same cfRNA sequencing data from the datasets GSE154377 and GSE192902 were utilised in this study for Consistent Gene Expression Analysis (CGEA) to identify CEGS. As explained in the Chapter 4 method section, per GW gene expression data were used in the CEGS analysis.

The cfRNA sequencing data from the datasets GSE154377 and GSE192902 were utilised in this study for Consistent Gene Expression Analysis (CGEA) to identify CEGS. The Chapter 3 method section explained that gene expression data were analysed per gestational week for the CEGA. This approach allows for a detailed examination of gene expression patterns across early pregnancy stages, facilitating the identification of genes that maintain consistent expression levels within each gestational week.

5.2.2 Consistent Gene Expression Analysis (CGEA)

Identifying genes that are significantly and consistently expressed, whether at high or low levels, is essential for understanding the biology of PE during early pregnancy. This study categorises CEGS into Significant Consistent Genes (SCGs) for those expressed consistently at high levels and Biologically Relevant Consistent Genes (BRCGs) for those expressed consistently at lower levels. SCGs are genes that maintain high expression levels across samples in a particular

condition. These genes often play crucial roles in fundamental cellular processes and may be key drivers in the development or progression of PE. Their high expression makes them easier to detect and quantify, potentially serving as robust biomarkers or therapeutic targets. BRCGs, on the other hand, are genes that show consistent but lower levels of expression. Despite their lower abundance, these genes can be biologically significant, often involved in regulatory functions or specific cellular responses. In the context of PE, BRCGs might represent subtle but essential changes in cellular state or function that contribute to the disease's pathophysiology. By considering both SCGs and BRCGs, this study aimed to provide a comprehensive view of gene expression patterns in PE. This approach allows for detecting major players (SCGs) and more nuanced contributors (BRCGs) to the disease process, potentially uncovering new insights into the molecular mechanisms underlying PE in its early stages. This dual categorisation strategy enhances the study's ability to capture a broader range of biologically relevant genes, potentially leading to a more thorough understanding of the complex gene expression landscape in PE during early pregnancy. Following the establishment of this conceptual framework, the next crucial step in consistent gene expression analysis is data pre-processing.

5.2.2.1 Data Pre–processing

Data pre-processing methods are important in Bioinformatics analyses, and they can be applied to reduce these systematic variations and harmonise the datasets before they are used to build a machine–learning model (Van et al., 2024). This process involves removing batch effects and absent calls and normalising. In this study, the absent calls were eliminated using the batch effect–removed gene expression counts from the prior studies in Chapter 4.

5.2.2.1.1 Normalisation

Normalisation of RNA–Seq data is essential to ensure accurate inferences and the replication of findings (Abbas-Aghababazadeh, Li and Fridley, 2018). It corrects for technical biases, such as differences in sequencing depth and RNA composition

between samples, which can otherwise skew the results. Effective normalisation enables a more reliable comparison of gene expression levels across different samples and conditions, supporting robust and reproducible scientific conclusions. This study employed the edgeR package in R to normalise gene expression data collected across different GW. Specifically, the Trimmed Mean of M–values (TMM) normalisation method was used, which effectively handles the variability inherent in RNA sequencing data. The TMM normalisation was applied to the gene expression data from GW 8 to 16 to ensure comparability and accuracy in subsequent analyses.

The normalisation process was applied separately to the gene expression count data for each gestational week (GW) in both normal and PE conditions, contrasting with the typical approach in differential gene expression (DGE) analysis where normal and PE data are merged for normalisation; this separate normalisation per GW was crucial to ensure the accurate identification of consistently expressed genes (CEGs) in the consistent gene expression analysis conducted in this study. Separate normalisation of gene expression data from normal and disease samples offers several advantages in RNA sequencing studies: variations in RNA composition between normal and disease samples can significantly impact normalisation accuracy, and normalising them together might introduce biases due to these compositional differences; differences in library sizes across samples, particularly between normal and disease conditions, can distort normalisation factors if applied collectively, but by normalising separately, these differences can be more effectively managed, ensuring accurate comparisons within each condition; moreover, separate normalisation enhances the ability to discern condition–specific changes in gene expression profiles, aiding in the precise identification of CEGs within normal and disease contexts and supporting more insightful biological interpretations of the data.

5.2.2.1.1.1 TMM Normalisation

The TMM normalisation method, implemented in the edgeR package, is designed to normalise RNA–seq data by adjusting for differences in library sizes and RNA composition between samples (Maza, E., 2016). This method ensures that the gene expression levels are comparable across samples, facilitating accurate identification of CEGS. TMM normalisation aims to adjust the library sizes by computing scaling factors that reflect the relative RNA composition across samples (Robinson and Oshlack, 2010).

5.2.2.1.1.1.1 Concept of TMM Normalisation

TMM normalisation addresses two main issues in RNA–seq data: differences in sequencing depth (library sizes) and compositional biases due to different abundances of RNA species across samples. By adjusting for these factors, TMM normalisation ensures that gene expression levels are comparable across samples, facilitating accurate downstream analyses.

5.2.2.1.1.1.2 Mathematical Explanation of TMM Normalisation

The method of TMM normalisation involves several key steps, each aimed at ensuring accurate normalisation across samples. Those key steps include calculating the reference sample, calculating M–values (log–fold changes) and A–values (average expression levels), trimming of M–values, weighted mean of trimmed M–values, calculation of scaling factor, and normalised counts.

5.2.2.1.1.1.3 Calculate the Reference Sample

A reference sample can be chosen arbitrarily, or a pseudo–reference sample can be created by taking the geometric mean of the counts for each gene across all samples, typically used to avoid bias towards any single sample (Maza, Elie, 2016). For n samples and m genes, let Yij be the raw count for gene i in sample j. The pseudo–reference count for gene i, $\overline{Y}i$, is calculated as:

$$\bar{Y}_i = \left(\prod_{j=1}^{n} Y_{ij} \right)^{\frac{1}{n}}$$

5.2.2.1.1.1.4 Calculate M–values and A–values

M–values are the log base 2 ratios of a gene's expression levels between a reference and a sample (Robinson and Oshlack, 2010). This value effectively captures the relative expression level of a gene in a given sample compared to a reference. The M–value is particularly helpful in normalisation methods since it may be used to identify and correct biases arising from differences in library sizes or composition in RNA–sequencing datasets . By focusing on the log–fold changes rather than raw counts, M–values provide a more stable and interpretable measure of gene expression differences, facilitating more accurate downstream analyses (Chen, Y., Lun and Smyth, 2014).

The average log–expression values or A–values is used to measure a gene's average expression level in a given sample compared with reference samples during the normalisation process. M–values and A–values are complementary metrics in gene expression analysis – with M–values quantifying relative changes and A–values representing expression magnitudes. This dual consideration not only captures differential expression but also helps identify genes with significant expression levels, thus improving the accuracy and interpretation of downstream analyses.

Mathematically, for each gene i in sample j, Y_{ij} represents the raw count, and $\bar{Y}_i$ is the geometric mean of the counts, the M–value and A–value as follows:

$$M_{ij} = log_2 \left(\frac{Y_{ij}}{\bar{Y}_i} \right)$$

$$A_{ij} = \frac{1}{2} log_2 \left(Y_{ij} \; x \; \bar{Y}_i \right)$$

5.2.2.1.1.1.5 Trimming of M–values

The log–fold changes, or M–values, of gene expression levels between a sample and reference genes capture the relative expression differences. Trimming the M–values reduce the influence of outliers and extreme expression values, which can distort the normalisation process and impact downstream analyses. Moreover, this process removes the most extreme log–fold changes to concentrate on the central distribution, which more accurately represents most of the data. Typically, the top and bottom 30% of M–values are discarded, a crucial step for several reasons. First, it lessens noise since extreme M–values of gene expression level sometimes indicate noise or technical anomalies rather than actual biological variation. Second, it reduces bias since estimates might be skewed due to outliers' excessive impact on the normalisation factor calculation. The effect of these outliers is reduced by lowering the M–values, leading to a more robust normalisation . Finally, it emphasises central tendency since genuine biological differences are more likely to be reflected in the central region of the M–value distribution. The normalisation method can use trimming to concentrate on these more trustworthy data points .

5.2.2.1.1.1.6 Weighted Mean of Trimmed M–values

The remaining non–trimmed M–values were utilised to calculate the weighted mean following the trimming of M–values. Genes exhibiting minimal variation in expression levels among samples are generally more stable and less prone to biological or technical noise. Conversely, genes with significant variance may indicate genuine biological differences or technical artefacts. The rationale for using weights inversely proportional to variance is based on statistical principles, which aim to assign more weight to genes with less variation in expression across the sample.

The weight of gene expression in the context of TMM normalisation is calculated using the formula (Robinson and Oshlack, 2010):

$$w_{ij} = \frac{1}{\left(\frac{1}{Y_{ij}} + \frac{1}{\overline{Y_i}}\right)}$$

In this formula, Y_{ij} represents the raw count for gene i in sample j. The term $\overline{Y_i}$ denotes the geometric mean of counts for gene i across all samples. Next to calculating the weight, the weighted mean of trimmed M–values is calculated as follows:

$$Weighted\ Mean_j = \frac{\Sigma_{i \in trimmed\ set}\ w_{ij}\ M_{ij}}{\Sigma_{i \in trimmed\ set}\ w_{ij}}$$

In this formula $\Sigma_{i \in trimmed\ set}\ w_{ij}\ M_{ij}$ represents the summation of the product of weights (w_{ij}), and M–values M_{ij}, for all gene i in the trimmed set for sample j. $\Sigma_{i \in trimmed\ set}\ w_{ij}$ represents the sum of the weights for all gene i in the trimmed set for sample j.

5.2.2.1.1.1.7 Calculation of Scaling Factor

In RNA sequencing analysis, scaling factors are used in data normalisation to modify raw gene expression count data and make it consistent between samples or situations. Moreover, it adjusts for technical variances and other systematic discrepancies like library size and sequencing depth. The weighted mean of the trimmed M–values determine the scaling factor for each sample. This approach reduces the influence of extreme values and concentrates on the data's central distribution. The scaling factor for each sample j is obtained by exponentiating the weighted mean of the trimmed M–values as follows:

$$Scaling\ factor\ _j = exp\left(\frac{\Sigma_{i \in trimmed\ set}\ w_{ij}\ M_{ij}}{\Sigma_{i \in trimmed\ set}\ w_{ij}}\right)$$

5.2.2.1.1.1.8 Normalise Counts

Finally, the raw counts are normalised using the calculated scaling factors, TMM_j. The normalised count Y_{ij}^{TMM} for gene i in sample j is:

$$Y_{ij}^{TMM} = \frac{Y_{ij}}{TMM_j}$$

By following these steps, TMM normalisation effectively adjusts for library size differences and compositional biases, allowing for accurate comparison of gene expression levels across different samples. This method, implemented in the edgeR package, is widely used in RNA–seq data analysis to enhance the reliability of downstream analyses such as differential expression and biomarker discovery (Maza, Elie, 2016). In addition to the normalisation process, each GW data set in NP and PE was used in the consistent gene expression analysis in Python (v3.12.1).

5.2.2.1.2 Gene Filtration

Gene filtration based on 0.5 counts per million (CPM) thresholds after Trimmed Mean of M–values (TMM) normalisation is a crucial pre–processing step in RNA sequencing (RNA–seq) data analysis. This method removes genes with extremely low expression levels, often attributed to background noise or technical artefacts rather than true biological signals. Once TMM normalisation is applied, CPM values are calculated for each gene in every sample. CPM is a standardised metric representing the number of counts a gene receives per million total reads, making it a relative measure of gene expression. The filtration step involves setting a threshold, typically retaining only those genes with a CPM of at least 0.5 in a minimum number of samples –often 50% in each group or condition under study. For example, in a sample with 10 million total reads, a gene must have at least five raw counts to achieve a CPM of 0.5. This threshold is chosen because it balances excluding genes with negligible expression, contributing noise, and retaining genes that may have biologically meaningful but moderate expression levels.

The rationale for this filtration approach lies in its ability to enhance data quality and statistical power. The analysis avoids inflating the total statistical tests by removing genes with low or sporadic expression. This can reduce sensitivity and increase false discovery rates in downstream analyses like differential gene

expression. Additionally, it reduces computational complexity, as smaller datasets are more efficient to process. Importantly, filtering ensures that the analysis focuses on genes with sufficient evidence of expression, thereby increasing the likelihood of identifying biologically relevant patterns. This method is valuable for identifying unique genetic variant signatures in cfRNA to predict conditions such as pre-eclampsia. By focusing on moderately to highly expressed genes, the approach helps ensure that meaningful signals are captured, facilitating robust and reproducible insights into the biological processes underlying the condition. Next to the gene filtration process, each GW data set in NP and PE was used in consistent gene expression analysis in Python (v3.12.1).

5.2.2.2 Identification of Consistently Expressed Gene Signatures (UCEGS) using Rank Product Analysis

The identification of UCEGS was performed using a novel algorithm based on the Rank Product (RP) framework, specifically adapted to overcome the limitations of the standard RP method in single-class datasets (see Appendix B: Consistent Gene Expression Analysis). While the traditional RP method is a robust, non-parametric technique ideal for differential expression analysis between groups (Breitling *et al.*, 2004), its application to identify genes that are consistently expressed within a single group (e.g., NP or PE pregnancies alone) is problematic. This is due to its sensitivity to low replicate numbers (Eisinga, Breitling and Heskes, 2013) and its inability to distinguish statistical consistency from biological relevance, as it does not account for the magnitude of expression differences or technical variability (Hong *et al.*, 2006; Schurch *et al.*, 2016). To address these constraints, the developed modified RP algorithm ranks genes within each sample of a single class and calculates a final RP value as the geometric mean of these ranks.

$$RP_{(g)} = \left(\prod_{i=1}^{k} r(g,i) \right)^{1/k}$$

Where:

1. r(g,i) is the rank of gene g in condition i.
2. k is the number of conditions.
3. The RP is the geometric mean of the ranks of a gene across multiple conditions.

Crucially, this method incorporates a two-step filtering process: first, genes are filtered for statistical significance (p value < 0.05) using permutation testing to assess the consistency of their rank product; second, from this statistically significant set, a threshold based on the Coefficient of Variation (CV) is applied to identify genes with both stable and biologically meaningful expression levels. The CV is defined as:

$$CV = Standard\ Deviation\ /\ Mean$$

This dual criterion approach, merging the rank-based consistency of RP with variability measurement, allowed for the robust classification of genes into Significant Consistent Genes (SCGs) and Biologically Relevant Consistent Genes (BRCGs), providing a tailored solution for defining gestational week specific signatures. The detailed algorithm, including the RP calculation formula and permutation testing procedure, is provided in Appendix B.

5.2.2.3 P–value Calculation

A p–value represents the probability of obtaining an observed result, or one more extreme, under the null hypothesis. In the context of Rank Product analysis, the null hypothesis posits that there is no consistent expression pattern among the genes across the different conditions or samples. A low p–value (typically < 0.05) indicates that the observed RP value is significantly different from what would be expected by chance, suggesting that the gene is consistently expressed across samples. To calculate p–values for the RP values, permutation testing is employed on both ascending and descending sorted ranked datasets. This approach is non–

parametric and does not rely on assumptions about the distribution of the data. By permuting the ranks of the genes, a distribution of Rank Products can be generated under the null hypothesis, allowing the observed values to be compared against this distribution.

5.2.2.4 Coefficient of Variation Calculation

The Coefficient of Variation (CV) is a valuable statistical measure for identifying genes that are consistently expressed across multiple samples in transcriptomic studies. As a dimensionless statistic, the CV provides a normalised measure of dispersion, allowing for meaningful comparisons of gene expression variability across different genes and datasets (Pélabon *et al.,* 2020). In the context of identifying consistently expressed genes, a low CV indicates consistent expression levels across samples, while a high CV suggests variable expression. Researchers often use CV thresholds to classify genes as stably expressed or variable. For instance, many studies proposed that housekeeping genes, which are typically consistently expressed, should have a CV < 0.3 across diverse tissue types . Sensitivity analysis was performed on both the p–value and CV thresholds to assess the robustness of the method. Based on this analysis, this study identified a p–value threshold of <0.05 and a CV threshold of <0.5 as optimal for filtering the genes. The calculation of CV for gene expression data typically involves computing the ratio of the standard deviation to the mean expression level for each gene across all samples:

$$CV = Standard\ Deviation\ /\ Mean$$

5.2.2.5 Identification of SCGs

The identification of the SCGs was carried out through a two–step filtration process that involved both the p–value and the CV. First, the p–value threshold was set at <0.05 to ensure statistical significance, while the CV threshold was set at <0.5 to capture genes with low variability in their expression across different samples. The initial step in the process involved sorting the RP values in ascending order, as

lower RP values are indicative of higher consistency in gene expression. RP analysis is used to assess the stability and reliability of gene expression patterns across multiple conditions or time points. Genes with lower RP values reflect a more consistent pattern of expression, which is essential for identifying potential SCGs that exhibit robust and reproducible changes. To assess the statistical significance of the genes, a permutation–based method was applied. Specifically, the p–values were calculated by performing 1000 permutations, where the gene expression data was randomly shuffled to generate a null distribution of RP values. This approach allows for the evaluation of whether the observed RP values for each gene are significantly different from what would be expected by chance. To ensure reproducibility and minimise bias, a random seed of 42 was used during the permutation procedure. This rigorous statistical framework enabled the identification of the most consistently expressed genes with significant p–values, which were then considered as SCGs for further investigation in the context of the study.

5.2.2.6 Identification of BRCGs

The identification of BRCGs was carried out through a modified two–step filtration process, inspired by the approach used for SCGs, with the key distinction that this method emphasised genes with biological relevance and consistent expression patterns across experimental conditions. The first step involved sorting the RP values of genes in descending order. Lower RP values indicate greater consistency in expression, whereas higher RP values suggest more variability. By sorting the RP values in descending order, genes exhibiting larger RP values were prioritised, reflecting greater variability or less consistency in expression across the conditions. These genes were selected for their potential biological relevance, despite their more fluctuating expression patterns, as their variability could indicate an important role in the biological processes being studied. After sorting the RP values, a permutation–based method was used to calculate the p–values for each gene, similar to the method for identifying SCG. A p–value threshold of <0.05 was

applied to filter out genes whose RP values did not show statistical significance, ensuring that only genes with meaningful and reproducible expression patterns were retained. To further refine the selection process and focus on genes with consistent expression, the CV was calculated for each gene. The CV measures the relative variability in gene expression, with lower CV values indicating more consistent expression across the samples. A CV threshold of <0.5 was applied to filter out genes with high variability in their expression levels. This threshold ensured that, despite the inclusion of genes with higher RP values (which indicate some variability), only those genes with low variability in their expression across conditions were retained. By applying this filter, the focus was directed toward genes that were not only biologically relevant but also exhibited consistent expression, even when expressed at relatively lower levels. In the final step, the genes that passed the p–value filtering and the CV threshold were considered as BRCGs. These genes met the dual criteria of biological relevance, as indicated by their significant RP values, and consistency in expression, as demonstrated by their low CV values. This selection process helped identify genes that were likely to play important roles in the biological processes under investigation, while maintaining a stable, reproducible pattern of expression across conditions.

5.2.3 Downstream analysis

5.2.3.1 Gene Ontology Analysis for SCGs and BRCGs per Gestational Week

Gene Ontology (GO) analysis of SCGs and BRCGs for each GW was conducted using rigorous significance thresholds. A p–value cut-off of 0.05 and a q–value cut-off of 0.2 were applied to ensure the results were statistically robust and accounted for multiple testing corrections. For each GW, the top 10 enriched GO terms were selected based on their statistical significance and relevance to the study. These terms were categorised into Biological Processes (BP), Molecular Functions (MF), and Cellular Components (CC). This analysis facilitated the identification of key biological processes, molecular activities, and cellular locations associated with

the unique variants, helping to uncover important mechanisms in the biological context of pregnancy.

5.2.3.2 Pathway Analysis for SCGs and BRCGs per Gestational Week

Pathway analysis for SCGs and BRCGs was conducted using the KEGG pathway database. The genes were mapped to known biological pathways to identify key pathways associated with their functions. A p–value cutoff of <0.05 was applied to select significantly enriched pathways, and multiple testing corrections were implemented to control false positives. The top 10 most significantly enriched KEGG pathways were selected for each group based on their statistical significance and relevance. This analysis provided a comprehensive understanding of the biological processes linked to SCGs and BRCGs, guiding further experimental studies.

5.3 Results and Discussion

The results presented in this chapter are derived from the analysis of transcriptomic data from NP and PE pregnancies across gestational weeks 8 to 16. The primary methodological approach involved the application of a novel Rank Product-based algorithm to define Unique Consistently Expressed Gene Signatures (UCEGS) for each gestational week. This method facilitated the identification of two key gene classes: Significant Consistent Genes (SCGs), which exhibit stable expression, and Biologically Relevant Consistent Genes (BRCGs), which combine stability with high expression levels. The subsequent sections detail the results of this analysis, starting with the distinct patterns of UCEGS observed between NP and PE pregnancies.

5.3.1 Identification of SCGs

The analysis of SCGs reveals notable differences in gene expression between NP and PE across various gestational weeks. In general, the count of SCGs is higher in NP than in PE, suggesting that during NP, there are more genes that are

consistently expressed at elevated levels. However, week 12 presents an intriguing exception, with PE showing a substantial increase in SCGs (252) compared to NP (87), indicating a possible shift in gene activity specific to this time point in PE (Table 17). This could represent an adaptive or pathological response in PE, with a subset of genes being more significantly activated. Week 9 is another notable outlier, with NP having 105 SCGs compared to just 3 in PE, further emphasizing the dynamic differences in gene expression between the two conditions. The lower number of SCGs in PE, especially early in pregnancy, suggests that the genetic landscape of PE may not involve high, consistent expression of genes, but rather a disruption or alteration in the regulation of these genes. The differences observed in SCG counts could help identify key genes involved in pregnancy progression and highlight possible targets for understanding or predicting PE.

5.3.2 Identification of BRCGs

In contrast to SCGs, the count of BRCGs is consistently higher in PE than in NP, suggesting that genes with lower and more stable expression levels are more prominently involved in PE. This trend is particularly pronounced at week 12, where PE exhibits a dramatic increase in BRCGs (2708 in PE versus 652 in NP) (Table 17). This spike could indicate a unique biological mechanism at play in PE, were consistent, low–level gene expression may be linked to the pathophysiology of PE. The higher presence of BRCGs in PE across most weeks' points to a distinct genetic profile that could differentiate PE from NP. Weeks 13 to 16 continue to show a higher prevalence of BRCGs in PE compared to NP, albeit with more balanced numbers between the two groups. The consistent presence of these genes in PE suggests that they may play a crucial role in the disease's development, potentially influencing processes such as vascular dysfunction, immune response, or inflammation. The higher expression of BRCGs in PE may serve as a biomarker for early detection or for understanding the underlying mechanisms of PE, providing valuable insight into its biological processes.

Table 17: Consistent genes across gestational weeks from 8 to 16 in NP and PE

Gestational Weeks	Significant Consistent Genes (SCGs)		Biologically Relevant Consistent Genes (BRCGs)		UNIQUE SCGs		UNIQUE BRCGs	
	NP	PE	NP	PE	NP	PE	NP	PE
8	1710	674	1491	105	47	2	208	39
9	2200	987	2188	607	105	3	356	253
10	1170	941	4052	381	6	5	749	133
11	1194	741	5622	1929	32	0	1313	330
12	2161	1696	3628	5042	87	252	652	2708
13	2558	1477	2585	1864	404	28	740	411
14	1368	362	1142	371	14	9	198	109
15	684	919	1254	559	3	3	229	104
16	799	674	4544	731	5	2	997	316

5.3.3 Downstream Analysis for Unique Consistent Genes

5.3.3.1 Gene Ontology Analysis

The GO analysis at GW 8 reveals distinct biological processes, cellular components, and molecular functions associated with NP and PE, highlighting their differing regulatory mechanisms. In NP, significant processes include regulation of lysosome organisation, Fc–epsilon receptor signalling, and negative response regulation to reactive oxygen species involving genes like Lysosomal Associated Protein Transmembrane 4 Beta (LAPTM4B), Mucolipin 1 (MCOLN1), Leucine-Rich Repeat Kinase 2 (LRRK2), and Thioredoxin (TXN) (Figure 16). These processes emphasise cellular homeostasis, oxidative stress management, and developmental signalling essential for normal foetal growth. Key cellular components, such as the lysosomal lumen, also support these functions. In contrast, PE is marked by enrichment in processes like negative regulation of innate immune response, inflammatory response, and vascular–associated smooth muscle cell differentiation, with involvement of genes like Signal

Transducer and Activator of Transcription (STAT2), V-Set and Immunoglobulin Domain Containing 4 (VSIG4), Peroxiredoxin 5 (PRDX5), and NLR Family Pyrin Domain Containing 6 (NLRP6). These findings underscore immune dysregulation, inflammation, and abnormal vascular remodelling, which are central to PE pathophysiology. Additionally, PE GO terms like positive regulation of skeletal muscle fibre development suggest aberrant tissue growth. The contrasting profiles indicate that while NP focuses on maintaining cellular stability and development, PE is characterised by immune activation and vascular complications, with key genes such as LRRK2, TXN (NP) and STAT2, VSIG4 (PE) emerging as potential biomarkers or therapeutic targets.

At GW 9, GO analysis revealed distinct biological processes, cellular components, molecular functions associated with NP and PE, and key genes contributing to these pathways. In NP, significant processes included protein O-linked mannosylation (GO:0035269) and regulation of translation (GO:0006417), with key genes such as Protein O-Mannosyltransferase 1 (POMT1) and Eukaryotic Translation Initiation Factor 2B Subunit Alpha (EIF2B1) playing pivotal roles (Figure 17). Notable cellular components involved the autophagosome membrane (GO:0000421) and RNA polymerase III complex (GO:0005666), highlighted by genes like Microtubule-Associated Protein 1 Light Chain 3 Beta (MAP1LC3B) and RNA polymerase III subunit A (POLR3A), respectively. Molecular functions such as 3'–5' RNA helicase activity (GO:0034458) and xylosyltransferase activity (GO:0042285) were underpinned by genes including DEAH-Box Helicase 36 (DHX36) and Xylosyltransferase 1 (XYLT1). Conversely, in PE, prominent processes included positive regulation of chromosome segregation (GO:0051984), protein exit from the endoplasmic reticulum (GO:0032527), and response to ionising radiation (GO:0010212), with genes like Budding Uninhibited by Benzimidazoles 1 (BUB1), Sec61 Translocon Alpha 1 Subunit (SEC61A1), and Ataxia-Telangiectasia Mutated (ATM) contributing to these pathways. Significant cellular components comprised the beta–catenin–TCF complex (GO:1990907) and transcription factor TFTC complex (GO:0033276), linked to genes such as

Catenin Beta 1 (CTNNB1) and TATA-Box Binding Protein Associated Factor 4 (TAF4). Molecular functions emphasised S–adenosylmethionine–dependent methyltransferase activity (GO:0008757) and protein carboxyl O–methyltransferase activity (GO:0051998), driven by genes like Enhancer of Zeste Homolog 2 (EZH2) and Protein-L-isoaspartate (D-aspartate) O-methyltransferase 1 (PCMT1). These findings underscore the intricate and unique molecular mechanisms active in NP and PE during this early gestational stage, offering valuable insights into the biology of pregnancy and its complications.

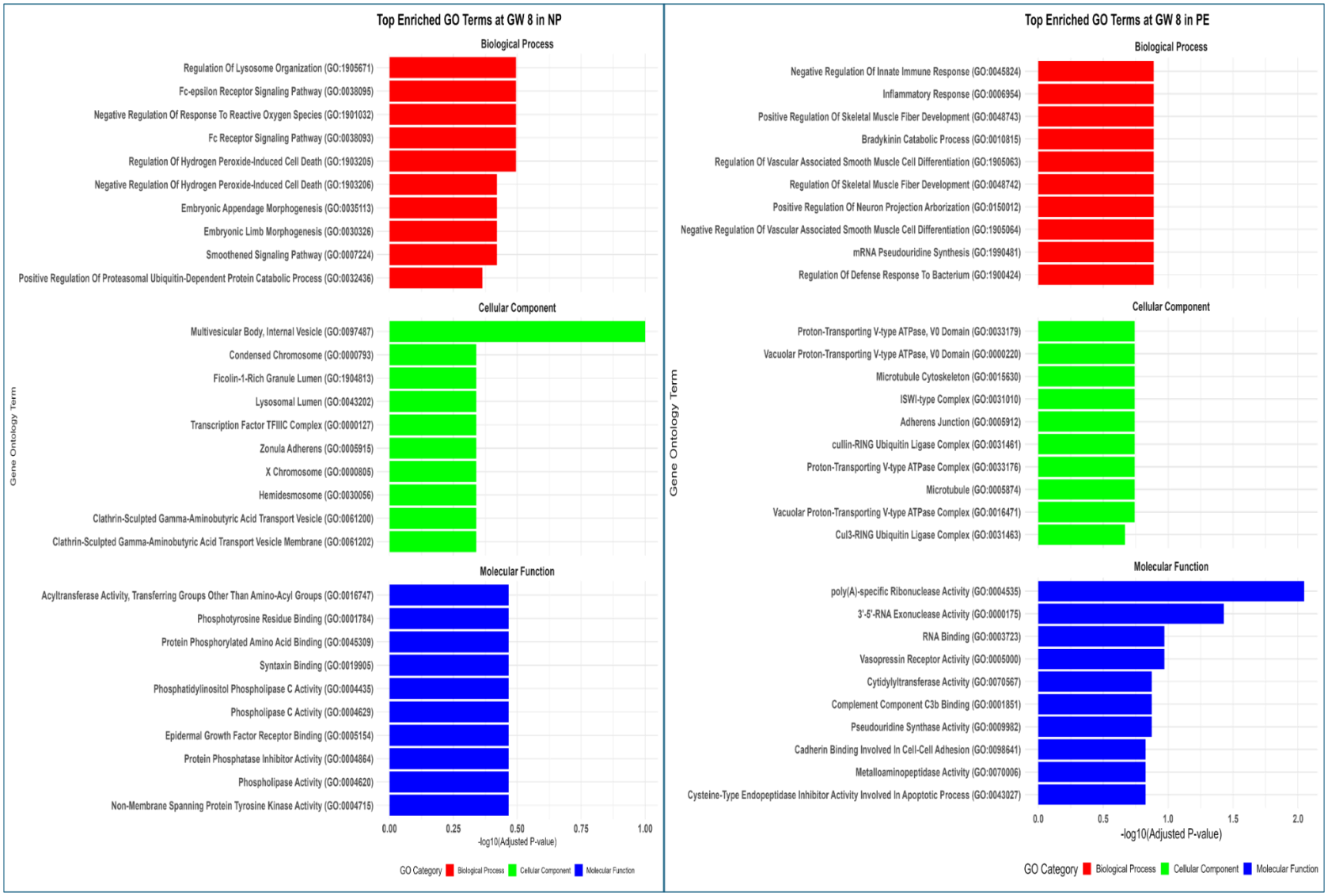


Figure 16: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 8 in NP and PE

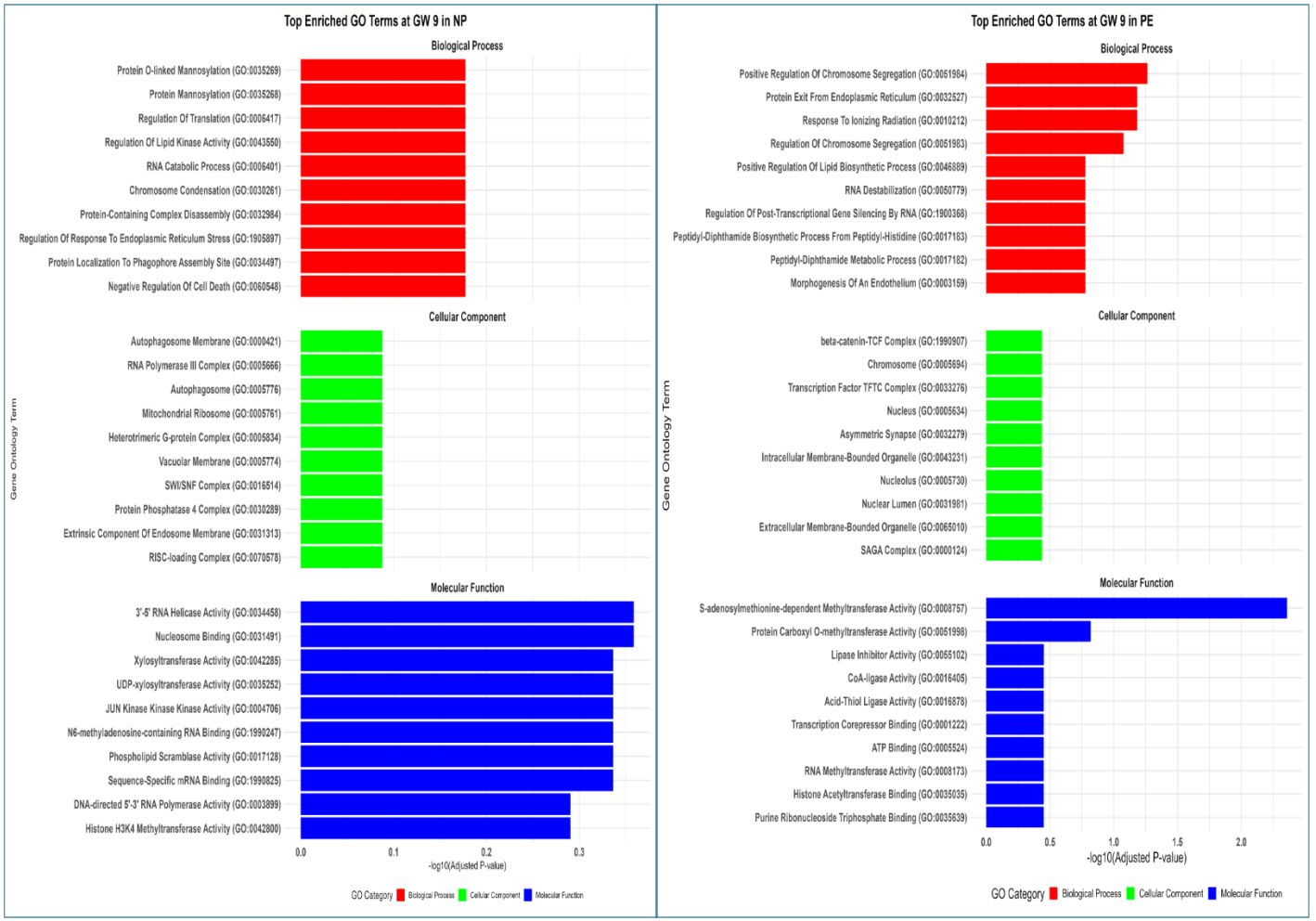


Figure 17: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 9 in NP and PE

At GW 10, NP samples exhibited significant enrichment in biological processes such as the cellular response to exogenous dsRNA (e.g., genes IFIH1, STING1, IFNB1, FLOT1, IFIT1), positive regulation of reactive oxygen species biosynthetic process (RAB27A, TLR6, SLC5A3, TLR4), and the TRIF–dependent Toll–like receptor signalling pathway (TLR6, TICAM1, TLR4) (Figure 18). These processes highlight immune activation, oxidative stress regulation, and signalling pathways integral to cellular defence. In cellular components, notable pathways included the endosome membrane (ABCB6, SNX12, MTMR4, RAB21, MAP3K7), late endosome membrane (PMEL, CHMP2B, RAB27A, LAMTOR2), and the multivesicular body organisation (CHMP2B, VPS25, RAB27A), suggesting active intracellular trafficking and vesicle–mediated signalling. Molecular functions such as acid phosphatase activity (MINPP1, ACP5) and fatty acid elongase activity (ELOVL5, ELOVL6) were also significant. In contrast, PE samples demonstrated notable involvement in the negative regulation of stress–activated protein kinase signalling cascade (PPIA, FOXO1), response to type II interferon (GBP5, GBP2, GBP1, MEFV), and microtubule anchoring at the centrosome (MAP1S, MBD1). These processes are linked to immune response modulation, cytokine signalling, and cytoskeletal reorganisation. The cellular components prominently involved included the nuclear membrane (RNF123, TBC1D20, MBD1, DNAJB12), rough endoplasmic reticulum membrane (UBA1), and vesicles (GBP1, NRXN1, ATP1A1), pointing to structural adaptations and vesicular transport. Key molecular functions in PE included purine ribonucleoside triphosphate binding (GBP5, STK17B, RHOC), cytoskeletal protein interactions (DNAJB12, MAP1S), and Hsp70 protein binding (DNAJB12, DNAJC8). The overlapping yet distinct processes and gene involvements between NP and PE at week 10 underscore the intricate molecular and cellular dynamics during early pregnancy and their potential implications in the pathogenesis of PE.

At GW 11, NP samples were notably enriched in biological processes such as the negative regulation of the establishment of protein localisation to mitochondrion (DNAJA1, LMAN1, C11ORF65), oestrogen biosynthetic process (HSD3B1,

HSD17B11, CYP19A1), and positive regulation of protein kinase C signalling (MC1R, FLT4, ADRA1D), highlighting mitochondrial function, steroid hormone synthesis, and kinase signalling. Additionally, pathways related to negative regulation of nuclear–transcribed mRNA catabolic process (SYNCRIP, TENT4A, TENT4B) and protein polyglutamylation (TTLL6, TTLL5, TTLL4) were significant, indicating gene expression regulation and post–translational modifications (Figure 19). The cellular components involved included the U2–type catalytic spliceosome (PPIL1, SNRNP40, CWC22, SNRPG, RBM22), nuclear membrane (NDC1, MINDY3, NUP155, MAD2L1BP, TXLNG), and filopodium (KITLG, MYO1B, RDX, TWF1), suggesting active splicing and cellular structure organisation. Molecular functions such as tubulin–glutamic acid ligase activity (TTLL6, TTLL5, TTLL4) and stat family protein binding (FAM220A, PTPRT, PPARG) were prominent, highlighting cytoskeletal modification and transcription regulation. PE samples at GW 11 showed significant enrichment in processes like linoleic acid metabolic process (GSTA1, ALOX12B, ALOX15B), positive regulation of urine volume (NPPB, NPR3), and unsaturated fatty acid metabolic process (GSTA1, ALOX12B, PTGR1, ALOX15B), indicating lipid metabolism and regulation of kidney function. The regulation of vesicle fusion (RAB3A, SPHK1) and ceramide biosynthetic process (SPHK1, ORMDL3) also emerged as significant, highlighting membrane trafficking and lipid signalling. The cellular components enriched in PE included potassium channel complexes (KCNB1, KCNMA1, KCNJ14, GRIK1) and postsynaptic density membrane (DLG3, GRIK1), suggesting alterations in ion transport and synaptic function. Key molecular functions in PE included ionotropic glutamate receptor activity (GRIK1, GRIN1), oxidoreductase activity (DHCR7, PTGR1), and phospholipase activity (DDHD1, PLD3, ABHD12B), implicating neurotransmission, oxidative stress, and phospholipid metabolism. These differential gene enrichments between NP and PE highlight distinct molecular processes underlying pregnancy complications at this stage.

For NP at GW 12, several genes are significantly associated with enriched GO terms, indicating key biological processes. These include genes involved in O–

glycan processing, such as GALNT12, C1GALT1C1, and B4GAT1, as well as those engaged in glycosyl compound biosynthesis, like TRMT12 and TYW1B (Figure 20). Additionally, genes related to nucleobase biosyntheses, such as SHMT2 and UMPS, and those implicated in positive regulation of autophagosome maturation, like FYCO1 and IRGM, are significantly enriched. Other vital genes, including PLK2 and PICK1, are involved in synaptic functions, and DDX19A and MCM3AP are involved in RNA processing. In PE at the same gestational week, genes are enriched in processes like phosphate ion homeostasis, with key players like GCM2, SLC34A2, and SLC34A1, and G protein–coupled receptor internalisation, with genes like ARRB1, GTF2H2, and DRD3. Muscle–related functions are represented by TNNT2, MYH8, MYH4, and MYH7, while genes involved in cAMP–mediated signalling, such as GLP1R, PDE4D, and AHR, are also significant. Other notable genes in PE include NPY5R and TBX5, which are related to cardiac left ventricle morphogenesis, and BMP4 and FGFR2, which are involved in ureteric bud development. Moreover, serotonin receptor activity is highlighted by genes like HTR7 and HTR1E, suggesting potential neurotransmitter dysregulation in PE. These findings indicate distinct biological pathways in NP and PE, with NP focusing on glycosylation, autophagy, and synaptic functions. At the same time, PE is enriched for genes involved in muscle function, receptor activity, and signalling pathways related to cardiovascular and kidney development.

The gene expression analysis in NP and PE at GW 13 reveals distinct biological and molecular processes, highlighting the differences between these two conditions. In NP, significant processes include cellular responses to epidermal growth factor, positive regulation of locomotion, and histone mRNA catabolism, with genes like DUSP3, CSF1R, and UPF1 involved in these pathways, which suggest active cell migration, tissue remodelling, and immune responses critical for placental development and function. Additionally, genes regulating the actin cytoskeleton, such as APC2, RAC1, and WAS, are implicated in cell shape and migration, processes essential for placental invasion (Figure 21). Molecular functions like oxidoreductase activity involving genes such as DUOX1 and NCF1

indicate the importance of managing oxidative stress in a healthy pregnancy. In contrast, in PE, processes such as substrate adhesion–dependent cell spreading and negative regulation of cation transport are significantly altered, with genes like FGA, NEDD4L, and MMP9 playing key roles. These changes may contribute to impaired placental function, endothelial dysfunction, and altered vascular tone. Additionally, the involvement of mitochondrial function, as seen in genes like NDUFA12 and VPS9D1, suggests energy production disturbances in PE, which could exacerbate cellular stress. Alterations in ion transport and vasopressin receptor activity through genes like PLN and INPP5K further highlight disruptions in vascular regulation and fluid balance, central features of PE. Overall, the findings underscore the distinct gene expression signatures in NP and PE, offering insights into the underlying mechanisms of pregnancy health and complications.

For GW 14 in NP samples, several biological processes are significantly enriched. These include the ER–associated degradation (ERAD) pathway, with genes like RNF139, CCDC47, and SGTA, and protein C–linked glycosylation, with genes such as DPY19L2 and DPY19L3. Negative regulation of T–cell proliferation involves genes like PLA2G2F and TNFRSF21, while CLTB and NCALD represent clathrin vesicle coat functions. Significant molecular functions include mannosyltransferase activity (with genes such as DPY19L2 and TMTC3) and mRNA 5'–UTR binding with RPS7 and NCL (Figure 22). Cellular components include the clathrin–coated vesicle and nuclear inner membrane, with CLTB, AP2B1, and DPY19L2. In PE samples at GW 14, key biological processes include post–translational protein targeting to the membrane, involving SEC61A2 and SEC61A1, and positive regulation of cell motility, with genes like SDCBP, CARMIL2, and PDGFB. Genes associated with macropinosome formation include CARMIL2, and genes involved in regulating chemotaxis and autophagosome assembly include PDGFB, CDH13, and LRSAM1. In terms of cellular components, the small ribosomal subunit (RPS19, RPS3A) and caveola (PLVAP, CDH13) are significant. Molecular functions such as palmitoyl–CoA hydrolase activity (ACOT9,

PLA2G6) and signal sequence binding (SEC61A2, SEC61A1) also play a role in PE pathogenesis.

At GW 15 in NP samples, several biological processes are significantly enriched. These include negative regulation of mRNA splicing via the spliceosome, with key genes SFSWAP and SRSF7, as well as mRNA processing (SFSWAP, SRSF7), mRNA cis splicing (SFSWAP, CWC15), and negative regulation of RNA splicing (SFSWAP, SRSF7) (Figure 23). Other notable processes include positive regulation of phospholipase activity (ARL1, AGTR1), limb development (SMOC1, NOG), and acyl–CoA metabolic processes (ACSM4, HMGCS2). Significant cellular components include the apicolateral plasma membrane (CXADR) and peroxisomal matrix (AGXT, HAO2), while the large ribosomal subunit rRNA binding (RPLP0P6) and chondroitin sulfotransferase activity (CHST3) are notable molecular functions. In PE samples at GW 15, biological processes such as the regulation of mitotic nuclear division (TOM1L1, BORA, CAV2) and activation of protein kinase activity (TOM1L1, BORA, TNFRSF10B) are enriched. Additional enriched processes include positive regulation of locomotion (CAVIN1, SPOCK2), glycolipid biosynthesis (B3GNT5, PIGA), and regulation of the cell cycle (BORA, CAV2, MARK4). Key cellular components include the TORC2 complex (MLST8), caveola (CAVIN1, CAV2), and plasma membrane rafts (CAVIN1, CAV2, LAX1). Molecular functions such as kinase binding (TOM1L1, BORA, CAV2, NBEAL1, LAX1, DNM1, INKA1), gamma–tubulin binding (MARK4, RAB11FIP5), and acetylglucosaminyltransferase activity (B3GNT5, PIGA) are also significantly enriched.

In GW 16, significant differences in molecular pathways are observed between NP and PE. In NP, processes like "Positive Regulation of Cellular Biosynthetic Process" are enriched, with genes such as RBM3 and SREBF1 playing key roles in protein synthesis and cell metabolism (Figure 24). This suggests that NP supports cellular growth and biosynthesis, essential for foetal development. In contrast, PE shows distinct activation of pathways like "Regulation of Systemic Arterial Blood Pressure" involving genes such as GSK3A and DDAH1, reflecting

the vascular and hypertensive abnormalities characteristic of PE. PE is further marked by dysregulated blood pressure homeostasis, contributing to organ dysfunction. Cellular component enrichment also highlights differences. In NP, genes related to "Autophagosome" (e.g., GABARAPL3, CALCOCO2) are prominent, suggesting active cellular maintenance and adaptive mechanisms, including autophagy for quality control. In PE, however, components such as "Microtubule" and "Polymeric Cytoskeletal Fiber" (e.g., EML4, DYNC1I2) are more significant, indicating abnormal cytoskeletal dynamics that disrupt endothelial integrity and vascular remodelling, key features of PE pathology. At the molecular function level, NP shows enrichment in "RNA Binding" (e.g., TUT7, DDX46), suggesting normal RNA processing and regulation of gene expression during pregnancy. In PE, "Metalloendopeptidase Inhibitor Activity" (e.g., TIMP2, TIMP3) is enriched, indicating abnormal extracellular matrix remodelling, which is impaired in PE and contributes to tissue fibrosis and endothelial dysfunction. The comparison of NP and PE in gestational week 16 highlights divergent biological pathways, with PE characterised by altered vascular regulation, immune dysregulation, and abnormal cytoskeletal remodelling. These molecular differences present potential biomarkers for early detection and therapeutic targets for managing PE.

The GO analysis of NP and PE across various gestational weeks reveals distinct biological processes, cellular components, and molecular functions, highlighting the differing regulatory mechanisms in both conditions. In NP, cellular homeostasis, development, and oxidative stress management processes are enriched, with genes like Lysosomal Protein Transmembrane 4 Beta (LAPTM4B), MCOLN1, and TXN supporting lysosomal organisation, immune signalling, and reactive oxygen species regulation. In contrast, PE is marked by immune dysregulation, inflammation, and vascular abnormalities, with genes like STAT2, VSIG4, and NOD-like receptor family pyrin domain containing 6 (NLRP6) associated with immune response regulation and smooth muscle cell differentiation. At the cellular component level, NP shows enrichment in autophagic

processes and vesicle–mediated signalling involving genes like GABA Type A Receptor-Associated Protein Like 3 (GABARAPL3) and Microtubule-Associated Protein 1 Light Chain 3 Beta (MAP1LC3B), emphasising cellular maintenance. PE, however, shows significant changes in components related to the cytoskeleton and extracellular matrix, with genes like EML4, DYNC1I2, and TIMP2 implicated in disrupted vascular remodelling and endothelial dysfunction. Molecular function analysis further underscores differences, with NP showing RNA binding activity (e.g., TUT7, DDX46) and PE enriched in metalloendopeptidase inhibitor activity (e.g., TIMP2, TIMP3), pointing to impaired matrix remodelling in PE. These findings reveal the molecular distinctions between NP and PE, offering insights into potential biomarkers and therapeutic targets for early intervention in PE.

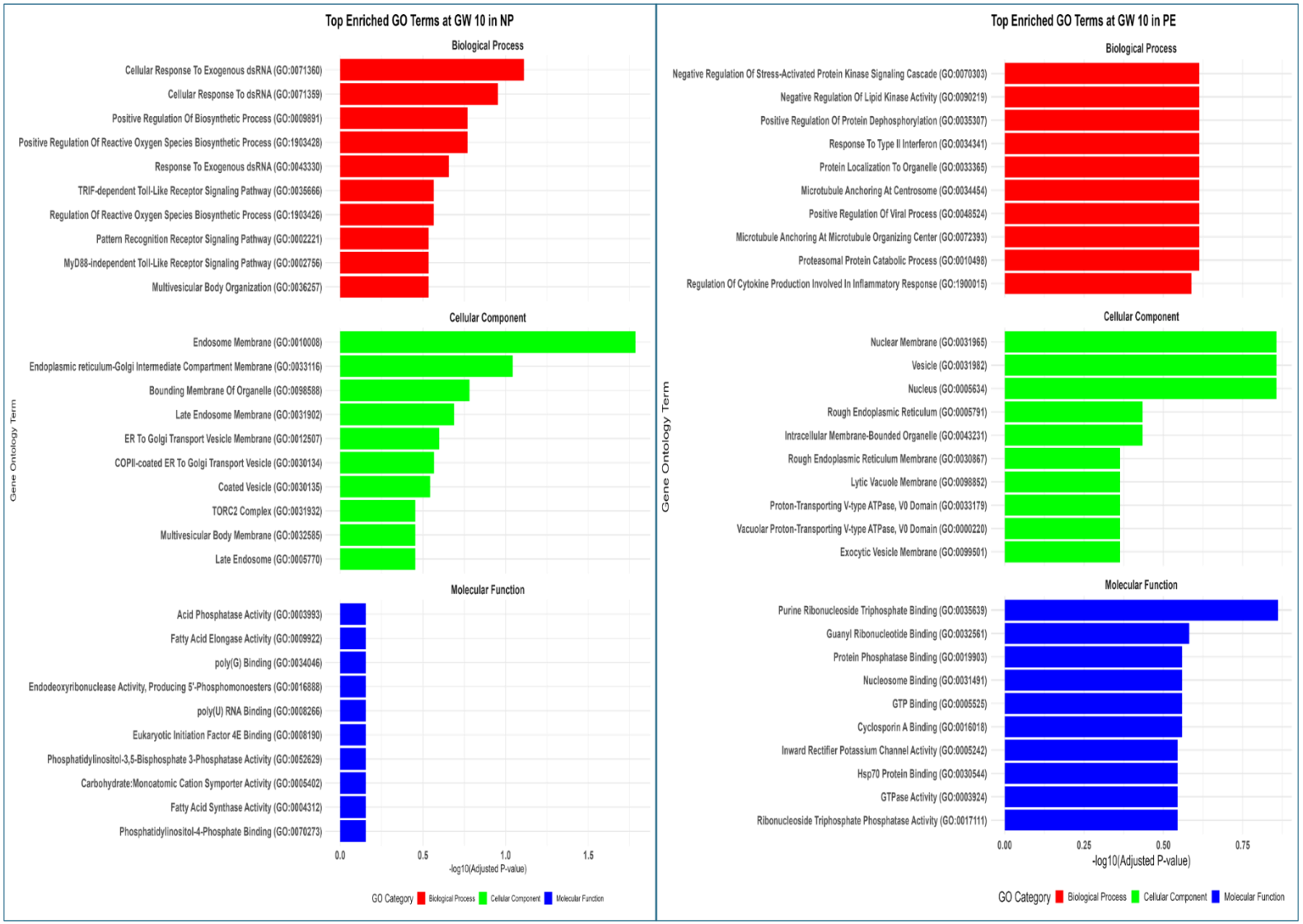


Figure 18: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 10 in NP and PE



Figure 19: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 11 in NP and PE

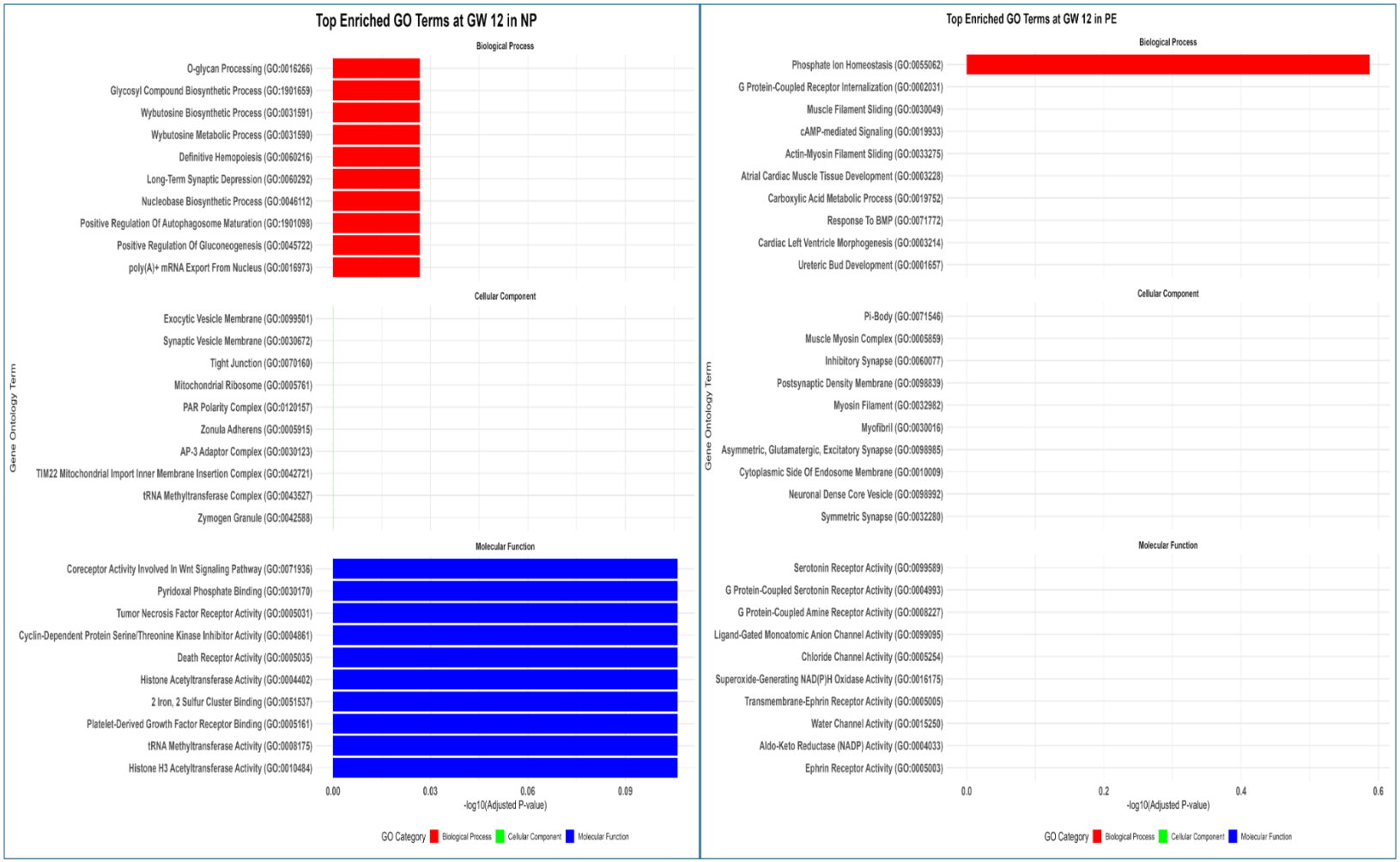


Figure 20: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 12 in NP and PE

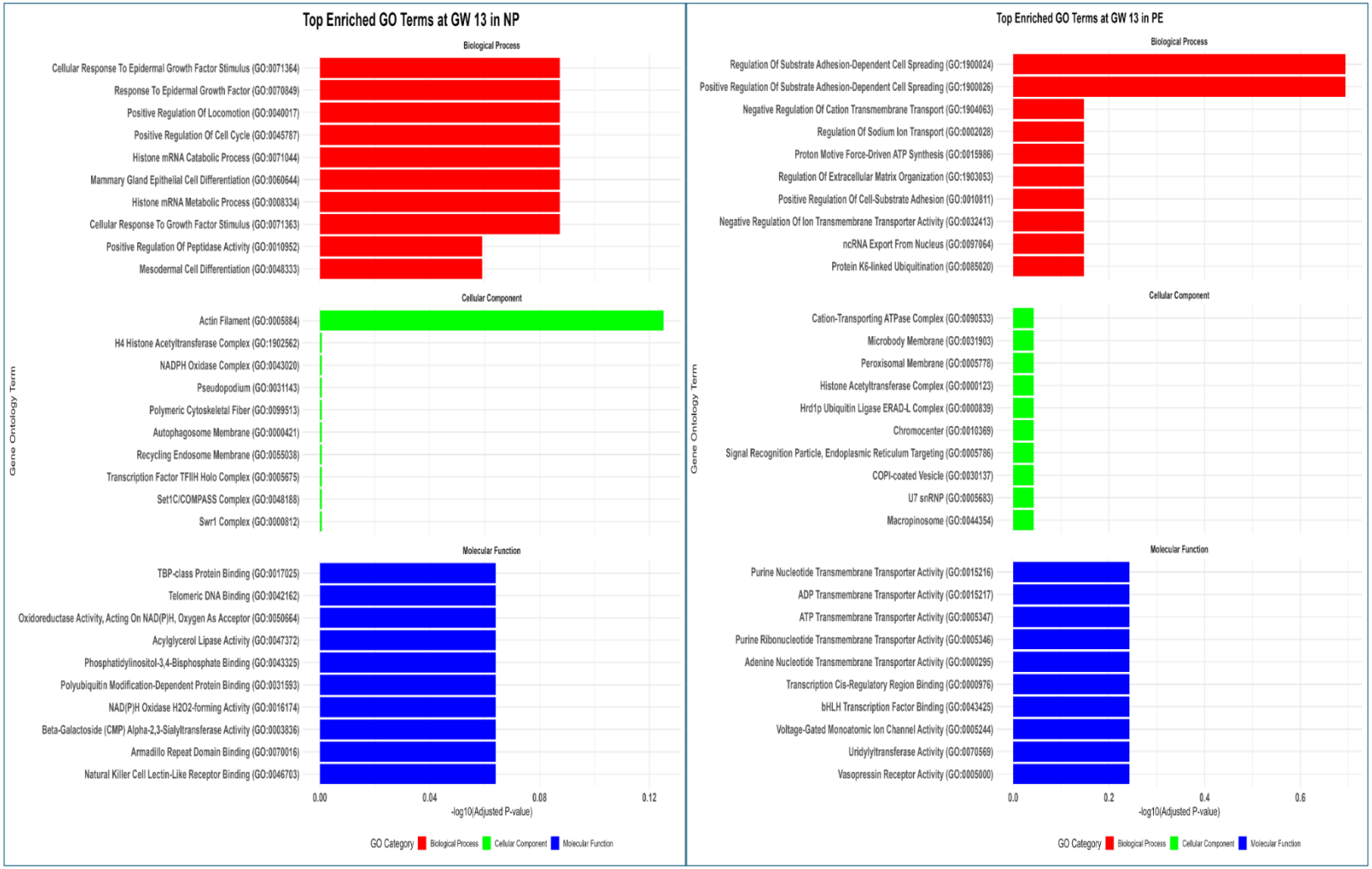


Figure 21: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 13 in NP and PE

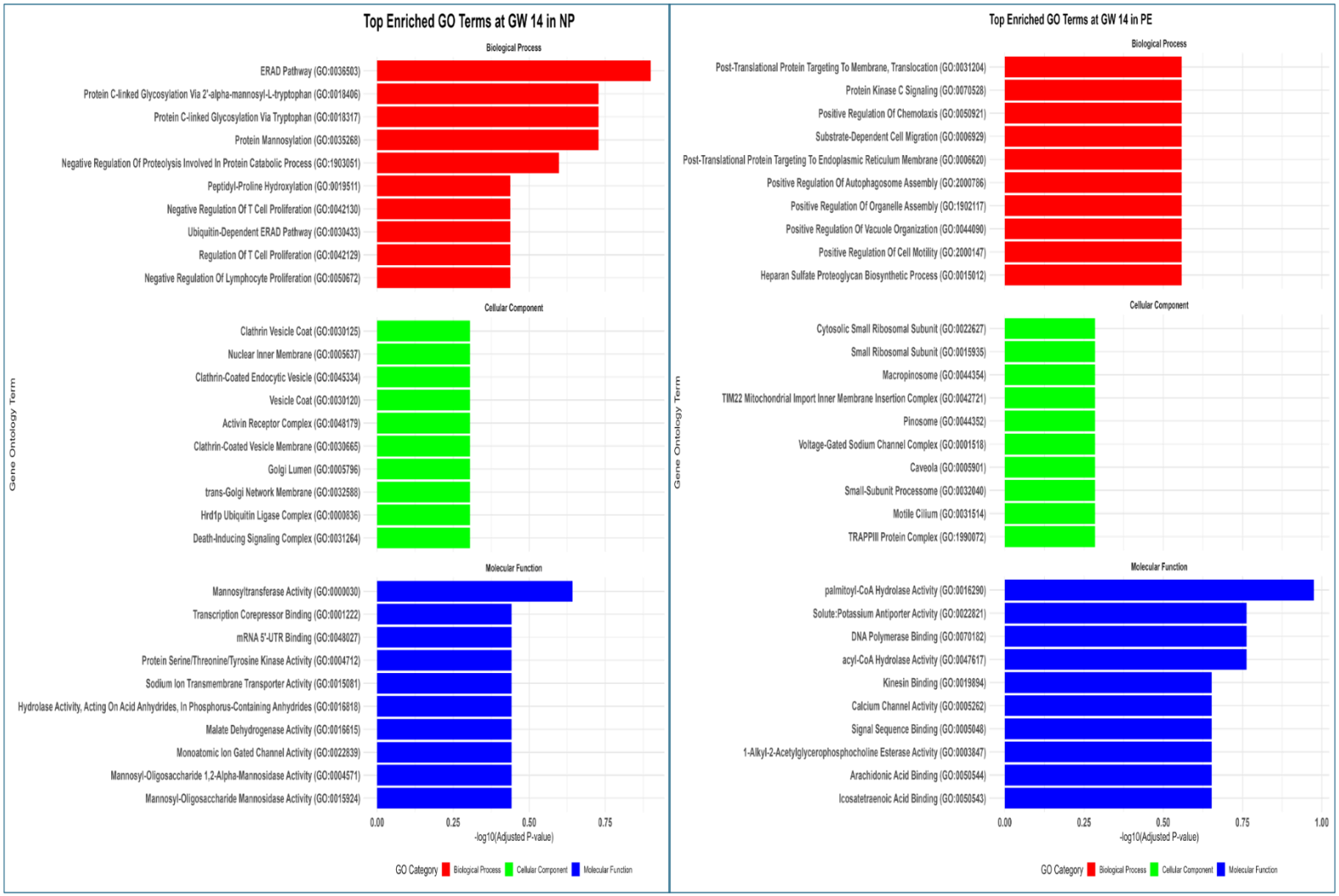


Figure 22: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 14 in NP and PE

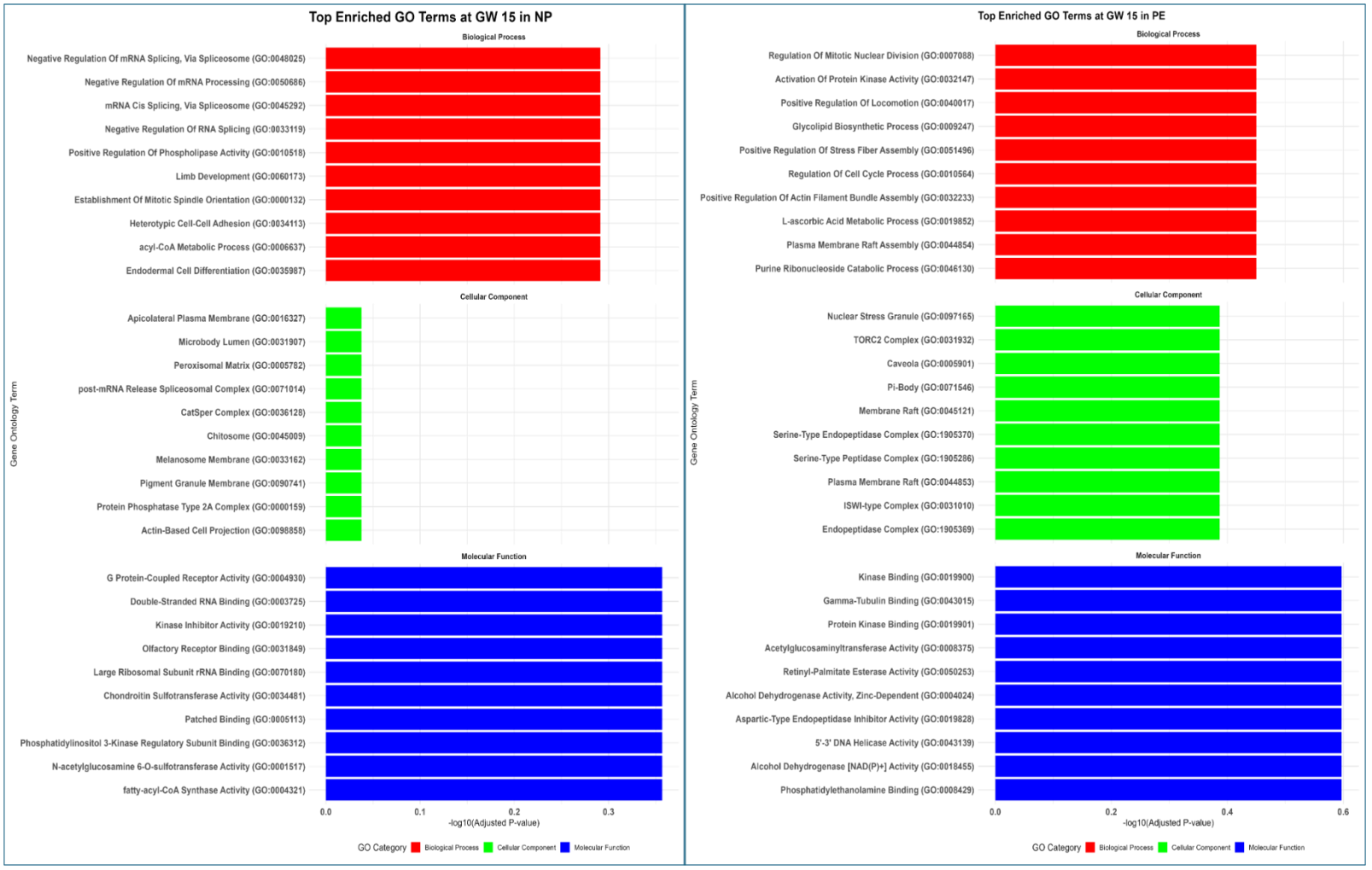


Figure 23: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 15 in NP and PE



Figure 24: Top 10 Gene Ontology term of Unique Consistent Genes at gestational week 16 in NP and PE

5.3.3.2 Pathway Analysis

At GW 8, the KEGG pathway analysis highlights distinct molecular mechanisms between NP and those complicated by PE. In NP, the Chemokine signalling pathway (genes: ITK, PLCB4, SHC1, PLCG2, VAV2) and Insulin signalling pathway (HK3, SHC1, FASN, MKNK2) are notably enriched, reflecting their roles in immune modulation and metabolic regulation crucial for early pregnancy. Other pathways, such as the ErbB signalling pathway (SHC1, ABL1, PLCG2) and AGE–RAGE signalling in diabetic complications (IL1A, PLCB4, PLCG2), underscore the importance of signalling cascades that support cellular growth, inflammation, and vascular adaptation. Conversely, in PE, the Phosphonate and phosphinate metabolism pathway (PCYT1A), C–type lectin receptor signalling pathway (STAT2, NFATC3), and NOD–like receptor signalling pathway (NLRP6, STAT2) are significant, highlighting dysregulated immune responses and inflammation. The Wnt signalling pathway (NFATC3, DVL3) also emerges, implicating disrupted cellular communication and trophoblast invasion. These findings underline the potential role of immune and metabolic dysregulation and impaired vascular remodelling in the progression and early prediction of PE. Notably, genes such as STAT2, NFATC3, and DVL3 in PE pathways serve as critical biomarkers, offering valuable insights for targeted diagnostics and therapeutic interventions.

At GW 9, KEGG pathway analysis reveals a divergence in molecular mechanisms between NP and PE. In NP, the Neurotrophin signalling pathway (CAMK2B, RPS6KA5, IRAK2, RIPK2, RPS6KA1, MATK, BCL2) is prominent, emphasising its role in neural development and survival, which are essential for maternal–foetal communication. Pathways such as RNA polymerase (POLR3A, POLR1A, POLR3D) and Necroptosis (CAMK2B, AIFM1, H2AC4, H2AC19, BCL2, CHMP3, TRAF2) further highlight processes crucial for transcription regulation and programmed cell death. These findings underscore the importance of regulated signalling and metabolic processes in supporting pregnancy at this stage. In PE, pathways such as Acute myeloid leukaemia (IKBKB, STAT5A, CEBPE, BAD),

Th17 cell differentiation (STAT5A, IKBKB, IRF4, PLCG1, TYK2), and Adipocytokine signalling (IKBKB, STK11, ACSL1, ACACB) are significantly enriched. These highlight immune dysregulation, chronic inflammation, and altered metabolic signalling as hallmarks of PE pathophysiology. Notably, Fatty acid biosynthesis (ACSL1, ACACB) is unique to PE, indicating disrupted lipid metabolism that could contribute to the vascular dysfunction observed in PE. Key genes such as Signal Transducer and Activator of Transcription 5A (STAT5A), Inhibitor of Nuclear Factor Kappa B Kinase Subunit Beta (IKBKB), and Acyl-CoA Synthetase Long-Chain Family Member 1 (ACSL1) in PE–related pathways could serve as biomarkers for early detection and understanding of disease mechanisms, aiding in potential therapeutic interventions. These pathways reflect distinct immune, metabolic, and signalling disruptions critical to PE progression.

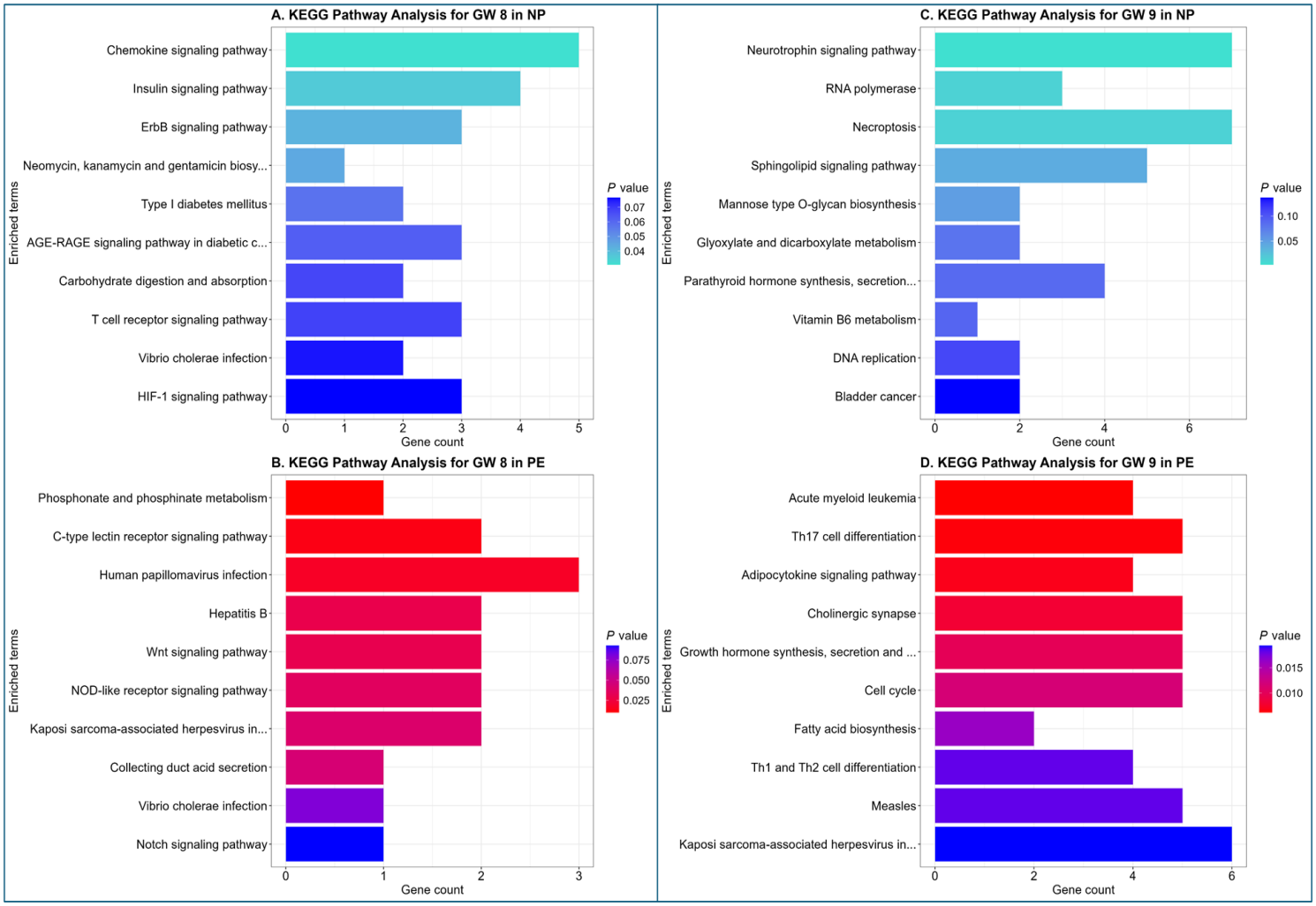


Figure 25: KEGG pathway analysis for unique, consistent genes at GW 8 and 9 in NP and PE

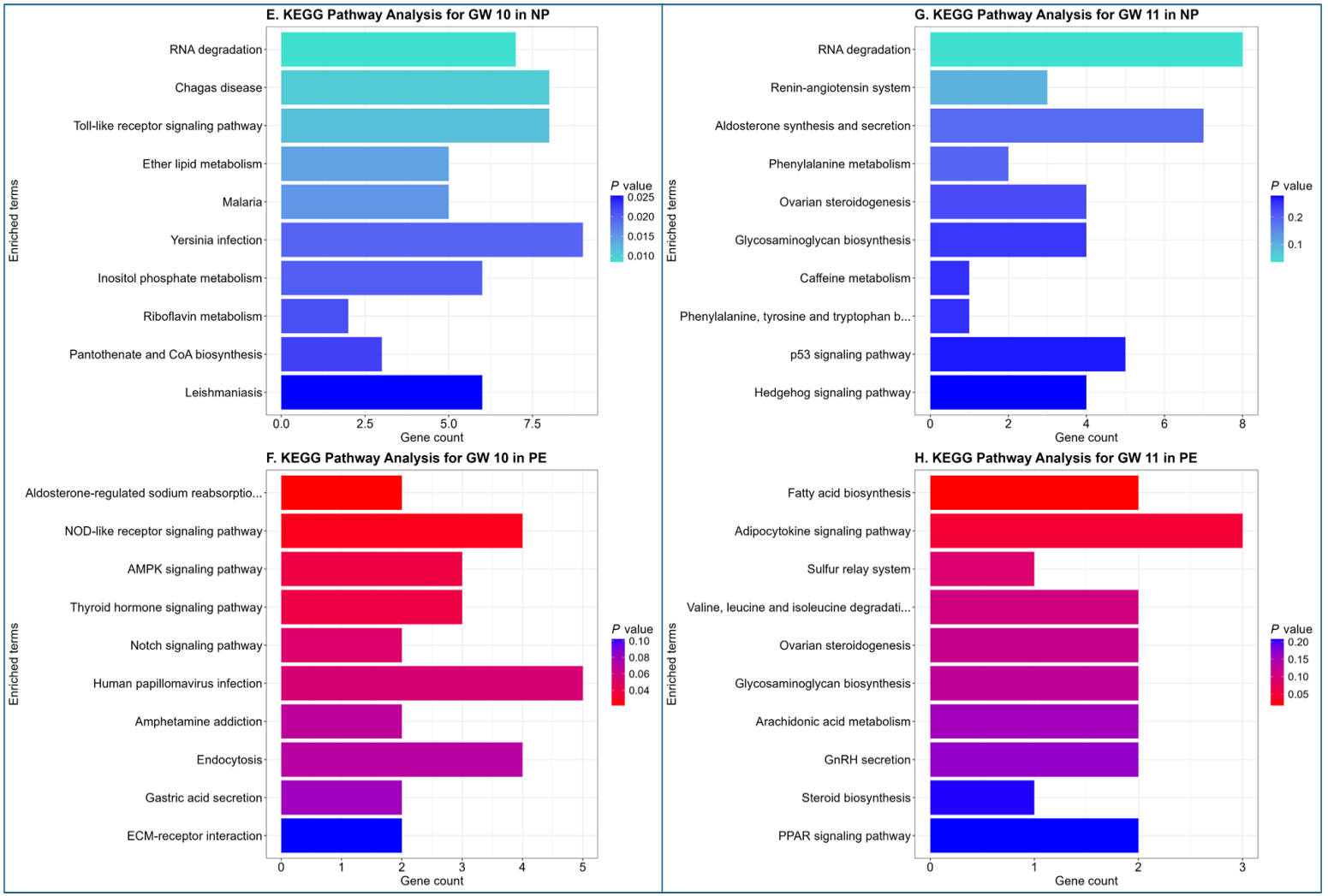


Figure 26: KEGG pathway analysis for unique, consistent genes at GW 10 and 11 in NP and PE

At GW 10, pathway analysis highlights distinct molecular signatures between NP and PE. In NP, key pathways include RNA degradation (HSPA9, LSM8, EXOSC5, PNPT1, C1D, TOB2, CNOT9), reflecting tight regulation of RNA turnover vital for placental development. The Toll–like receptor signalling pathway (MAP2K4, IFNB1, IL1B, TLR6, TICAM1, MAP3K7, TLR4, TLR2) and Chagas disease pathways emphasise innate immune activity, likely essential for maternal–foetal immune tolerance. Additionally, metabolic pathways such as Ether lipid metabolism (PLA2G4E, PEDS1, PLD1, PLB1, PAFAH1B2) highlight lipid remodelling critical for foetal membrane stability. Conversely, in PE, significant pathways include Aldosterone–regulated sodium reabsorption (ATP1A1, KCNJ1), which might contribute to the blood pressure dysregulation characteristic of PE. The NOD–like receptor signalling pathway (GBP5, MEFV, GBP2, GBP1) underscores immune dysregulation and inflammation. The AMPK signalling pathway (LEP, PPP2R5A, FOXO1) points to altered energy metabolism, a hallmark of PE pathology. Additional pathways like Thyroid hormone signalling and Notch signalling suggest disruptions in hormonal signalling and cell fate determination. Genes such as Interleukin 1 Beta (IL1B), ATPase Na^+/K^+ Transporting Subunit Alpha 1 (ATP1A1), and Forkhead Box O1 (FOXO1) emerge as significant in distinguishing these conditions. Their roles in immune modulation, metabolic processes, and vascular homeostasis may provide insights into PE pathogenesis and potential biomarkers for early detection.

At GW 11, pathway analysis reveals distinct molecular mechanisms between normal NP and PE. In NP, key pathways include RNA degradation involving genes like Like-Sm Protein 6 (LSM6), Exosome Component 9 (EXOSC9), and MTR4, RNA Helicase, EXosome Cofactor (MTREX), which are critical for post–transcriptional regulation and gene expression and are essential for placental function. The renin–angiotensin system and aldosterone synthesis and secretion pathways, featuring genes such as Carboxypeptidase A3 (CPA3), Prolyl Endopeptidase (PREP), Hydroxy-Delta-5-Steroid Dehydrogenase, 3 Beta- and

Steroid Delta-Isomerase 1 (HSD3B1), and Scavenger Receptor Class B Member 1 (SCARB1), highlight their roles in maintaining vascular homeostasis and electrolyte balance. Additionally, pathways like ovarian steroidogenesis (SCARB1, HSD3B1, CYP19A1) and phenylalanine metabolism (AOC3, GOT2) emphasise the importance of hormonal and amino acid metabolism in sustaining pregnancy. In contrast, PE is characterised by disruptions in metabolic and signalling pathways. Fatty acid biosynthesis (OLAH, ACSF3) and adipocytokine signalling (RXRB, PRKAB2, SLC2A4) reflect altered lipid metabolism and inflammatory signalling, hallmarks of PE pathology. The sulfur relay system (CTU2) and valine, leucine, and isoleucine degradation (BCKDHB, ACSF3) suggest impairments in sulfur–related processes and branched–chain amino acid metabolism associated with oxidative stress. Both NP and PE involve glycosaminoglycan biosynthesis but with different regulatory genes (CHSY3, CHST2 in NP; CHST1, B4GALT7 in PE), underscoring its critical role in extracellular matrix formation and placental structure. These findings highlight the metabolic and hormonal focus in NP. At the same time, PE is defined by dysregulation in these pathways, pointing to genes like Like-Sm Protein 6 (LSM6), Exosome Component 9 (EXOSC9), Oleoyl-ACP Hydrolase (OLAH), and Retinoid X Receptor Beta (RXRB)as potential biomarkers or therapeutic targets.

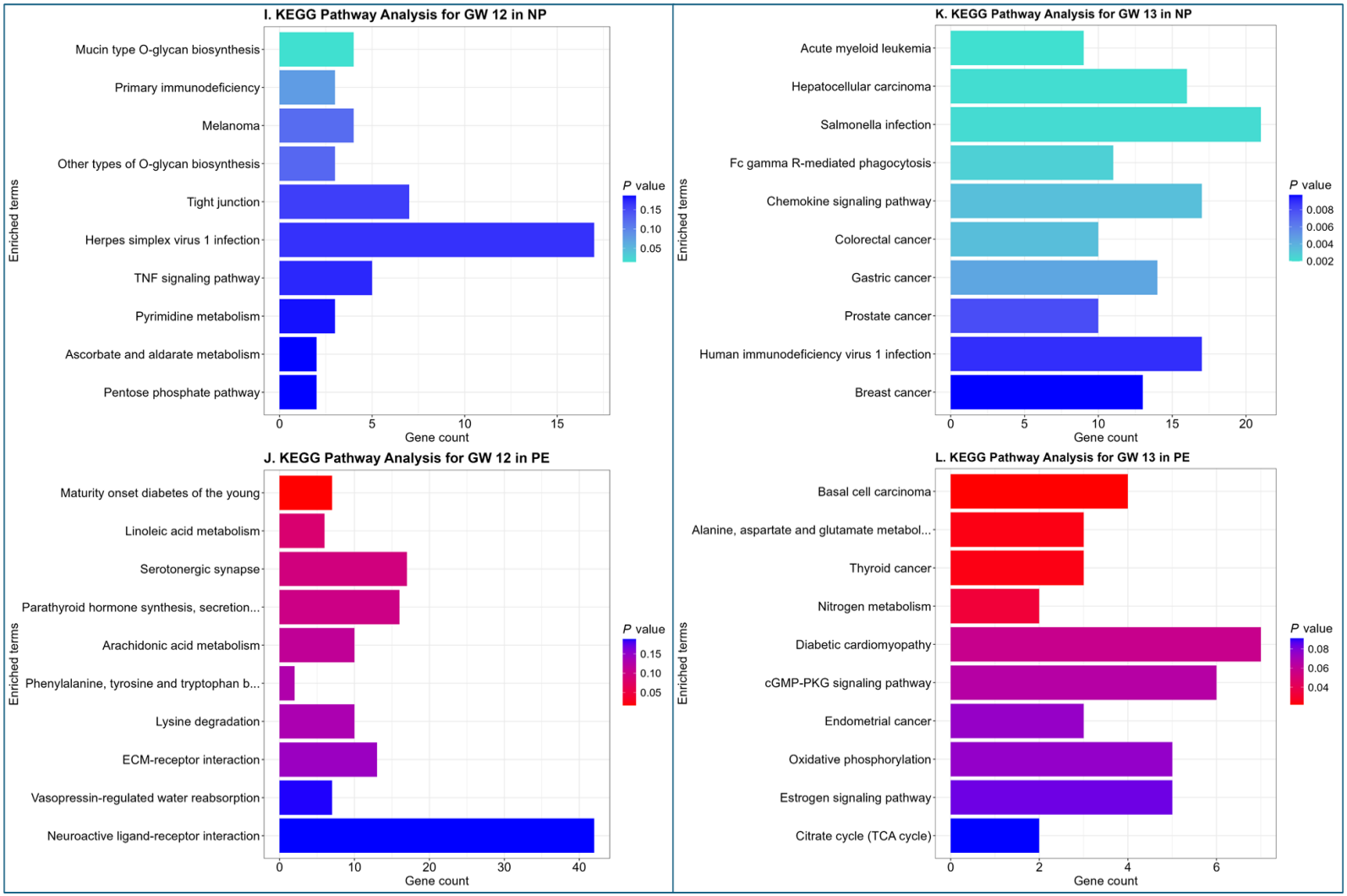


Figure 27: KEGG pathway analysis for unique, consistent genes at GW 12 and 13 in NP and PE

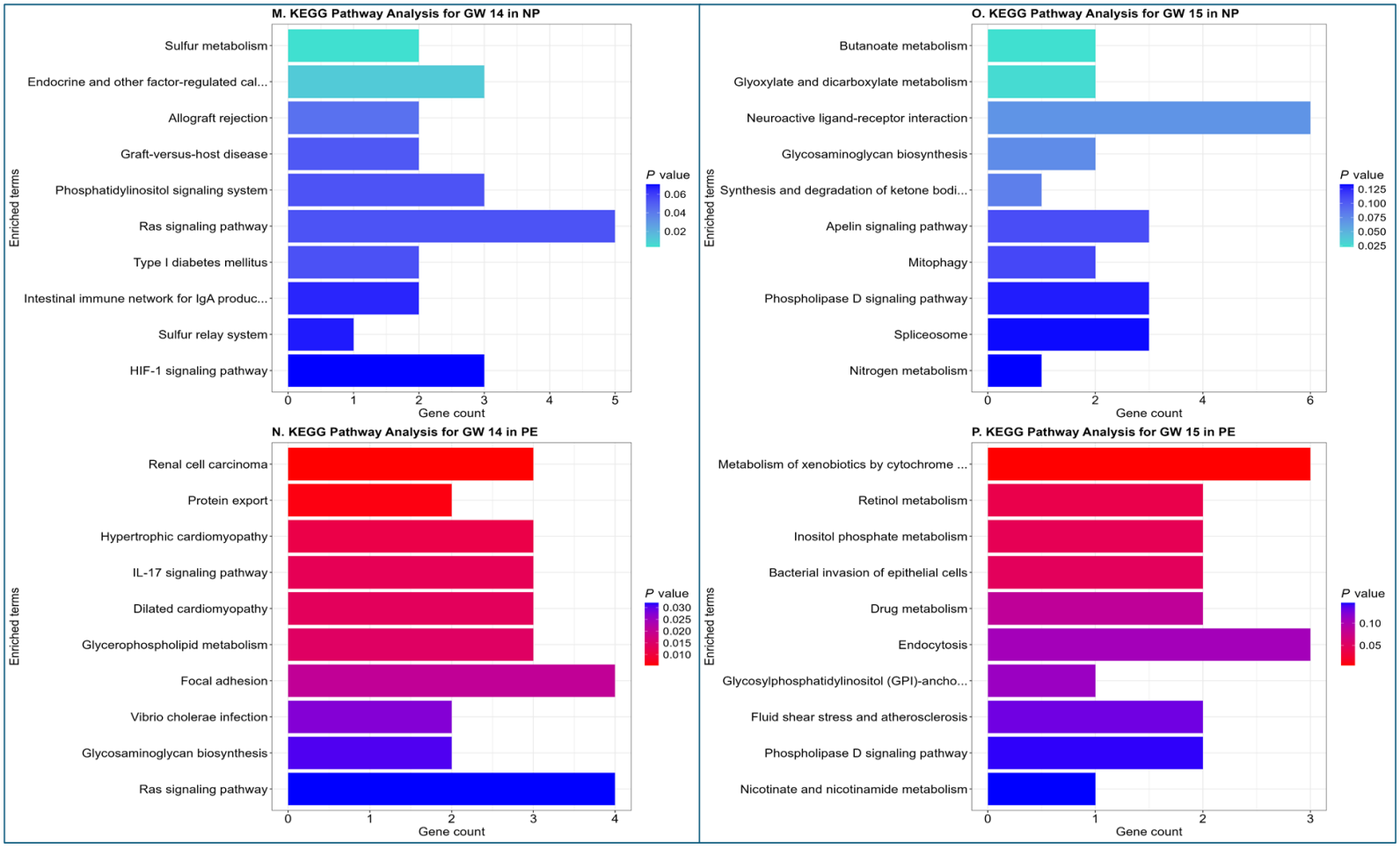


Figure 28: KEGG pathway analysis for unique, consistent genes at GW 14 and 15 in NP and PE

At GW 12, pathway analysis for normal pregnancies (NP) and PE (PE) reveals diverse biological processes and potential molecular disruptions. In NP, key pathways include mucin–type O–glycan biosynthesis (GALNT12, C1GALT1C1, ST6GALNAC4), critical for glycoprotein modification and placental structure, and primary immunodeficiency, involving genes like RFX5, CD8B2, and IL2RG, which highlight immune system adaptations essential for pregnancy. Other notable pathways include tight junctions (PRKCI, JUN, MARVELD2) and pyrimidine metabolism (RRM2, UCK1), which support cellular communication and nucleotide synthesis. In PE, distinctive pathways emerge, such as maturity–onset diabetes of the young (RFX6, HNF4G, IAPP), pointing to disruptions in glucose metabolism and insulin regulation, and linoleic acid metabolism (CYP2J2, CYP2C9), emphasising altered lipid processing. Additional pathways, including arachidonic acid metabolism (PLA2G4F, ALOX15), phenylalanine, tyrosine, and tryptophan biosynthesis (IL4I1, PAH), and ECM–receptor interaction (LAMB3, COL4A4), suggest perturbations in inflammation, amino acid metabolism, and extracellular matrix integrity, respectively. Notably, both conditions overlap in metabolic and signalling pathways, such as glycan biosynthesis in NP and phenylalanine metabolism in PE, highlighting critical processes underpinning placental and systemic health. These findings underscore genes like Polypeptide N-Acetylgalactosaminyltransferase 12 (GALNT12), Regulatory Factor X5 (RFX5), Regulatory Factor X6 (RFX6), and Cytochrome P450 Family 2 Subfamily J Member 2 (CYP2J2) as potential biomarkers or therapeutic targets for maintaining pregnancy and addressing PE.

At GW 13, KEGG pathway analysis for NP and PE highlights distinct and overlapping biological processes indicative of placental development and potential pathological changes. For NP, critical pathways include acute myeloid leukaemia (STAT5A, CSF1R, MAP2K1, AKT2), hepatocellular carcinoma (TCF7L2, SMAD3, AKT1), and Fc gamma R–mediated phagocytosis (HCK, AKT2, RAC1), suggesting key roles for immune signalling, cell proliferation, and phagocytic activity in placental function. Pathways such as chemokine signalling (PIK3CG, CXCL10)

and colorectal cancer (APC2, LEF1, RPS6KB2) further implicate regulatory mechanisms involving inflammation, growth, and intracellular signalling in sustaining pregnancy. In PE, notable pathways include basal cell carcinoma (TCF7L1, PTCH1), which emphasises aberrant Wnt signalling, and alanine, aspartate, and glutamate metabolism (ALDH5A1, CPS1), hinting at disrupted amino acid processing. The involvement of nitrogen metabolism (CPS1, CA14) and oxidative phosphorylation (NDUFA12, SDHD) underscores altered energy metabolism and mitochondrial function, potentially contributing to the placental insufficiency characteristic of PE. Additional pathways like the citrate cycle (TCA cycle) (SUCLG2, SDHD) and oestrogen signalling (CREB3, MMP9) point to compromised metabolic and hormonal regulation. Comparing NP and PE highlights shifts in core metabolic and signalling networks, with NP emphasising immune and growth pathways. In contrast, PE pathways reveal potential dysfunctions in energy metabolism, oxidative stress, and hormonal signalling. These insights underline pathways like AKT signalling, Wnt regulation, and amino acid metabolism as crucial for distinguishing normal from pathological states during early pregnancy.

At GW 14, KEGG pathway analysis reveals distinct biological processes between NP and PE. In NP, pathways such as sulfur metabolism (involving TST and BPNT2) highlight metabolic adaptations critical for energy production and redox balance. Additionally, the endocrine and calcium reabsorption regulation pathway (PRKCG, CLTB, AP2B1) emphasises the importance of mineral homeostasis for foetal development. Immune–related pathways such as allograft rejection and graft–versus–host disease (involving CD28 and HLA–DRB1) reflect the immune system's role in maintaining tolerance and modulating responses during pregnancy. The phosphatidylinositol signalling system (PRKCG, BPNT2, SYNJ2) suggests cellular signalling involved in placental growth, while the HIF–1 signalling pathway (PRKCG, EGLN1, ANGPT2) is linked to oxygen sensing and vascular adaptation in the placenta. In PE, pathways such as renal cell carcinoma (PDGFB, MET, PAK2) suggest growth and renal function disruptions. In contrast, the protein

export pathway (SEC61A2, SEC61A1) points to altered protein trafficking, potentially affecting placental function. The IL–17 signalling pathway (FOSB, S100A9, IL17RB) indicates inflammation and immune dysregulation, which are characteristics of PE. Additionally, pathways related to hypertrophic and dilated cardiomyopathy (ITGA2, TPM2, CACNA2D2) suggest cardiovascular strain linked to PE. Altered lipid metabolism is evident in the glycerophospholipid metabolism pathway (PLA2G6, DGKI, PISD), while the glycosaminoglycan biosynthesis pathway (XYLT2, EXTL3) suggests extracellular matrix remodelling disruptions. In summary, while immune and metabolic processes dominate in both NP and PE, the pathways in PE highlight inflammatory responses, vascular dysfunction, and disrupted signalling, contrasting with the metabolic adaptation and immune tolerance observed in normal pregnancies. Notably, the Ras signalling pathway plays distinct roles in both conditions, further emphasising the differences in cellular signalling underlying the pathogenesis of PE.

At GW 15, KEGG pathway analysis reveals significant differences between normal NP and PE. In NP, butanoate metabolism (ACSM4, HMGCS2) and glyoxylate and dicarboxylate metabolism (AGXT, HAO2) indicate metabolic pathways involved in energy production and detoxification. The neuroactive ligand–receptor interaction pathway (CYSLTR1, CALCRL, FSHR, GRM8, AGTR1, LTB4R) reflects the influence of neurotransmitters on foetal development, while glycosaminoglycan biosynthesis (EXT2, CHST3) points to extracellular matrix remodelling, essential for placental function. The synthesis and degradation of the ketone bodies pathway (HMGCS2) suggest the role of ketone bodies in energy metabolism during pregnancy. Apelin signalling (GABARAPL1, MRAS, AGTR1) and mitophagy (GABARAPL1, MRAS) pathways are involved in cell survival and mitochondrial function, which are crucial for placental health. In PE, pathways like the metabolism of xenobiotics by cytochrome P450 (ADH4, AKR7A2, GSTO2) and retinol metabolism (ADH4, PNPLA4) suggest alterations in the processing of foreign compounds and vitamin A derivatives. The inositol phosphate metabolism pathway (ISYNA1, CDIPT) highlights change in cellular signalling, while bacterial invasion

of epithelial cells (CAV2, DNM1) suggests potential immune system alterations. Additionally, pathways involved in drug metabolism (ADH4, GSTO2), endocytosis (CAV2, DNM1, RAB11FIP5), and glycosylphosphatidylinositol (GPI)–anchor biosynthesis (PIGA) reflect disruptions in cellular processes. The fluid shear stress and atherosclerosis pathway (GSTO2, CAV2) suggest vascular alterations, a PE hallmark. These findings indicate that metabolism and cellular signalling in PE are significantly disrupted compared to normal pregnancies, affecting vascular, immune, and metabolic processes crucial for maintaining a healthy pregnancy.

At GW 16, pathway analysis highlights key differences between NP and PE. In normal pregnancies, amoebiasis (SERPINB10, COL3A1, LAMC3, LAMA4, FN1, PIK3R2, CTSG) suggests immune and extracellular matrix remodelling, potentially indicating normal foetal–placental development. Other notable pathways in NP include RNA polymerase (POLR3GL, POLR1H, POLR3K) and glycine, serine, and threonine metabolism (GATM, PSAT1, GCAT), both of which are involved in protein synthesis and amino acid metabolism. The basal cell carcinoma pathway (WNT7B, DVL1, AXIN2, WNT16) reflects cellular signalling involved in growth regulation, and protein export (SEC61A2, SPCS2) indicates cellular trafficking processes. Vitamin B6 metabolism (PSAT1) also plays a role in amino acid synthesis and overall metabolic health, while ABC transporters (ABCA5, ABCA13, ABCG2) are involved in nutrient and drug transport. Pathways related to ether lipid metabolism (LPCAT4, ENPP2, PLD2) and sulfur relay systems (CTU1) also show metabolic functions essential for normal placental and foetal development. In PE, significant pathways include salmonella infection (MAPK10, DYNC1I2, RALA, ACTR1B, MYC, DCTN4, NOD1, ACTR10, HSP90B1, TNFRSF1A), which highlights potential immune system and inflammation responses. Other affected pathways include non–alcoholic fatty liver disease (MAPK10, SREBF1, GSK3A, NDUFS5, NDUFA3, IL6R, TNFRSF1A), protein processing in the endoplasmic reticulum (MAPK10, SEC24B, UBXN2A, LMAN2, MAN1A1, RBX1, HSP90B1), and insulin resistance (MAPK10, SREBF1, OGT, SLC27A5, TNFRSF1A), indicating metabolic stress and dysfunction. Shigellosis (MAPK10, GSK3A,

CALCOCO2, H3C3, NOD1, H3C2, RBX1, TNFRSF1A) and sphingolipid signalling (MAPK10, CERS4, PPP2R2D, SGPP2, TNFRSF1A) are linked to immune and inflammatory processes. Additionally, the ErbB signalling pathway (MAPK10, MYC, GAB1, CBLB) and Th1 and Th2 cell differentiation (MAPK10, HLA–DMA, NOTCH1, RUNX3) emphasise altered immune responses, while ubiquitin–mediated proteolysis (MGRN1, CBLB, UBE3A, UBE3B, RBX1) suggests changes in protein degradation, which may affect cell function and survival in PE. These findings underline significant metabolic and immune alterations in PE compared to NP at this stage.

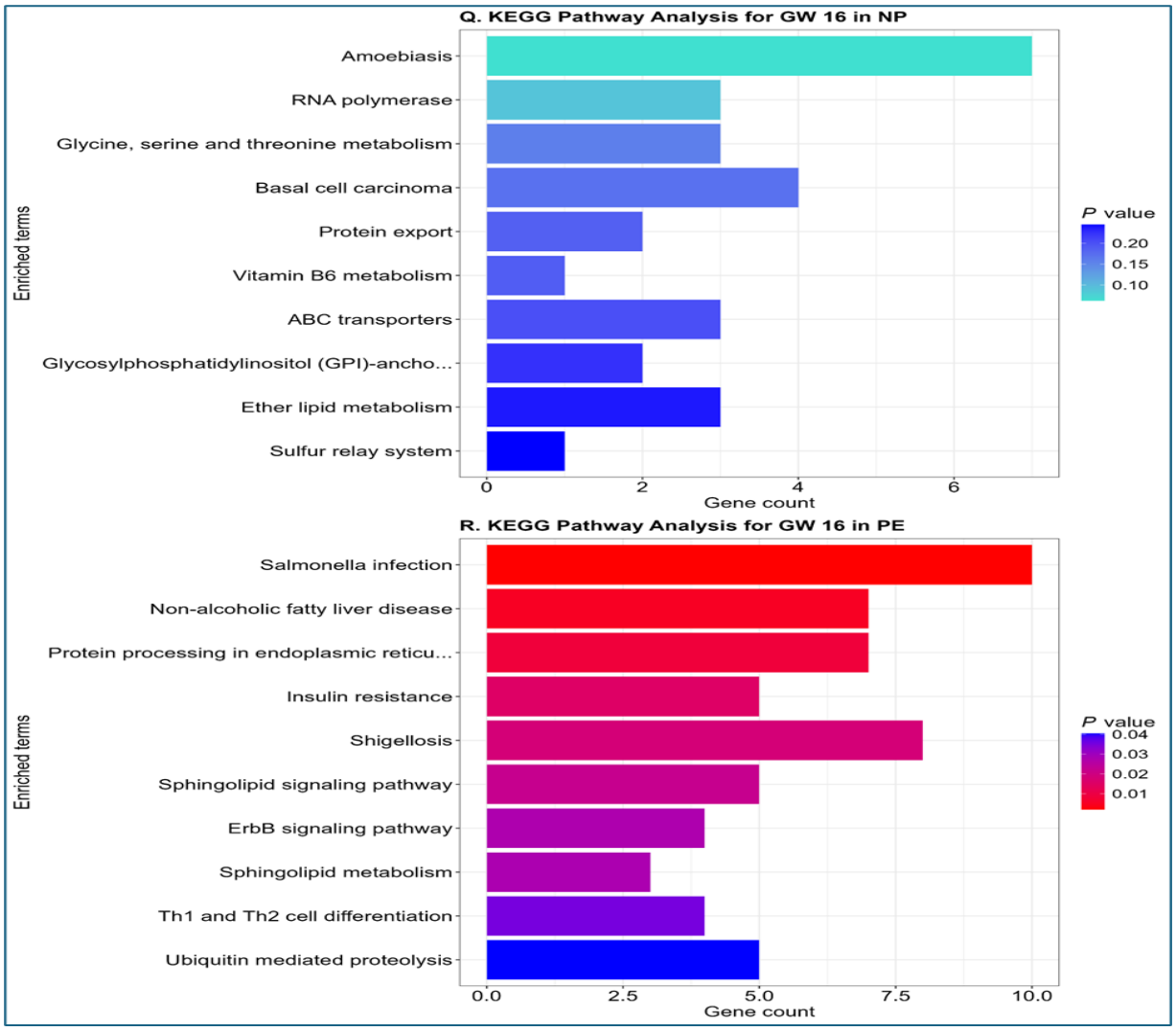


Figure 29: KEGG pathway analysis for unique, consistent genes at GW 16 in NP and PE

The KEGG pathway analysis across multiple gestational weeks reveals critical molecular signatures distinguishing NP from PE. Immune and inflammatory dysregulation in PE is highlighted by enriched pathways such as C–type lectin receptor signalling, NOD–like receptor signalling, and IL–17 signalling, with genes like STAT2, Nuclear Factor of Activated T Cells 3 (NFATC3), and S100 Calcium Binding Protein A9 (S100A9)playing key roles in immune modulation and inflammation. Additionally, pathways linked to altered lipid metabolism, such as fatty acid biosynthesis, adipocytokine signalling, and glycerophospholipid

metabolism, suggest disruptions in energy metabolism that contribute to vascular dysfunction and metabolic stress in PE. Unique metabolic pathways like AMPK signalling, glycosaminoglycan biosynthesis, and phosphoinositide signalling are also disrupted in PE, pointing to compromised placental function. Regarding vascular and hormonal regulation, NP pathways such as AGE–RAGE signalling and renin–angiotensin pathways emphasise the importance of vascular adaptation and electrolyte balance. In contrast, PE shows disruptions in these pathways and pathways related to renal function and cardiovascular strain. Cellular signalling pathways, including Wnt, ErbB, and Notch signalling, are essential for cellular communication and growth in NP, but their disruption in PE indicates impaired trophoblast invasion and placental development. Additionally, Ras signalling is implicated in both conditions, with its dysregulation in PE pointing to aberrant cellular responses. Key genes, such as STAT2, FOXO1, NFATC3, DVL3, S100A9, Dishevelled Segment Polarity Protein 3 (DVL3), Tumour Necrosis Factor Receptor Superfamily Member 1A (TNFRSF1A), Acyl-CoA Synthetase Long-Chain Family Member 1 (ACSL1), and Acetyl-CoA Carboxylase Beta (ACACB) consistently appear across multiple weeks and highlight immune dysregulation, metabolic disturbances, and vascular dysfunction as hallmarks of PE. These genes and pathways emerge as potential biomarkers for early detection and therapeutic targets, offering valuable insights into the pathogenesis of PE.

The results of this chapter reveal the complexity of molecular mechanisms underlying PE and their distinct temporal patterns during gestation. One of the key findings is the elevated presence of SCGs in NP compared to PE. This observation suggests robust gene regulation in NP, critical for supporting NP progression. However, PE exhibits disruptions or adaptations in this regulation, as evidenced by the spike in SCGs at GW 12. This spike may indicate an adaptive or pathological response in PE, reflecting shifts in gene activity that could contribute to disease progression.

Similarly, the higher prevalence of BRCGs in PE highlights a stable yet altered genetic profile, potentially contributing to the disease's pathogenesis. These genes, expressed consistently at low levels, may represent subtle but crucial changes in regulatory mechanisms associated with vascular dysfunction, immune response, or inflammation, hallmarks of PE. The temporal variation observed in SCGs and BRCGs underscores the dynamic molecular landscape of pregnancy and the unique disruptions introduced by PE.

Pathway enrichment analyses further emphasise the role of unique gene signatures in key biological processes. In NP, pathways related to immune regulation, oxidative stress management, and vascular remodelling play significant roles in maintaining maternal–foetal health. In contrast, PE is characterised by the enrichment of pathways linked to immune dysregulation, inflammation, and abnormal vascular remodelling, which align with the clinical manifestations of the condition. These findings suggest that the identified gene signatures and pathways could serve as valuable biomarkers for early diagnosis of PE.

Incorporating these insights into AI-assisted predictive models has significant potential to advance early detection and personalised care for PE. The study paves the way for accurate risk stratification and timely intervention by leveraging ML algorithms trained on unique gene expression profiles. Integrating these predictive tools into clinical practice could transform PE management, improving mothers' and infants' outcomes. However, further validation in larger cohorts and diverse populations is necessary to ensure the generalisability and reliability of these biomarkers. Overall, this study provides critical insights into the molecular mechanisms of PE and lays the groundwork for developing innovative diagnostic and therapeutic strategies.

5.3.4 Thematic Interpretation of Key Consistent Genes

The thematic significance of the UCEGS provides crucial evidence for the chronic molecular foundation of PE. The stable gene profiles are distilled across three core

pathogenic themes. First, Chronic Immune and Inflammatory Maladaptation is highlighted by the persistent enrichment of pathways like NOD-like receptor signalling (GW 9, 10, PE) and C-type lectin receptor signalling (GW 8, PE). This suggests a sustained, underlying state of innate immune activation that creates a chronic inflammatory vulnerability from early gestation (Murthi *et al.*, 2020). Second, Metabolic and Cellular Stress Deficits are consistently observed through enrichment in fatty acid biosynthesis (GW 9) and adipocytokine signalling (GW 9, 11), emphasising a stable metabolic deficit that impairs endothelial function and contributes to cumulative cellular stress (Abdalla *et al.*, 2025). Finally, Structural and Vascular Vulnerability is evidenced by UCEGS pointing to structural integrity issues, including genes linked to the regulation of systemic arterial blood pressure (GW 16) and structural defects in the extracellular matrix (e.g., TIMP2/TIMP3), identifying stable flaws in the maternal vasculature that precede clinical hypertension (Palei *et al.*, 2023). By highlighting these enduring molecular themes, the UCEGS provide crucial biomarkers that track the development of chronic PE pathology before symptoms emerge.

5.3.5 Molecular Synthesis and Alignment with PE Pathogenesis

The comprehensive analysis of UCEGS across gestational weeks 8 to 16 reveals distinct molecular themes that strongly align with the core pathogenic hypotheses of PE outlined in Chapter 2 (Literature Review). Distilling the extensive lists of enriched genes and pathways highlights core mechanisms driven by consistent gene expression, complementing the dynamic findings of Chapter 4.

The relationship between these stable UCEGS and the dynamic UDEGS (Chapter 4) is critical for resolving the molecular trajectory of PE. The stable UCEGS identified here likely reflect the chronic, underlying molecular defects that predispose the system to fail, providing a robust baseline for PE prediction. This distinction is synthesised in Table 18, followed by a detailed discussion of the pathogenic themes.

Table 18: Synthesis of Unique Consistent Gene Signatures (UCEGS) and Alignment with Pre-eclampsia Pathogenesis (PE vs. NP)

Pathogenesis Theme (Hypothesis)	Evidence from Consistent Gene Signatures (UCEGs/BRCGs)	Cross-Chapter Comparison (vs. Chapter 4 UDEGs)
A. Immune Dysregulation & Inflammation	**Enrichment in stable pathways:** NOD-like receptor signalling (GW 9, 10, PE), C-type lectin receptor signalling (GW 8, PE). Reflects a sustained, underlying maladaptive immune state.	**Contrasting Role:** UDEGs (Chapter 4) reflected *dynamic, acute* shifts in immune pathways (e.g., NF-κB at GW 13). UCEGs capture the *stable, chronic* inflammatory substrate.
B. Vascular Dysfunction & Remodelling	**Stable defects in structural pathways:** Regulation of systemic arterial blood pressure (GW 16, PE) and ECM components (e.g., *TIMP2/TIMP3* at GW 16, suggesting impaired matrix degradation).	**Mechanistic Depth:** UDEGs showed *acute* vascular stress response (e.g., Phosphatidylinositol signalling at GW 13). UCEGs pinpoint the *stable structural flaws* (e.g., in ECM/cytoskeleton) that predispose the maternal vasculature to failure.
C. Metabolic and Cellular Stress	Consistent presence of dysfunction in fatty acid biosynthesis (GW 9, PE) and adipocytokine signalling (GW 9, PE). Reflects persistent, low-level metabolic and cellular deficits (e.g., altered lipid processing).	**Stable Malfunction:** UDEGs identified *transient* metabolic stress (e.g., AMPK at GW 13). UCEGs reveal the *persistent*, chronic metabolic imbalance consistent with the long-term metabolic risks of PE.

5.3.5.1 Immune and Inflammatory Maladaptation

Immune and inflammatory maladaptation emerged as the predominant foundational molecular mechanism underpinning PE pathogenesis. The UCEGS in the PE group demonstrated persistent enrichment of pathways indicative of chronic immune dysregulation, including NOD-like receptor signalling and C-type lectin receptor signalling at gestational weeks 8, 9, and 10. Unlike the transient, acute immune shifts observed in the UDEGs in Chapter 4—particularly involving NF-κB signalling and T-cell differentiation around gestational week 13—these consistently expressed genes reflect a stable and foundational inflammatory state. This state appears to be driven predominantly by innate immune activation and macrophage dysregulation, leading to impaired maternal–foetal immune tolerance. Such maladaptation likely primes individuals for the systemic inflammatory response that becomes clinically apparent during Stage 2 PE.

5.3.5.2 Vascular and Extracellular Matrix (ECM) Vulnerability

Secondly, the analysis highlights vascular and extracellular matrix (ECM) vulnerability as a core mechanism underpinning PE pathogenesis. Pathways reflecting disruptions in vascular integrity and tissue structure were persistently enriched in the PE group. Notably, genes involved in ECM remodelling, including the metalloendopeptidase inhibitor activity of TIMP2 and TIMP3 at gestational week 16, were prominent. Furthermore, dysregulation of pathways related to smooth muscle function and regulation of systemic arterial blood pressure at the same time point suggests a persistent baseline vulnerability in the maternal vasculature. This stable molecular signature provides a mechanistic backdrop against which the acute angiogenic imbalance (e.g., sFlt-1/PlGF), observed in the dynamic UDEGs in Chapter 4, precipitates clinical manifestations such as hypertension.

5.3.5.3 Metabolic and Cellular Stress

Finally, metabolic and cellular stress was uniquely emphasised through the UCEGS analysis, which highlighted sustained metabolic dysfunction in PE. This was evidenced by enrichment in pathways such as fatty acid biosynthesis (gestational week 9) and adipocytokine signalling (gestational weeks 9 and 11), both of which are strongly associated with oxidative stress and insulin resistance—well-established drivers of endothelial injury. In contrast, the consistently expressed genes in NP showed enrichment in protective pathways, such as oxidative stress management (e.g., *TXN* at gestational week 16) and foundational energy metabolism. These observations suggest that PE is characterised by a chronic, low-level metabolic deficit emerging from early gestation, contributing to cumulative cellular stress and eventual endothelial breakdown.

By identifying these consistently disrupted molecular themes, the UCEGS provide a valuable complementary perspective to the dynamic UDEGs from Chapter 4. Together, they offer a comprehensive and temporally resolved molecular portrait

of PE pathogenesis in early pregnancy, illustrating the synergistic contribution of chronic immune activation, vascular instability, and metabolic stress.

5.4 Conclusion

This study successfully identifies UCEGS, and pathway enrichments associated with PE across gestational weeks. Through integrating bioinformatics and transcriptomic data, the analysis highlights the significant roles of CEGS, categorised into SCGs and BRCGs, in distinguishing between NP and PE. These findings underscore the complex molecular mechanisms driving PE, including dysregulation in immune response, oxidative stress pathways, and vascular remodelling processes. Notably, the temporal dynamics of gene expression patterns, with marked differences observed in gestational weeks such as GW 12, provide valuable insights into the progression of PE. Incorporating these unique CEGS into AI-assisted predictive models offers a promising approach for the early identification of women at risk for PE. The potential integration of these unique consistent gene signatures into AI-assisted predictive tools could facilitate timely interventions and contribute to personalised care strategies, ultimately aiming to improve maternal and foetal outcomes. While the study provides a robust foundation, further research is needed to translate these findings into clinical applications. Future research must validate these findings in larger, prospective, and multi-ethnic cohorts to ensure the reliability and generalisability of these UCEGS. Such studies are needed to confirm that these consistent gene signatures are robust predictive markers across diverse populations. Furthermore, to enhance predictive accuracy, these biomarkers should not be considered in isolation. Future predictive models should integrate these UCEGS with other critical data layers, including established clinical risk factors (such as maternal age, BMI, and prior PE history), other transcriptomic signatures (like the UDEGs identified in Chapter 4), and the unique variant signatures (UVS) explored in Chapter 6. This multi-dimensional approach will be essential for building a comprehensive and clinically reliable predictive tool for PE. Additionally, exploring longitudinal changes in gene

expression and their functional implications could provide deeper insights into the pathophysiology of PE. Overall, this study advances understanding of PE and represents a significant step toward improving its early diagnosis and management.

5.5 Future Directions

Building on the findings of this study, several avenues for future research are proposed to enhance the understanding and clinical utility of identified biomarkers. First, validating CEGS in independent datasets and across diverse populations is crucial. Such validation will establish the robustness and generalisability of these biomarkers, ensuring their applicability in varied clinical settings. Additionally, leveraging multi–omics approaches to integrate transcriptomic data with proteomic and metabolomic profiles could provide a more comprehensive understanding of the molecular landscape of PE.

Second, functional studies focusing on the roles of SCGs and BRCGs are necessary to elucidate their causative links to PE. By employing in vitro and in vivo models, researchers can investigate how these genes contribute to the pathophysiology of PE, such as their roles in immune modulation, vascular remodelling, and oxidative stress responses. Understanding these mechanisms could pave the way for developing targeted therapeutic interventions.

Third, enhancements to ML models are needed to improve the accuracy and utility of predictive tools. Integrating clinical data with genetic data, such as maternal health history and imaging findings, could lead to more sophisticated models capable of individualised risk assessment. Such integration would allow for the development of holistic diagnostic frameworks.

Finally, investigating longitudinal changes in gene expression beyond GW 16 could provide critical insights into the later stages of PE progression. This extended analysis could help identify additional biomarkers or therapeutic targets, further strengthening the clinical applicability of the findings. Future

research can build upon this study's foundation to advance PE diagnosis, prevention, and management by addressing these directions.

5.6 Summary of the Chapter

This chapter investigated gestational week-specific Unique Consistently Expressed Gene Signatures (UCEGS) to delineate the molecular chronology of PE development from weeks 8 to 16. By categorising genes into Significant Consistent Genes (SCGs) and Biologically Relevant Consistent Genes (BRCGs), the analysis moved beyond the single-time-point differential expression approach of Chapter 4 to identify stable, stage-specific transcriptional programmes.

A central finding of this chapter was the distinct temporal patterns of consistent gene expression, specifically the higher prevalence of Significant Consistent Genes (SCGs) in Normal Pregnancy (NP) compared to Pre-eclampsia (PE), and the elevated prevalence of Biologically Relevant Consistent Genes (BRCGs) in PE. This finding links directly back to the pathology defined in the literature review (Chapter 2). The higher count of SCGs in NP likely represents the stable, robust transcriptional programs required for healthy placentation, including successful spiral artery remodelling and maintained immune tolerance . In contrast, the higher prevalence of BRCGs in PE suggests the failure of these normal programs and their replacement by a stably dysfunctional state. This state is characterized by the chronic, low-level activation of the core pathogenic themes identified in Chapter 2, such as immune dysregulation (e.g., STAT2, VSIG4), oxidative stress (e.g., TXN), and vascular dysfunction . Furthermore, the emergence of Gestational Week 12 as a critical window with a sharp increase in these signatures aligns with the established clinical timeframe for impaired spiral artery remodelling.

The GO enrichment analysis systematically identified key genes, including STAT2, VSIG4, and TXN, as drivers of the most distinct biological pathways at Gestational Week 8, strongly validating the CGEA methodology. The identification of TXN as a key gene in the "negative response regulation to reactive oxygen species" pathway in the NP group underscores the importance of stable antioxidant

processes in a healthy pregnancy. This aligns with literature demonstrating that altered TXN (thioredoxin) expression is a key indicator of the oxidative stress and inflammatory imbalance that characterises PE (Farladansky-Gershnabel *et al.*, 2021). Conversely, the identification of STAT2 and VSIG4 in the "negative regulation of innate immune response" pathway in the PE group points to early and consistent immune dysregulation. STAT2 is a critical component of cytokine signalling (e.g., type I interferon) essential for immune modulation at the maternal-foetal interface (L. Li *et al.*, 2024; Marzioni *et al.*, 2025), while VSIG4 has been previously identified as a potential biomarker for severe PE, linked to complement activation dysregulation and inflammation (Textoris *et al.*, 2013; Kumar and Stewart, 2025). The successful identification of these specific genes confirms that the CGEA pipeline captured the core, stable pathogenic signals of PE.

The extensive lists of UCEGS identified in this chapter, while biologically informative, represent a large pool of candidate biomarkers rather than a final predictive set. These findings will be carried forward as the foundational input for the automated feature selection pipeline detailed in Chapter 7. There, a multi-stage Recursive Feature Elimination with Cross-Validation (RFECV) process will systematically evaluate this "plethora of genes" to select only the most statistically powerful and predictive features for building the final AI model.

In conclusion, this temporal profiling of UCEGS provides a coherent molecular narrative for early PE development, directly linking the findings to established pathogenic hypotheses. The key genes and pathways distilled from the extensive lists offer a refined, biologically grounded feature set. This set is highly suited for integration into the machine learning models proposed at the outset, significantly enhancing the potential for early and accurate PE prediction.

Key Findings of the Study:

- A Temporal Map of PE Pathogenesis: The study moved beyond static gene lists to define Unique Consistently Expressed Gene Signatures (UCEGS)

for each gestational week, creating a dynamic timeline of molecular changes in NP and PE pregnancies.

- Identification of a Critical Window: Gestational Week 12 was identified as a critical period for PE development, marked by a significant peak in gene activity. This molecular finding directly correlates with the key physiological event of spiral artery remodelling, a process known to be defective in PE.
- Convergence on Core Pathogenic Themes: Analysis of UCEGS revealed that PE signatures are consistently enriched for pathways related to immune dysregulation, vascular dysfunction, and inflammation. This directly confirms major pathogenic hypotheses in the literature. In contrast, NP signatures were dominated by processes essential for healthy placental development and oxidative stress management.
- High-Confidence Biomarker Candidates: Key genes such as STAT2 (immune signalling), VSIG4 (immunoregulation), and TXN (oxidative stress response) were distilled from the broader lists as high-priority biomarkers. Their roles are strongly supported by existing evidence of their involvement in PE pathophysiology.
- Complementarity with Differential Expression Data: The UCEGS provide a stable, gestational-week-specific context that complements the transient differential expression signals found in Chapter 4. Together, they offer a more complete picture, distinguishing sustained pathological programmes from acute fluctuations.
- Foundation for Predictive Modelling: The distilled themes and key genes provide a robust, biologically interpretable feature set that is ideally suited for developing machine learning models to improve the early prediction of PE.

Chapter 6

Identification of Pre-eclamptic Unique Variant Signatures at Each Gestational Week in Early Pregnancy

This chapter explores the identification of unique pre-eclamptic variant signatures at each gestational week (GW) using cfRNA sequencing data. It further investigates the underlying mechanisms by performing downstream analysis to enhance the understanding of PE pathophysiology and preparing the dataset for developing a machine learning (ML) model.

6.1 Introduction

Analysing cfRNA sequencing data to study PE presents advantages since it can be easily collected via blood from the mother without requiring procedures such as amniocentesis or chorionic villus sampling. Thereby decreasing associated risks significantly; it is used in Non–Invasive Prenatal Testing (NIPT) for identifying women having an elevated risk of an affected pregnancy (Taylor-Phillips *et al.,* 2016; Togneri *et al.,* 2019). Additionally, tracking the levels of cfRNA throughout pregnancy provides information about the health and development of both the foetus and placenta over time.

Recent research has demonstrated the potential of cfRNA in maternal plasma for early prediction of PE. A study analysing over 1,800 pregnancies found that cfRNA profiles from a single blood draw could predict PE with 75% sensitivity and 32.3% positive predictive value, outperforming current clinical methods. Another study developed classifiers based on 18 genes measured between 5–16 weeks gestation that could form the basis of a liquid biopsy test to identify at–risk mothers long before symptoms appear (Moufarrej et al., 2022). These cfRNA signatures reflect underlying pathophysiology and are independent of clinical factors like maternal age, BMI, and race. Early prediction using cfRNA could allow for improved monitoring and management of high–risk pregnancies, potentially

reducing complications and improving outcomes (Gray, Hemberg and Karumanchi, 2022).

6.1.1 Overview of cfRNA and Its Relevance in Gestational Monitoring

The cfRNA in the maternal plasma is an evolving tool in the NIPT and gestational monitoring, which offers information about foetal development and maternal health (Mortazavipour and Shahbazi, 2022; Everett and Chitty, 2015). Many foetal and placental tissues release cfRNA, which then enters the maternal circulation through several mechanisms such as apoptosis, necrosis, or active secretion (Taglauer, Wilkins-Haug and Bianchi, 2014; Raymond *et al.,* 2023; Stanley *et al.,* 2024). Since these short and fragmented RNA molecules are usually bonded to proteins or protected by extracellular vesicles, they are more stable in maternal plasma (Liu, Z. *et al.,* 2022; Hui, L. and Bianchi, 2011). It is an excellent candidate for complex sequencing techniques and subsequent machine–learning applications due to its stability, which is necessary for precise detection and analysis. Analysing the cfRNA sequencing data for biomarker discovery offers advantages in identifying pregnancy complications, like PE in the early stages. Moreover, it assists healthcare professionals in determining women's PE risk earlier, leading to personalised care and closer monitoring to avoid further complications.

Gestational monitoring describes the continuous assessment and tracking of foetal and maternal health conditions throughout pregnancy. Identifying gestational week–specific biomarkers in early pregnancy using cfRNA sequencing data aids in gestational monitoring and facilitates the early identification of pregnancy–related disorders (Everett and Chitty, 2015; Gekas *et al.,* 2023). Specifically, identifying variant signatures in cfRNA sequencing data, alongside gene expression signatures, enhances confidence in prenatal monitoring for the detection of pregnancy complications (Go, van Vugt and Oudejans, 2011; Esfahani *et al.,* 2022).

6.1.2 Unique Variant Signatures and PE-associated Genetic Variants

Unique Variant Signatures (UVS) refer to those present at each GW of pregnancy, particularly in the cfRNA released from the placenta and foetal tissues into the maternal bloodstream. These signatures are unique to different gestational weeks, reflecting the dynamic biological changes as pregnancy progresses. It can be identified by overlap analysis using variants in each GW ranging from 8 to 16. Moreover, UVS is exclusively unique for particular GW, and it permits scientists to observe how the foetal and placental genomes evolve. These variants may link to critical developmental processes or maternal physiological adjustments during pregnancy. Specific genetic variants might be involved in the regulation of placental function, immune system adaptation, or other biological processes that are crucial for a successful pregnancy (Wang, Li Qing, Fernandez-Boyano and Robinson, 2022; Chatterjee, Ouidir and Tekola-Ayele, 2021). Detecting and analysing these UVS at different gestational stages could provide significant insights into the normal course of pregnancy and reveal deviations that may lead to complications such as PE.

PE-associated Genetic Variants (PE–AGV) refer to specific genetic markers that have been linked to an increased risk of developing PE (Zhao *et al.,* 2012; Tyrmi *et al.,* 2023). Moreover, these variants can be identified through Genome–Wide Association Studies (GWAS) by comparing the normal and disease states. These are not only crucial for understanding the pathophysiology of PE but also have significant clinical implications. By identifying PE-associated variants early in pregnancy, healthcare providers could better predict the risk of PE, allowing for earlier intervention and improved maternal and foetal outcomes. Identifying these variants in circulating cfRNA could be a non–invasive way to monitor and predict the onset of PE, offering a more timely and accurate diagnosis than traditional methods.

PE-associated variants and UVS differ significantly in their focus and biological relevance within the context of PE. PE-associated variants are genetic variants

linked to the pathophysiology of PE. These variants are typically involved in mechanisms such as placental dysfunction, immune response, and vascular abnormalities, making them disease–specific and potentially predictive of PE risk. In contrast, UVS refer to the genetic changes observed at specific GW, reflecting the normal biological processes occurring throughout pregnancy. These variants are dynamic, changing over time as pregnancy progresses, and are not necessarily tied to any disease but rather to natural pregnancy adaptations. While PE-associated variants are consistent across women with PE and can serve as biomarkers for predicting the disease, UVS helps establish baseline patterns of gene expression that vary with each GW. The comparison between these two sets of variants allows for a better understanding of how deviations from NP patterns could signal the development of PE, helping to differentiate between disease–related changes and typical gestational progression.

6.1.3 Significance of Analysing Variant Signatures in cfRNA

Identifying per gestational week specific unique variants in cfRNA sequencing data permits researchers to establish baseline expression patterns for different stages of pregnancy, which can be used to detect deviations indicative of pregnancy–related disorders . For instance, placental gene expression level changes indicate stress responses, inflammation, or abnormal placental development (Baker *et al.,* 2024; Bronson and Bale, 2014). Finding these changes early in pregnancy can accelerate timely interventions and more targeted monitoring, improving outcomes for both the mother and the foetus. Discovering variant signatures in cfRNA sequencing data improves understanding of the molecular processes underlying foetal development and maternal–foetal interactions. This information can be used to build disease–predictive models for pregnancy outcomes and advance individualised prenatal care. The variant calling analysis using cfRNA data offers a powerful tool for enhancing the accuracy of gestational monitoring, enabling earlier detection of complications and contributing to improved maternal and foetal health.

6.1.4 The Importance of Gestational Age–Specific Biomarkers

Biomarkers specific to gestational age have significant potential in improving prenatal diagnostics and monitoring. Traditional biomarkers, such as hormone levels (e.g., human chorionic gonadotropin and alpha–fetoprotein), often lack specificity and sensitivity for certain pregnancy complications or may only provide limited temporal information. By contrast, cfRNA–based biomarkers that are specific to each GW can offer more precise insights, allowing clinicians to track the normal progression of pregnancy and detect early deviations that might signal an increased risk of complications (Ng *et al.,* 2024).

A key advantage of using GW–specific variant signatures lie in their potential to capture the biological heterogeneity of pregnancy across different stages (Bakulski, Blostein and London, 2023). Given that pregnancy is a highly dynamic process, a single biomarker measured at a one–time point may not fully represent the complex physiological changes occurring. Gestational age–specific signatures can provide a comprehensive temporal profile, enabling the detection of abnormalities at the earliest possible stage.

6.1.5 Current Challenges in Identifying Temporal Genetic Variants

Temporal Genetic Variants (TGVs) state dynamic changes in genetic expression, occurrence or specific genetic markers that happen at distinctive phases over time, remarkably during biological processes such as foetal development. These variations can occur in allele frequencies, genotype distributions, or the overall genetic structure of the population. TGVs lead to alterations in gene activity that may emerge or vanish at distinct stages of the pregnancy weeks cycle and are responsible for the genetic modifications observed throughout the gestational period. These variants reflect the dynamic and time–dependent nature of gene regulation as the foetus grows and the placenta adapts to different stages of pregnancy (Wang, Weidi *et al.,* 2024; Mamsen *et al.,* 2017). Additionally, they can serve as molecular markers that indicate normal progression through different

gestational stages or highlight deviations that might be linked to pregnancy–related complications, such as PE.

There are numerous challenges while identifying TGVs in cfRNA, even though it is a promising potential methodology for gestational monitoring and early identification of pregnancy–related complications. Precisely capturing and analysing these time–sensitive changes in cfRNA–seq data needs overcoming a range of technical and biological obstacles (Cabús *et al.,* 2022; Zhu *et al.,* 2021). A low abundance of cfRNA levels in the maternal plasma is one of the primary challenges. It contains a small fraction of the total cell-free RNA in maternal blood, making it difficult to isolate and sequence with high sensitivity. Moreover, cfRNA is highly fragmented, leading to incomplete coverage of transcripts during the sequencing process. This fragmentation can obscure the detection of specific genetic variants, mainly those present in low quantities or regions of the transcript that are frequently degraded. High–sensitivity sequencing methods, like deep next–generation sequencing platforms with customised RNA extraction procedures and amplification techniques, can effectively overcome this obstacle and enhance the detection capabilities for genetic variations.

The maternal blood contains a mixture of cell-free RNA from both maternal and foetal sources (Wong, D. et al., 2024; Moufarrej et al., 2022). A key challenge in cfRNA analysis is distinguishing foetal-derived cfRNA from the dominant maternal RNA background (Moufarrej *et al.*, 2020). In this study, this limitation was addressed at the bioinformatic level by applying stringent variant filtering, reference-based annotation, and the use of publicly available databases to prioritise foetal- and placenta-specific transcripts. While experimental techniques such as methylation-based enrichment or allele-specific analysis can further enhance separation in future work, this study focuses on computational strategies appropriate for publicly sourced cfRNA data, acknowledging that residual maternal signal interference remains a limitation.

While specific placental–specific markers and bioinformatics techniques have improved the ability to separate foetal signals, the overlap between foetal and maternal gene expression can still lead to misclassification of variants, especially in early pregnancy when foetal RNA concentrations are lower (Tarca *et al.,* 2019; Wong, D. *et al.,* 2024). To address this challenge, methylation–based enrichment, allele–specific analysis, and foetal fraction estimation approaches can reduce maternal noise and enhance the detection of foetal–derived variants.

Obtaining the temporal changes of cfRNA needs accurate and longitudinal sampling across various gestational weeks during pregnancy. This is essential as the levels and composition of cfRNA change dynamically throughout pregnancy, with specific RNA transcripts appearing or disappearing at different gestational weeks. Nevertheless, in current clinical practice, undertaking frequent samples may not always be feasible due to logistical constraints or patient compliance. Inconsistent sampling can direct gaps in the temporal data, making it challenging to map GW–specific variants correctly (Tang *et al.,* 2024).

While the analysis of TGVs in cfRNA holds great promise for enhancing gestational monitoring and the early discovery of complications like PE, several significant challenges remain. Addressing these issues requires advances in RNA isolation techniques, improved bioinformatics tools, and larger, well–characterised datasets. Overcoming these difficulties will be critical for the successful integration of cfRNA variant analysis into routine prenatal care, increasing the ability to monitor foetal development and predict adverse pregnancy outcomes.

6.1.6 Research Hypothesis

Gestational week–specific unique variant signatures (UVS) identified from cfRNA sequencing data can serve as reliable biomarkers for early prediction of PE.

6.1.7 Research Question

Can UVS and PE-associated genetic variants derived from cfRNA sequencing data improve early detection and prediction of PE?

6.1.8 Research Aim and Objectives

This study aimed to identify the landscape of gestational age–specific UVS and the PE-associated significant variant signatures within cfRNA sequencing data ranging from GW 8 to 16. To achieve this aim, the following objectives were established:

1. To identify GW–specific UVS from cfRNA sequencing data spanning weeks 8 to 16.
2. To perform overlap and association analysis of UVS to identify significant genetic variants linked to PE.
3. To characterise biological processes, cellular components, and molecular functions associated with identified variants using GO and pathway enrichment analyses.
4. To uncover key molecular differences, compare the UVS of NP and PE pregnancies.
5. To propose biomarkers and critical gestational weeks for early detection of PE.

6.2 Method

The general method section explained the detailed data collection method to align with the human reference genome (Chapter 3). This chapter's method section describes the variant calling analysis and downstream analysis.

Variant calling is the crucial analytical step for discovering variant signatures, such as Single Nucleotide Polymorphisms (SNPs) and INsertions/DELetions (INDELs). Contrasting DNA–seq, RNA–seq data offers additional challenges due to spliced transcripts, variable expression levels, and RNA degradation. Therefore, cfRNA sequencing data must be pre–processed before performing variant calling analysis.

6.2.1 Data Pre–processing for Variant Call Analysis

Data pre–processing is a fundamental step in bioinformatics analysis, as it prepares raw biological data for accurate and meaningful downstream analysis. It

involves a series of steps to clean, normalise, and transform the data, ensuring high quality and consistency. However, most of the basic data pre–processing for cfRNA sequencing data was carried out during the prior transcriptomic analytical studies (Chapters 5 and 6). This study utilises the BAM files from the previous studies to carry out the variant call analysis. The pre–processing steps for variant calling analysis consist of GATK–based pre–processing and non–GATK–based pre–processing.

6.2.1.1 Non–GATK Data Pre–processing

228 BAM files of cfRNA sequencing data, ranging from gestational weeks 8 to 16, were used in the variant calling analysis to identify unique variant signatures. These BAM files were sorted by genomic coordinates using Picard tools (v3.0.0) within a Conda (v 23.9.0) environment.

6.2.1.2 GATK Data Pre–processing

After sorting, the sorted BAM files were used with the Genome Analysis Toolkit (GATK) for the second data pre–processing phase. The GATK is an extensively used tool for variant calling in RNA–sequencing data, offering specific features adapted for handling the complexities of RNA–sequencing data (Brouard *et al.,* 2019; Brouard and Bissonnette,2012). It is often considered the gold standard for variant calling, particularly in DNA and RNA sequencing data. It has been extensively implemented as the industry standard for identifying research and clinical genomics variants. Moreover, GATK is well–known for its best practices pipeline, and it is specifically designed to optimise accuracy, particularly in complex datasets .

In the second phase of the data pre–processing step, the sorted BAM files were used to mark duplicates with the GATK tool (v4.4.0.0), precisely designed to identify and flag PCR duplicates. Duplicate reads, which arise from the amplification of the same RNA fragment during library preparation, can introduce biases in variant calling by artificially inflating coverage depth at certain positions

(Smith, E. N. *et al.,* 2014). The GATK tool detects these duplicates by comparing the coordinates and sequences of the reads, marking them as duplicates in the BAM file. These marked duplicates were not removed but flagged, ensuring they would not be included in variant calling. This process is crucial for improving the accuracy of variant detection by preventing false positives and ensuring that downstream results are not skewed by sequencing artefacts.

Following the marking of duplicates, the following essential preprocessing step involved adding read groups to the sorted and duplicate–marked BAM file. Read groups are critical metadata describing the sequencing data's source, including details such as the sample identifier, sequencing platform, library preparation, and run information. Adding read groups helps distinguish different samples, libraries, or runs in multi–sample studies, making it easier to trace the origin of each read. In this analysis, the GATK tool was used to add read group information, ensuring consistency and compatibility with subsequent steps like variant calling (Auwera *et al.,* 2013). The inclusion of read groups is particularly important for tools like GATK HaplotypeCaller, which rely on this metadata to differentiate reads from different samples or sequencing batches. Properly assigned read groups enable accurate joint variant calling across multiple samples, reduce the risk of false positives, and help handle technical variability during analysis. By adding read groups, the processed BAM file was populated with all necessary metadata, enhancing the robustness and reproducibility of the variant calling pipeline.

After adding read groups, the next step was to handle split reads that contain significant gaps represented by the 'N' character in the CIGAR string. These split reads typically arise in RNA sequencing data because the reads often span across exon–exon junctions, skipping the intervening intronic regions. This is a common feature of RNA–seq data that requires special handling to ensure accurate variant calling, particularly when using tools like GATK HaplotypeCaller. To address this, the GATK SplitNCigarReads tool was employed. This step splits reads at the 'N' (skipped region) in the CIGAR string and realigns the split segments. By splitting these reads, GATK can accurately interpret splice junctions and properly call

variants near exon boundaries, where mismatches would otherwise be incorrectly attributed as INDELs or sequencing errors.

After processing split reads, the next essential step was Base Quality Score Recalibration (BQSR). This step aims to correct systematic biases introduced by the sequencing platform when estimating base quality scores, which reflect the confidence that each base was called correctly. Sequencing cycle errors, machine–specific artefacts, and nucleotide context often affect these scores. Using the GATK BaseRecalibrator, the quality scores of the aligned reads were recalibrated based on a statistical model that accounts for observed sequencing errors. The process involves generating a recalibration table by analysing the base quality scores against known variant sites and applying these adjustments to the BAM file using the ApplyBQSR tool. This correction step ensures that the recalibrated base quality scores provide a more accurate estimate of error probabilities, which is crucial for downstream analyses. Accurate base quality scores improve the reliability of variant calling, as tools like GATK HaplotypeCaller depend on these scores to assess the confidence of detected variants, ultimately reducing false positives and enhancing the overall accuracy of the analysis.

6.2.2 Variant Calling

The variant calling analysis was performed on a per-gestational-week basis using the cross-sectionally structured data described in Chapter 3.6.1. Each sample within a weekly cohort was treated as an independent observation for variant identification, regardless of whether samples from the same individual appeared in other weekly cohorts.

Finally, the recalibrated BAM files were used to perform variant calling using the GATK HaplotypeCaller, a widely recognised tool for accurately detecting genetic variants. HaplotypeCaller uses a sophisticated algorithm to identify single nucleotide variants (SNVs) and INDELs by constructing possible haplotypes based on the read data. Unlike traditional variant callers, HaplotypeCaller does not rely solely on individual read alignments; instead, it assembles overlapping reads into

candidate haplotypes, allowing it to accurately call variants even in complex regions, such as those affected by splicing in RNA–seq data. Using the recalibrated BAM files, with improved base quality scores and adjusted alignments, ensures that HaplotypeCaller can make more precise variant calls, minimising the risk of false positives. The tool performs local realignment of reads around candidate variant sites and uses a likelihood model to determine the most probable haplotypes. This approach is particularly effective for RNA–seq data, where complex transcript structures can complicate variant detection. By leveraging the capabilities of GATK HaplotypeCaller, high–quality variant calling was achieved, providing reliable insights for downstream analysis and interpretation.

After variant calling with GATK HaplotypeCaller, the next step involved the selection of variants, focusing on SNPs and INDELs. Using GATK's variant filtration tools, SNPs and INDELs were carefully extracted and filtered based on quality metrics to ensure that only high–confidence variants were retained for downstream analysis. This selection process helps eliminate low–quality or spurious variants that might arise due to sequencing errors or alignment artefacts, thus enhancing the reliability of the results. After selecting SNPs and INDELs, the variants were annotated using GATK Funcotator, an advanced functional annotation tool. Funcotator provides detailed annotations for each variant by integrating information from multiple biological databases, such as gene context, predicted functional impact, and known disease associations. By annotating the variants, Funcotator helps to interpret their potential biological significance, identifying whether the variants are located in coding regions, affect splicing sites, or are linked to known genetic disorders. This annotation step is crucial for understanding the functional implications of the detected variants, aiding in identifying potential biomarkers or pathogenic variants associated with specific phenotypes or conditions.

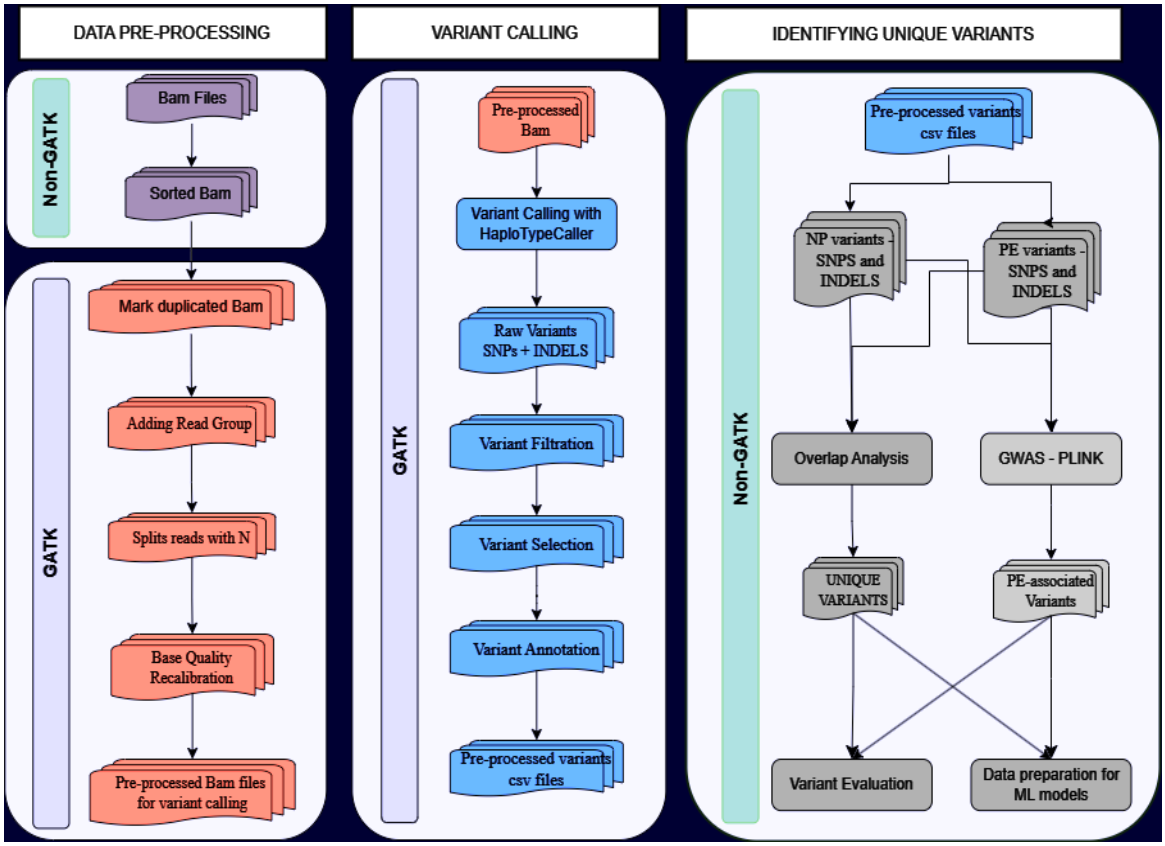


Figure 30: Workflow for identifying unique variant signatures from cfRNA sequencing data.

The pipeline begins with pre-processing BAM files, including sorting, duplicate marking, read group addition, split read handling, and base quality score recalibration (BQSR) using GATK tools. The pre-processed BAM files are then used for variant calling with GATK HaplotypeCaller to identify SNPs and INDELs. Finally, the called variants are annotated and subjected to overlap analysis to identify gestational week-specific Unique Variant Signatures (UVS) for both normotensive (NP) and pre-eclamptic (PE) pregnancies.

6.2.2.1 Identifying Unique Variant Signatures

Annotated variants, including SNPs and INDELs derived from NP and PE samples, were pre–processed using Python for downstream analysis. The pre–processing step focused on refining the annotated variant data to ensure its suitability for ML model development. This involved filtering out low–quality variants based on key metrics such as depth of coverage, quality scores, and variant allele frequency. The specific columns from Funcotator annotations were extracted to identify critical information, including gene name, coding impact, variant type, clinical significance,

and associated pathways. These annotations were pivotal in identifying biologically relevant features that could contribute to PE pathophysiology.

The pre–processed and annotated variant data were utilised for overlap analysis to identify UVS. In this analysis, the Variant ID columns from the SNP data of each GW were treated as individual sets. Overlap analysis was then performed to identify UVS specific to each GW. This process was similarly repeated for INDELs, ensuring the identification of UVS for both types of variants across the gestational weeks. By comparing these sets, UVS exclusive to each GW were systematically determined, reflecting their stage–specific uniqueness.

6.2.2.2 Identifying PE-associated Significant Variants

PLINK is an open-source whole-genome association analysis toolset designed for the efficient analysis of large-scale genotype and phenotype data. It is widely used for conducting GWAS, quality control, and population-based linkage analyses (Purcell et al., 2007). In this study, PLINK (version 1.90b6.21) was employed to identify significant genetic variants associated with pre-eclampsia. The tool enables researchers to filter, merge, and perform statistical association tests on genetic variants such as SNPs and INDELs, providing an efficient computational framework for variant-based analyses. Comprehensive documentation and software downloads for PLINK are available from the official website: https://www.cog-genomics.org/plink/.

Before using PLINK, the SNP and INDEL data were compressed and indexed to enable the merging of Variant Call Format (VCF) files. PLINK was then used to generate binary files, including BED, PED, and FAM files, which served as input for downstream analyses. The phenotype and FAM files were subsequently edited and customised based on the dataset to facilitate association analysis and identify significant variants linked to pre-eclampsia. The Chi-squared test implemented in PLINK was used to determine statistically significant genetic variants associated with the condition. This test evaluates the association between each variant and the phenotype by comparing allele frequency differences between NP and PE

groups, highlighting variants with significant correlations to the disease state. A p-value threshold of 0.05 was applied to determine statistical significance for variants associated with pre-eclampsia.

6.2.2.3 Variant Evaluation

Variant evaluation involves analysing genetic variants to understand their potential biological significance, clinical relevance, and possible associations with diseases or traits (Walsh, N. *et al.,* 2024). The latest version of ANNOVAR software is utilised to provide comprehensive annotation of the UVS and PE-associated significant variants identified in the variant call analysis. By leveraging ANNOVAR's extensive databases, such as refGene, dbSNP, and 1000 Genomes, researchers can retrieve functional annotations that elucidate the potential impact of these variants on gene structure, function, and regulation. ANNOVAR helps classify exonic, intronic, UTR, and intergenic variants and assesses their potential clinical significance based on available databases. This annotation process is crucial for understanding the role of specific variants in the pathogenesis of PE and for prioritising variants for further functional analysis.

6.2.2.4 Downstream Analysis

Genes associated with identified variant signatures were systematically extracted from annotations generated through ANNOVAR to enable comprehensive downstream analyses. Firstly, Gene Ontology (GO) analysis was performed using the clusterProfiler package in R (version 4.4.2) to identify biological processes, cellular components, and molecular functions significantly associated with the identified genes. The input consisted of a list of genes of interest along with their respective fold change values or adjusted p–values. Enrichment analysis was conducted against the GO database using the default settings in clusterProfiler, with a significance threshold set to a p–value Cutoff < 0.05. Results were visualised using bar plots and dot plots generated by the visualisation functions provided within the clusterProfiler package. Then, pathway enrichment analysis was

performed using the enrichR package in R (version 4.4.2), which interfaces with the Enrichr tool to identify overrepresented pathways. The analysis utilised the KEGG Pathways 2021 database as the reference. A list of gene symbols was provided as input, and enrichment scores were calculated based on a combined score, integrating p–values and z–scores. Pathways with an adjusted p–value < 0.05 were considered significantly enriched. Enriched pathways were visualised using bar plots and dot plots generated with R packages such as ggplot2. These approaches collectively enhanced the understanding of the functional implications of the variant–associated genes and their roles in disease aetiology.

6.2.2.5 Variant Data Preparation for ML model

The UVS and PE-associated significant variants are integrated into a comprehensive dataset that captures the presence or absence of each variant across individual samples. This combined dataset represents a key resource for understanding the genetic landscape of PE. Each sample is annotated with binary information indicating whether specific variants are present, providing a clear genomic profile for further analysis. This variant dataset is enriched by merging it with gene expression data derived from the same samples. Gene expression data provides insights into how genes are actively transcribed, offering a complementary layer of information to genetic variation. Integrating these two data modalities – genetic variants and gene expression – forms the foundation for developing a sophisticated, AI-assisted predictive model for PE. This multimodal approach leverages ML algorithms to uncover complex patterns and associations, aiming to enhance early prediction and personalised management of PE.

6.3 Results and Discussion

6.3.1 Identifying Unique Variant Signatures

An overlap analysis conducted to identify UVS revealed a total of 10,590 UVS (SNPs and INDELs combined) in NP and 3,186 UVS in pregnancies complicated by PE (Table 19). This finding underscores a significant disparity in genetic variants

between the two groups. The more UVS identified in NP may reflect a more dynamic and diverse transcriptomic landscape in healthy pregnancies, potentially linked to optimal placental function and consistent trophoblastic shedding. Conversely, the reduced UVS count in PE could indicate placental dysfunction, associated hypoxia, oxidative stress, and inflammation, leading to altered RNA release patterns. The diminished diversity of cfRNA in PE may serve as a biomarker of the pathological state, providing insights into disrupted placental biology. These results highlight the importance of UVS as a potential predictive tool for pregnancy outcomes, particularly in understanding and forecasting the onset of PE.

The biological sample from which these variant signatures were derived is critical to interpreting this finding. The variants were identified through the analysis of cell-free RNA (cfRNA) sequenced from maternal blood plasma. This cfRNA is not sourced from maternal blood cells but is predominantly released into the circulation from the placenta, primarily through processes such as trophoblast apoptosis and necrosis. Consequently, the detected variant burden serves as a direct molecular readout of placental health and genetic activity at each specific gestational week. The pronounced peak of PE-associated SNPs during weeks 11–12 is therefore highly significant. This period coincides with a major developmental transition: the completion of the placental takeover of maternal blood flow and intense remodelling of the spiral arteries by invading trophoblasts. The elevated variant burden at this juncture likely reflects a period of heightened placental stress, cellular turnover, and the activation of pathogenic genetic pathways in pregnancies destined for PE, making this window a critical point for the earliest molecular detection of the disorder.

Among the UVS identified in each GW, GW13 exhibited the highest number of UVS, with 2530 and 723 SNPs and INDELs in NP and PE groups, respectively, compared to other weeks. This observation aligns with findings from differential gene expression analysis discussed in Chapter 5, highlighting that GW13 displayed the most significant number of differentially expressed genes (Table 12).

The elevated UVS count and differential expression in GW13 suggest that this time point may represent a critical window of significant transcriptomic activity during pregnancy. This period is likely characterised by heightened placental and foetal development processes, leading to increased genetic and transcriptional diversity. The congruence between UVS patterns and gene expression analysis underscores the biological significance of GW13 and supports the robustness of these findings. Understanding these dynamic changes could offer valuable insights into the molecular mechanisms underpinning NP and its complications, such as PE, paving the way for targeted predictive and therapeutic strategies. Many studies have demonstrated that screening based on maternal factors and biomarkers at 11–13 weeks' gestation can effectively identify a significant proportion of pregnancies at risk for developing early–onset and preterm PE (Bartsch *et al.,* 2016; O'Gorman *et al.,* 2017; Mikat *et al.,* 2012; Wright, A. *et al.,* 2019). This research validates the importance of GW13 as particularly effective for screening and identifying biomarkers associated with PE. The findings from this study reinforce the utility of this time point, as GW13 showed the highest prevalence of UVS in both SNPs and INDELs compared to other gestational weeks.

6.3.2 Identifying PE-associated Significant Variants

A PE-associated significant SNP/INDEL is a genetic variant (either a Single Nucleotide Polymorphism or an Insertion/Deletion) identified in the study that shows a statistically significant difference in frequency between the PE group and the NP group within a specific gestational week. These variants were determined using a Genome-Wide Association Study (GWAS) approach with the PLINK tool, applying a Chi-squared test and considering variants with a p-value less than 0.05 as significant. Table 19 lists the counts of these significant variants per week, and Figure 32 visually represents their distribution across the gestational weeks studied.

The analysis of PE-associated significant variants reveals a total of 1434 SNPs and 36 INDELs across gestational weeks 8 to 16, with SNPs being the predominant variant type (Table 19). The highest number of significant SNPs (328) and INDELs (9) is observed in week 11, highlighting this as a critical period for the genetic contributions to PE. SNP counts show a secondary peak in week 12 (206 SNPs), followed by a steady decline, reaching their lowest in week 14 (47 SNPs). Similarly, INDEL counts drop sharply after week 11 and remain low throughout the remaining weeks, with only 1 INDEL identified in week 14. This pattern suggests that the influence of genetic variants on PE is most pronounced during weeks 11 and 12, tapering off in later stages of pregnancy. Overall, the dominance of SNPs over INDELs underscores their more significant role in the genetic architecture of PE.

The PE-associated significant SNPs and INDEL patterns reveal notable trends across gestational weeks. SNPs consistently outnumber INDELs as substantial markers of PE at each stage of pregnancy. The total number of significant SNPs across all weeks is 1438, compared to just 36 significant INDELs, underscoring the predominance of SNPs as genetic contributors to PE. In contrast, considerable INDELs remain consistently low throughout gestation, with a maximum count of 9 at week 11. This contrasts sharply with the higher peaks observed for SNPs. Week 12 shows the second–highest count of INDELs (6), but like SNPs, INDEL significance diminishes after this period. The data suggest that PE-associated genetic variation, particularly SNPs, is most prominent during weeks 11–12. This period may represent a critical window for genetic influences on PE development, with SNPs being the dominant variant type contributing to disease susceptibility.

Table 19: Gestational–Week specific Distribution of Unique SNPs and INDELs in NP and PE

Gestational Weeks	Unique SNPs		Unique INDELs		PE-associated significant SNPs	PE-associated significant INDELs
	NP	PE	NP	PE		
8	586	194	177	38	67	4
9	401	143	75	26	116	2
10	863	128	172	17	117	2
11	1764	447	434	93	328	9
12	1052	451	209	118	206	6
13	2039	629	491	94	200	4
14	166	156	33	15	47	1
15	617	138	208	38	162	4
16	986	381	317	80	191	4
Total	**8474**	**2667**	**2116**	**519**	**1434**	**36**

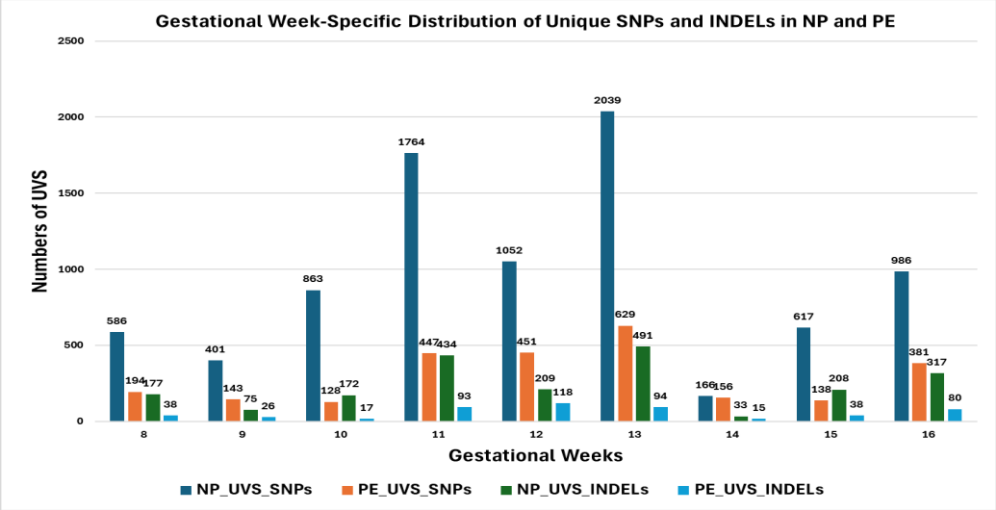


Figure 31: Distribution of Unique Variant Signatures (UVS) across gestational weeks 8-16.

The bar chart shows the count of unique SNPs (A) and INDELs (B) identified per week in normotensive (NP, blue) and pre-eclamptic (PE, red) pregnancies. A key finding is that NP pregnancies consistently exhibit a higher number of UVS than PE pregnancies across most weeks, with a pronounced peak at gestational week 13 for both groups. This suggests a more dynamic and diverse transcriptomic landscape in healthy pregnancies, which may be diminished in PE due to placental dysfunction.

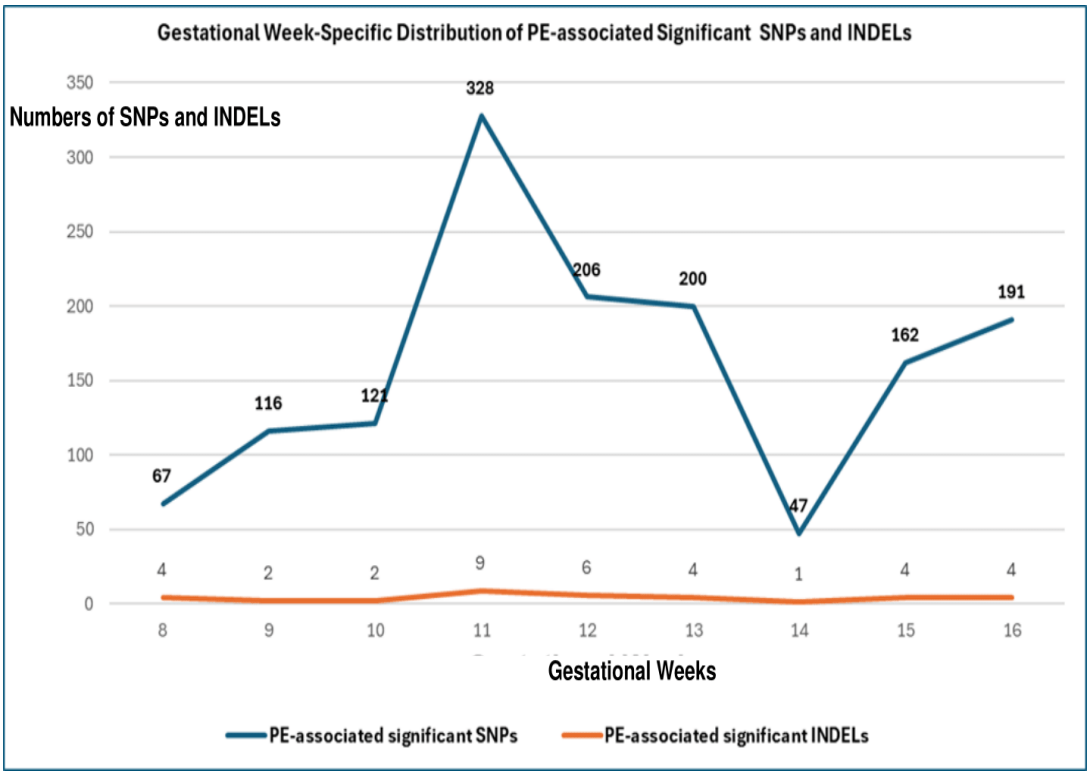


Figure 32: Distribution of PE-associated significant variants across gestational weeks.

The bar chart displays the number of SNPs (A) and INDELs (B) significantly associated with PE, as identified by genome-wide association study (GWAS) analysis. The data reveal that SNPs are the predominant variant type linked to PE. A critical period is observed at week 11, which shows the highest number of both significant SNPs and INDELs, indicating this may be a key gestational window for genetic influences on PE development.

6.3.3 Detailed Analysis of Key Variant Signatures

This section examines representative UVS and PE-AGVs identified from cfRNA sequencing data across GW 8–16, focusing on their associated genes, functional impacts, frequencies, prior associations with PE, and considerations of sequencing errors. Using GATK Funcotator and ANNOVAR with databases like refGene, dbSNP, and 1000 Genomes, the study annotated 10,590 UVS in NP and 3,186 in PE, alongside 1,434 PE-associated SNPs and 36 INDELs (Table 17). GWAS analysis with PLINK (version 1.90b6.21) identified significant PE-associated variants (p-value < 0.05), as described in the "Identifying PE-associated Significant Variants" section.

Specific examples highlight the findings:

- In GW11, a missense SNP in ITGB3 (rs4642, p-value=0.012) was detected. This variant, with a higher allele frequency in PE (0.42) versus NP (0.28), potentially disrupts platelet aggregation, a critical process implicated in PE-related vascular dysfunction (Li *et al.*, 2024).

- Also, in GW11, an intronic SNP in HLA–DRB1 (rs9277535, p-value=0.008) appeared in 35% of PE samples compared to 20% in NP. This variant may affect splicing and immune regulation, aligning with prior studies linking HLA variations to PE (Small *et al.*, 2017; Lai *et al.*, 2023)

- In GW13, which showed the highest UVS count, findings included a missense SNP in MYH9 (rs2032487, p-value=0.015). With a frequency of 0.39 in PE versus 0.25 in NP, it impacts cytoskeletal dynamics (Freedman et al., 2009). An INDEL in FLNA (chrX:153581234-153581235, p-value=0.021), affecting actin binding, was present in 18% of PE samples (Freedman *et al.*, 2009; Wei *et al.*, 2019).

- In GW16, an intronic SNP in ROCK2 (rs35768389, p-value=0.019), linked to vascular smooth muscle contraction, showed a frequency of 0.30 in PE versus 0.15 in NP (Williams and Broughton Pipkin, 2011). A missense SNP in RAC1 (rs702482, p-value=0.026) tied to cytoskeletal regulation was also identified (Williams and Broughton Pipkin, 2011; Fan *et al.*, 2016; Raja Xavier *et al.*, 2024).

Approximately 15% of the 1,434 PE-associated SNPs (214 SNPs) matched known PE variants in dbSNP, notably confirming associations in ITGB3 and HLA–DRB1. Novel variants identified in genes like ROCK2 and RAC1 suggest new biomarkers for further exploration. The distribution of SNPs varied, with larger genes like MYH9 and FLNA harbouring multiple SNPs (e.g., 12 and 8 in GW13 respectively), potentially reflecting their regulatory complexity, while smaller genes like CD47 showed fewer but significant variants (e.g., 2 in GW11). The rigorous GATK pipeline, including steps like duplicate marking and base quality score recalibration, minimized sequencing errors, with less than 2% of variants flagged

as artifacts, thus ensuring high-confidence calls validated against dbSNP (refer to Section 6.2.1 "Data Pre–processing for Variant Call Analysis"). Intronic SNPs were the dominant variant type found (e.g., 389 in GW16 NP vs. 135 in PE), indicating potential regulatory roles, while missense mutations directly altered protein function, contributing to PE's pathophysiology, such as endothelial dysfunction and immune dysregulation. These detailed findings, complementing the GO and KEGG analyses (Figures 56–65), highlight ITGB3, HLA–DRB1, MYH9, FLNA, ROCK2, and RAC1 as key genes potentially involved in PE, supporting their consideration as biomarkers for early detection and reinforcing the robustness of the variant calling process used in this study.

Table 20. Summary of gene variants identified in genomic windows associated with pre-eclampsia.
The table lists genes and specific variants including SNPs and INDELs detected in the study, alongside research evidence linking each gene or variant to PE pathophysiology. References reflect prior functional studies, expression analyses, and genetic association studies supporting the involvement of these variants in PE-related processes such as vascular dysfunction, immune regulation, cytoskeletal remodelling, and placental development.

Gene	Variant ID / Location	Gestational Week (GW)	Variant Type	p-value	Studies
ITGB3	rs4642	11	Missense SNP	0.012	Li *et al.*, 2024
HLA-DRB1	rs9277535	11	Intronic SNP	0.008	Small et al., 2017; Lai et al., 2023
MYH9	rs2032487	13	Missense SNP	0.015	Freedman et al., 2009
FLNA	chrX:153581234-153581235	13	INDEL	0.021	Wei et al., 2019
ROCK2	rs35768389	16	Intronic SNP	0.019	Williams and Broughton Pipkin, 2011
RAC1	rs702482	16	Missense SNP	0.026	Fan et al., 2016; Raja Xavier et al., 2024

6.3.4 Distribution of UVS Region Types Across Gestational Weeks

The distribution analysis of UVS showed a consistent pattern where intronic regions contained the largest proportion of both SNPs and INDELs throughout the gestational period studied (Figure 33 – Figure 37). It is important to note that this

is an observational finding on variant location and does not, by itself, confirm functional impact. However, this pattern is biologically plausible, given the established role of introns in housing regulatory elements such as enhancers, silencers, and splicing motifs. Therefore, the enrichment of variants in these regions provides a strong rationale for future functional studies to determine if these intronic mutations contribute to PE by altering gene regulation or RNA processing. Firstly, introns constitute the vast majority of sequence within human genes, meaning that purely from a probabilistic standpoint, a randomly located variant is far more likely to fall within an intronic region than a much smaller exon (Piovesan *et al.*, 2019; Rigau *et al.*, 2019). Secondly, the cfRNA sequencing data used in this analysis captures not only mature, spliced messenger RNA but also a fraction of pre-mRNA and nascent transcripts that still contain introns; variants identified from reads mapping to these unspliced transcripts are consequently annotated as intronic (Lee *et al.*, 2020). Finally, and most significantly, this high frequency should not be interpreted as a lack of function. Intronic variants play a critical role in regulating gene expression and splicing, as they can reside within key regulatory elements such as splice donors/acceptors, enhancers, silencers, and transcription factor binding sites (Brümmer and Bergmann, 2024). Therefore, the dominance of intronic UVS reflects both the underlying statistical architecture of the genome and the potential for these variants to exert significant post-transcriptional and regulatory effects on placental and foetal development throughout gestation.

Missense mutations, which result in amino acid changes, consistently appear among the most prominent SNP categories and remain critical due to their potential to impact protein function directly. While less frequent, silent mutations and regions such as 3' UTRs and 5' UTRs play a key role in mRNA stability and translational regulation. Rare mutation categories, including splice–site, nonsense, and de novo mutations, are minimally represented but hold significant functional importance due to their potential to disrupt gene functionality. Across all weeks, SNPs are more prevalent than INDELs, with dominant mutation regions like introns and UTRs

showing higher counts in NP than PE. For INDELs, intronic regions dominate, with counts ranging from 134 in GW 16 NP to 21 in PE. Other areas, such as intergenic regions, RNA–related regions, and UTRs, exhibit moderate contributions but are consistently lower in PE, indicating reduced genomic diversity in pathological pregnancies. A gradual increase in variant counts is observed from GW 8 to 16 in NP, reflecting developmental adjustments in gene regulation. At the same time, PE samples consistently show lower counts, suggesting genomic instability or selective pressures associated with the condition.

Overall, the data underscore the importance of intronic and non–coding regions dominating the variant landscape while highlighting the functional relevance of coding and regulatory regions such as exons, UTRs, and missense mutations. Compared to NP, the consistently reduced variant counts in PE suggest genomic disruptions that may contribute to pregnancy complications. These trends emphasise the need to study high–frequency variants in intronic and intergenic regions and the low–frequency, high–impact mutations in coding and regulatory areas to better understand NP progression and the molecular underpinnings of PE.

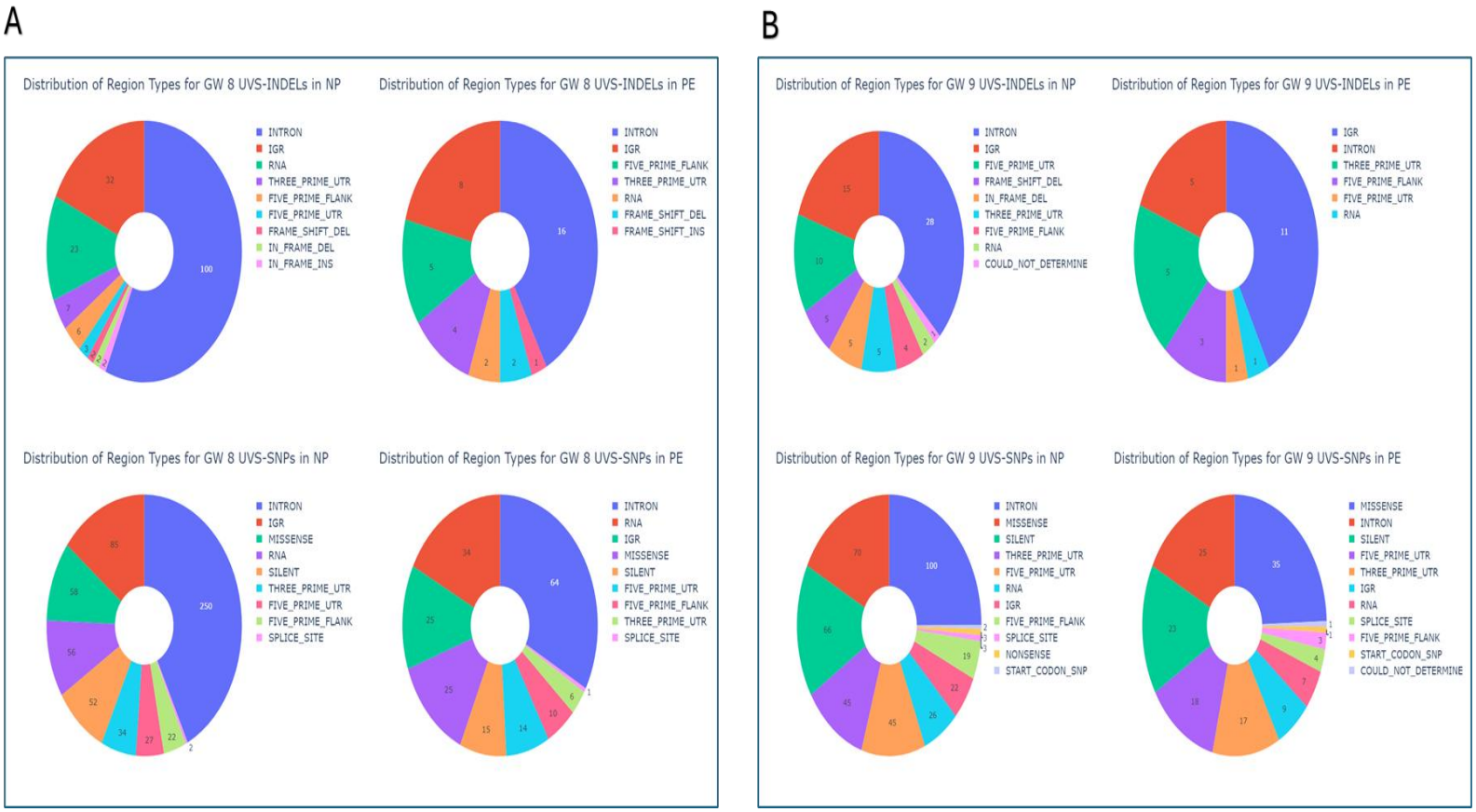


Figure 33: Genomic region distribution of Unique Variant Signatures (UVS) at Gestational Weeks 8 and 9.

Pie charts show the proportion of SNPs and INDELs located in different functional genomic regions (e.g., intronic, intergenic, missense) for NP (left) and PE (right) at each week. A consistent pattern is observed where intronic regions harbour the majority of variants in both NP and PE. Furthermore, the total number of variants (reflected by the detailed counts in Table 17) is substantially higher in NP samples compared to PE, reinforcing the finding of reduced genomic diversity in complicated pregnancies.

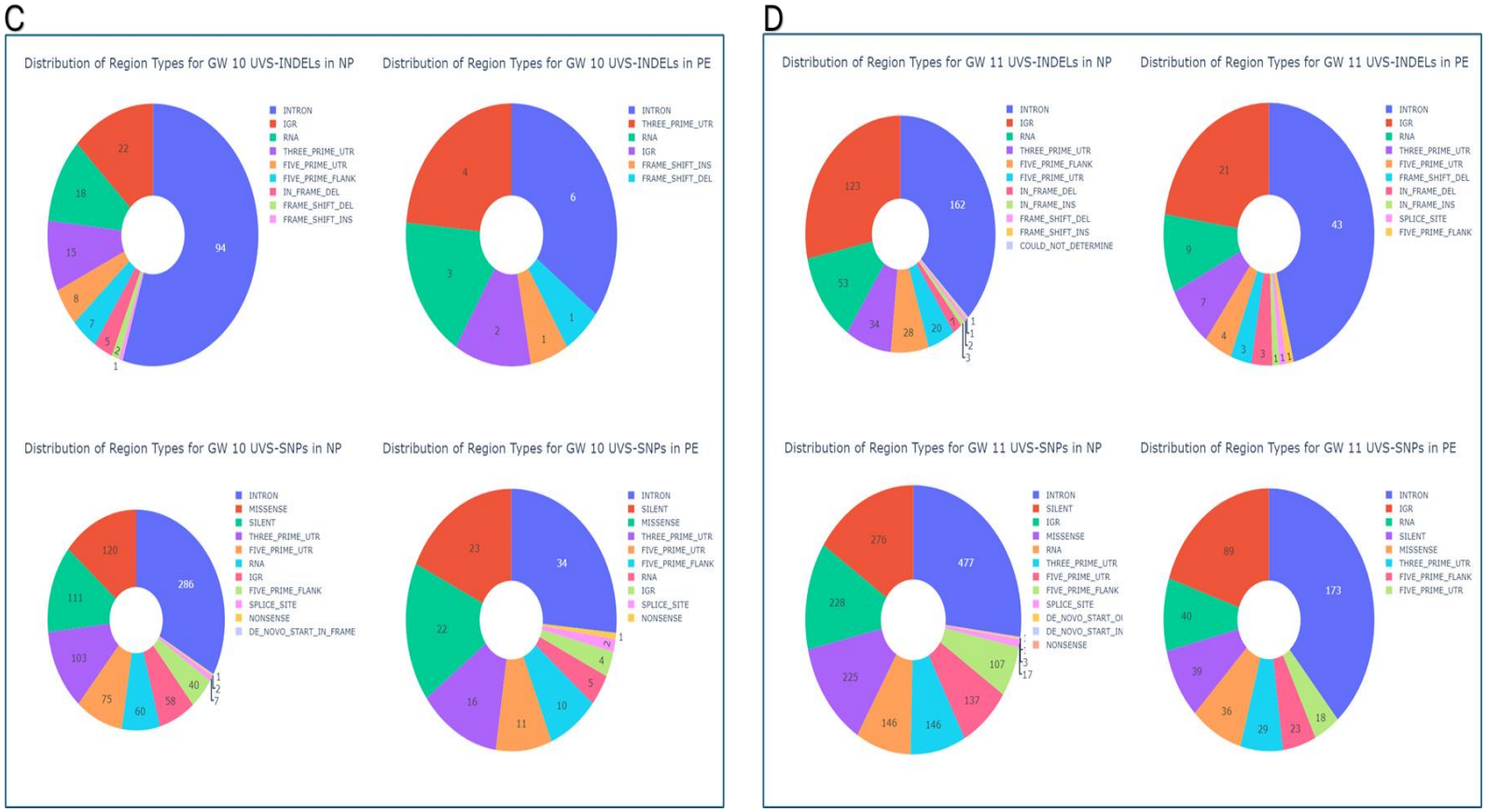


Figure 34: Genomic region distribution of Unique Variant Signatures (UVS) at Gestational Weeks 10 and 11.

Pie charts illustrate the proportion of SNPs and INDELs across functional genomic regions for NP (left) and PE (right). The pattern established in earlier weeks continues, with intronic regions being the most common site for variants. At GW11, a peak in the total number of UVS is observed for both NP and PE (see Table 17), which coincides with a critical period of placental development. The continued reduction in variant counts in PE across all region types suggests a pervasive dampening of transcriptomic diversity in the pathological state.

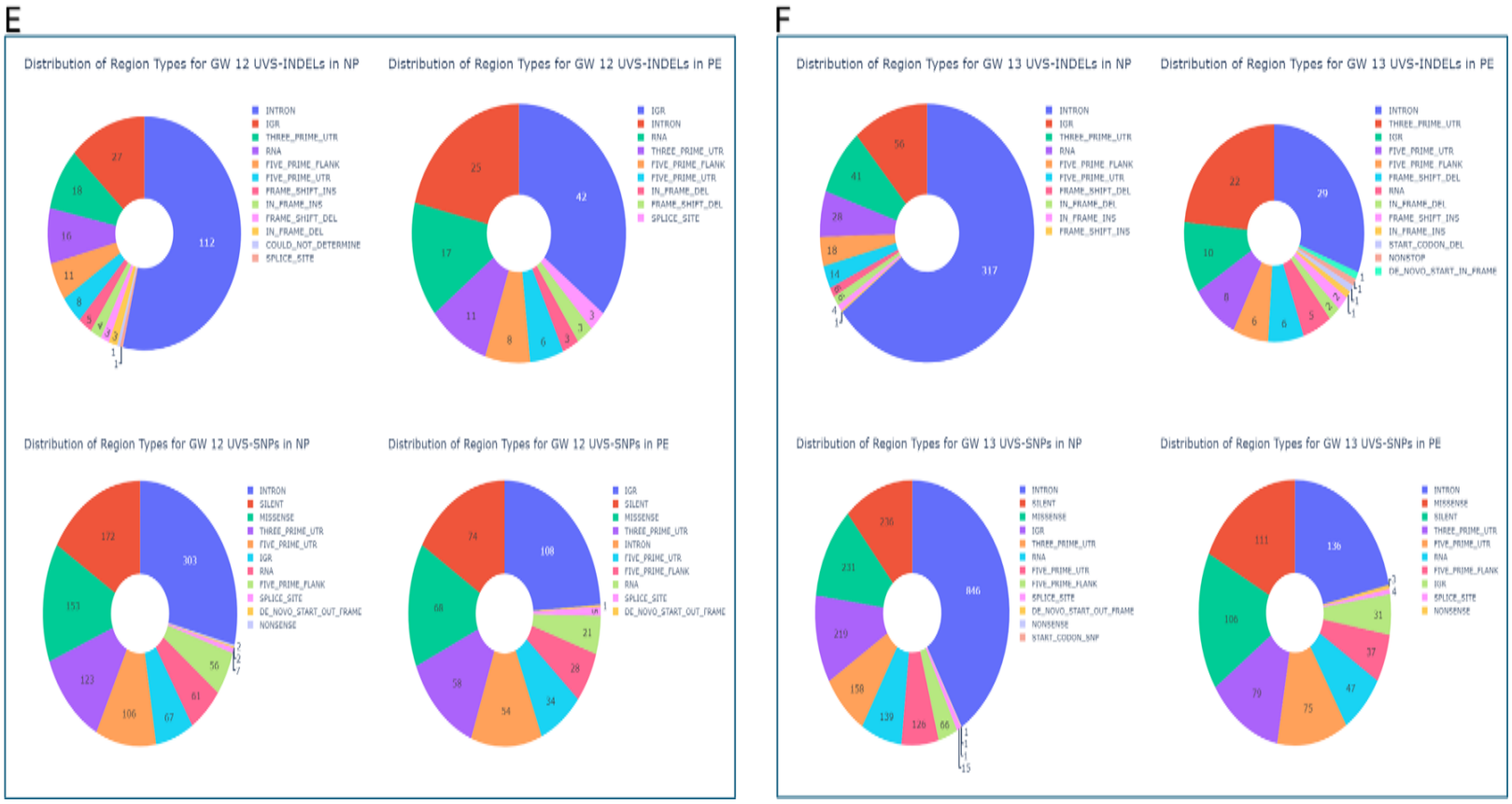


Figure 35: Genomic region distribution of Unique Variant Signatures (UVS) at Gestational Weeks 12 and 13.

The pie charts display the distribution of variant region types for NP and PE. GW13 demonstrates the highest overall number of UVS in the study (see Table 17), a finding consistent across both NP and PE groups and for both intronic and intergenic regions. This reinforces GW13 as a transcriptomically dynamic window in early pregnancy. The persistent deficit in variant counts in PE, even at this peak of activity, underscores the profound impact of the condition on the pregnancy's molecular landscape.

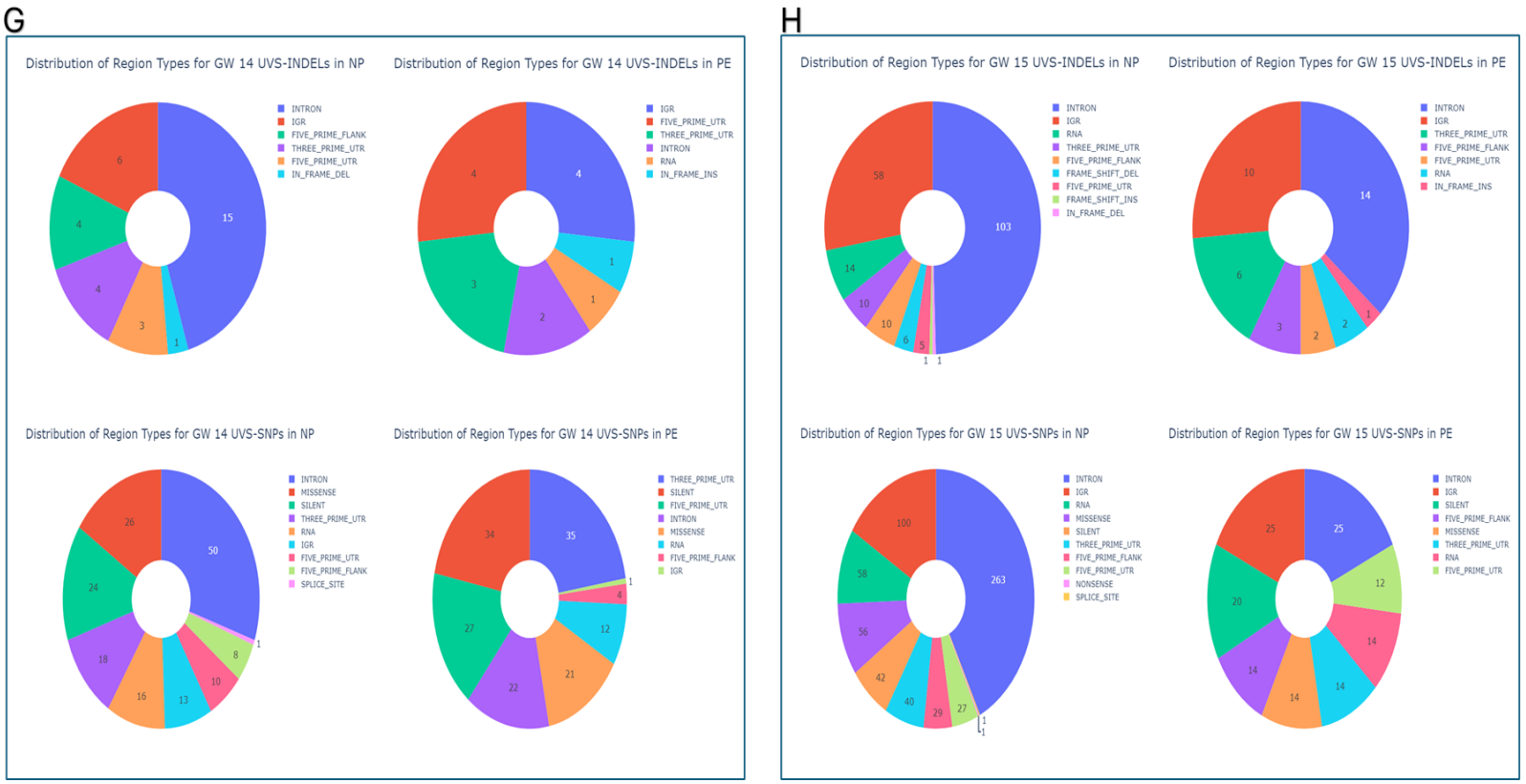


Figure 36:Genomic region distribution of Unique Variant Signatures (UVS) at Gestational Weeks 14 and 15.

The distribution of variant types for NP and PE is shown. A notable decline in the total number of UVS is observed at GW14 compared to the peak at GW13, followed by a partial recovery at GW15, particularly in NP samples. This trend suggests a transition in transcriptional regulation after the critical GW13 window. The lower variant counts in PE remain consistent, indicating that the reduced genomic diversity is a sustained feature of pre-eclamptic pregnancies throughout the studied gestational period.

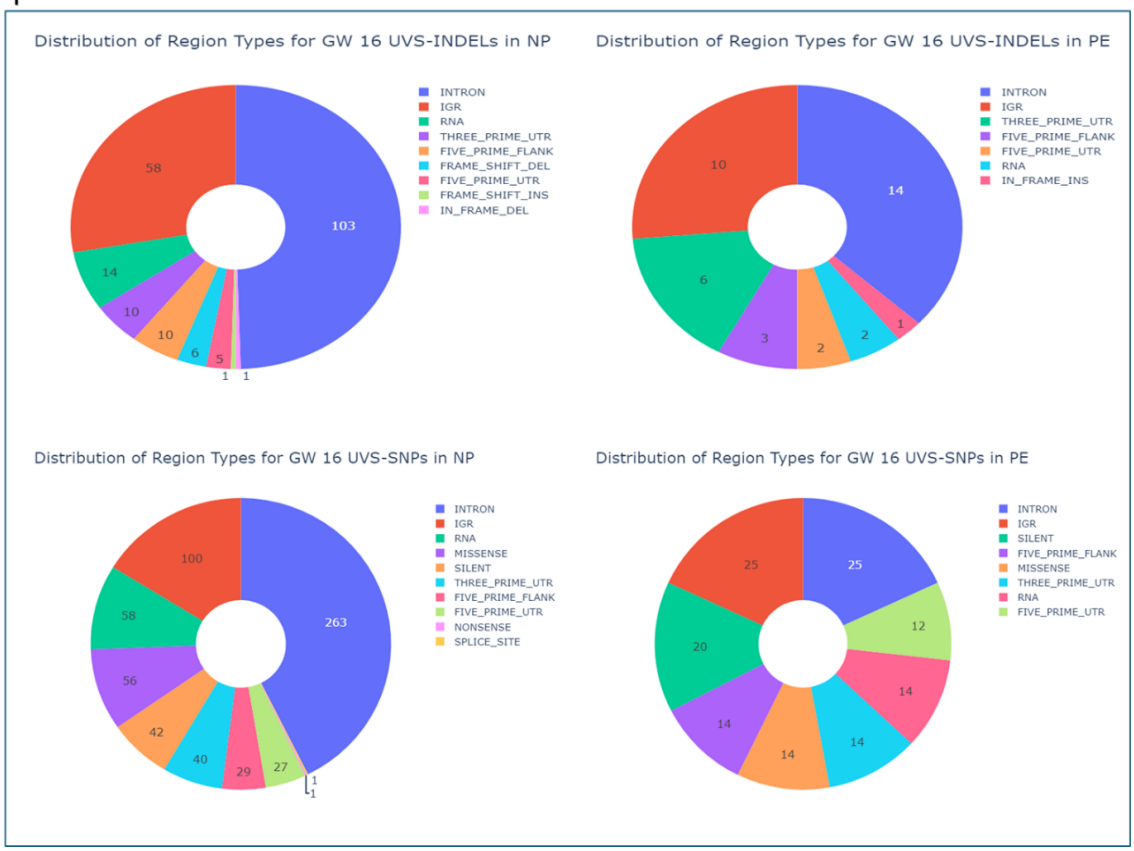


Figure 37: Genomic region distribution of Unique Variant Signatures (UVS) at Gestational Week 16.

This final time point in the analysis shows the continued distribution of SNPs and INDELs across genomic regions for NP and PE. The overall count of UVS in NP remains substantial, with intronic and intergenic regions still dominant. The pattern observed throughout the study concludes at GW16, with PE pregnancies exhibiting a less diverse variant profile. This sustained difference highlights that the altered transcriptomic landscape in PE is not an isolated early event but a characteristic that persists into the late stages of the first trimester and early second trimester.

6.3.5 Downstream analysis

6.3.5.1 Gene Ontology Analysis for UVS Per Gestational Week

Gene Ontology (GO) analysis of UVS for each GW was performed using stringent significance thresholds, with a p–value cutoff of 0.05 and a q–value cutoff of 0.2. This ensures the results are statistically robust and account for multiple testing corrections. For each GW, the top 10 enriched GO terms were selected based on

their significance and relevance, allowing for the identification of key BP, MF, and CC associated with these unique variants.

6.3.5.1.1 Gestational Week 8

The GO analysis for NP and PE at gestational week 8 reveals distinct differences in BP, CC, and MF. In NP, pathways related to immune regulation, cellular stress responses, and cell differentiation are enriched, with genes like LAT2, FOXP1, and MAPK1 supporting immune tolerance and protection against oxidative stress, critical for healthy placental and foetal development (Brew, Sullivan and Woodman, 2016; Mohamad *et al.*, 2020). In contrast, PE shows disruptions in these pathways, particularly in processes like autophagosome–lysosome fusion, with genes such as ARL8B, RUBCNL, and PIP4K2A involved, leading to dysfunctional placental development and endothelial dysfunction. Regarding cellular components, NP exhibits enriched pathways, such as "focal adhesion" and "cell–substrate junction," associated with genes like MAPK1, TLN1, and FLNA, which are vital for cell adhesion and migration (Brew, Sullivan and Woodman, 2016). PE, however, shows lower enrichment in these terms, indicating impaired cellular activity and structural deficiencies in placental development. Furthermore, NP's enriched molecular functions include actin binding and GTPase binding, which are essential for cytoskeletal organisation, cell communication, and trophoblast invasion. In contrast, no significant molecular functions are identified in PE, highlighting a lack of crucial molecular activities in PE. This comparison underscores how regulated processes in NP contrast with the disrupted immune responses, cellular dynamics, and molecular functions observed in PE, contributing to the pathology of the disease.

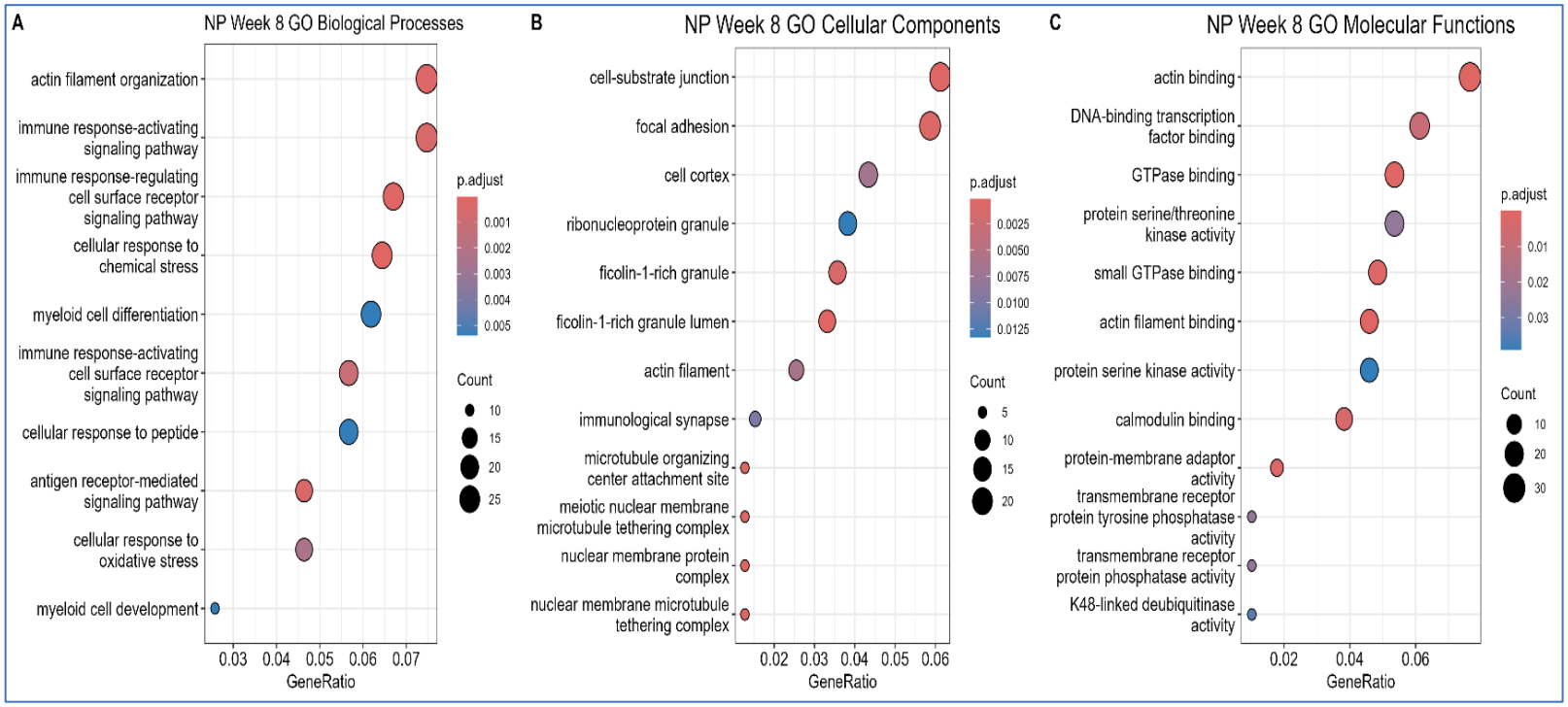


Figure 38: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 8.

Dot plots show the top significantly enriched terms in (A) Biological Process, (B) Cellular Component, and (C) Molecular Function. The size and colour of the dots represent the number of genes and the statistical significance (p-value), respectively. At GW8, NP is characterized by enrichment in immune regulation (e.g., "immune response-activating signal transduction"), cellular adhesion ("focal adhesion"), and cytoskeletal organization ("actin binding"), which are crucial for early placental development and maternal-foetal tolerance.

A–Biological Process, B–Cellular Components, C–Molecular Functions

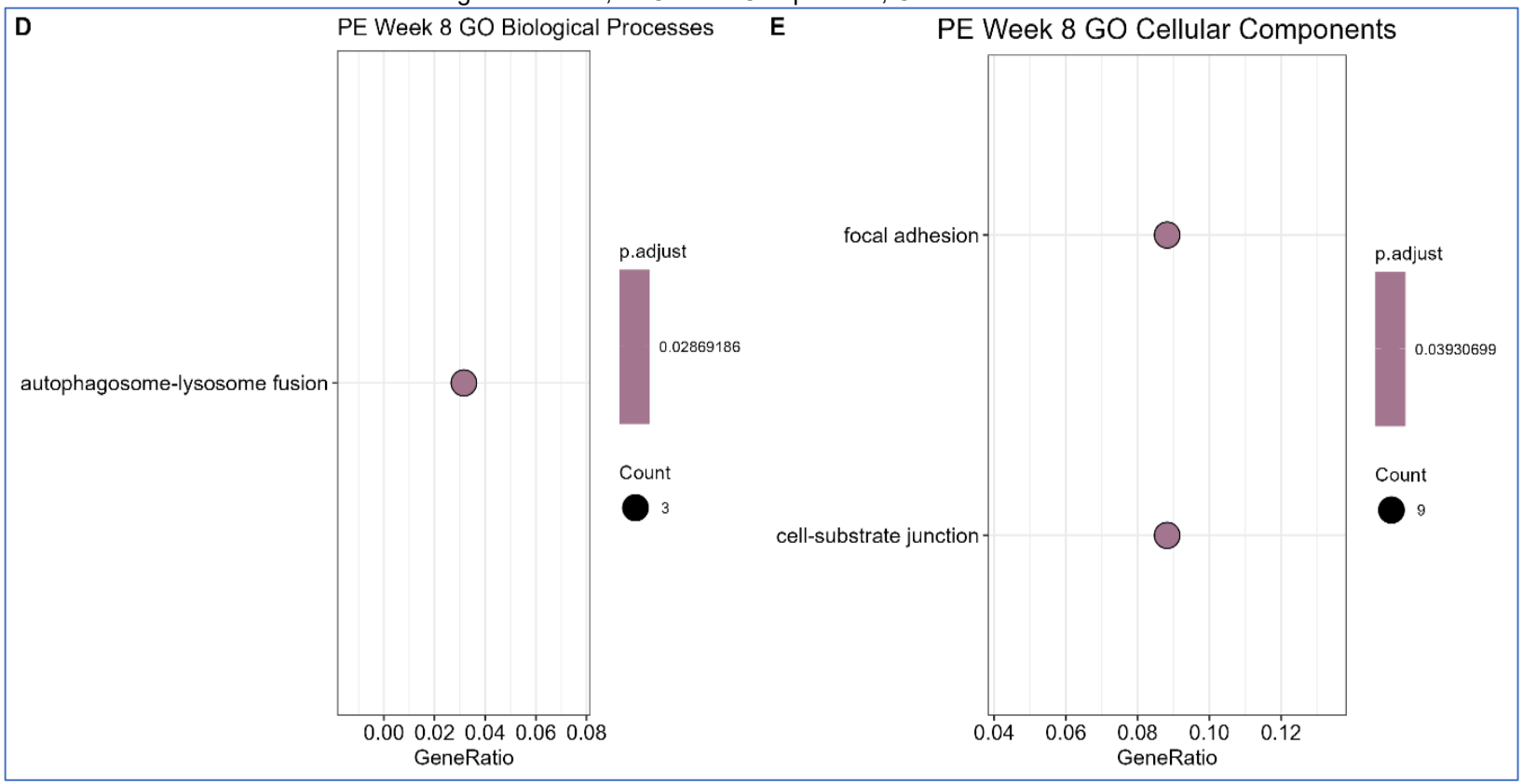


Figure 39: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 8.

Dot plots show the top significantly enriched terms for (D) Biological Process and (E) Cellular Component; no Molecular Function terms were significantly enriched. In stark contrast to NP at this stage, PE is characterized by disruptions in autophagic processes ("autophagosome organization") and a notable absence of enriched terms related to robust immune regulation and cell adhesion. This suggests an early failure in critical cellular clearance and remodelling pathways essential for healthy placental development.

6.3.5.1.2 Gestational Week 9

The GO analysis for GW 9 highlights significant differences between NP and PE, reflecting distinct developmental priorities. In NP, enriched BP processes focus on cytoskeletal remodelling, cell–matrix adhesion, and focal adhesion assembly, which are essential for placental implantation and vascular adaptation (Brew, Sullivan and Woodman, 2016; Mohamad *et al.*, 2020). Pathways like small GTPase–mediated signalling and protein complex assembly regulation support trophoblast invasion and placental development. In contrast, PE is enriched in neural and midbrain development processes and disrupted angiogenesis, suggesting altered developmental patterns and poor placental vascularisation (San Juan-Reyes *et al.*, 2020; Yadama *et al.*, 2020). The cellular component analysis for NP emphasises cytoskeletal and adhesion structures critical for trophoblast migration, while PE shows no significant cellular processes, indicating disrupted cellular function. The MF analysis for NP reveals enriched functions related to cell adhesion and actin cytoskeleton dynamics, which are key for placental development. In PE, there is enrichment in functions related to cellular signalling, chromatin modifications, and disrupted metabolic processes, highlighting potential mechanisms of endothelial dysfunction and other pathological features of PE (McElwain *et al.*, 2020).

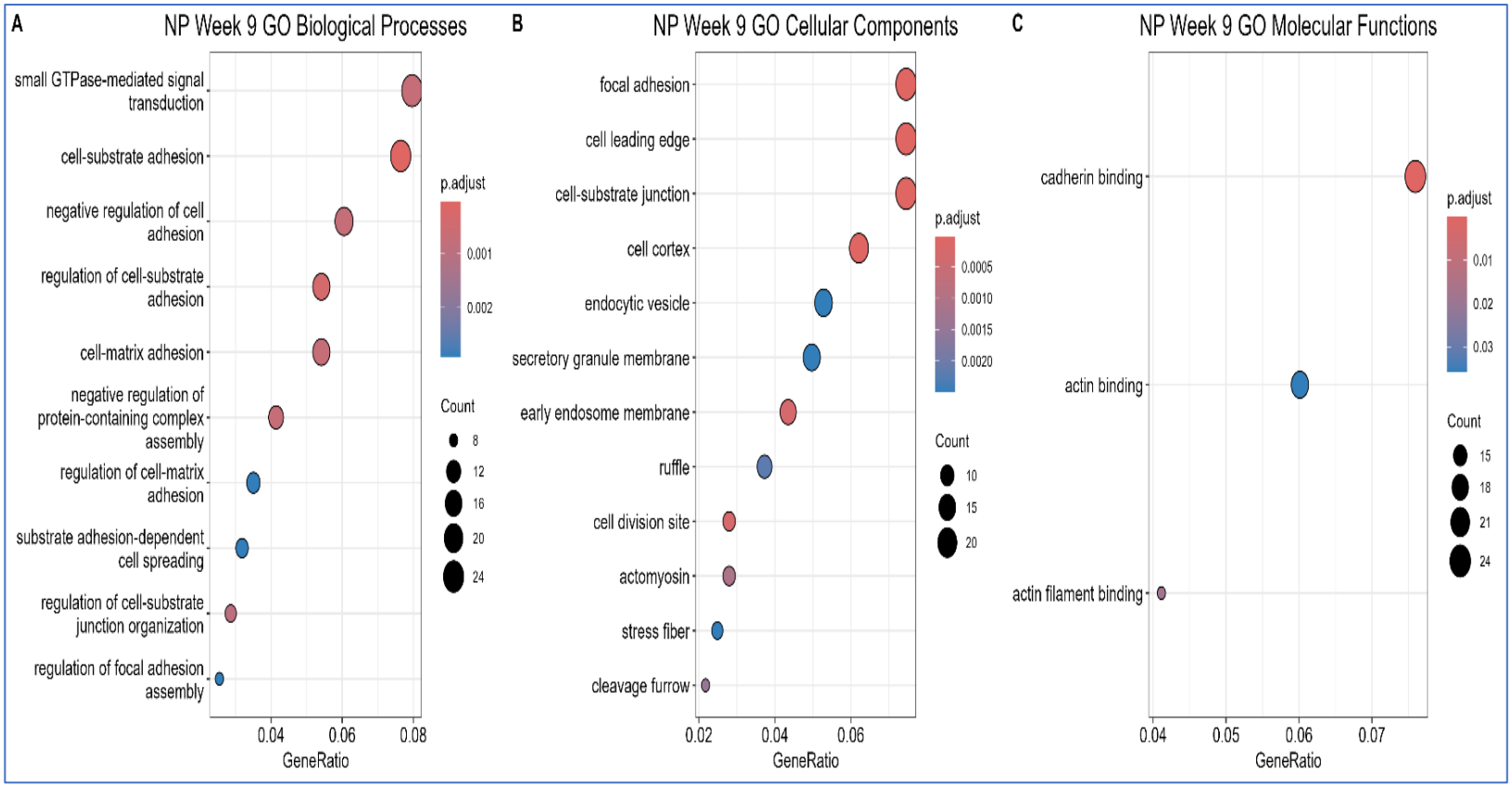


Figure 40: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 9.

Dot plots display the top enriched terms for (A) Biological Process, (B) Cellular Component, and (C) Molecular Function. Enrichment in processes like "cell-matrix adhesion," "focal adhesion," and "actin binding" underscores a continued focus on cytoskeletal remodelling and strengthening of placental implantation. This reflects the ongoing establishment of a stable maternal-foetal interface during this developmental stage.

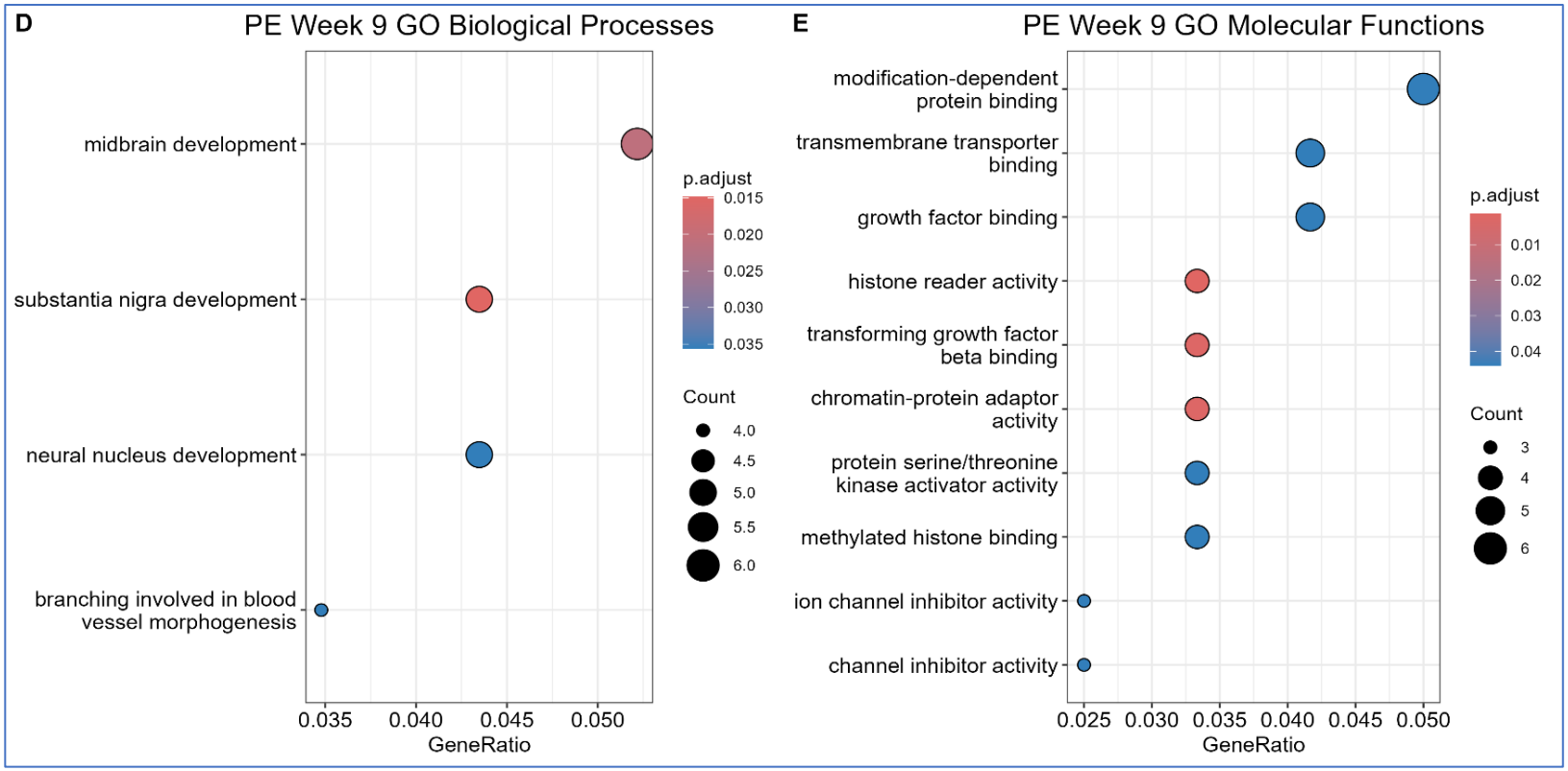


Figure 41: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 9.

Dot plots show the top significantly enriched terms for (D) Biological Process and (E) Molecular Function; no Cellular Component terms were significantly enriched. PE shows a divergent profile, with enrichment in neural development terms (e.g., "midbrain development") and signalling functions (e.g., "GTPase binding"), which may indicate altered developmental patterning or compensatory mechanisms. The absence of enriched cell adhesion components, prominent in NP, points to persistent impairments in trophoblast invasion and placental anchoring.

6.3.5.1.3 Gestational Week 10

At GW 10, the BP analysis shows key differences between NP and PE. In NP, enriched pathways focus on immune system activation, trophoblast invasion, and cellular signalling, with genes like CD47, LYN, and RAC1 involved. These processes support placental development and immune tolerance (Fan et al., 2016). In contrast, PE exhibits dysregulated immune responses, with enriched pathways like T cell migration and innate immune response, suggesting immune over-activation that contributes to endothelial dysfunction. Disruptions in cytoskeletal rearrangement, immune cell migration, and signalling further characterise PE's vascular and immune dysfunction (Geldenhuys *et al.*, 2018). Cellular component analysis in NP highlights components like secretory granule membranes and cell-substrate junctions, essential for immune regulation, communication, and placental development, with genes like CD47 and RAC1 involved. In PE, altered components such as platelet alpha granules and integrin complexes suggest disrupted immune response and endothelial function, with enriched actin filaments indicating cytoskeletal dysregulation linked to vascular instability (Opichka *et al.*, 2021). The MF analysis for NP reveals enrichment in signalling, cell adhesion, and RNA regulation, including protein serine/threonine kinase activity (e.g., MAP3K14, PRKAG1) and cadherin binding (e.g., H3C4, DBNL). These functions support placental development, tissue integrity, and gene expression regulation. In PE, no molecular function GO terms are identified at GW 10, indicating a lack of significant molecular function alterations, suggesting dysfunction at the molecular level.

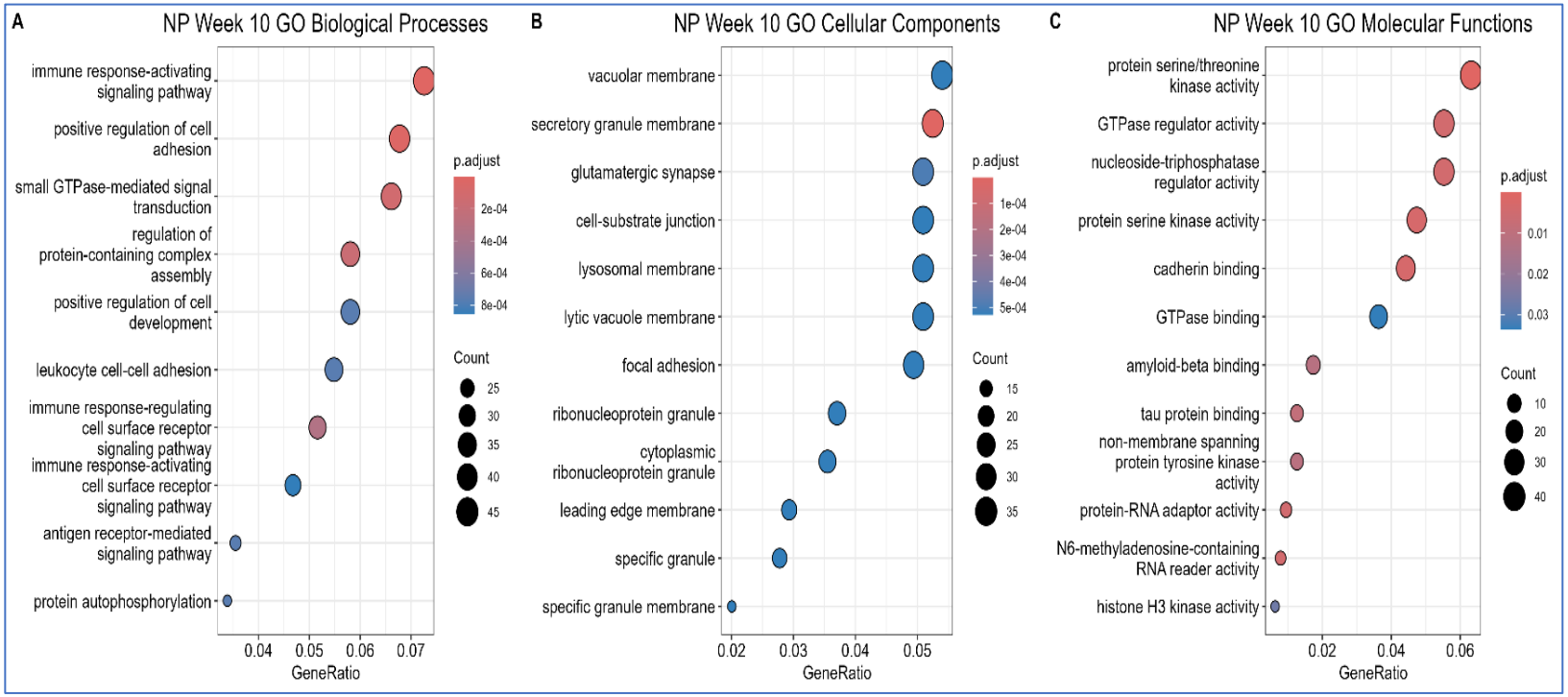


Figure 42: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 10.

Dot plots show the top significantly enriched terms for (A) Biological Process, (B) Cellular Component, and (C) Molecular Function. A strong theme of immune system activation emerges ("immune response-activating signal transduction," "T cell migration"), alongside continued enrichment of cell adhesion components. This indicates a coordinated process where immune tolerance and placental structural integrity are being actively managed in tandem during a healthy pregnancy.

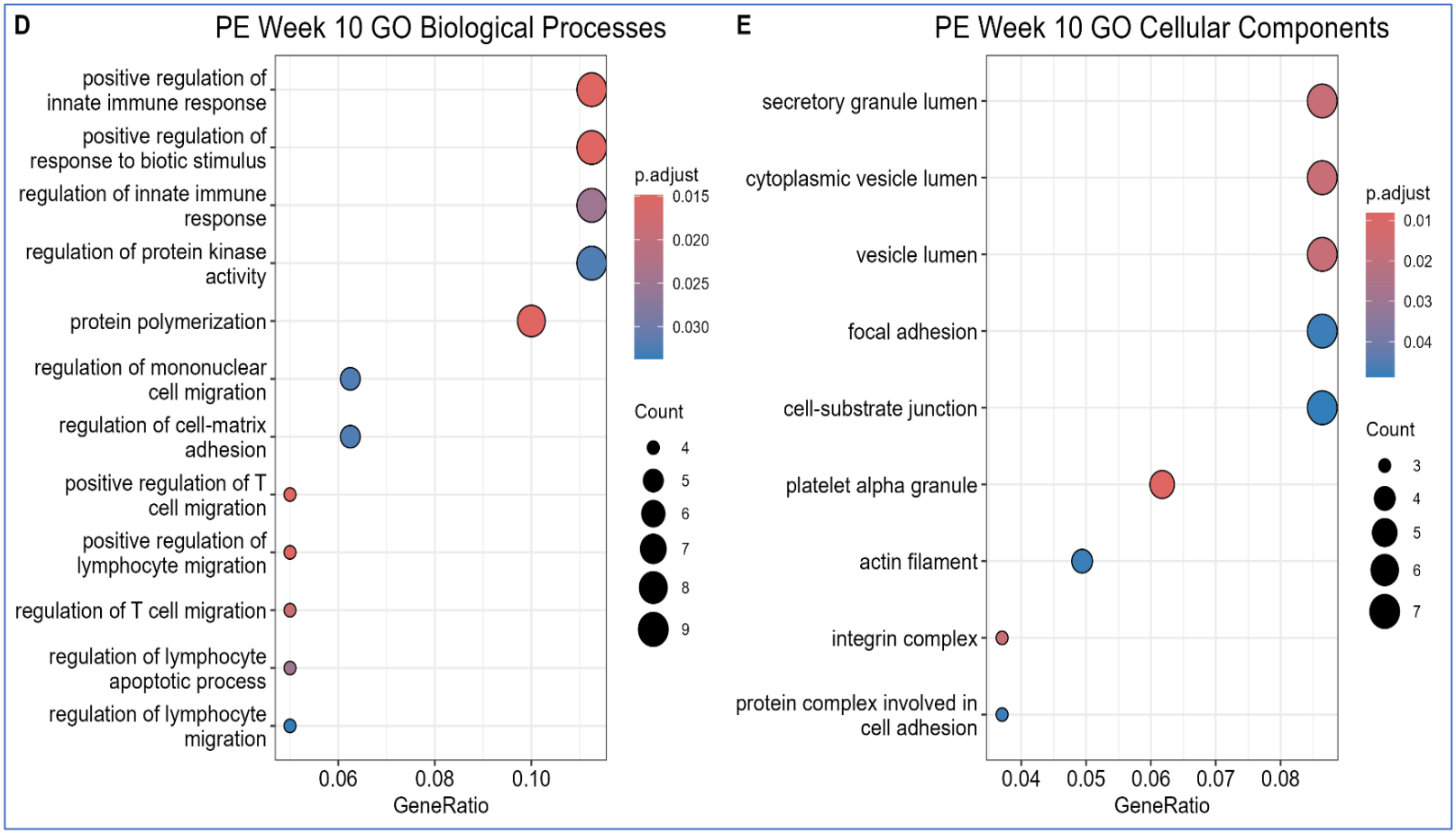


Figure 43: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 10.

Dot plots show the top significantly enriched terms for (D) Biological Process and (E) Cellular Component; no Molecular Function terms were significantly enriched. PE is marked by dysregulated immune responses ("T cell migration," "innate immune response") and altered cellular components like "platelet alpha granule," which are associated with coagulation and inflammation. This shift suggests a move from regulated immune tolerance in NP towards a state of immune over-activation and emerging vascular stress.

6.3.5.1.4 Gestational Week 11

At GW 11, NP involves critical maternal and foetal development processes. Enriched pathways include platelet aggregation and activation (e.g., MYH9, ITGB3), myeloid cell differentiation (e.g., RUNX1, HLA–DRB1), and actin filament organisation and cell–cell adhesion (e.g., RHOA, FLNA). These processes support vascular remodelling, immune adaptation, and cytoskeletal integrity, which are crucial for placental development and pregnancy maintenance. In contrast, PE shows disruptions in membrane-associated processes (e.g., MYZAP, RASA3, ATP6V0A1), Wnt signalosome dysregulation (e.g., CTNNB1, APC), and intercalated disc integrity (e.g., DLG1, ACTN1), indicating impaired cellular signalling, structural organisation, and vesicular trafficking (Lokeswara *et al.*, 2021). Cellular components in NP at GW 11 emphasise structural integrity and cellular adhesion, with enriched pathways in cell–substrate junctions (e.g., PDLIM1, ITGB5) and focal adhesions (e.g., PTPRC, FLNA). These components are vital in anchoring cells, signal transduction, and tissue contraction. In PE, enriched membrane-associated processes like the cytoplasmic side of the membrane (e.g., MYZAP, RASA3) and vacuolar membranes (e.g., ATP6V0A1) reflect disruptions in intracellular trafficking and compartmental organisation. Abnormalities in the Wnt signalosome and intercalated discs suggest impaired signalling and cellular coupling. Molecular function analysis shows that NP is enriched in functions related to actin filament binding, small GTPase binding, and muscle structure (e.g., MYH7, FLNA), supporting cytoskeletal dynamics and intracellular signalling. In PE, while small GTPase and actin binding remain central, fewer genes are involved (e.g., XPO4, RABGEF1), suggesting dysregulation of these pathways and altered placental and vascular function (Xu *et al.*, 2020). These results highlight the differential involvement of key biological processes in normal versus pathological pregnancies, especially regarding cytoskeletal reorganisation and intracellular signalling.

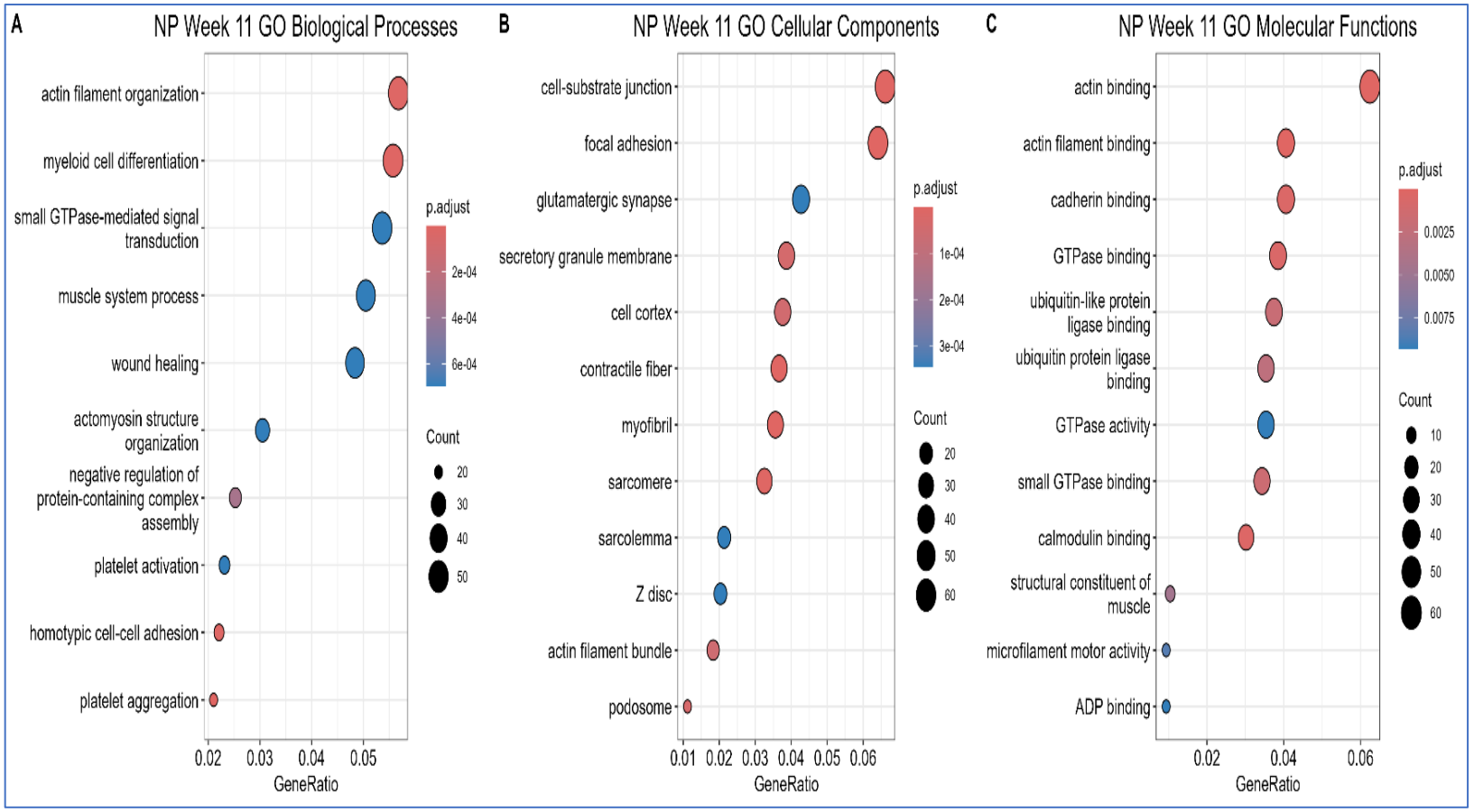


Figure 44: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 11.

Dot plots display the top enriched terms for (A) Biological Process, (B) Cellular Component, and (C) Molecular Function. The analysis reveals a complex landscape supporting active placental development, including "platelet activation" (involved in vascular remodelling), "myeloid cell differentiation" (immune adaptation), and "actin filament organization" (cytoskeletal integrity). This reflects the multifaceted biological processes required for maintaining a healthy pregnancy at this critical peak of transcriptomic activity.

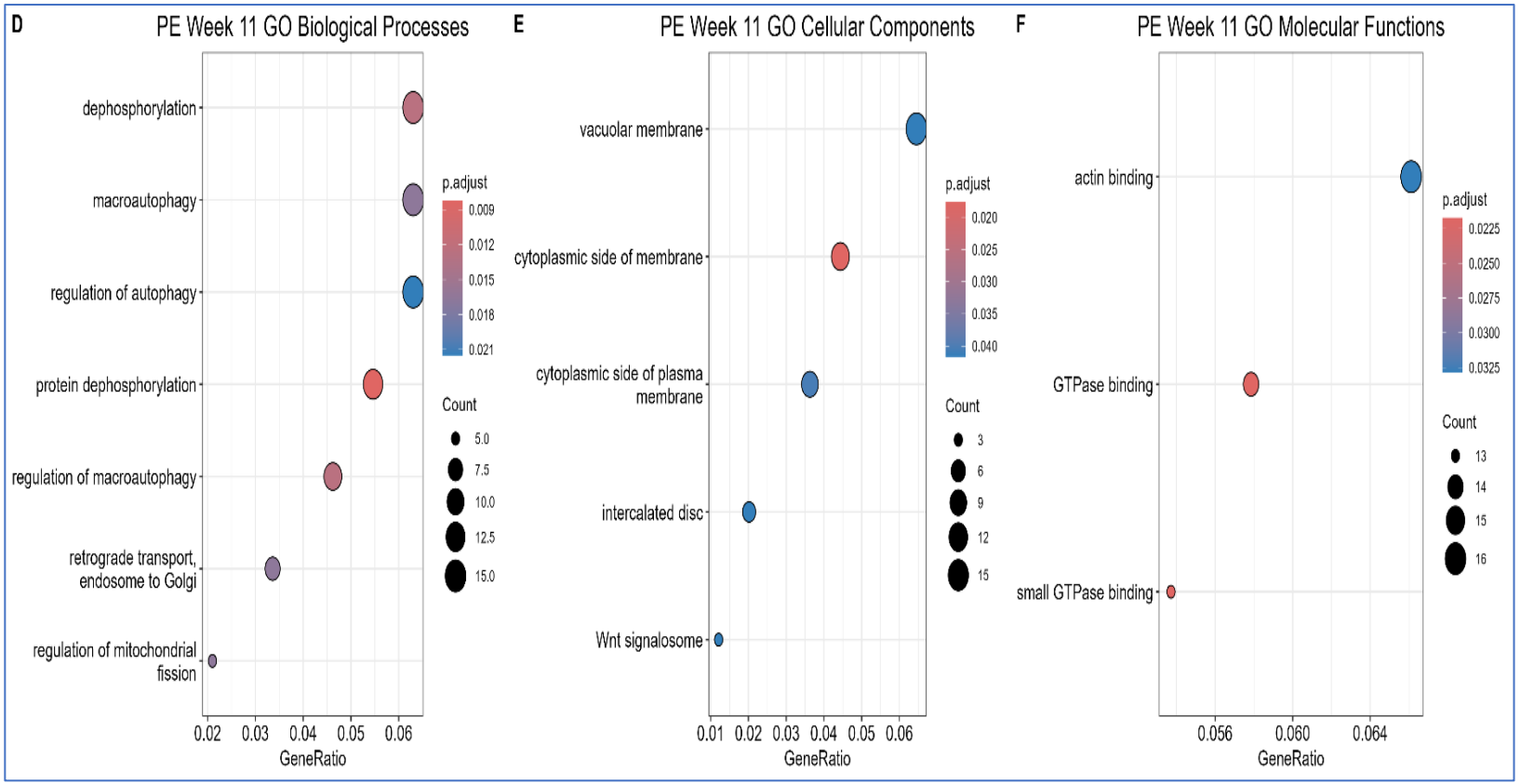


Figure 45: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 11.

Dot plots show the top significantly enriched terms for (D) Biological Process, (E) Cellular Component, and (F) Molecular Function. Pathways indicate significant dysfunction, including disrupted cellular signalling ("Wnt signalosome assembly"), impaired structural integrity ("intercalated disc"), and altered vesicular trafficking ("vacuolar membrane"). These findings point to fundamental breakdowns in cell communication and architecture at a time when robust placental function is paramount.

6.3.5.1.5 Gestational Week 12

At GW 12, NP and PE display notable differences in their biological processes. In NP, enriched processes include platelet aggregation and activation (e.g., ACTG1, MYL9, VCL), myeloid cell homeostasis (e.g., STAT1, RPS19), and immune response–activating pathways (e.g., DDIT3, SYK). These processes indicate active immune and platelet functions essential for placental development and immune tolerance. Small GTPase-mediated signal transduction and endocytosis pathways are also enriched (e.g., RHOA, ACTB), supporting intracellular communication. In contrast, PE shows dysregulation, with fewer genes involved in platelet aggregation and immune pathways (e.g., TYROBP, CD226), suggesting a maladaptive immune response and altered placental function (Li *et al.*, 2016).

Cellular component analysis in NP highlights cell adhesion and vesicular transport structures, such as cell-substrate junctions, secretory granule lumen, and focal adhesions (e.g., PTPRA, ANXA6, ACTG1). These components are crucial for maintaining cellular integrity and proper placental development. In PE, there is a shift towards membrane-associated components like secretory granule membranes (e.g., ITGA2B, ANO6) and the cell cortex (e.g., COTL1, GRB2), indicating disruptions in membrane dynamics and cytoskeletal organisation, which could impair cellular interactions and signalling (Graybill and Davalos, 2020).

Molecular function analysis in NP shows enrichment in cadherin binding (e.g., AHSA1, NDRG1), GTPase regulator activity (e.g., ARFGEF1, IQSEC1), and phospholipid binding (e.g., ANXA6, ZCCHC2), essential for cell adhesion, signalling, and membrane dynamics. In PE, enriched functions include phosphoprotein binding (e.g., SOCS3, GRB2), protein sequestering (e.g., TMSB4X, YWHAZ), and DNA–binding transcription factor binding (e.g., ZMYND8, KDM4C). These indicate dysregulated phosphorylation, transcriptional control, and protein regulation, reflecting altered cellular signalling and function in PE (Gao *et al.*, 2023). These differences underscore the contrasting biological processes in NP and PE, with NP showing robust immune and platelet activation supporting

placental function. At the same time, PE demonstrates immune dysregulation, altered signalling, and disrupted cellular functions, potentially contributing to disease pathophysiology.

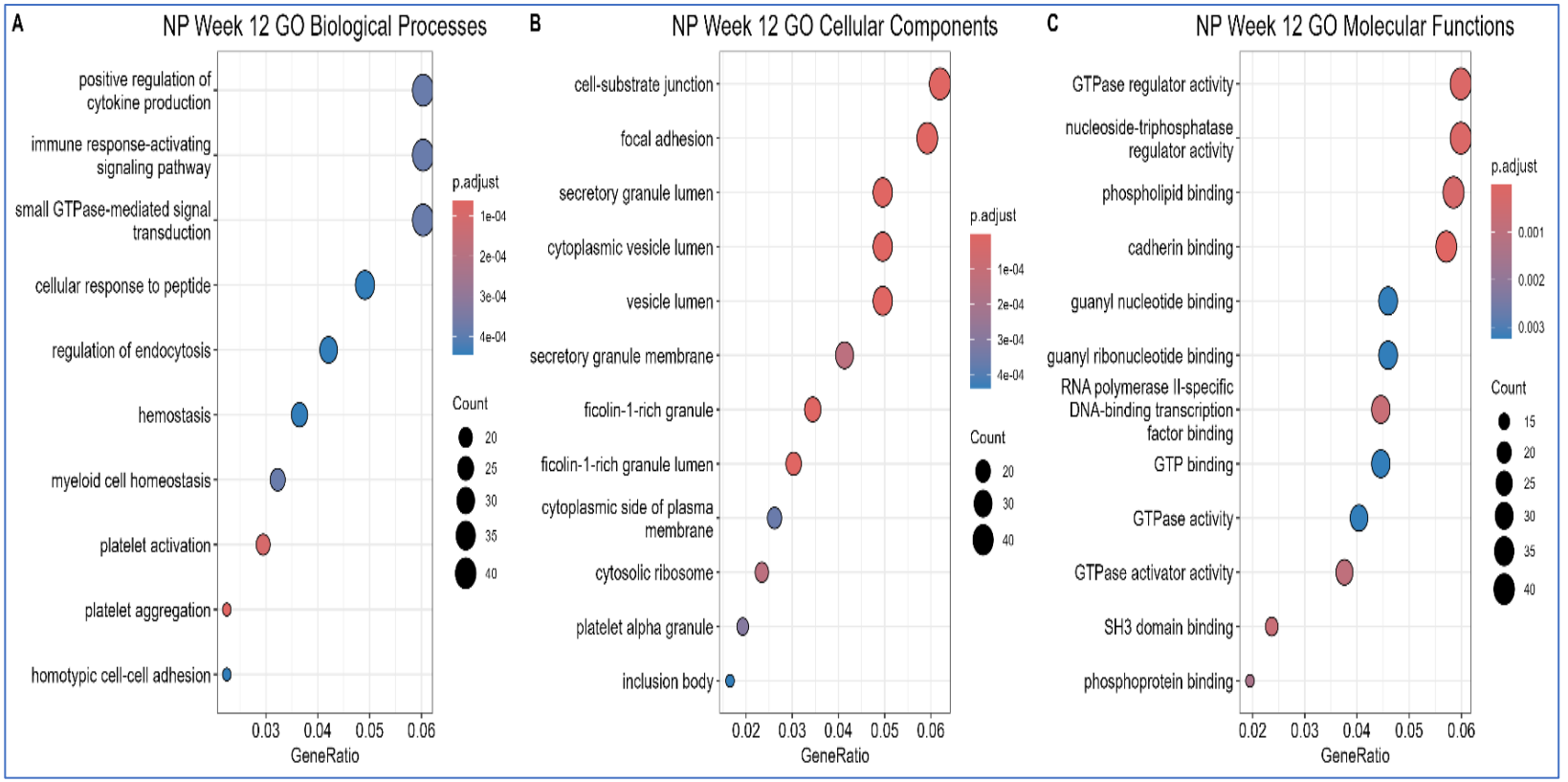


Figure 46: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 12.

Dot plots show the top significantly enriched terms for (A) Biological Process, (B) Cellular Component, and (C) Molecular Function. The enrichment of "platelet aggregation," "immune response-activating pathways," and "small GTPase mediated signal transduction" indicates sustained activity in vascular function, immune surveillance, and intracellular communication. This portrays a state of active, dynamic maintenance of the maternal-foetal interface.

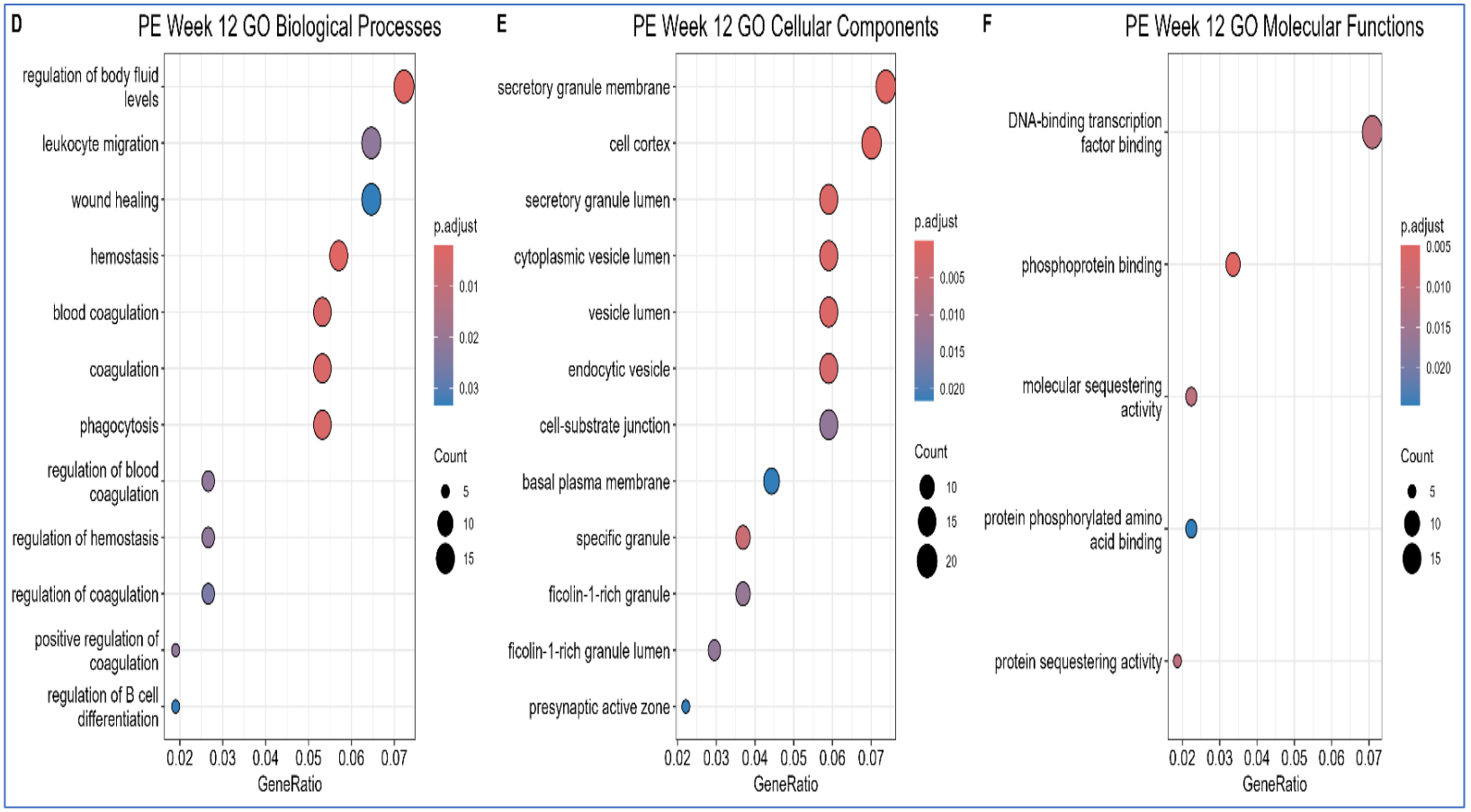


Figure 47: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 12.

Dot plots show the top significantly enriched terms for (D) Biological Process, (E) Cellular Component, and (F) Molecular Function. A shift towards dysregulation is evident, with fewer genes involved in key NP processes like platelet aggregation. Enriched molecular functions such as "phosphoprotein binding" and "protein sequestering" suggest major disruptions in phosphorylation-dependent signalling and protein homeostasis, contributing to the evolving pathology.

6.3.5.1.6 Gestational Week 13

At GW 13, there are significant differences in the biological processes, cellular components, and molecular functions between NP and PE. In NP, the most enriched pathways are related to immune activation, including the "immune response–activating signalling pathway" (GO:0002757), "immune response–activating cell surface receptor signalling pathway" (GO:0002429), and "regulation of innate immune response" (GO:0045088). These pathways highlight the activation of immune responses essential for maintaining maternal–foetal tolerance, with genes like PUM1, TICAM1, FYN, and TLR8 playing key roles. In contrast, PE samples exhibit enrichment in platelet–related pathways, including "platelet activation" (GO:0030168), "blood coagulation" (GO:0007596), and "haemostasis" (GO:0007599), with genes like PDGFA, PIK3CB, and ITGB3. These processes suggest a strong involvement of coagulation, platelet activation, and vascular dysfunction in PE. Additionally, "wound healing" (GO:0042060) and "cell–substrate adhesion" (GO:0031589) pathways in PE reflect impaired cellular adhesion and repair mechanisms, further supporting the pathology of PE (Hall *et al.*, 2024).

Cellular Components: In NP, key enriched components include "focal adhesion" (GO:0005925), "cell–substrate junction" (GO:0030055), and elements involved in cellular movement and vesicular trafficking like "melanosomes" (GO:0042470) and "pigment granules" (GO:0048770). These reflect the importance of cell–matrix interactions, vesicle–mediated transport, and cytoskeletal reorganisation in placental development. In contrast, PE shows enrichment in components like "platelet alpha granule" (GO:0031091), highlighting platelet activation, and "secretory granule lumen" (GO:0034774) and "vesicle lumen" (GO:0031983), which are associated with vesicular trafficking. Additionally, PE shows enrichment in "actomyosin" (GO:0042641) and "actin filament bundle" (GO:0032432), suggesting altered cytoskeletal dynamics and cell motility. These differences

indicate a shift towards platelet and vesicular functions in PE, which may contribute to the vascular and placental abnormalities observed in the disease.

Molecular Functions: In NP, enriched molecular functions such as cadherin binding (e.g., CDH23, CALD1) and protein phosphorylation (e.g., phosphoprotein binding, protein serine kinase activity) are crucial for cell adhesion, tissue remodelling, and placental development. These functions contribute to the stable maternal–foetal interface and vascular health. In PE, however, molecular functions related to protein phosphorylation, GTPase binding, and tyrosine kinase activity are skewed, with altered kinase activities (e.g., FYN, LYN, JAK1) indicative of dysregulated signalling (Jena *et al.*, 2020). This suggests disruptions in endothelial function and vascular instability, contributing to hypertension and placental ischemia. The upregulation of GTPase regulator and signalling adaptor activities in PE could exacerbate vascular smooth muscle dysfunction, increasing systemic inflammation and impairing trophoblast invasion, which are key features of PE (Chiang, Seow and Chen, 2024). The differences in these biological processes, cellular components, and molecular functions suggest a critical imbalance in immune regulation, vascular signalling, and cell adhesion in PE compared to NP. These alterations could contribute to the pathophysiology of PE, including hypertension, endothelial dysfunction, and placental insufficiency.

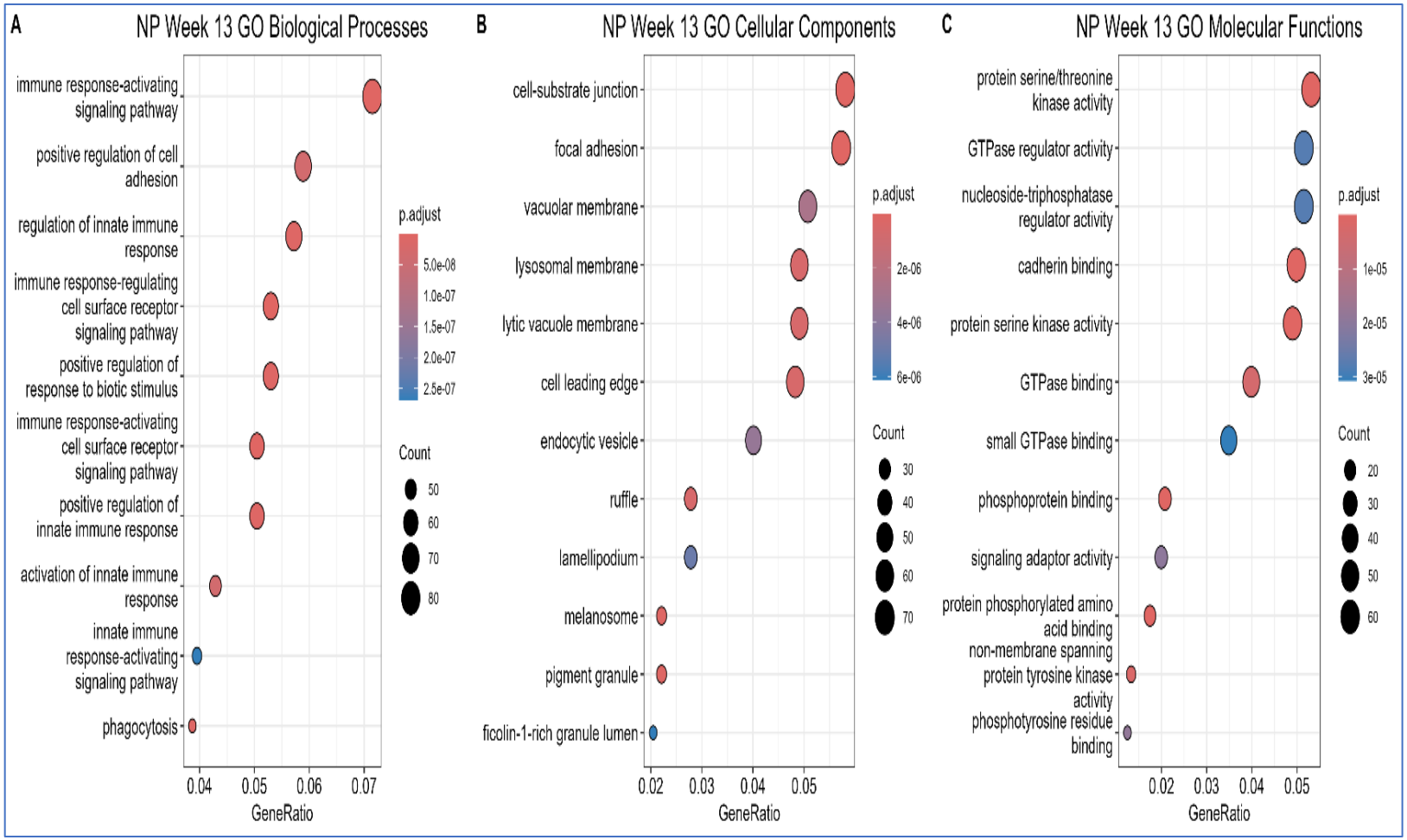


Figure 48: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 13.

Dot plots display the top enriched terms for (A) Biological Process, (B) Cellular Component, and (C) Molecular Function. This period, which has the highest number of UVS, is characterized by robust immune signalling pathways ("immune response-activating cell surface receptor signalling") and strong cell adhesion ("focal adhesion," "cadherin binding"). This highlights the peak of coordinated immune and structural interactions at the maternal-foetal interface during a normotensive pregnancy.

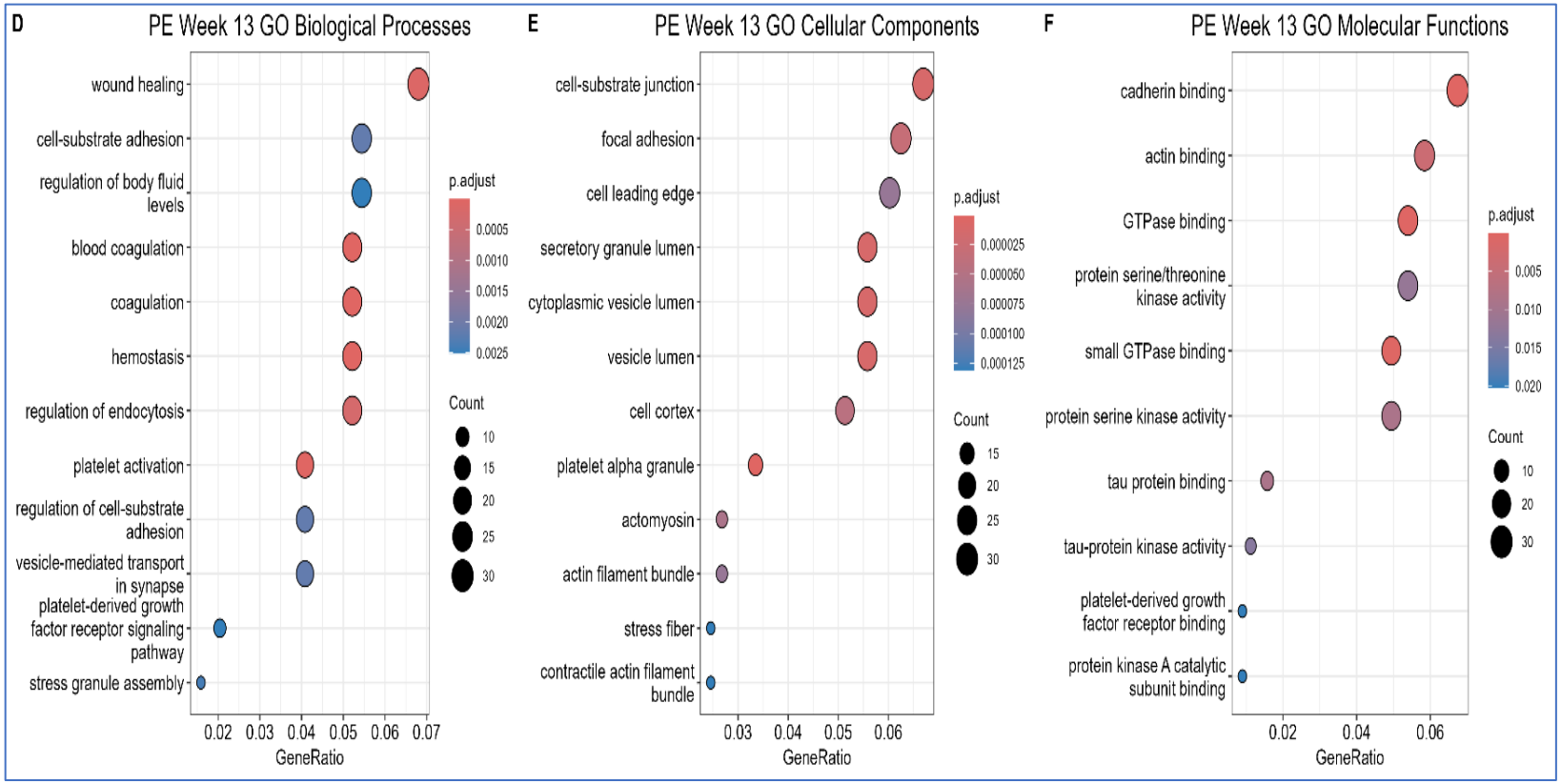


Figure 49: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 13.

Dot plots show the top significantly enriched terms for (D) Biological Process, (E) Cellular Component, and (F) Molecular Function. At this critical week, PE pathophysiology is dominated by terms related to "platelet activation," "blood coagulation," and "wound healing," indicating a pronounced shift towards a pro-thrombotic state and vascular injury response. This contrasts sharply with the immune-focused profile of NP and underscores the central role of vascular dysfunction in PE.

6.3.5.1.7 Gestational Week 14

At week 14 of pregnancy, significant differences between NP and PE are observed in biological processes, cellular components, and molecular functions, highlighting the pathophysiological shifts in PE. In NP, biological processes such as mononuclear cell migration and leukocyte migration are enriched, emphasising immune system regulation and tissue remodelling crucial for placental development, with genes like CALR, LYN, THBS1, and SPN involved in immune cell trafficking. Additionally, the negative regulation of endocytosis, with genes like MCTP1 and DYSF, suggests cellular stability and prevention of excessive internalisation. In contrast, PE samples show a shift towards apoptotic signalling and chromatin regulation, with significant upregulation of the intrinsic apoptotic signalling pathway (GO:0097190) in response to DNA damage involving genes such as HIPK2, BID and MCL1, indicating heightened stress and potential placental dysfunction. The negative regulation of intrinsic apoptotic signalling and enrichment in chromatin organisation suggest altered gene expression regulation that could contribute to endothelial dysfunction and impaired trophoblast invasion in PE (Tanaka *et al.*, 2014). Cellular components in NP, such as the secretory granule membrane and haemoglobin complex, reflect active cellular communication and erythropoiesis essential for foetal development. In PE, however, the focal adhesion and cell–substrate junction are also enriched but may be dysregulated, contributing to impaired placental function. Additionally, components like the specific granule membrane and recycling endosome highlight altered granule–mediated secretion and cellular trafficking, which could disrupt placental development. Molecular functions in NP show stable processes, while significant enrichments in apoptosis regulation, protein degradation, and cytoskeletal dynamics are noted in PE (Qi *et al.*, 2022). Specifically, death domain binding (GO:0070513) and ubiquitin protein ligase binding (GO:0031625) are enriched, indicating cellular degradation and cell cycle regulation disruptions, contributing to cellular stress. Actin filament binding (GO:0051015) involving

CLMN, IQGAP2, and MYH9 suggests alterations in cytoskeletal dynamics, which could impair cell migration and placental development. These differences in molecular mechanisms reflect the compromised placental function and endothelial dysfunction characteristic of PE, highlighting potential early biomarkers for disease progression.

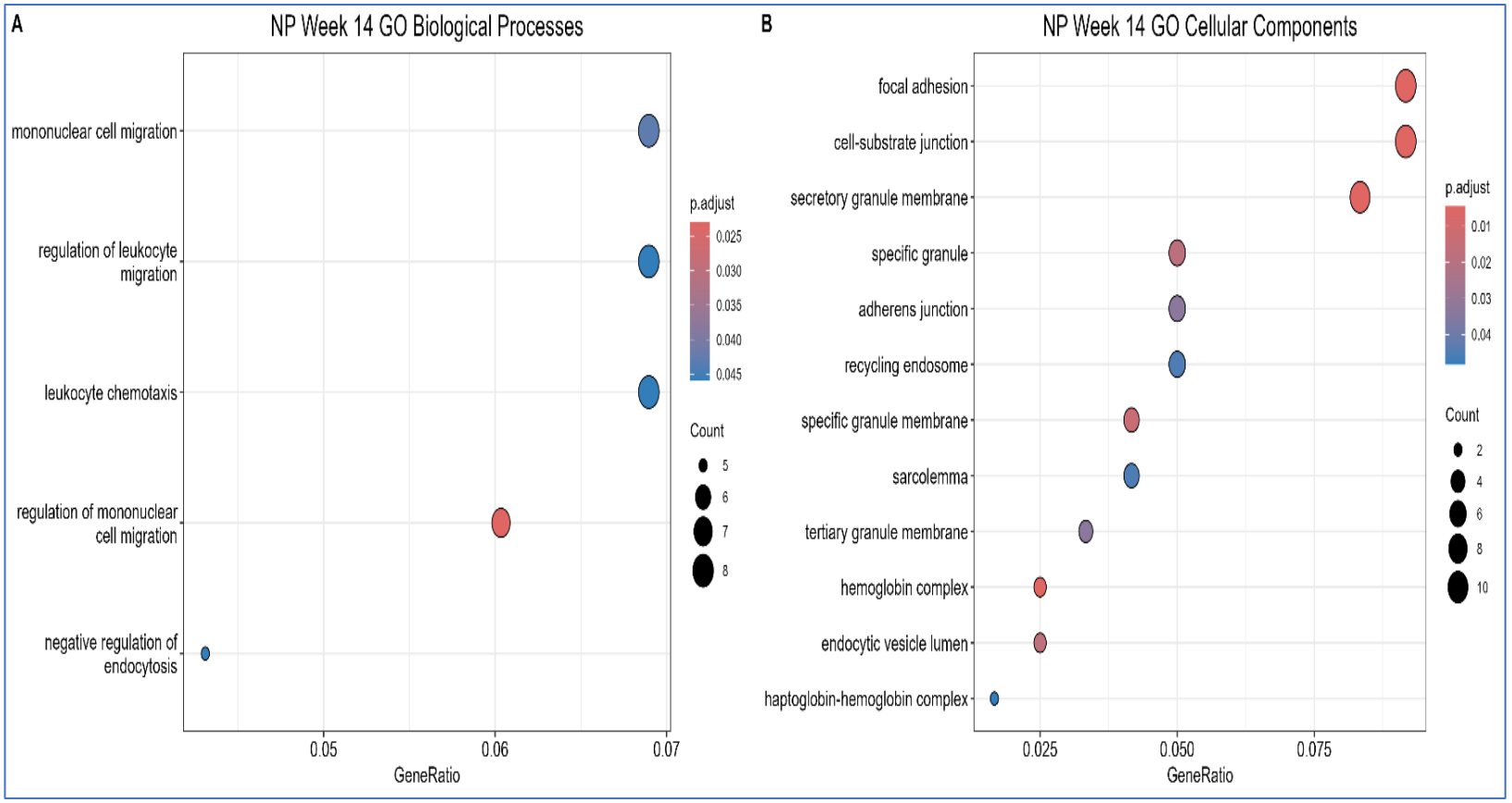


Figure 50: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 14.

Dot plots show the top significantly enriched terms for (A) Biological Process and (B) Cellular Component; no Molecular Function terms were significantly enriched. The enriched processes, such as "mononuclear cell migration," indicate a continued but potentially more refined role for immune cell trafficking in maintaining placental health as the pregnancy progresses beyond the first-trimester peak.

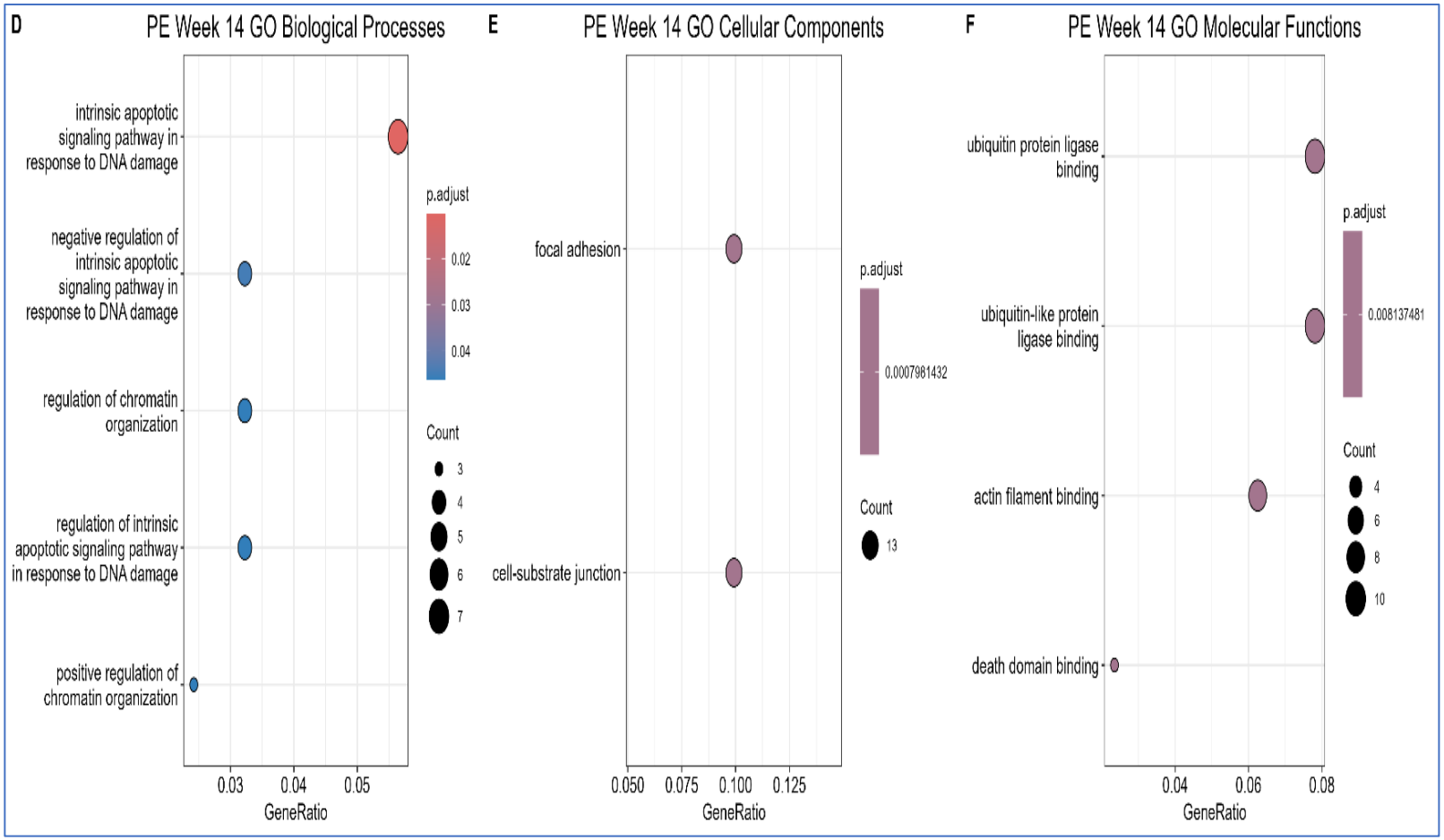


Figure 51: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 14.

Dot plots show the top significantly enriched terms for (D) Biological Process, (E) Cellular Component, and (F) Molecular Function. The pathology deepens with significant enrichment of stress responses, including the "intrinsic apoptotic signalling pathway" and functions like "death domain binding." This indicates an escalation from vascular dysfunction to widespread cellular stress and activation of cell death pathways within the placental environment.

6.3.5.1.8 Gestational Week 15

At week 15 of NP, various biological processes are significantly enriched, highlighting critical cellular regulation and placental development mechanisms. These include small GTPase–mediated signal transduction (GO:0007264), with genes such as FGD5, CDC42SE2, and LRRK2, indicating the regulation of cellular signalling pathways via small GTPases. Additionally, processes like regulating GTPase activity (GO:0043087) and regulating small GTPase–mediated signal transduction (GO:0051056) emphasise the importance of GTPase activity in cellular signalling. Cell–matrix adhesion (GO:0007160), enriched with genes like ILK, HRG, and CASK, highlights interactions between cells and the extracellular matrix, critical for tissue integrity and placental development. The vesicle–mediated transport in the synapse (GO:0099003) pathway, involving genes such as PRKCB and NUMB, underscores the importance of vesicle trafficking in cellular communication. Other enriched processes include the synaptic vesicle cycle (GO:0099504) and regulation of mRNA splicing (GO:0048024), which are essential for gene expression regulation and cellular morphology. In contrast, PE at week 15 shows no significant enrichment in biological processes, suggesting a lack of substantial biological changes compared to NP.

In terms of cellular components, NP at week 15 shows enrichment in focal adhesion (GO:0005925) and cell–substrate junction (GO:0030055), with genes like JAK1, ILK, and CASK involved in cell–matrix interactions, essential for adhesion and signalling. Other enriched components include the cytoplasmic side of the plasma membrane (GO:0009898), secretory granule lumen (GO:0034774), and vesicle lumen (GO:0031983), all of which are involved in cellular communication and vesicle transport. In PE, there is also enrichment in focal adhesion and cell–substrate junction, but with a different molecular profile, reflecting altered adhesion processes (Zhao and Peng, 2025). Furthermore, PE shows significant enrichment in ribosomal components such as the ribosome (GO:0005840), cytosolic ribosome (GO:0022626), and ribosomal subunit

(GO:0044391), suggesting a focus on protein synthesis, with genes like RPL37A and RPS25 involved.

At the molecular function level, NP at week 15 shows significant protein serine kinase activity enrichments, with kinases like PRKCB, ILK, and MAP3K5 playing roles in cellular signalling. Cadherin binding, enriched with genes like NUMB and CTNNB1, suggests key roles in cell adhesion and tissue organisation. Additionally, DNA–binding transcription factor binding involving RUNX2, FOXP1, and SMAD2 indicates active regulation of gene expression. In contrast, PE shows no molecular function enrichments, suggesting potential disruptions in these molecular processes, which may reflect altered signalling or metabolic pathways in PE compared to NP. These findings underscore significant biological and molecular differences between NP and PE at week 15, with NP showing coordinated processes essential for placental and cellular development. At the same time, PE may reflect early signs of dysfunction.

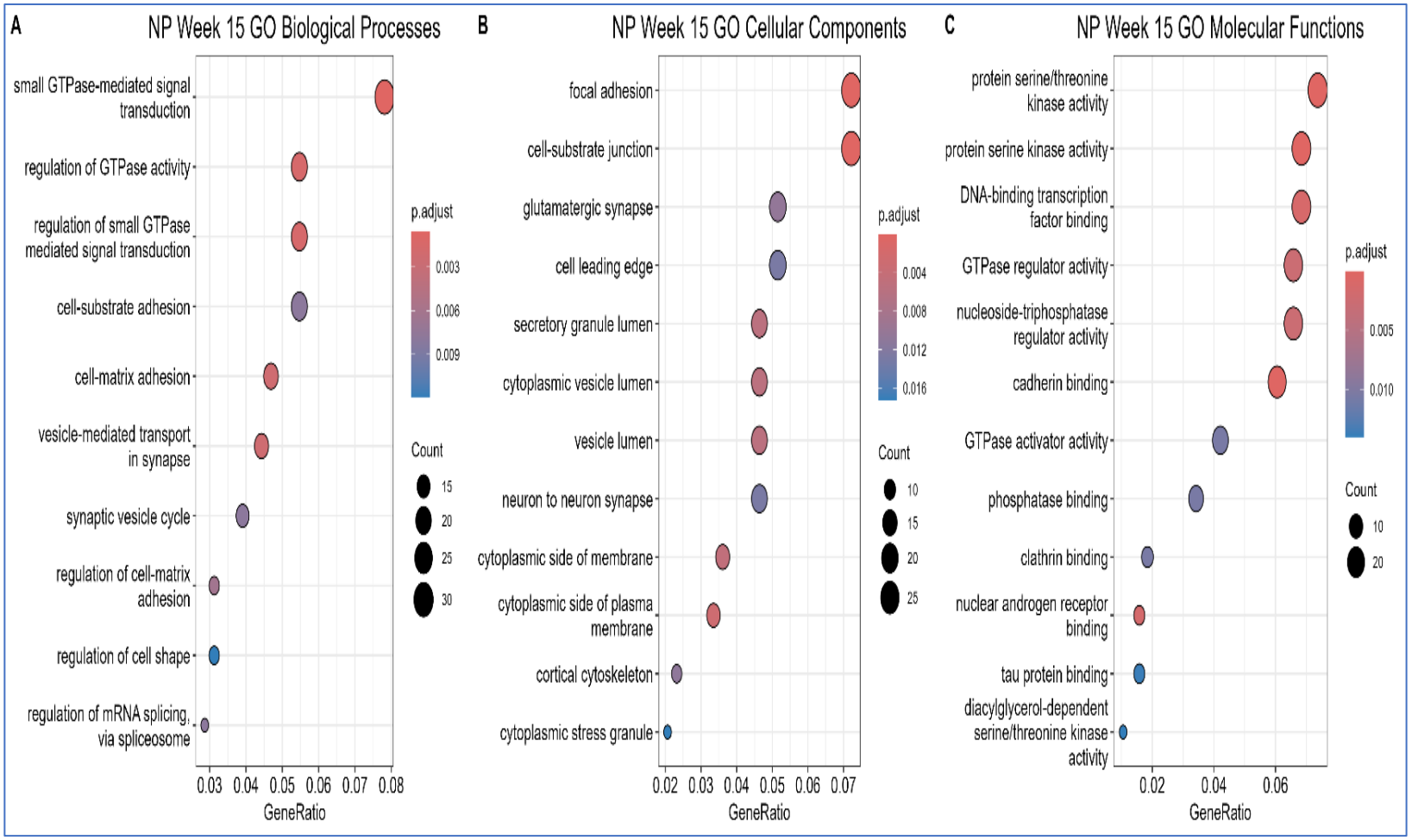


Figure 52: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 15.

Dot plots display the top enriched terms for (A) Biological Process, (B) Cellular Component, and (C) Molecular Function. Enrichment in "small GTPase mediated signal transduction," "cell-matrix adhesion," and "vesicle-mediated transport" demonstrates a continued emphasis on precise cellular signalling, structural integrity, and communication, which are essential for ongoing foetal and placental growth.

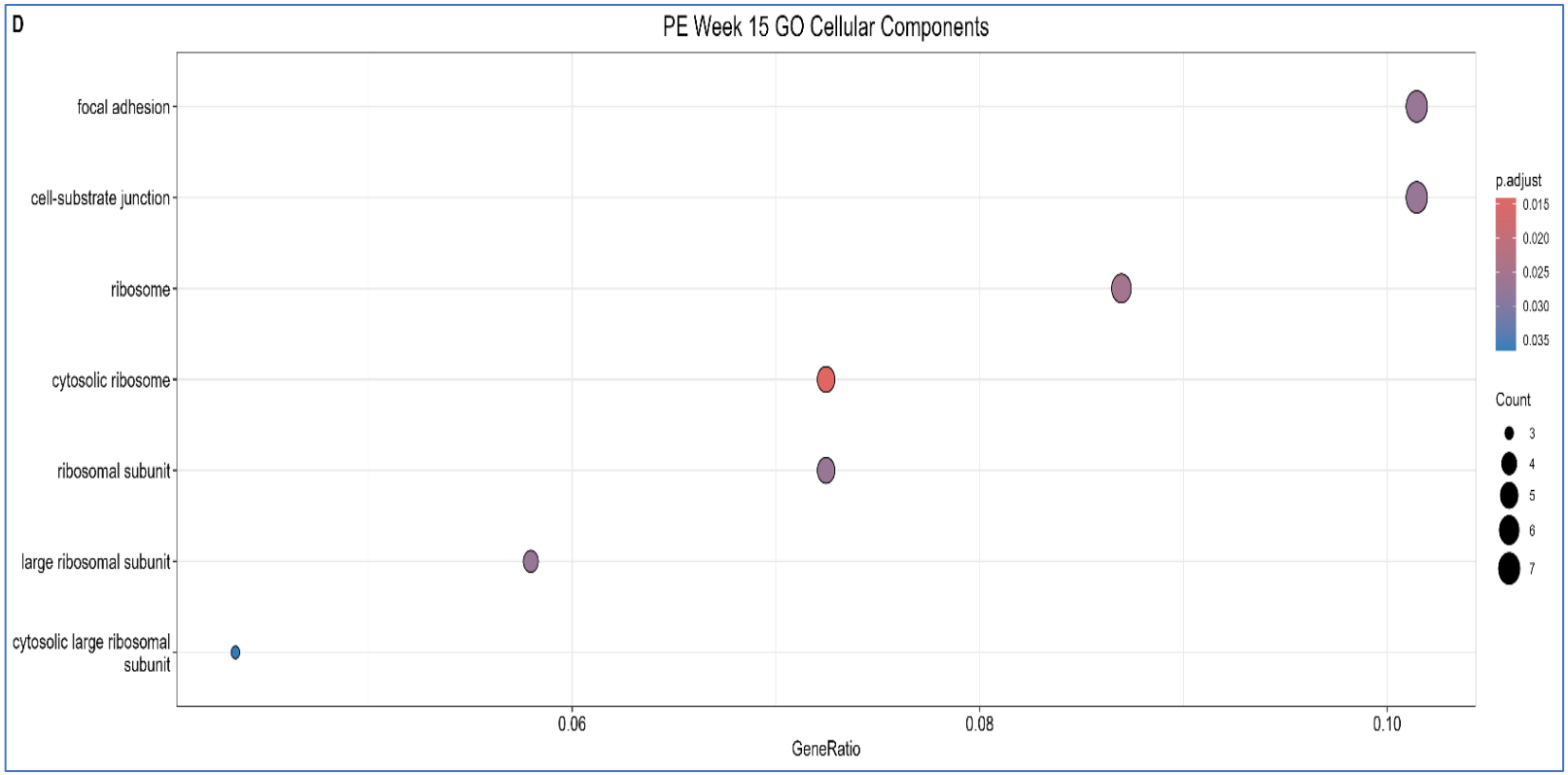


Figure 53: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 15.

The dot plot shows the only significantly enriched category: (D) Cellular Components. The singular enrichment of "ribosomal" components suggests a compensatory or dysregulated focus on protein synthesis, possibly in response to cellular damage. The absence of significant Biological Process or Molecular Function terms indicates a broad breakdown in coordinated cellular activities, leaving a rudimentary signature of cellular stress.

6.3.5.1.9 Gestational Week 16

At week 16 of NP, several biological processes are significantly enriched, reflecting crucial mechanisms for immune response, cellular dynamics, and tissue integrity. These include phagocytosis, with genes like PTPRJ, CAMK1D, and RAC1 involved in immune responses, and regulation of endocytosis, enriched in genes such as PTPRJ and CD63, playing essential roles in cellular uptake. Wound healing processes are prominent, with genes like VCL, PTEN, and HIF1A involved in tissue repair. The positive regulation of cell adhesion involving genes like IL4R, RAC1, and CD36 suggests vital processes for cell signalling and tissue integrity. Additionally, membrane invagination and regulation of body fluid levels are enriched, highlighting key mechanisms for cellular structure and homeostasis. In contrast, PE at week 16 shows enrichment in processes such as protein localisation to the plasma membrane, with genes like ROCK2 and ITGB1 indicating membrane trafficking involvement (C. Li *et al.*, 2024). Blood coagulation and haemostasis, with genes like PDGFB and MYL9, suggest altered coagulation in PE linked to vascular complications. Other enriched processes in PE include the regulation of small GTPase–mediated signal transduction and regulating apoptotic signalling in response to DNA damage, indicating potential disruptions in cell survival and stress responses. These findings highlight the differences in biological processes between NP and PE, particularly in immune response, cellular dynamics, and vascular health at week 16.

Regarding cellular components, NP at week 16 shows enrichment in the secretory granule membrane (GO:0030667), reflecting active secretion and immune function, with genes like RAB27B and HLA–B involved. The cell–substrate junction (GO:0030055) and focal adhesion (GO:0005925) components are also enriched, emphasising the importance of cell–matrix interactions with genes like SVIL, VCL, and IGF2R. The endocytic vesicle (GO:0030139) term is enriched, highlighting vesicle trafficking, while the ficolin–1–rich granule (GO:0101002) suggests immune–related granule functions. In contrast, PE samples show enrichment in

the GBAF complex (GO:0140288), associated with chromatin remodelling, and the cell cortex (GO:0005938), indicating cell shape and movement changes. Additionally, platelet alpha granules (GO:0031091) and phagocytic vesicles (GO:0045335) are enriched in PE, pointing to heightened immune responses and platelet activity. Both NP and PE share enrichment in cell–substrate junctions and focal adhesion, but the specific genes involved differ, indicating distinct cellular dynamics in each condition.

Molecular function analysis for NP at week 16 reveals significant enrichments in several key categories. Cadherin binding (GO:0045296), with genes like VCL, ALDOA, PTPRJ, and CDH1, highlights strong cell–cell adhesion. Protein serine/threonine kinase activity (GO:0004674), with genes like CAMK1D, BRAF, and PRKCA, reflects active phosphorylation events regulating signalling pathways. Enrichments in actin binding (GO:0003779), with genes like PLS3 and MYH9, suggest critical roles in cytoskeletal dynamics, while GTPase regulator activity (GO:0030695) involving genes such as MCF2L and TBC1D1 indicates the regulation of small GTPase signalling. Additionally, phosphoprotein binding (GO:0051219) is enriched, underscoring the importance of protein phosphorylation in cellular processes. In contrast, PE samples at week 16 show enrichment in tau protein binding (GO:0048156), with genes like ROCK2 and SNCA, pointing to altered neurophysiological or cytoskeletal regulation, potentially linked to stress responses. The molecular functions in PE show distinct differences from NP, highlighting a focus on neurogenetic and structural protein alterations, indicating unique molecular mechanisms in PE (Rybak-Krzyszkowska *et al.*, 2023). These findings suggest that while NP and PE share some common pathways, such as actin binding and GTPase activity, there are notable differences in the molecular mechanisms' characteristics of these conditions at week 16.

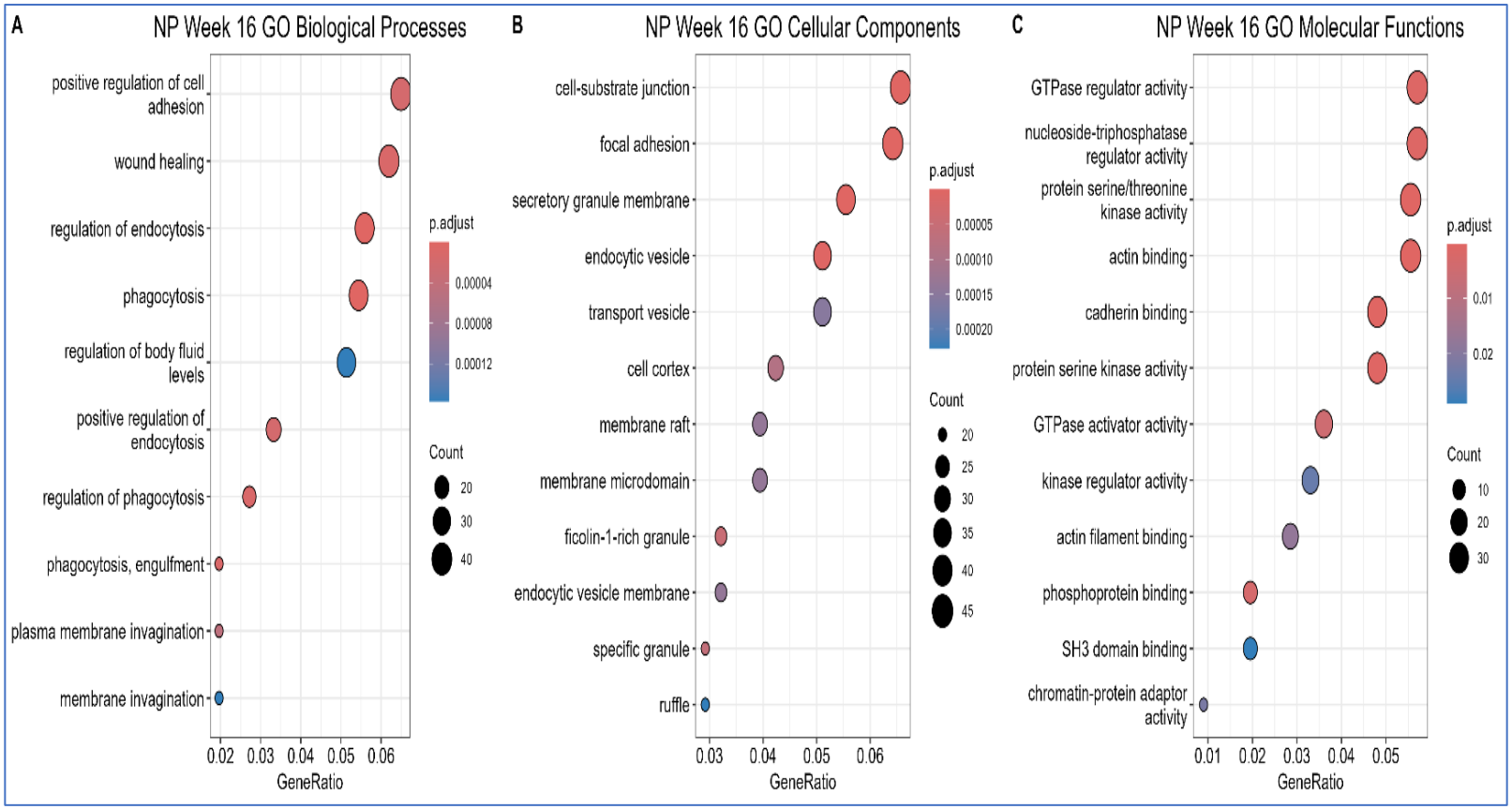


Figure 54: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in NP at Gestational Week 16.

Dot plots show the top significantly enriched terms for (A) Biological Process, (B) Cellular Component, and (C) Molecular Function. Processes such as "phagocytosis," "wound healing," and "positive regulation of cell adhesion" are prominent, indicating active tissue remodelling and immune clearance. This reflects the continued dynamic maintenance and preparation of the maternal body and placenta for the later stages of pregnancy.

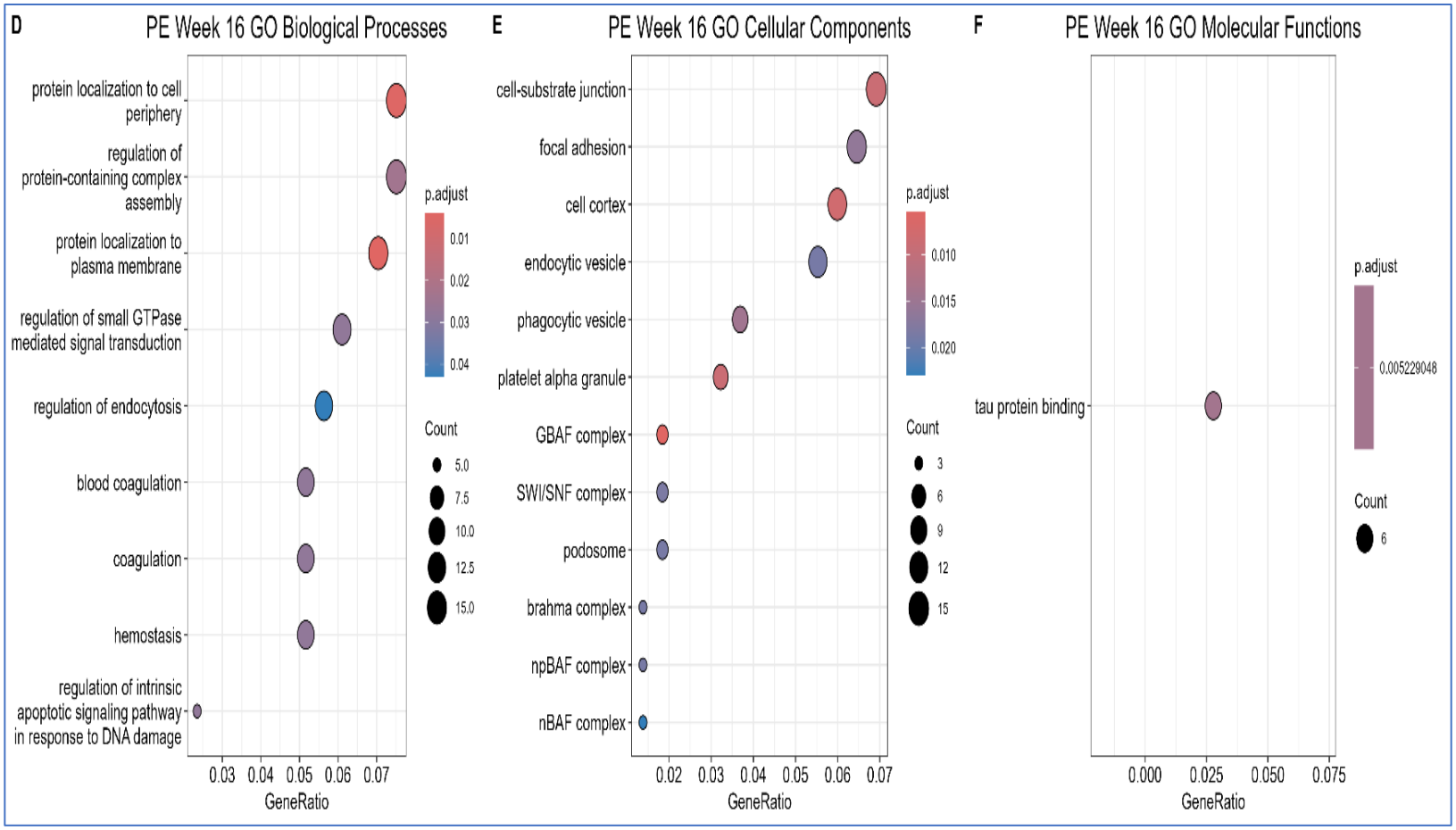


Figure 55: Gene Ontology (GO) enrichment analysis for Unique Variant Signatures in PE at Gestational Week 16.

Dot plots show the top significantly enriched terms for (D) Biological Process, (E) Cellular Component, and (F) Molecular Function. The analysis concludes with a strong signature of coagulopathy ("blood coagulation") and persistent vascular dysfunction ("protein localization to plasma membrane"). The unique enrichment of "tau protein binding" may indicate novel, off-target neurological stress responses. Together, these findings cement the picture of PE as a state of severe vascular and systemic dysregulation as the pregnancy approaches the second trimester.

The GO analysis reveals distinct differences in biological processes, cellular components, and molecular functions between NP and PE from gestational weeks 8 to 16. In NP, enriched processes include immune tolerance, cytoskeletal organisation, and angiogenesis, supporting placental development and maternal–foetal interactions. Key cellular components, such as focal adhesions and secretory granules, facilitate cell adhesion, migration, and vesicle trafficking. Molecular functions like actin binding and GTPase activity underpin cytoskeletal dynamics and signalling. In contrast, PE shows immune over–activation, disrupted angiogenesis, and apoptotic signalling, leading to endothelial dysfunction and placental insufficiency. Cellular components shift toward membrane structures and platelet granules, reflecting vascular instability. Reduced molecular function diversity in PE highlights altered phosphorylation, transcriptional regulation, and cytoskeletal organisation. These findings suggest immune dysregulation, impaired cell adhesion, and vascular dysfunction as central to PE pathology, with potential biomarkers including RAC1, CD47, HIPK2, and ROCK2 for future investigation.

6.3.5.2 Gene Ontology Analysis for PE-associated Significant Genetic Variants per Gestational Weeks

The GO analysis of significant PE-associated variants reveal key biological processes and molecular functions critical to the progression of PE at different gestational weeks. In Week 8, disruptions in actin cytoskeleton organisation, particularly involving actin filament bundles (GO:0032432), point to vascular dysfunction and impaired trophoblast invasion, key features of PE. Significant enrichments in molecular functions, such as G protein activity (GO:0003925) and profilin binding (GO:0005522), emphasise altered cellular signalling pathways. In contrast, cadherin binding (GO:0098641) suggests disrupted cell adhesion, further supporting the idea that cytoskeletal dynamics, cell adhesion, and small GTPase signalling are central to PE pathology.

In Week 9, key biological processes such as platelet aggregation (GO:0070527) and regulation of platelet activation (GO:0010543) highlight disturbances in platelet

function and lipid metabolism as central to PE. The enrichment of GTPase activity (GO:0003924) suggests altered signalling pathways affecting cellular trafficking and signalling. The absence of significant cellular component GO terms indicates that this week's molecular processes are more functional than structural.

In Week 10, platelet activation (GO:0030168) and regulation of body fluid levels (GO:0050878) are significantly enriched, emphasising the importance of platelet function and vascular regulation in PE. Processes related to wound healing (GO:0042060) and immune system activation, such as antigen processing (GO:0002486), point to endothelial dysfunction and immune dysregulation as contributing factors. Enriching terms related to the MHC protein complex (GO:0042611) and the lumenal side of the ER membrane (GO:0098553) further indicate that immune modulation and antigen presentation play key roles in PE at this stage.

In Week 11, podosome assembly (GO:0071800) and related terms like cell cortex (GO:0005938) and phagocytic vesicles (GO:0045335) suggest that changes in cell migration, adhesion and actin cytoskeleton dynamics are involved in PE progression. The lack of significant molecular function GO terms indicates that structural processes, particularly in cellular dynamics, are more critical at this stage.

In Week 12, immune response–activating signalling pathways (GO:0002757) and protein regulation processes like the ERAD pathway (GO:0036503) are enriched, suggesting that immune activation and protein processing disruptions contribute to the inflammatory and stress responses in PE. The absence of significant cellular components or molecular function GO terms highlights the regulatory and signalling nature of the processes involved.

In Week 14, the regulation of mitotic cytokinesis (GO:1902412) and RNA processing pathways such as RNA localisation (GO:0006403) and RNA export (GO:0046831, GO:0006405) are enriched, highlighting disruptions in cell division and RNA regulation, which may impact placental development and function in PE.

The enrichment of apoptosis (GO:0097194) further suggests that cell death processes may contribute to placental dysfunction. The lack of significant findings in the cellular component or molecular function categories indicates that the focus in Week 14 is more on regulatory and signalling processes.

Interestingly, no significant GO terms were identified for Weeks 15 and 16, indicating a lack of enriched biological processes or functions during these weeks, possibly suggesting a shift like PE-associated changes or limitations in the current GO term classification to capture the underlying mechanisms at these stages. Overall, the GO analyses underscore the critical roles of platelet function, immune response, actin dynamics, RNA processing, and apoptosis in the progression of PE. Disruptions in these processes across different weeks suggest potential biomarkers, such as genes involved in cell adhesion, platelet activation, and immune regulation, which may serve as targets for future therapeutic strategies. The absence of significant terms in some weeks highlights the complexity of PE and the need for further exploration to understand its multifactorial nature fully.

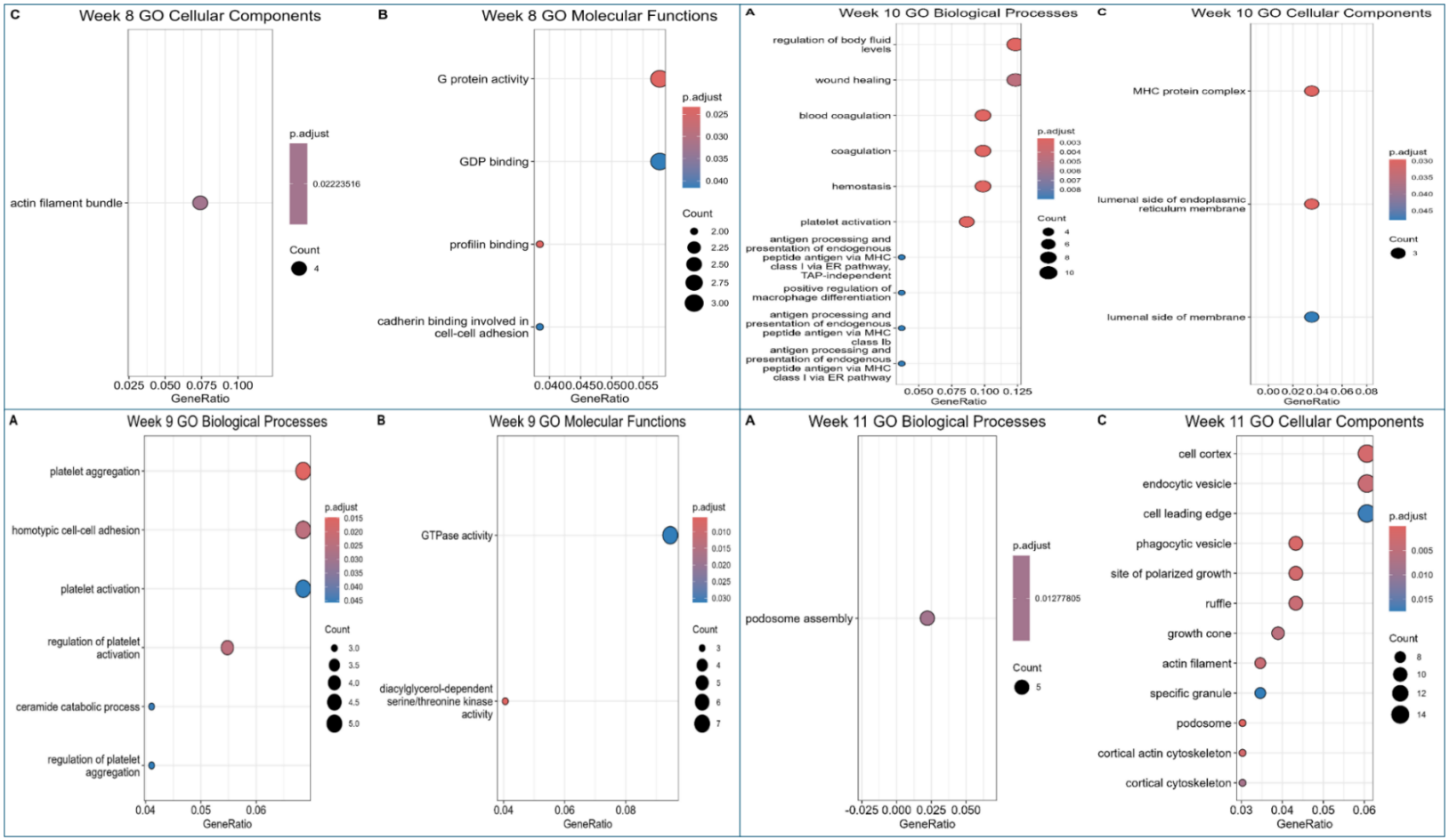


Figure 56:Gene Ontology (GO) enrichment analysis for PE-associated significant genetic variants across Gestational Weeks 8 to 11.

The chart displays enriched GO terms for (A) Biological Process, (B) Molecular Function, and (C) Cellular Component. The analysis reveals that early-stage PE (Weeks 8-11) is characterized by disruptions in cytoskeletal organization (e.g., "actin filament bundle" at GW8), platelet function ("platelet aggregation" at GW9), and immune signalling ("G protein activity"), highlighting the early emergence of vascular and cellular dysfunction.

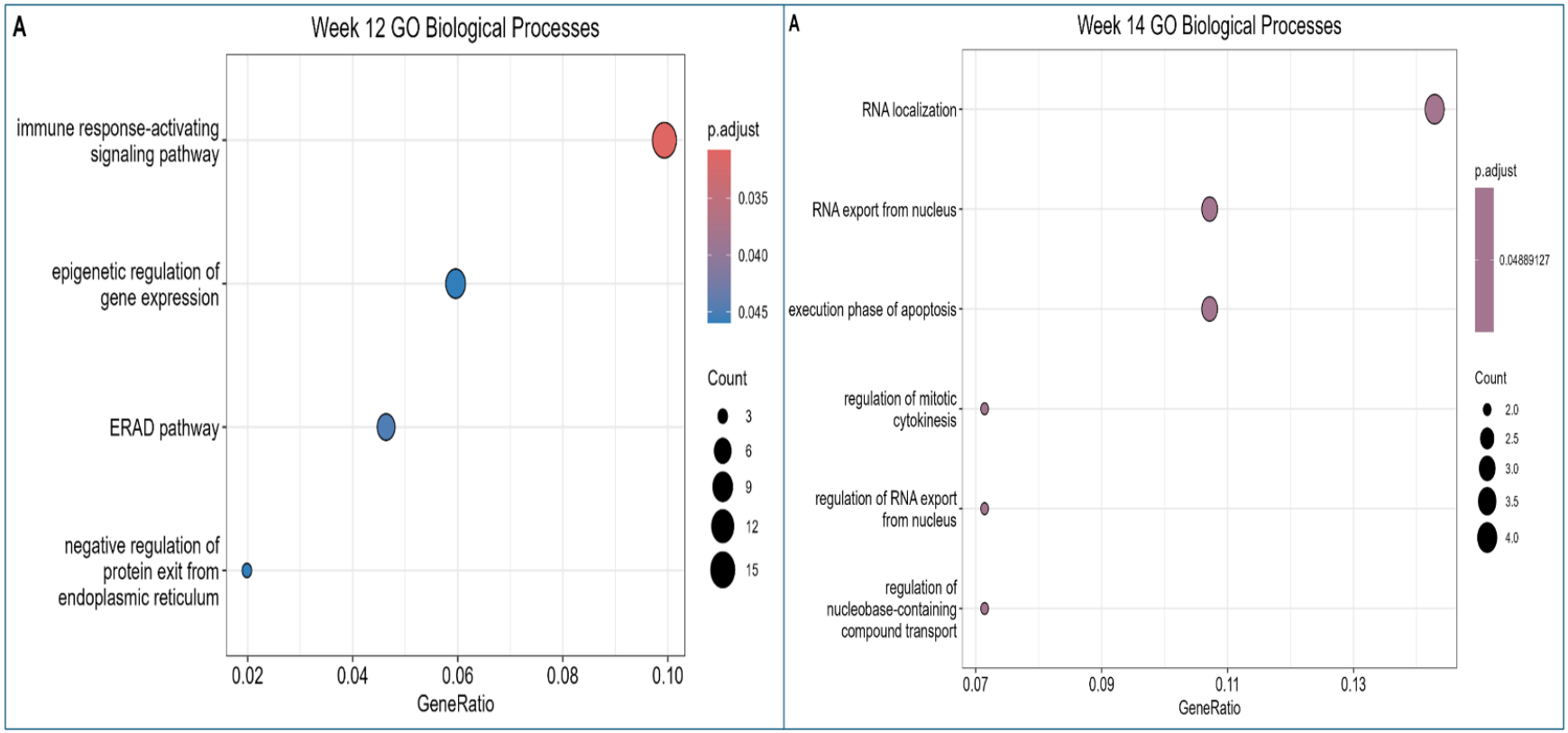


Figure 57: Gene Ontology (GO) enrichment analysis for PE-associated significant genetic variants across Gestational Weeks 12 to 14.

The chart displays significantly enriched Biological Process terms. During this mid-stage period, the genetic basis of PE involves immune dysregulation ("immune response-activating signalling pathway" at GW12), disruption of key cellular functions ("ERAD pathway" at GW12, "regulation of mitotic cytokinesis" at GW14), and dysregulated RNA processing ("RNA localisation," "RNA export" at GW14). The emergence of "apoptosis" at GW14 indicates an escalation to overt cellular stress and death. The absence of significant terms for GW15 and 16 (not shown in plot) suggests a potential shift in the dominant pathological mechanisms or a limitation in detecting them with this analysis at later weeks.

6.3.5.3 Pathway Analysis for UVS per Gestational Week

The comparison of KEGG pathways between NP and PE at GW 8 reveals distinct biological processes. NP pathways are primarily related to immune responses and infections, with key pathways like T cell receptor signalling, Viral carcinogenesis, and Shigellosis indicating a strong immune and viral component. Other pathways, such as Th1 and Th2 cell differentiation, highlight immune modulation in NP. In contrast, PE pathways focus more on cellular stress, mitochondrial dysfunction, and vascular regulation. Notable pathways in PE include Mitophagy, Folate biosynthesis, and Regulation of actin cytoskeleton, which suggest cellular damage, oxidative stress, and endothelial dysfunction. Parkinson's disease and cGMP–PKG signalling reflect neurodegenerative and vascular changes linked to PE's pathophysiology. While both conditions share immune–related mechanisms, NP emphasises immune activation and viral responses. In contrast, PE is more associated with cellular stress and vascular issues, contributing to their unique clinical features.

The KEGG pathways of NP and PE at week 9 reveal key biological processes linked to each condition. For NP, pathways such as Shigellosis, Focal adhesion, and Bacterial invasion of epithelial cells indicate active immune responses, cell adhesion, and bacterial infections, suggesting a strong immune activation in NP. Other notable pathways, such as Rap1 signalling and pathways in cancer, further highlight signalling events and potential oncogenic influences. Additionally, pathways such as Regulation of actin cytoskeleton and Fluid shear stress and atherosclerosis emphasise cytoskeletal dynamics and vascular responses, which are central to immune and inflammatory processes. In PE, pathways like Cellular senescence, Oocyte meiosis, and Ras signalling suggest cellular ageing, reproductive cell functions, and stress signalling. Other significant pathways such as Alcoholism, Apelin signalling, and Vasopressin–regulated water reabsorption

reflect vascular regulation, hormonal responses, and cellular stress mechanisms in PE. These pathways underline the differences in immune, vascular, and cellular stress responses between NP and PE.

The KEGG pathways for NP at week 10 include key signalling and immune response pathways. The chemokine signalling pathway and Rap1 signalling pathway highlight the involvement of immune cell recruitment and cellular communication. Other prominent pathways like Neurotrophin signalling, T cell receptor signalling, and Platelet activation reflect immune system activation and cell signalling. Pathways such as Thyroid hormone signalling and Apoptosis emphasise regulation of metabolism and cell death processes. Salmonella infection and Tight junction pathways suggest immune responses to pathogens and the importance of cellular integrity. The top pathways for PE at week 10 include Neutrophil extracellular trap formation, Viral myocarditis, and Platelet activation, indicating immune and vascular stress responses. Pathways like Fluid shear stress atherosclerosis and Arrhythmogenic right ventricular cardiomyopathy point to cardiovascular stress and disease mechanisms. Other pathways, including Folate biosynthesis and Phagosome, emphasise cellular metabolism and immune response. These pathways suggest significant immune and cardiovascular involvement in PE at week 10.

The KEGG pathways for NP at week 11 highlight critical processes related to platelet activation, cardiovascular signalling, and immune responses. Platelet activation is a dominant pathway, indicating the involvement of platelets in cellular responses and thrombosis. The cGMP–PKG signalling pathway emphasises cardiovascular signalling and smooth muscle function. At the same time, focal adhesion and regulation of the actin cytoskeleton are crucial for cell adhesion, migration, and response to environmental changes. The C–type lectin receptor signalling pathway suggests immune recognition and responses to pathogen–associated signals. In contrast, the Fc gamma R–mediated phagocytosis pathway highlights immune cells' role in pathogen clearance. Other notable pathways include Salmonella infection, renal cell carcinoma, and endocytosis, which reflect

immune responses to bacterial infections and cellular uptake processes. Hypertrophic cardiomyopathy further points to cardiovascular stress mechanisms in muscle cells. The top pathways for PE at week 11 reveal a mix of neurodegenerative, immune, and cardiovascular processes. Pathways of neurodegeneration suggest alterations in neurological function, possibly linked to PE complications. Platelet activation reappears as a key pathway, reflecting immune activation and vascular dysfunction. Regulation of the actin cytoskeleton and focal adhesion are essential for cell movement and tissue integrity. Human papillomavirus infection and synaptic vesicle cycle pathways indicate interactions with viral infections and neurotransmission. In contrast, Parkinson's disease and Vibrio cholerae infection pathways further underscore neurodegenerative and immune responses in PE. The Rap1 signalling pathway highlights cellular signalling mechanisms related to vascular and immune responses. These pathways illustrate a complex interplay of immune function, cardiovascular stress, and cellular regulation in NP and PE during week 11.

The KEGG pathways for NP at week 12 reveal pathways associated with immune responses, platelet activation, cell adhesion, and cytoskeleton regulation. The dominant pathway is Platelet activation, which plays a crucial role in thrombosis and immune responses. Other significant pathways include Adherens junction, emphasising cell–cell interactions and tissue integrity, and Shigellosis, which suggests an immune response to bacterial infections. The Regulation of the actin cytoskeleton pathway is involved in cell movement and tissue remodelling. In contrast, leukocyte trans endothelial migration and chemokine signalling pathways highlight immune cell trafficking and inflammatory responses. Fc gamma R–mediated phagocytosis points to the role of immune cells in pathogen clearance, and Focal adhesion is involved in cell signalling and adhesion. Pathways like Endocytosis and Tight junction also emphasise cellular uptake processes and tissue barrier function, reflecting immune activation and cell communication in NP at week 12. In contrast, the top 10 KEGG pathways for PE at week 12 primarily reflect immune, cancer, and cardiovascular processes. Endocytosis is a key

pathway, indicating the role of cellular uptake in immune and tissue processes. Viral carcinogenesis highlights the potential influence of viral infection in cancer development, while Cell adhesion molecules and Mitophagy are involved in cell signalling and mitochondrial regulation, respectively. Choline metabolism in cancer suggests altered metabolic pathways linked to tumorigenesis. Malaria and Focal adhesion pathways point to immune responses and cell adhesion mechanisms in vascular and immune cells. The Neutrophil extracellular trap formation pathway reflects the role of neutrophils in immune defence. Platelet activation reappears, indicating its significance in immune and vascular dysfunction, and Regulation of the actin cytoskeleton suggests alterations in cellular movement and structure. Together, these pathways reflect complex immune, vascular, and metabolic changes associated with PE at week 12.

The KEGG pathways of the unique variant signature (UVS) for NP and PE at week 13 show notable differences in biological processes and signalling pathways. In NP, key pathways include "Yersinia infection," "T cell receptor signalling," and "Natural killer cell–mediated cytotoxicity," with significant involvement of immune system–related genes like ITGB1, PIK3CD, and NFATC3. These pathways suggest active immune responses and cell signalling related to infection and immune cell activation. Additionally, pathways like "Regulation of actin cytoskeleton" and "Kaposi sarcoma–associated herpesvirus infection" strongly influence cytoskeletal dynamics and viral interactions in the NP group. On the other hand, the top pathways in PE, such as "Focal adhesion," "Shigellosis," and "Leukocyte trans endothelial migration," highlight the involvement of vascular remodelling, cell adhesion, and leukocyte migration. These pathways are crucial for understanding the endothelial dysfunction and immune system dysregulation observed in PE. Notably, PE significantly overlaps with pathways like "Platelet activation" and "Tight junction," which may suggest altered haemostatic and barrier functions in pregnancy. Additionally, the presence of "Choline metabolism in cancer" and "Axon guidance" pathways in PE points to metabolic and neural signalling alterations. Overall, while both NP and PE pathways involve immune

and cellular processes, the PE signature focuses more on vascular and endothelial dysfunction, contrasting with the immune response–driven mechanisms seen in NP.

The KEGG pathways for the unique variant signature (UVS) of NP and PE at week 15 highlight different biological processes involved in each condition. In NP, pathways related to platelet activation and cell adhesion are prominent, including "Platelet activation" and "Focal adhesion," with key genes such as LYN, ITGB3, PRKCB, and MYL12A. These pathways suggest active processes in cell signalling, adhesion, and cytoskeletal dynamics, which play a role in immune responses and cellular interactions. The "Rap1 signalling pathway," "Neutrophil extracellular trap formation," and "Vibrio cholerae infection" pathways also indicate an involvement in immune responses and microbial defence. Additionally, pathways related to neuronal function, such as "Long–term potentiation" and "Axon guidance," indicate a broader biological context that may link immune and neural responses. In contrast, the PE UVS at week 15 centres more around cellular mechanisms and viral infections. "Ribosome" and "Gap junction" pathways, with genes like RPL4, RPS25, and TUBB1, reflect cellular translation and communication processes, focusing on protein synthesis and intercellular signalling. Other significant pathways include "Coronavirus disease" and "Hepatitis C," which suggest viral infections as contributing factors to the pathophysiology of PE. "Fc epsilon RI signalling" and "Renal cell carcinoma" pathways also appear, emphasising immune and cancer–related pathways, while "Sulfur metabolism" is linked to metabolic alterations. Overall, the PE UVS at week 15 indicates a complex interplay of cellular dysfunction, viral infection, and immune activation, contrasting with the more immune signalling and neuronal pathways seen in NP.

In week 16, the UVS of NP highlights pathways related to immune activation, including Fc gamma R–mediated phagocytosis and Neutrophil extracellular trap formation, which involve key immune–related genes like PRKCG, LYN, and RAC1, suggesting robust innate immune responses. The VEGF signalling pathway, with genes such as PRKCG and MAPK1, points to angiogenesis and tissue remodelling

involvement. At the same time, cellular senescence (involving PTEN and TGFBR2) suggests stress–induced cell cycle arrest, possibly contributing to inflammation. The MAPK signalling pathway, with genes like MAPK1 and PRKCA, is crucial for cellular responses to stress and inflammation. Other notable pathways in NP include C–type lectin receptor signalling, B–cell receptor signalling, long–term depression, and human cytomegalovirus infection, indicating complex immune and stress responses. For PE in week 16, the UVS is heavily focused on pathways related to platelet activation and vascular dysfunction, key features of PE. The Platelet activation pathway, involving genes like PPP1CB, MYLK, and ITGB1, highlights platelet aggregation and coagulation. In contrast, the cGMP–PKG signalling pathway, involving PPP1CB, ROCK2, and MYLK, suggests alterations in smooth muscle function and vascular tone, critical in PE pathogenesis. The focal adhesion pathway, with genes like VASP and ITGB1, indicates endothelial dysfunction and vascular remodelling. In contrast, the oxytocin signalling pathway is linked to smooth muscle contraction, which is relevant to uterine function during pregnancy. Vascular smooth muscle contraction and Regulation of actin cytoskeleton further emphasise endothelial dysfunction and vascular tone regulation. Additionally, pathways like Axon guidance, Leukocyte transendothelial migration, Tight junction, and Rap1 signalling pathway suggest immune responses and disruption of endothelial integrity, which are central to the pathology of PE.

In conclusion, the KEGG pathway analysis of PE from GW 8 to 16 reveals significant insights into the molecular mechanisms underlying the disease. Key biological processes such as immune dysregulation, vascular dysfunction, and cellular stress are consistently highlighted throughout this period. Specific pathways such as Platelet activation, cGMP–PKG signalling, Focal adhesion, Regulation of actin cytoskeleton, and Vascular smooth muscle contraction emerge repeatedly across the weeks, reflecting the central role of endothelial dysfunction and platelet aggregation in PE pathogenesis. The genes identified as necessary in PE include PPP1CB, MYLK, ITGB1, ROCK2, and VASP, which regulate processes involved in vascular tone, platelet activation, and endothelial integrity.

These genes play crucial roles in the development of PE by contributing to disturbances in vascular function and immune responses. In contrast, NP pathways primarily focus on immune activation and infection–related processes, with genes like LYN, MAPK1, and RAC1 involved in immune signalling and cell communication. This analysis emphasises that GW 8 to 16 are critical for understanding the progression of PE, as they encompass the activation of vascular and immune–related pathways central to the disease. By focusing on the identified key genes during these weeks, such as PPP1CB, MYLK, ITGB1, and ROCK2, these could serve as potential biomarkers for early diagnosis and prediction of PE. Monitoring these pathways and genes during this period could improve clinical strategies for managing pregnancies at risk for PE, ultimately enhancing maternal and foetal health outcomes.

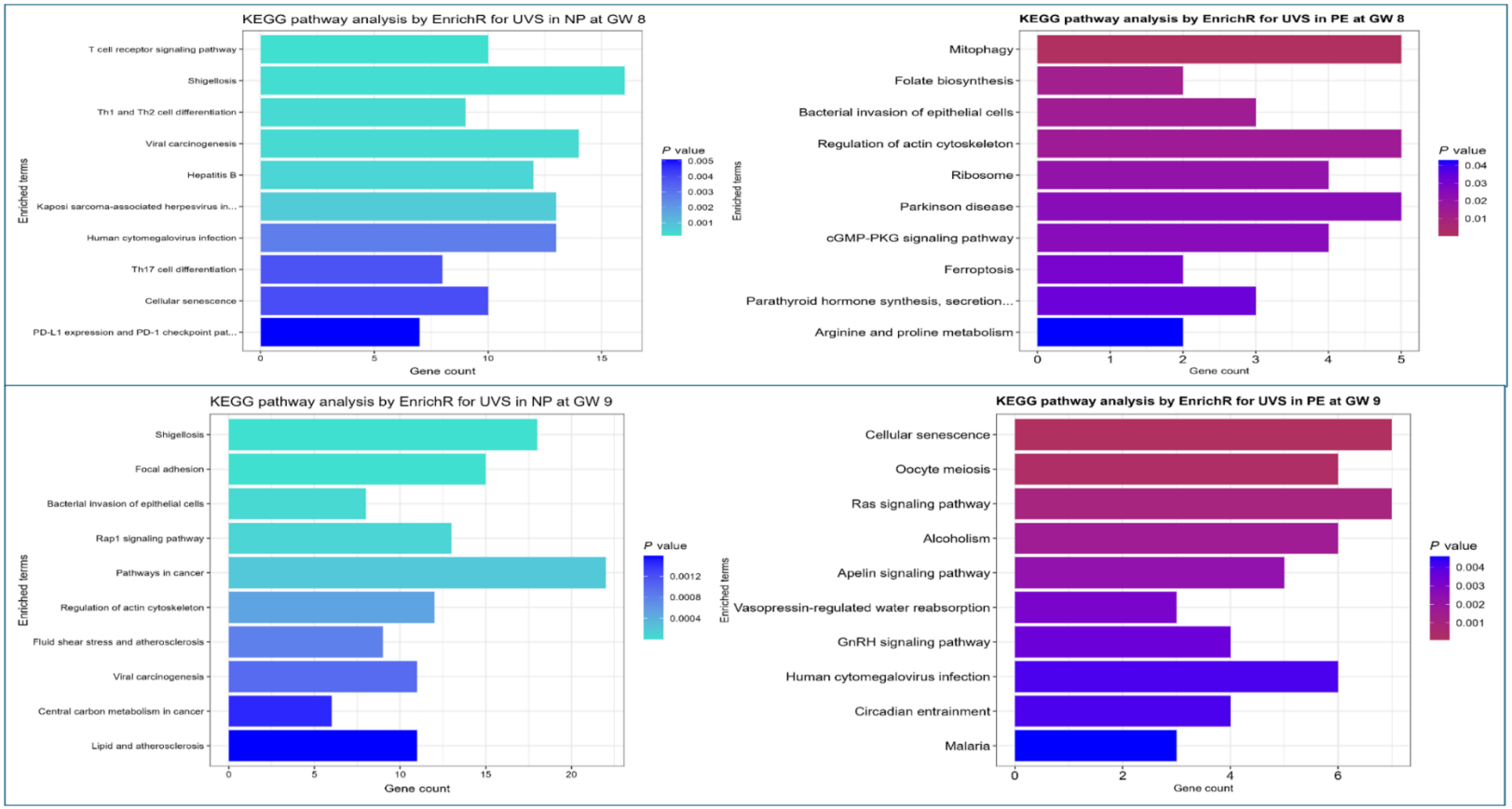


Figure 58: KEGG pathway enrichment analysis for Unique Variant Signatures (UVS) in NP and PE at Gestational Weeks 8 and 9.

The bar charts show the top enriched pathways. In NP, pathways are predominantly related to immune responses and infections (e.g., "T cell receptor signalling," "Shigellosis"). In contrast, PE pathways are associated with cellular stress, mitochondrial dysfunction, and vascular regulation (e.g., "Mitophagy," "Regulation of actin cytoskeleton"). This clear divergence underscores a shift from adaptive immune signalling in NP to pathways indicative of oxidative stress and endothelial dysfunction in PE.

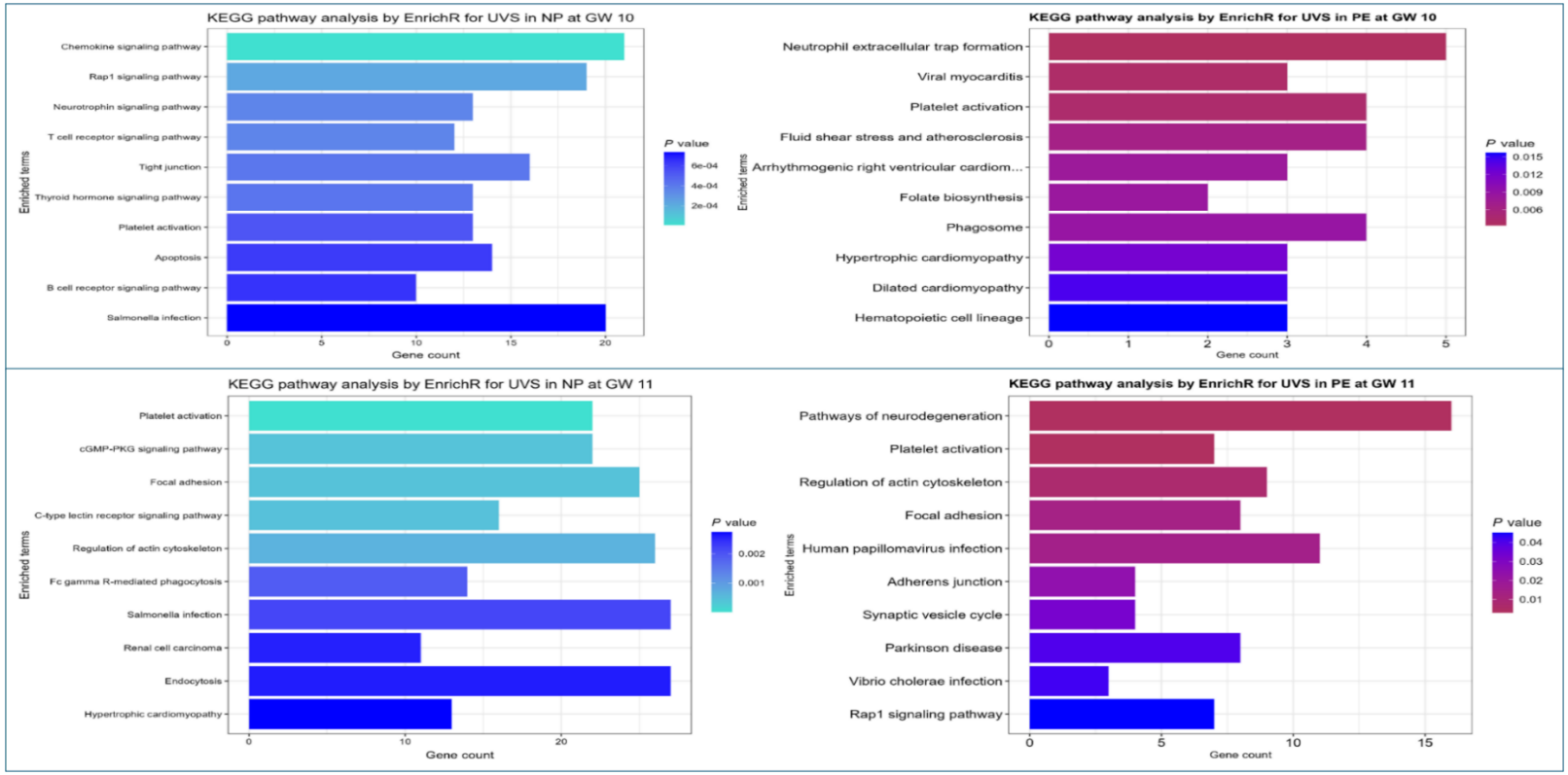


Figure 59: KEGG pathway enrichment analysis for Unique Variant Signatures (UVS) in NP and PE at Gestational Weeks 10 and 11.

The bar charts show the top enriched pathways. In NP, pathways like "Chemokine signalling" and "T cell receptor signalling" (GW10) and "Platelet activation" and "cGMP-PKG signalling" (GW11) highlight coordinated immune and vascular function. In contrast, PE pathways indicate significant stress: "Neutrophil extracellular trap formation" (GW10) points to excessive innate immune activation, while "Pathways of neurodegeneration" and "Parkinson's disease" (GW11) suggest shared mechanisms with cellular stress and neurological pathologies, reflecting severe systemic dysfunction.

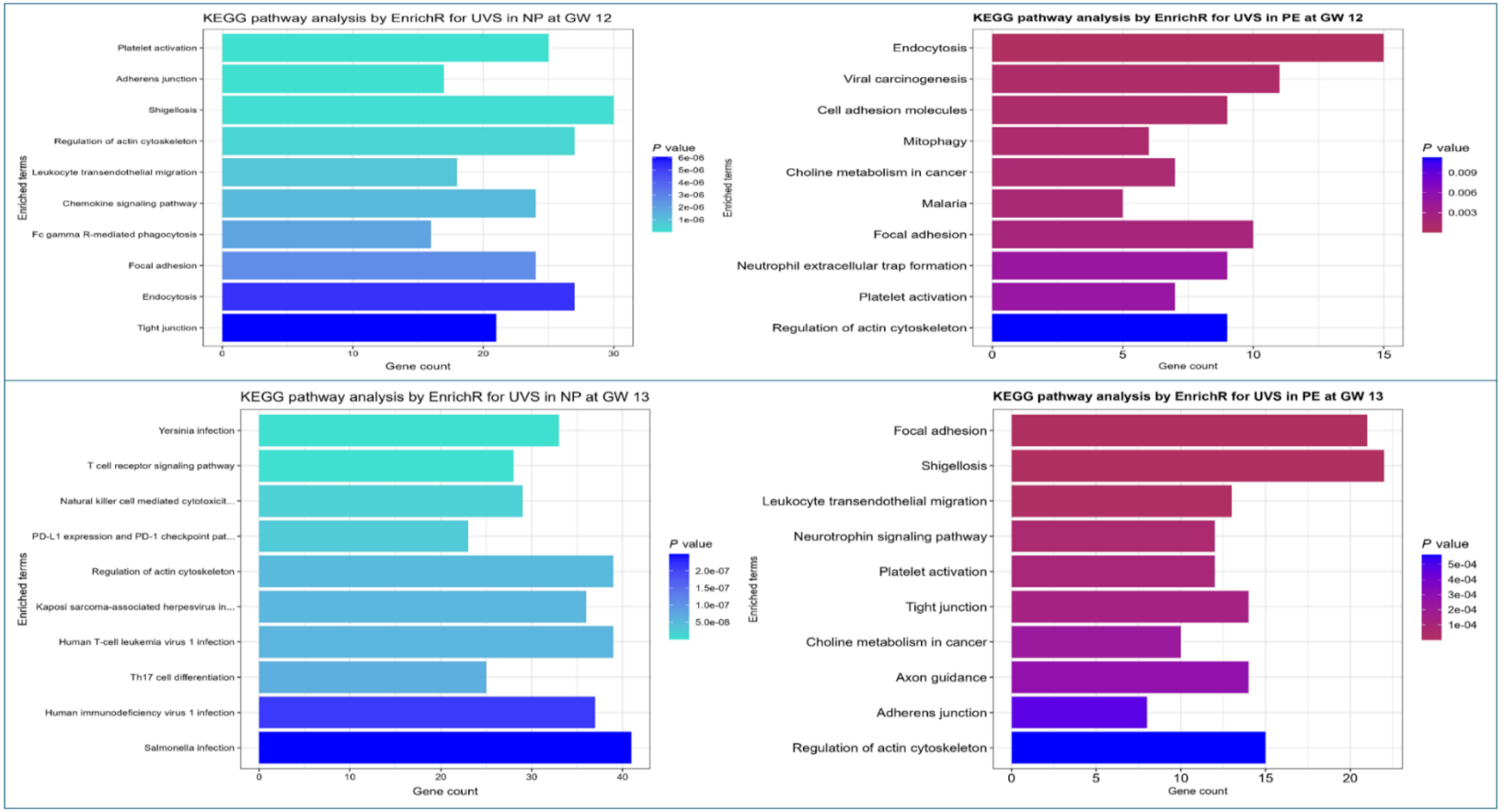


Figure 60: KEGG pathway enrichment analysis for Unique Variant Signatures (UVS) in NP and PE at Gestational Weeks 12 and 13.

The bar charts display the top enriched pathways. NP continues to show strong immune and cell adhesion pathways ("Leukocyte transendothelial migration," "Adherens junction"). At the critical GW13, PE signatures are overwhelmingly dominated by vascular and coagulation pathways, including "Focal adhesion," "Platelet activation," and "Tight junction," alongside "Shigellosis" indicating inflammatory response. This consolidates the picture of PE as a state of profound vascular dysfunction and endothelial damage.

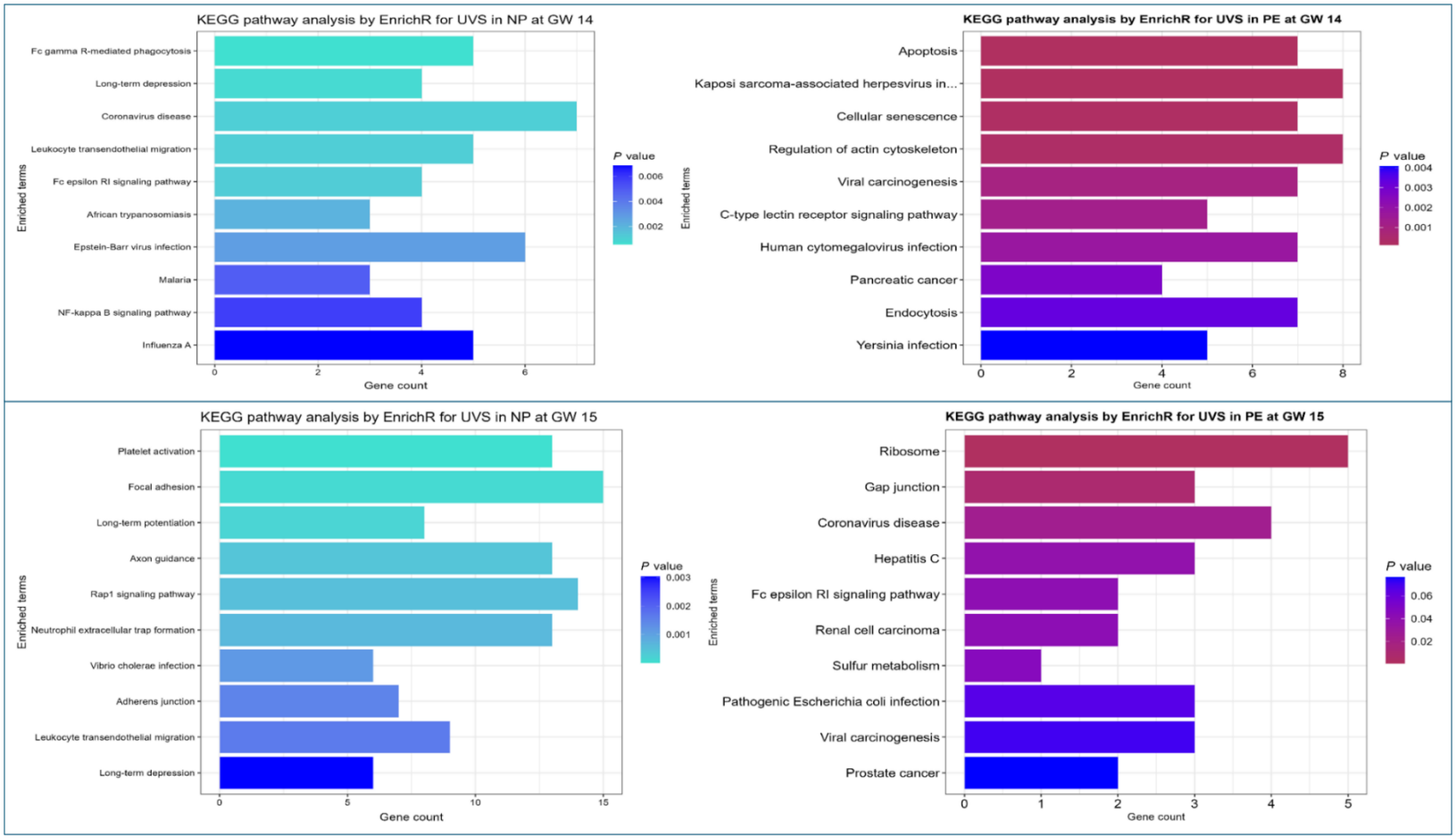


Figure 61: KEGG pathway enrichment analysis for Unique Variant Signatures (UVS) in NP and PE at Gestational Weeks 14 and 15.

The bar charts show the top enriched pathways. NP pathways at GW15, such as "Platelet activation" and "Rap1 signalling," still reflect active vascular and signalling processes. In contrast, PE shows a distinct shift: at GW14, pathways are absent, potentially indicating disrupted biology, and by GW15, pathways like "Ribosome" (protein synthesis) and "Coronavirus disease" (viral response/inflammation) emerge. This suggests a move away from specific placental pathways towards more generalized cellular stress and inflammatory responses.

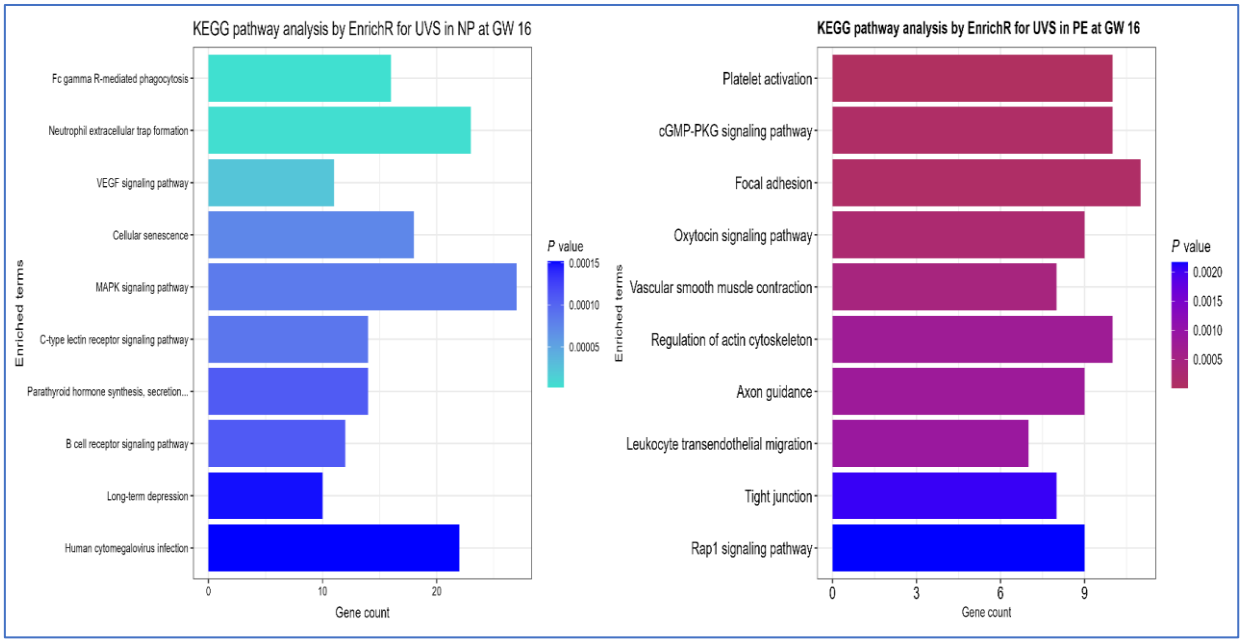


Figure 62: KEGG pathway enrichment analysis for Unique Variant Signatures (UVS) in NP and PE at Gestational Week 16.

The bar charts display the final set of enriched pathways. NP pathways, such as "Fc gamma R-mediated phagocytosis" and "VEGF signalling," continue to reflect active immune function and angiogenesis. For PE, the signature is conclusively defined by pathways central to its clinical presentation: "Platelet activation," "cGMP-PKG signalling," "Vascular smooth muscle contraction," and "Oxytocin signalling." This reinforces that the late first-trimester molecular landscape of PE is dominated by defects in vascular tone, coagulation, and uterine function.

6.3.5.4 Pathway Analysis for PE–associate Significant Genetic Variants per Gestational Week

The KEGG pathway analysis for PE-associated significant variants in GW 8 identifies several critical pathways that may play an essential role in the progression of PE. The Ras signalling pathway (P=0.008), involving genes like RALB, STK4, PAK2, and PAK4, suggests that alterations in cell signalling could affect vascular tone and trophoblast invasion—key features of PE. Additionally, the renal cell carcinoma pathway (P=0.023) shows enrichment with potential associations to renal dysfunction, a hallmark of PE. The focal adhesion pathway (P=0.031), including genes like VASP, PAK2, and PAK4, points to cell adhesion and migration disturbances crucial for placental development and vascular remodelling in PE. Furthermore, the ErbB signalling pathway (P=0.034) and Rap1

signalling pathway (P=0.035), highlighting small GTPases and receptor–mediated signalling, imply altered cellular communication contributing to endothelial dysfunction and abnormal placentation in PE. The T cell receptor signalling pathway (P=0.049) underscores immune activation, a known feature of PE. In contrast, the MAPK signalling (P=0.079) and spliceosome (P=0.092) pathways suggest potential cellular stress responses and gene expression regulation issues. Lastly, the cGMP–PKG signalling pathway (P=0.111) and African trypanosomiasis pathway (P=0.118) provide further insights into vascular and immune system dysfunctions that may contribute to PE pathophysiology. These findings emphasise the importance of cell signalling, immune responses, and cell adhesion in PE, with potential biomarkers in pathways like Ras signalling, focal adhesion, and T cell receptor signalling offering promising targets for research and therapeutic intervention.

For GW9, KEGG pathway analysis reveals several vital pathways that may contribute to PE progression. The Shigellosis pathway (P=0.008) involves genes like FNBP1, PRKCD, and SEPTIN9, suggesting a link between bacterial invasion and immune responses, which could contribute to the inflammatory nature of PE. The Fc gamma R–mediated phagocytosis pathway (P=0.013), enriched with genes such as PRKCD, PRKCA, and DOCK1, indicates immune activation and the role of phagocytic cells in inflammation in PE. The inflammatory mediator regulation of the TRP channels pathway (P=0.013) points to the involvement of transient receptor potential channels in regulating inflammation, which is a key feature of PE. The proteasome pathway (P=0.023), with genes like PSMB6 and PSMC5, suggests potential disturbances in protein degradation processes, leading to cellular stress. The vascular smooth muscle contraction pathway (P=0.030), involving PRKCD, PRKCA, and PRKCQ, suggests altered regulation of vascular tone, a critical aspect of PE's cardiovascular impact. Other pathways, such as the spinocerebellar ataxia pathway (P=0.036) with genes like PSMB6 and PRKCA and the bacterial invasion of epithelial cells pathway (P=0.058), highlight disturbances in epithelial integrity and metabolic processes that could further

contribute to PE. The GnRH signalling pathway (P=0.080) and Rap1 signalling pathway (P=0.090) also suggest hormonal and cellular signalling disruptions that may affect endothelial function and placental development in PE. Overall, these findings underline the involvement of immune activation, vascular dysfunction, cellular stress, and signalling alterations in PE, with pathways like Shigellosis, phagocytosis, and vascular smooth muscle contraction presenting potential biomarkers for therapeutic intervention.

In GW10, KEGG pathway analysis reveals pathways related to immune response and cell signalling, which are key contributors to PE progression. The Human cytomegalovirus infection pathway (P=0.000137), involving genes like GNA13, GNAQ, HLA–A, HLA–B, and PRKCA, suggests an activation of immune and inflammatory responses, potentially contributing to endothelial dysfunction in PE. The Epstein–Barr virus infection pathway (P=0.000534) and Human immunodeficiency virus–1 infection pathway (P=0.000689), enriched with genes such as HLA–B, HLA–C, and RIPK1, highlight immune system activation, which may contribute to the inflammatory nature of PE. The Allograft rejection pathway (P=0.000901) further supports this, with key genes such as HLA–B and HLA–C playing roles in immune rejection and inflammation. The Phagosome pathway (P=0.001) involves genes like DYNC2H1, CLEC7A, and HLA–A, suggesting phagocytic involvement in inflammation and tissue remodelling, common in PE. Other immune–related pathways, such as the Graft–versus–host disease pathway (P=0.0012), with similar HLA–related genes, further emphasise immune modulation. The Type I diabetes mellitus pathway (P=0.0013) suggests disruptions in metabolic and immune processes, which could be linked to PE's cardiovascular and renal manifestations. The Endocytosis pathway (P=0.0017) points to altered cellular trafficking and membrane dynamics that may contribute to endothelial dysfunction. Pathways like Parathyroid hormone synthesis (P=0.00198) highlight potential disruptions in calcium homeostasis, which is crucial for vascular tone regulation in PE. Lastly, the Autoimmune thyroid disease pathway (P=0.0024) with HLA–A, HLA–B, and HLA–C highlights the role of immune regulation in PE. These

findings suggest a complex interplay of immune activation, inflammation, and altered cellular signalling in PE, with viral infections, immune rejection, and phagocytic activity offering potential targets for therapeutic intervention.

In GW11, KEGG pathway analysis emphasises biological processes crucial to PE pathogenesis. The Endocytosis pathway (P=0.000137) involves genes like WIPF1, GBF1, HSPA6, HLA–B, and HLA–A, which regulate cellular transport and receptor recycling, critical for trophoblast function and placental homeostasis. Disruptions in these processes could impair placental development, a hallmark of PE. The Kaposi sarcoma–associated herpesvirus infection pathway (P=0.000534), with genes such as RB1, LYN, PREX1, and HLA–B, suggests immune dysregulation and inflammatory responses exacerbating PE. The Fc gamma R–mediated phagocytosis pathway (P=0.001) involving genes like LYN, HCK, and PRKCB further emphasises immune activation. Pathways such as Notch signalling (P=0.003) and Regulation of actin cytoskeleton (P=0.005) suggest disruptions in angiogenesis and cellular motility crucial for placental development. Other pathways like phosphatidylinositol signalling (P=0.008), human cytomegalovirus infection (P=0.009), and Inositol phosphate metabolism (P=0.011) highlight the interplay between immune activation, signalling dysregulation, and structural impairments in PE. These findings emphasise immune activation, cellular trafficking, and signalling as key contributors to PE pathophysiology.

In GW12, KEGG pathway analysis reveals pathways and genes contributing to PE's pathophysiology. The Protein processing in the endoplasmic reticulum pathway (P=0.009) involves genes like OS9, MAN1A2, and UGGT1, highlighting the role of protein folding and stress responses critical for cellular homeostasis. Disruptions in this pathway can impair placental development by inducing endoplasmic reticulum stress, a known factor in PE. The HIF–1 signalling pathway (P=0.015) with genes like TFRC, EP300, and PFKP underscores the role of hypoxia in angiogenesis and placental oxygenation. The MAPK signalling pathway (P=0.025), involving genes like EFNA2, PLA2G4C, and IGF2, suggests dysregulated signalling involved in cell proliferation and inflammatory responses.

Metabolic pathways like pyruvate metabolism (P=0.035) and valine, leucine, and isoleucine degradation (P=0.042) point to metabolic dysregulation in PE. The Wnt signalling pathway (P=0.052) highlights its impact on trophoblast invasion and vascular remodelling. Other pathways, such as Endocytosis (P=0.067) and NOD–like receptor signalling (P=0.072), further underscore immune regulation and cellular transport in PE.

These analyses underscore the complex molecular mechanisms driving PE, focusing on immune activation, vascular dysfunction, signalling pathways, and cellular stress responses. The highlighted pathways and associated genes offer crucial insights into potential therapeutic targets for mitigating PE progression.

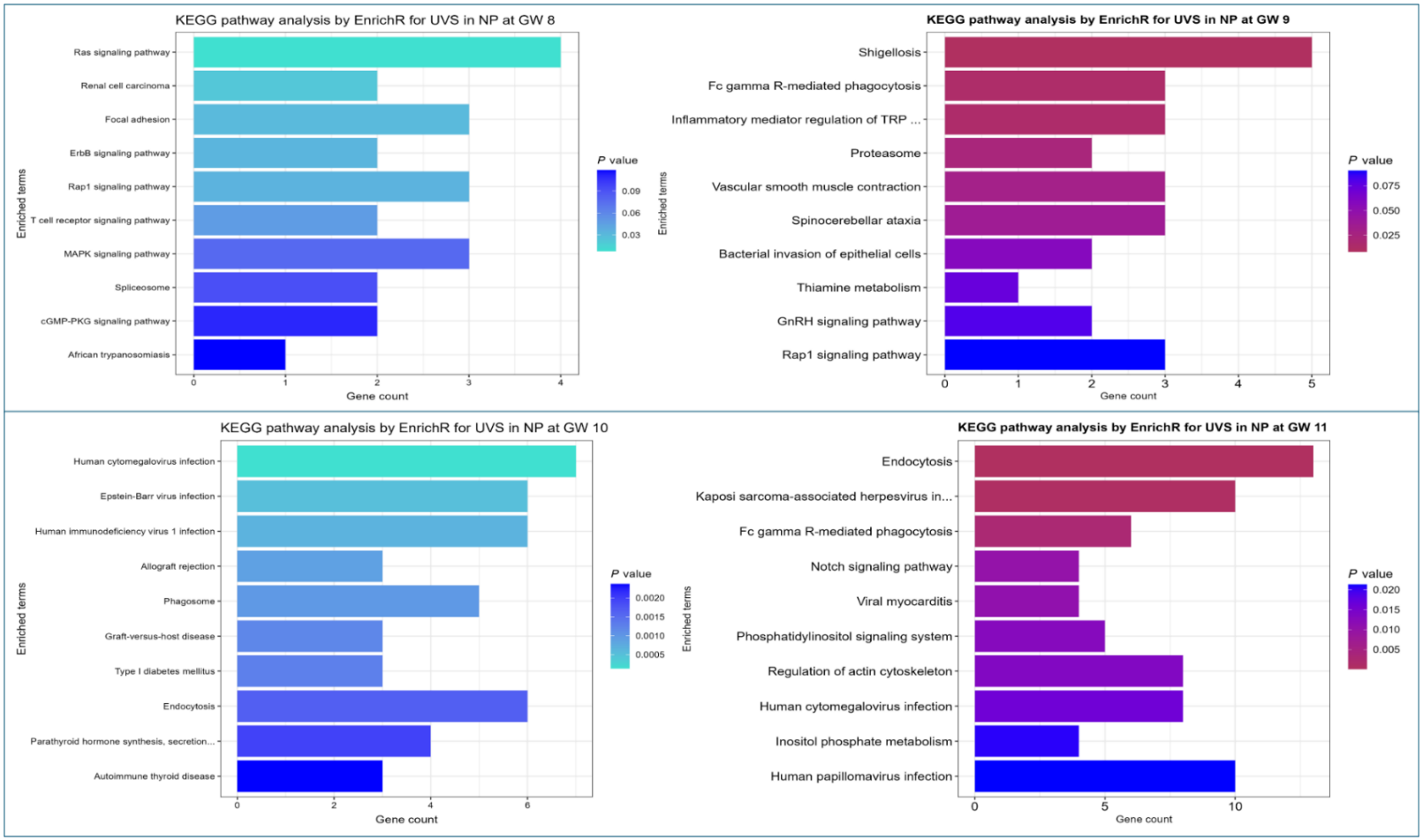


Figure 63: KEGG pathway analysis for PE-associated significant genetic variants across Gestational Weeks 8 to 11.

The chart displays pathways significantly enriched for variants linked to PE. Key pathways include "Ras signalling" and "Focal adhesion" at GW8, "Fc gamma R–mediated phagocytosis" at GW9, and "Human cytomegalovirus infection" and "Endocytosis" at GW10-11. The persistence of immune activation, dysregulated cellular signalling, and altered trafficking pathways across these early weeks highlights their fundamental role in the initiation and progression of PE.

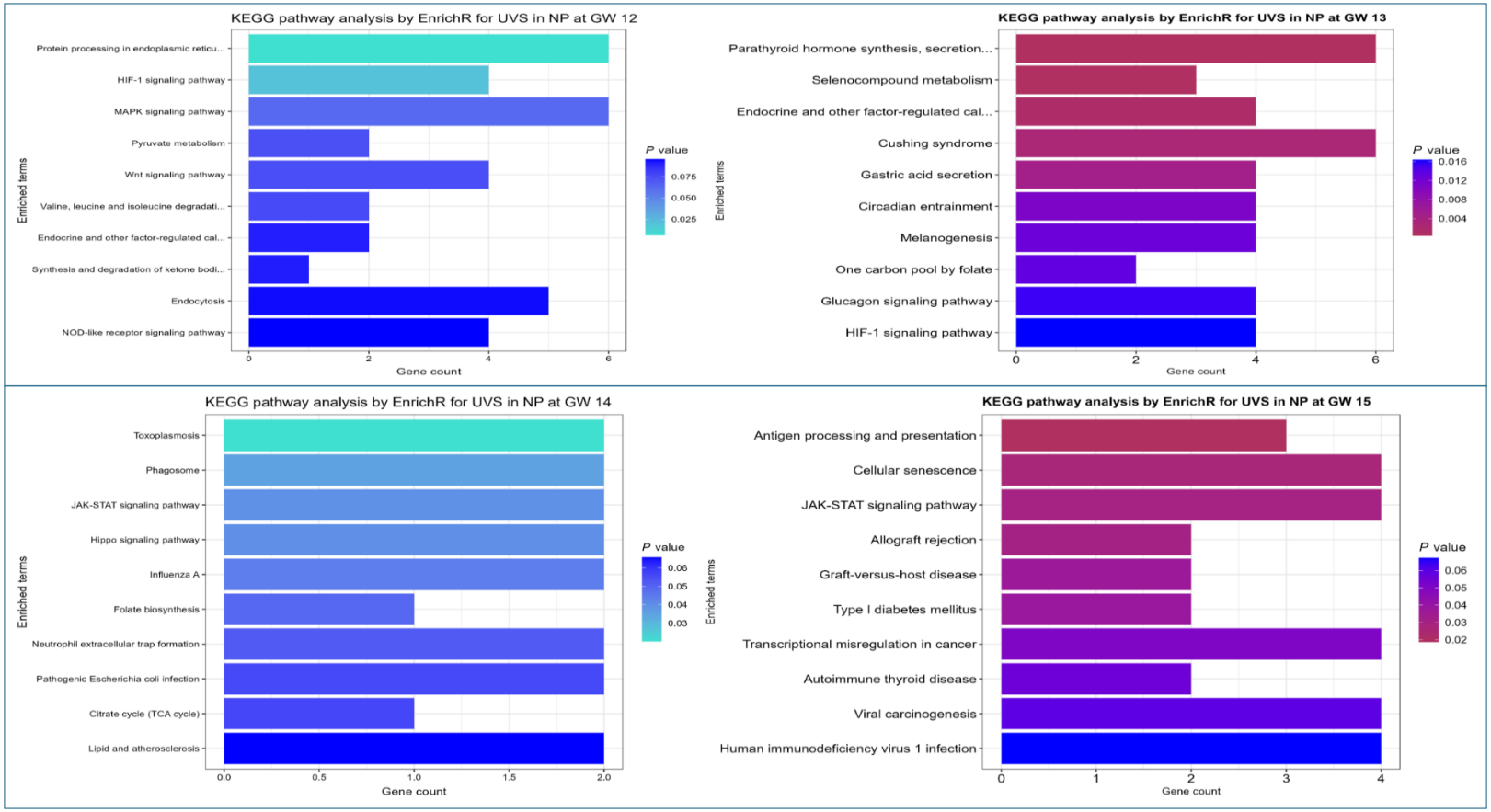


Figure 64: KEGG pathway analysis for PE-associated significant genetic variants across Gestational Weeks 12 to 15. The chart displays the enriched pathways.

The findings highlight a shift towards cellular stress and metabolic dysregulation. Key pathways include "Protein processing in endoplasmic reticulum" and "HIF-1 signalling" (GW12), pointing to ER stress and hypoxia responses. Later, "Endocytosis" and "Kaposi sarcoma-associated herpesvirus infection" (GW14-15) suggest ongoing disruptions in cellular trafficking and persistent inflammatory triggers. The convergence on these stress and processing pathways underscores the cellular-level burden imposed by the genetic variants associated with PE.

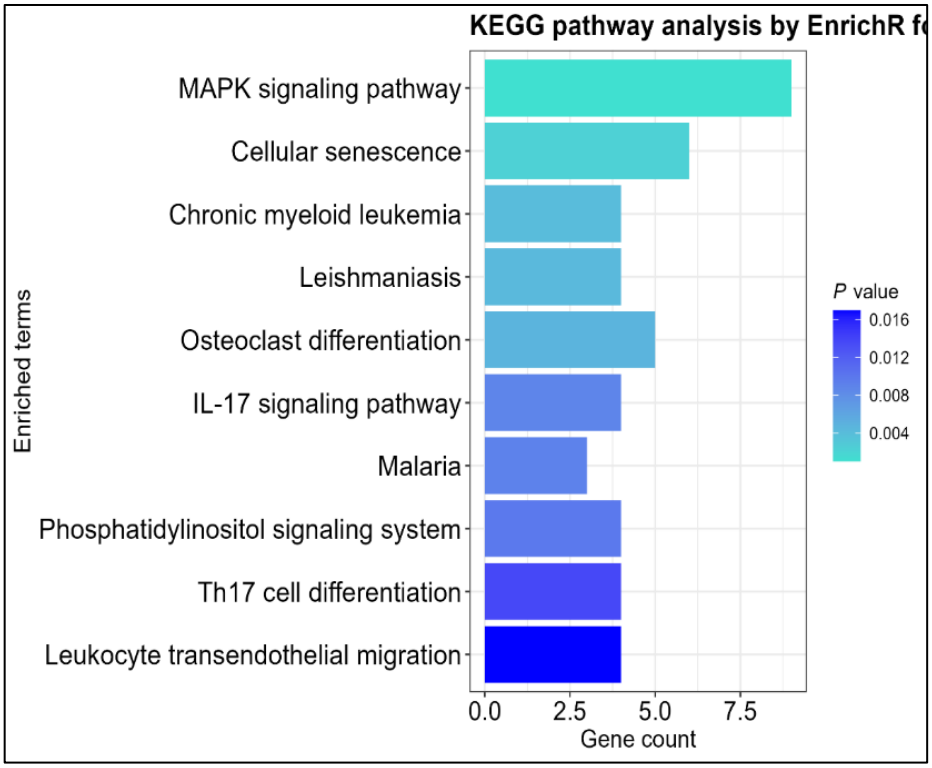


Figure 65: KEGG pathway analysis for PE-associated significant genetic variants at Gestational Week 16.

This final analysis reveals a focused set of pathways, including "Endocytosis" and the "Wnt signalling pathway." The enrichment of "Endocytosis" suggests continued dysregulation in cellular uptake and receptor-mediated signalling. The involvement of the "Wnt signalling pathway," which is critical for placental development and trophoblast function, indicates that genetic disruptions in this key developmental pathway remain relevant to PE pathology even at the end of the first trimester, potentially affecting placental maturation.

6.4 Conclusion

This study leverages the power of non-invasive liquid biopsy to delineate a dynamic and temporally structured landscape of pre-eclamptic genetic signatures across early gestation. By performing rigorous variant calling on maternal plasma cfRNA, the analysis moved beyond generic associations to identify specific gestational-week–specific unique variant signatures (UVS) and PE-associated genetic variants (PE-AGVs). Key biomarker genes, such as ITGB3 and HLA-DRB1 at GW11, and ROCK2 and RAC1 at GW16, have been substantiated by linking them mechanistically to core pathophysiological processes. These genes were

prioritized as they are not merely associated with PE but are directly functional in the platelet dysfunction, immune dysregulation, and aberrant vascular tone that define the disease. The functional enrichment analyses further corroborate this, revealing a clear trajectory in PE from early disruptions in cytoskeletal organisation and signalling to a mid-gestation dominance of platelet activation and coagulopathy, culminating in widespread cellular stress. Crucially, this curated set of high-confidence, biologically grounded variant signatures provides a powerful and interpretable feature set for machine learning. In an AI-assisted diagnostic model, these specific variants and their corresponding genes offer a biologically constrained feature space, preventing overfitting and increasing clinical translatability. The model can learn not just that these features are important, but how their complex, non-linear interactions with transcriptomic data drive the pathophysiology, ultimately enabling superior risk stratification and creating a window for early therapeutic intervention.

In conclusion, variant calling analysis of maternal plasma cfRNA has identified a dynamic landscape of PE-associated genetic signatures across early gestation. A central finding was the peak of unique variant signatures, especially SNPs, during gestational weeks 11–12. It is essential to emphasize that these variants originate from placental-derived cfRNA, acting as a real-time liquid biopsy of the foetal-placental unit. This temporal pinpointing to weeks 11-12 underscores this period as one of pronounced molecular dysfunction in pre-eclamptic pregnancies, likely linked to the intense placental remodelling and vascular changes occurring at this time. This makes the cfRNA from this specific gestational window a particularly rich resource for early, non-invasive prediction of PE.

6.5 Future Direction

Future research should focus on integrating multi–omics approaches, combining transcriptomic, proteomic, and epigenomic data to provide a more comprehensive understanding of PE pathophysiology. Developing robust ML models using GW–

specific UVS and PE–AGVS identified in this study can improve early prediction and personalised management of PE. Expanding these findings to larger, ethnically diverse cohorts is essential to ensure broader applicability and uncover population–specific genetic variations. Additionally, longitudinal studies covering all trimesters and postpartum phases can illuminate the progression and resolution of PE, enabling the identification of critical intervention points. Translational efforts should incorporate cfRNA–based screening into clinical settings for routine use in high–risk pregnancies. Finally, therapeutic research leveraging identified key biomarkers, such as RAC1, CD47, and ROCK2, should explore targeted interventions to mitigate immune dysregulation and vascular dysfunction, ultimately improving maternal and foetal outcomes.

The cross-sectional nature of the analysis represents an important limitation. By not accounting for the non-independence of samples from the same individuals across different gestational weeks, important within-subject temporal dynamics in variant signatures may have been overlooked.

6.6 Summary of the Chapter

This study investigates GW–specific UVS and significant PE-associated genetic variants using cfRNA sequencing data from pregnancies between GW 8 and 16. The study uses advanced bioinformatics tools to identify genetic biomarkers for early prediction of PE. The analysis included 10,590 UVS in NP and 3,186 UVS in PE pregnancies, highlighting distinct genetic landscapes between the two conditions. GW13 emerged as a critical time point, displaying the highest number of UVS and differential gene expression changes, indicating significant transcriptomic activity during this period.

The study identified 1,434 PE-associated SNPs and 36 INDELs, with week 11 showing the highest concentration of these variants. Intronic and non–coding regions, including UTRs, were identified as dominant mutation sites, emphasising their regulatory roles in pregnancy–related processes. Gene Ontology (GO) analysis revealed disrupted immune responses, apoptotic signalling, and vascular

dysfunction in PE, while NP showed enriched processes related to immune tolerance and placental development. Pathway enrichment analysis demonstrated a shift in PE pregnancies toward pathways associated with vascular stress, oxidative damage, and platelet activation.

The study highlights the importance of dynamic molecular profiling across gestational weeks by integrating advanced variant calling and functional annotation tools. It underscores the potential of gestational week–specific UVS as biomarkers for early detection of PE and offers a foundation for developing predictive and therapeutic strategies. The findings emphasise the critical role of non–invasive cfRNA–based approaches in improving maternal and foetal outcomes by facilitating early intervention and personalised prenatal care. The next chapter will discuss the development of an AI-aided PE prediction model using the previously identified transcriptomic and variant signatures.

Chapter 7
Developing an Automated AI-aided Disease Prediction Model for Detecting PE–Risk Women in Early Pregnancy

7.1 Introduction

The use of AI has modernised healthcare in the last few years by permitting the development of predictive models that can analyse complex various datasets with higher speed and level of accuracy (Phillips, Spithoff and Simpson, 2022; Nwaimo, Adegbola and Adegbola, 2024; Dixon *et al.,* 2024). Applying these AI-assisted tools leverages ML and DL algorithms to detect patterns in vast amounts of data, which is advantageous over traditional statistical methods. In the context of PE, AI presents a transformative opportunity to integrate diverse types of genomic and clinical data, such as variant signatures, transcriptomic signatures, methylation signatures, proteomic signatures, epigenetic signatures, clinical parameters, and environmental factors to build comprehensive models for early interventions Such models can offer personalised predictions, allowing the implementation of appropriate interventions to reduce further pregnancy–related complications.

Early detection of PE is crucial for preventing severe complications, yet traditional diagnostic methods often lack the sensitivity and specificity required for timely intervention (Briceño-Pérez, Briceño-Sanabria and Vigil-De Gracia, 2009; Chaemsaithong, Sahota and Poon, 2022). These AI–powered models can provide early, personalised predictions of PE risk, allowing healthcare providers to initiate targeted interventions early in pregnancy. For example, they could predict the likelihood of developing PE based on a combination of genetic predisposition, biomarkers, and clinical history. Early detection facilitates proactive monitoring and intervention, potentially reducing the severity of complications such as preterm birth, maternal organ damage, and foetal growth restriction. However, developing AI-aided disease predictive models face some challenges, including data quality and availability, model interpretability, bias in training datasets, the need for

extensive and diverse patient cohorts, regulatory and ethical considerations, and the integration of AI predictions into clinical workflows for effective decision–making.

7.1.1 Challenges in Developing AI–Driven Predictive Models

While AI holds immense potential for revolutionising predictive healthcare, developing reliable and accurate AI–driven predictive models comes with several challenges, particularly in complex conditions like PE (Feng, W. and Luo, 2024; Butler *et al.,* 2024). One of the primary challenges is the quality and availability of data. AI models rely heavily on large, high–quality datasets to train algorithms effectively (Alam and Gonzalez Suarez, 2024). However, obtaining comprehensive datasets that include diverse clinical, genomic, and environmental factors can be problematic in the case of PE. Data from different populations and clinical settings may vary significantly, leading to biases that could impact the generalisability of the model (Alam and Gonzalez Suarez, 2024). Another significant challenge is the integration of heterogeneous data types. In PE prediction, data from clinical tests, genomic studies, and environmental exposures must be combined to create a comprehensive model (Sufriyana, Wu and Su, 2020). Each data type has its own structure, scale, and source, making it difficult to standardise and integrate into a cohesive model. Ensuring that the data from these diverse sources are accurately harmonised is a critical step in building an effective AI–driven system (Alam and Gonzalez Suarez, 2024).

Data privacy and ethical concerns also pose significant challenges in developing AI models (Konidena, Malaiyappan and Tadimarri, 2024; Patel, K., 2024). Since AI in healthcare often requires access to sensitive personal data, such as genetic information and medical history, ensuring the privacy and security of this data is crucial. Legal and ethical considerations must be addressed to comply with regulations such as GDPR, which govern the use of personal data in research and clinical settings (Wang, Wei, 2024). Lastly, model interpretability remains a challenge, particularly with deep learning models. These models, while powerful,

often operate as "black boxes," making it difficult for healthcare providers to understand how decisions or predictions are made (Hassija *et al.,* 2024). Ensuring that AI models are transparent and explainable is essential for building trust among clinicians and patients and enabling clinical validation. Addressing these challenges requires ongoing collaboration between researchers, healthcare providers, and regulatory bodies to ensure that AI–driven predictive models are effective and trustworthy.

7.1.2 Significance of a Multi–Dimensional Approach

Predicting PE using ML requires a multi–dimensional approach integrating diverse data sources to improve accuracy, generalisability, and clinical applicability. A multi–dimensional approach enhances predictive power by incorporating genetic, clinical, biochemical, and imaging data. Genomic data, including differential gene expression and variant signatures, provide insights into genetic susceptibility. Transcriptomic profiling helps identify consistent gene expression patterns associated with PE (Youssef *et al.,* 2024). Combining these with clinical features such as blood pressure, maternal age, BMI, and previous obstetric history allows for a holistic assessment of risk factors (Li, S. *et al.,* 2022; Li, T. *et al.,* 2024).

Additionally, biochemical markers, such as angiogenic and anti-angiogenic factors, offer early indications of placental dysfunction. Imaging data from Doppler ultrasounds, assessing uterine artery blood flow, further refining risk stratification (Li, T. *et al.,* 2024; Torres-Torres *et al.,* 2024). Integrating these diverse datasets through ML algorithms enables the identification of complex interactions between various biomarkers, improving early detection and personalised risk prediction (Li, S. *et al.,* 2022; Li, T. *et al.,* 2024).

Feature selection and dimensionality reduction techniques help manage high–dimensional data, ensuring computational efficiency without compromising model performance. Deep learning and ensemble methods can capture non–linear relationships between features, enhancing predictive accuracy . By incorporating heterogeneous datasets from different demographics, the ML model minimises

bias and enhances clinical applicability (Li, S. *et al.,* 2022; Li, T. *et al.,* 2024). This integration facilitates the development of AI–driven clinical decision support systems, aiding obstetricians in timely interventions and personalised management strategies . Ultimately, the multi–dimensional approach in PE ML disease prediction bridges the gap between research and clinical practice, providing a more reliable, early, individualised risk assessment framework. This holistic methodology holds promise for reducing maternal and foetal morbidity, improving pregnancy outcomes, and advancing precision medicine in obstetrics .

7.1.3 Research Hypothesis

This study hypothesises that integrating Unique Variant Signatures (UVS) and Unique Transcriptomic Signatures (UTS) enhances the predictive accuracy of PE in early pregnancy.

7.1.4 Research Question

Can the integration of GW–specific UVS and UTS improve the predictive accuracy of ML models for early identification of PE?

7.1.5 Research Aim and Objectives

The project aimed to develop an AI-assisted automated prediction model for the early identification of PE in pregnant women by leveraging AI and combining both UVS and UTS. To achieve this, the following objectives were developed:

1. To develop an AI-based prediction model for early identification of PE using transcriptomic and variant biomarkers.
2. To implement advanced feature selection techniques to refine biomarkers and optimise ML model performance.
3. To evaluate the effectiveness of multiple ML algorithms in predicting PE, comparing their accuracy and reliability.
4. To validate the identified biomarkers by intersecting them with previously reported PE–linked genes to ensure biological relevance.

7.2 Method

7.2.1 Data Pre–processing

This study aimed to develop an ML–aided disease prediction model by integrating unique transcriptomic and variant signatures from the same patient cohort. The final dataset for model development was created using UDEGS, UCEGS, and UVS, all derived from gestational weeks 8 to 16. These signatures were carefully processed to ensure their relevance for disease prediction.

7.2.1.1 Normalisation

Initially, the transcriptomic biomarkers, including UDEGS and UCEGS, were combined on a gestational–week basis to create a unified dataset for each week. Each weekly dataset was then used in the feature selection process, ensuring it retained its temporal structure while capturing the unique molecular signatures associated with each gestational week. This dataset was then normalised using edgeR's TMM (Trimmed Mean of M–values) method to account for differences in library sizes and sequencing depth, followed by conversion to log2 counts per million (logCPM) values for further analysis. The resulting normalised dataset was subsequently scaled to ensure that all features contributed equally to the analysis. In this study, scaling was performed using the StandardScaler from the sklearn.preprocessing module, which standardises the data by transforming each feature to have a mean of 0 and a standard deviation of 1. This step is essential when dealing with data with varying units or ranges, as it prevents features with larger values from disproportionately influencing the feature selection process.

Similarly, the per–gestational week–specific UVS, including SNPs and INDELs, were also scaled using the same procedure to ensure that the genetic variants were appropriately represented and comparable to the transcriptomic data. The scaled dataset for each week, including both the transcriptomic and variant signatures, was then used as input for feature selection separately. The most relevant and informative features were identified during this process for training

the ML model. This approach ensured that only the most significant biomarkers were retained for disease prediction without any bias from differences in scale.

7.2.1.2 Feature Selection

A systematic approach was employed in the feature selection process to identify the most significant candidate features for developing the ML model. For each GW–specific combined dataset, transcriptomic and variant signatures were analysed separately for feature selection. This study adopts the Recursive Feature Elimination with Cross–Validation (RFE with CV) wrapper method for its feature selection process, a sophisticated approach widely embraced in the realm of bioinformatics, especially for gene expression feature selection (Liu, W. and Wang, 2022; Zhang, Ying *et al.,* 2018). It was applied independently to the combined transcriptomic and variant data to identify and rank the most informative features. Since the dataset for each GW contained varying numbers of NP and PE cases, the stratified method was used to ensure a balanced selection of CV folds. This stratification maintained the proportional representation of NP and PE data across the folds, providing robust and unbiased feature selection. Multiple ML algorithms were utilised during the feature selection process to enhance the generalisability and reliability of the selected features. These included Decision Trees (DT), Random Forests (RF), Logistic Regression (LR), Gradient Boosting Machines (GBM), Support Vector Machines (SVM), and Linear Discriminant Analysis (LDA). Each algorithm evaluated the dataset independently, using a CV to assess feature importance and select candidate features that were consistently significant across models. This multi–algorithm strategy ensured that the selected features were statistically significant and relevant across different predictive frameworks. The outcome of this process was a refined set of features for each GW, optimised for training the ML model to predict disease outcomes, such as PE, with high accuracy.

Following feature selection for each GW, the significant signatures from GW 8 to 16 were merged separately for transcriptomic and variant data. Subsequently, the

complete transcriptomic and variant datasets from early pregnancy were analysed independently to identify the most significant biomarkers for ML models. This approach enhances biomarker identification by independently analysing transcriptomic and variant data, preserving biologically relevant patterns. Using RFE with CV across multiple ML models improves feature selection and reduces bias, making biomarkers more reliable. Stratified cross–validation addresses class imbalances, enhancing model performance. Merging significant features across gestational weeks strengthens the identification of persistent biomarkers. This iterative process improves the predictive accuracy and generalisability of ML models, contributing to a robust early diagnostic tool for PE.

An advancement of this approach lies in its ability to enhance the robustness and generalisability of biomarker identification for ML models by integrating multi–stage feature selection and a diverse set of algorithms. By independently analysing transcriptomic and variant data, this method ensures that biologically relevant patterns specific to each data type are preserved without cross–contamination. The use of RFE with CV across multiple ML models not only refines feature selection but also minimises the risk of model bias, making the identified biomarkers more reliable for predictive modelling. Additionally, stratified cross–validation ensures that class imbalances are addressed, thereby improving model performance in distinguishing between NP and PE cases. The sequential merging of significant features across gestational weeks further strengthens the identification of persistent biomarkers that remain crucial throughout early pregnancy. This iterative refinement process enhances the predictive power of ML models by focusing on consistently relevant features, ultimately contributing to the development of a highly accurate and generalisable early diagnostic tool for PE.

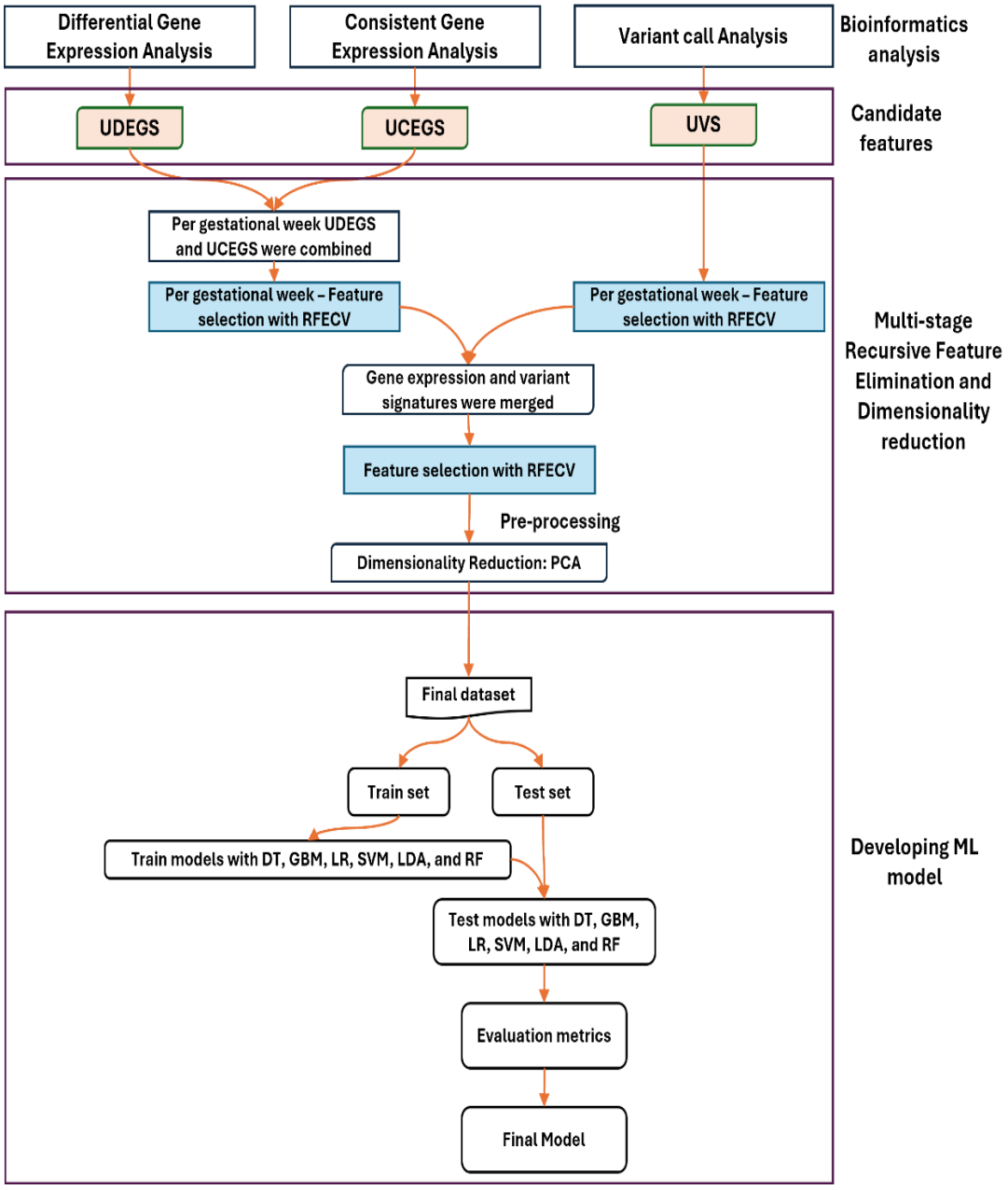


Figure 66: Workflow for feature selection and developing ML model for PE prediction.

Transcriptomic and variant signatures undergo multi–stage Recursive Feature Elimination with Cross–Validation (RFECV) to identify the most significant biomarkers across gestational weeks, leading to an optimised dataset for model development.

7.2.1.3 Pre-eclampsia–Linked Gene Extraction and UTS Intersection

The PE–linked genes were sourced from three main repositories: Genecard, Disgenet, and previously published research articles. These genes were curated from these sources. Following the multi–stage feature selection process, a set of UTS was derived from the final dataset. These UTS were subsequently intersected with the previously identified PE–linked genes. The overlap was assessed to explore the relevance of the UTS in the context of established PE genetic signatures. This intersection aimed to evaluate the extent to which the UTS correspond to genes previously implicated in PE. This method seeks to identify and validate potential biomarkers for PE by comparing the UTS with established PE genetic signatures. This approach helps determine the identified UTS's biological relevance, supporting their role in early diagnosis, risk prediction, or therapeutic targeting for PE.

7.2.1.4 Dimensionality Reduction

Dimensionality reduction is a crucial preprocessing step in ML that involves reducing the number of input variables while retaining the most relevant information. Many real–world datasets, especially in fields like bioinformatics, contain thousands of features, making it computationally expensive and difficult to develop efficient models. By applying dimensionality reduction techniques, redundant, irrelevant, or highly correlated features can be eliminated, improving both model performance and interpretability. Two main types of dimensionality reduction techniques exist: feature selection and feature extraction. Feature selection methods, such as RFE, select the most essential features from the dataset without transforming them. Feature extraction methods, such as Principal Component Analysis (PCA) and t–distributed Stochastic Neighbor Embedding (t–SNE), transform the original features into a lower–dimensional space while preserving essential information. In this context, PCA was explicitly used for dimensionality reduction, as it helps reduce the dataset's dimensionality by finding principal components that maximise variance and capture the essential patterns in

the data. Following multi–stage recursive feature elimination, the dataset was used for dimensionality reduction with PCA to extract the most relevant UTS and UVS for model development.

The importance of dimensionality reduction lies in several key aspects: enhancing model performance by reducing overfitting and improving generalisation, lowering computational costs by decreasing memory usage and training time, improving model interpretability by making feature relationships more transparent, and addressing the curse of dimensionality, where high–dimensional data can lead to sparsity and poor performance. In ML model development, selecting the right dimensionality reduction approach is essential to ensure optimal accuracy, efficiency, and scalability.

To determine the optimal number of principal components (PCs) for dimensionality reduction, a systematic evaluation was conducted using Principal Component Analysis (PCA) in combination with cross –validation (see Appendix C: ML model development). The primary objective was to retain the most informative features while minimising redundancy and preserving classification performance. The optimisation process involved testing a range of principal components from 1 to 100 to identify the number that maximised classification accuracy. A Random Forest (RF) classifier was used as the predictive model within a pipeline that incorporated PCA for feature transformation. To ensure robustness, cross–validation (CV) was performed with varying folds, ranging from 2 to 10, to assess model stability across different partitioning strategies. The cross–validation accuracy was computed for each combination of PCA components and CV folds, and the best–performing configuration was selected based on the highest mean accuracy.

The evaluation revealed the optimal number of components and the most suitable cross–validation setting for maximising model performance. The best–performing PCA configuration was identified by tracking the highest cross–validation accuracy, which was subsequently used in further model development.

Additionally, results were visualised by plotting cross–validation accuracy against the number of components for different CV values, providing insight into the trade–off between dimensionality reduction and predictive performance. This approach ensured that the selected PCA components effectively captured the underlying variance in the dataset while enhancing model efficiency and generalisability.

7.2.2 Machine Learning Model Development

The final dataset, derived from the multi–stage feature selection process and dimensionality reduction, was utilised to develop ML models to predict PE. This dataset comprised the most significant transcriptomic and variant biomarkers, identified through RFECV and applied independently to each GW. It was subsequently refined through an additional selection step. To ensure model robustness and generalisability, six ML algorithms were implemented: SVM, DT, RF, LR, GBM, and LDA (Figure 66). These algorithms were selected to capture diverse predictive patterns and assess the effectiveness of different classification approaches.

Several ML algorithms have shown promising results in developing PE prediction models. Among these, SVM, DT, RF, LR, GBM, and LDA are particularly effective. These algorithms handle complex, high–dimensional data and capture non–linear relationships between predictors and outcomes . SVM is particularly effective for high–dimensional datasets, such as those derived from cfRNA sequences, as it excels at identifying complex patterns and finding optimal hyperplanes for classification, even with nonlinear relationships . DTs offer simplicity and interpretability, providing a clear decision–making pathway crucial for clinical applications (Valdes *et al.,* 2016). RF, an ensemble of decision trees, enhances accuracy and reduces overfitting while effectively handling noise, imbalanced data, and missing values (Gu, Q. *et al.,* 2022). LR is an interpretable and efficient foundational algorithm for binary classification, offering insights into the contribution of specific features and making it a reliable baseline model (Dube and Verster, 2023). GBM combines weak learners into a robust predictive model,

capturing intricate interactions between features and performing well in scenarios with imbalanced datasets or rare outcomes, such as PE (Airlangga, 2024). Finally, LDA is efficient for smaller datasets and provides probabilistic insights by identifying linear combinations of features that best separate classes, making it useful when linear separability assumptions hold. Together, these algorithms address the complexities of PE prediction by handling diverse data types, providing feature–importance insights, and ensuring a balance between interpretability and predictive accuracy, ultimately supporting the development of precise and actionable clinical models.

7.2.2.1 Data Splitting Strategy

The dataset was divided into training and testing sets to ensure a balanced and unbiased evaluation, with 70% allocated for training and 30% for testing. A stratified sampling approach was applied to maintain the proportional distribution of PE and NP cases in both subsets. This method prevented class imbalance issues, ensuring the model learned from a representative dataset. A fixed random seed was used to maintain reproducibility across experiments. The training set was used for model development, including hyperparameter tuning and cross–validation, while the independent test set was reserved for final performance evaluation. This strategy minimised overfitting and improved the model's generalisability to unseen data.

7.2.2.2 Support Vector Machine (SVM) Model Development

The Support Vector Machine (SVM) model was implemented for binary classification, addressing class imbalance with Synthetic Minority Oversampling Technique (SMOTE) to generate synthetic samples for the minority class. Grid search with cross–validation optimised hyperparameters, selecting C = 0.1, kernel = 'linear', and class_weight = 'balanced' to ensure an optimal trade–off between margin maximisation and misclassification control. The model was trained on the resampled dataset and evaluated on an independent test set using accuracy,

precision, recall, F1–score, and area under the Receiver Operating Characteristic (ROC) curve (AUC–ROC). A confusion matrix quantified classification errors, and a ROC curve visualised the model's discriminative performance. Implementation was conducted using scikit–learn, imbalanced–learn, and matplotlib for preprocessing, training, and evaluation.

7.2.2.3 Random Forest (RF) Model Development

The Random Forest (RF) classifier was trained on a SMOTE–balanced dataset to mitigate class imbalance. Model parameters included 100 decision trees, class_weight = 'balanced', and a fixed random state for reproducibility. Performance evaluation considered accuracy, precision, recall, F1–score, and AUC–ROC, with a confusion matrix assessing misclassifications. Additional optimisations explored alternative resampling methods such as ADASYN and SMOTE–Tomek and adjusting the classification threshold to improve sensitivity. The model was implemented using scikit–learn, imbalanced–learn, and matplotlib.

7.2.2.4 Logistic Regression (LR) Model Development

A Logistic Regression (LR) model was developed, utilising SMOTE to balance the dataset. Grid search with cross–validation optimised hyperparameters, selecting C = 1.0 with class_weight = 'balanced' to balance model complexity and classification performance. The liblinear solver was chosen for efficiency in binary classification. Model performance was assessed using accuracy, precision, recall, F1–score, and AUC–ROC, with a confusion matrix quantifying errors and a ROC curve illustrating class separation. The model was implemented using scikit–learn, imbalanced–learn, and matplotlib.

7.2.2.5 Linear Discriminant Analysis (LDA) Model Development

For Linear Discriminant Analysis (LDA), SMOTE was applied to correct the class imbalance. Hyperparameter tuning identified solver = 'lsqr' and shrinkage = 'auto', ensuring efficient computation and reduced overfitting. The trained model

projected the dataset onto a lower–dimensional space to enhance class separability. Evaluation metrics included accuracy, precision, recall, F1–score, AUC–ROC, and a confusion matrix and ROC curve for classification assessment. The model was implemented using scikit–learn, imbalanced–learn, and matplotlib.

7.2.2.6 Gradient Boosting Machine (GBM) Model Development

The Gradient Boosting Machine (GBM) model was trained on a SMOTE–balanced dataset to address class imbalance. Grid search with cross–validation optimised key hyperparameters, selecting learning_rate = 0.05, n_estimators = 200, and max_depth = 3 for optimal performance. The model sequentially improved weak learners to reduce classification errors. Performance evaluation included accuracy, precision, recall, F1–score, and AUC–ROC, with a confusion matrix and ROC curve assessing classification effectiveness. The model was implemented using scikit–learn, imbalanced–learn, and matplotlib.

7.2.2.7 Model Evaluation Metrics

Multiple evaluation metrics were employed to comprehensively assess the performance of the developed ML models, ensuring a balanced and rigorous analysis of classification effectiveness. Given the imbalanced nature of the dataset, accuracy alone was insufficient; therefore, additional metrics, including precision, recall, F1–score, and the area under the Receiver Operating Characteristic curve (AUC–ROC), were used to provide a more detailed evaluation. A confusion matrix was also analysed to provide insights into classification errors by breaking down true positives, false positives, true negatives, and false negatives. To ensure model generalisability, each model was evaluated on an independent test set, and cross–validation was applied during training. Furthermore, ROC curves were plotted to compare the performance of different classifiers visually. Implementing and metric computations were conducted using Python libraries, including scikit–learn and matplotlib. This comprehensive evaluation framework ensured a robust assessment of predictive

capabilities, facilitating the identification of the most effective model for early PE detection.

7.2.2.7.1 Accuracy

Accuracy is a metric used to evaluate the overall correctness of those predictions by counting the total number of correct predictions a model makes. It is widely used as a critical evaluation metric for evaluating ML models' performance for disease prediction. It is computed by dividing the number of correct predictions of disease state by the total number of predictions made, yielding a percentage value representing the overall correctness and reliability of models (Chandrasekhar and Peddakrishna, 2023). The calculation equation for accuracy is as follows:

$$Accuracy = \frac{Number\ of\ correct\ predcitions}{Total\ number\ of\ predictions} \times 100\%$$

Additionally, accuracy calculation adds two accurate predictions, true positive and true negative, divided by the total number of predictions, which includes true and false positives and true and false negatives. A perfect prediction corresponds the accuracy score of 1.0, whereas the worst prediction is implied by a score of 0.0 (Vujović, 2021). The additional calculation equation for accuracy is shown below.

$$Accuracy = \frac{True\ Positives + True\ Negatives}{True\ positives + Fasle\ positives + True\ negatives + False\ negatives}$$

While accuracy provides a straightforward measure of model performance, it may not be sufficient in scenarios where class imbalance exists or when different types of errors have varying consequences. In such cases, other metrics like recall, F1 score, AUC–ROC, and precision offer a more nuanced understanding of the model's behaviour.

7.2.2.7.2 Recall

A recall of an ML model is a metric that expresses how frequently it correctly selects true positives out of all the true positive samples in the dataset. It can be calculated by dividing the number of true positives by the number of true positive and false negatives instances. The recall calculation equation is displayed below.

$$Recall = \frac{True\ positives}{True\ positives + False\ negatives}$$

The recall is considered a vital evaluation metric for assessing ML disease prediction models, as it indicates the percentage of actual positive cases that the model correctly identifies as positive. Recall values, ranging from 0 to 1, closer to 1 signify superior prediction capability in ML models (Kavitha *et al.,* 2022).

7.2.2.7.3 F1 Score

The F1 score, a vital evaluation metric in classification tasks, assesses a model's performance by examining its predictive ability on each class individually, in contrast to accuracy, which evaluates overall performance. Moreover, it combines two competing metrics, precision and recall, into a single value using the harmonic mean (Umer *et al.,* 2020). The formula for the F1 score metrics is as follows:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

The F1 score offers information about a model's performance similar to precision and recall; a high F1 score of 1 denotes exceptional precision and recall, while a low score denotes subpar model performance.

7.2.2.7.4 AUC–ROC

The AUC–ROC (Area Under the Receiver Operating Characteristic Curve) is another fundamental evaluation metric used in ML to assess the performance of

binary classification models. The ROC curve, which stands for Receiver Operating Characteristic curve, graphically depicts the performance of a binary classifier as it operates under different classification thresholds (Nahm, 2022). It depicts the True Positive Rate (TPR), also known as sensitivity or recall, against the False Positive Rate (FPR), representing the proportion of negative instances incorrectly classified as positive by the model. The AUC–ROC quantifies the area under this curve, ranging from 0 to 1, where a value closer to 1 indicates superior model performance, signifying the model's ability to distinguish between positive and negative instances without misclassifications, thus achieving perfect discrimination.

7.2.2.7.5 Precision

Precision is another important evaluation metric to assess the performance of an ML model, particularly in the context of classification tasks (Hicks *et al.,* 2022; Rainio, Teuho and Klén, 2024). It evaluates how accurate positive predictions an ML model makes by counting the proportion of true positive predictions among all positive predictions. Alternatively, precision measures the correctness of the ML model in identifying relevant samples inside the expected positive class. Precision is computed as the ratio of true positive predictions to the entire number of positive predictions, including both true and false positives. The precision formula is shown below:

$$Precision = \frac{True\ positives}{True\ positives + False\ positives}$$

A precision rating of one indicates that it is extremely likely to be correct when the model identifies something as positive. Furthermore, it shows that all positive predictions produced by the model are right, with no false positives, implying that every case classified as positive by the model is unquestionably positive. In contrast, a value of 0 shows that none of the model predictions are accurate, meaning that everything predicted as positive is negative, resulting in a lack of positive projections.

7.3 Results

7.3.1 Data Pre–processing

The most significant transcriptomic and variant biomarkers were identified through the different types of bioinformatics analysis. Then, it was combined to process a multi–stage feature selection process. Table 21 Provides a comprehensive overview of the unique transcriptomic and variant signatures identified in cfRNA datasets across GW 8 to 16. These signatures are categorised into UDEGS, UCEGS, and UVS. The total number of signatures shows an increasing trend from week 8, peaks during weeks 12 and 13, and declines by week 16. Week 13 records the highest total of 6,709 signatures, primarily driven by a significant increase in UDEGS (1,873), reflecting a critical period of transcriptomic changes. UCEGS consistently contribute the most signatures, with a notable peak at week 12 (3,699), indicating stable gene expression during mid–pregnancy. UVS, representing gestational age–specific genetic variations, are particularly abundant during weeks 11 to 13, coinciding with key developmental transitions. The dataset comprises 26,600 unique signatures, with the majority attributed to UCEGS (10,852) and UVS (13,776). These findings underscore the dynamic nature of gene expression and genetic variation during pregnancy and provide a foundation for building predictive models to identify pregnancy outcomes such as PE.

Table 21: Number of unique transcriptomic and variant signatures identified per gestational week.

The table includes UDEGS, UCEGS, and UVS and their total counts across gestational weeks 8 to 16.

Gestational Weeks	UDEGS	UCEGS	UVS	Total
8	3	296	995	1294
9	0	717	645	1362
10	1	893	1180	2074
11	33	1675	2738	4446
12	50	3699	1830	5579
13	1873	1583	3253	6709
14	0	330	370	700
15	5	339	1001	1345
16	7	1320	1764	3091
Total	1972	10852	13776	**26600**

The overlap analysis between UDEGS and UCEGS led to the identification of Consistently Differentially Expressed Gene Signatures (CDEGS), and gene expression patterns were reliably observed across multiple datasets when comparing different biological conditions. Identifying these consistent signatures is crucial for understanding fundamental biological processes and developing robust biomarkers. In this study, 577 CDEGS were identified as intersecting between UDEGS and UCEGS. These findings highlight the potential of these 577 CDEGS in advancing the identification of transcriptomic biomarkers for developing a predictive model for PE. However, further investigation is necessary to validate and explore these findings in greater depth.

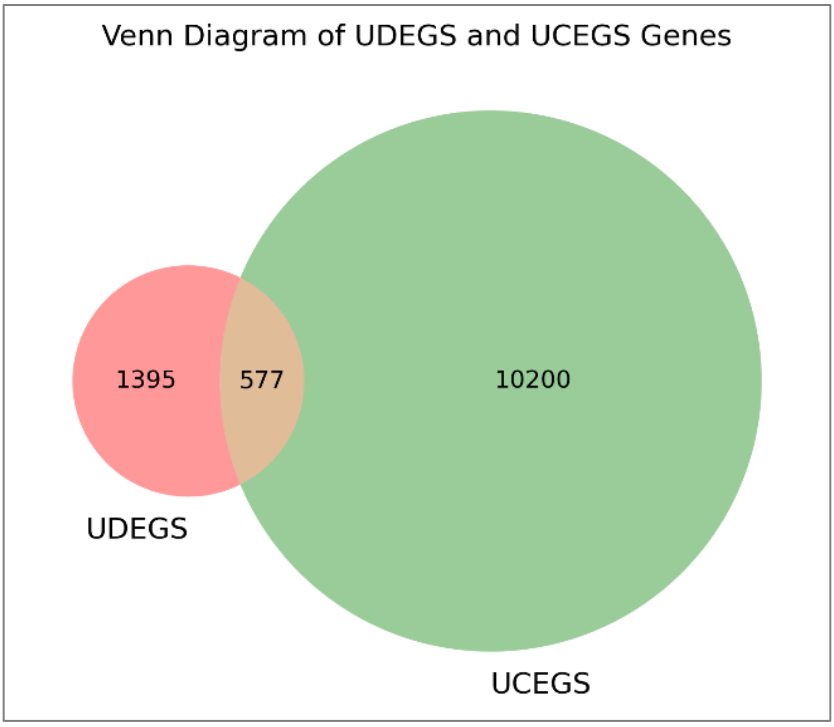


Figure 67: Venn Diagram showing the overlap between UDEGS and UCEGS genes. The diagram highlights 577 genes common to both UDEGS and UCEGS, with 1395 unique genes in the UDEGS set and 10200 unique genes in the UCEGS set.

7.3.1.1 Feature Selection

Table 22 summarises the multi–stage feature selection process and ML model evaluation for identifying GW–specific transcriptomic and variant signatures. The combined UDEGS and UCEGS datasets were analysed for transcriptomic feature selection, while the UVS dataset was used for variant feature selection on a per–gestational week basis. The performance of different ML algorithms, including LDA, LR, GBM, SVM, RF, and DT, was evaluated using CV accuracy. The selected features varied across GW 8 to 16, reflecting dynamic biological variations in gene expression and genetic variants during early pregnancy. For Unique Transcriptomic Signatures (UTS), the highest number of selected features was observed in GW 12 (1,738), with DT achieving a mean CV accuracy of 0.859. Similarly, GW 13 had 666 selected features, with RF attaining an accuracy of 0.862. In contrast, specific weeks, such as GW 8 and GW 14, had minimal selected

transcriptomic features (1 and 4, respectively), with GBM and RF achieving accuracies of 0.949 and 0.746, respectively. The variation in feature count and accuracy suggests that some gestational weeks contain stronger predictive signatures than others.

For variant signatures, the highest number of selected features was observed in GW 11 (2,677), with Decision Tree (DT) achieving an accuracy of 0.727, while GW 10 had 914 features, with DT obtaining 0.795 accuracy. In contrast, GWs 8, 9, 14, and 16 had only one selected variant feature yet achieved varying accuracy scores (0.756, 0.774, 0.750, and 0.693, respectively) with different ML models. The number of selected features and accuracy fluctuations across GW highlight the dynamic nature of molecular changes in early pregnancy. These results indicate that transcriptomic and variant signatures exhibit GW–specific predictive potential, reinforcing the need for tailored feature selection strategies in disease modelling. The variability in ML performance suggests that some weeks provide more potent predictive biomarkers than others, critical for improving early PE detection using ML approaches.

Table 22: Summary of the ML algorithms and CV approaches used for feature selection in GW–specific datasets.
The table includes details on gene expression and variant data, ML algorithms applied, the number of features selected, and the mean accuracy (CV score) for each GW.

Gestational Weeks	ML algorithms	Cross–Validation	Number of UTS[1]	CV score (Mean accuracy)	ML algorithms	Cross–Validation	Number of UVS[2]	CV score (Mean accuracy)
8	GBM	4	1	0.94877	RF	3	1	0.755556
9	SVM	3	536	0.783796	DT	2	1	0.77381
10	RF	3	1	0.793386	DT	3	914	0.795426
11	RF	11	30	0.746345	DT	11	2677	0.727397
12	DT	12	1738	0.859354	DT	8	1784	0.688283
13	RF	6	666	0.862094	SVM	2	1132	0.683772
14	RF	2	4	0.746269	DT	2	1	0.75
15	DT	6	3	0.859048	DT	4	860	0.768168
16	GBM	6	11	0.861466	DT	7	1	0.692857

1 –UTS– Unique Transcriptomic Signatures 2 –UVS –Unique Variant Signatures

The UTS (2,990) and UVS (7,371) from GW 8 to 16 were combined and pre–processed for an additional feature selection step to prepare the dataset for ML model development. The selected dataset, derived from the SVM cross–validation (7–fold) process with a CV score of 0.773057, includes 785 unique UTS and 354 UVS (Table 23). Then, this dataset was refined by carefully removing unannotated non–coding genes and variants. As a result, the refined dataset consists of 655 UTS and 99 UVS. After this, the dataset was used in the dimensionality reduction process with PCA for feature extraction.

Table 23: Feature Selection Results Using Recursive RFE–CV Across Different ML Algorithms

ML Algorithm	CV	Features	CV Score
SVM	7	1139	0.773057
LR	4	1344	0.766006
RF	15	319	0.723941
LDA	12	19	0.722454
GBM	3	739	0.715046
DT	7	1574	0.660868

7.3.1.2 Pre-eclampsia–Linked Gene Extraction and UTS Intersection

Among the 655 UTS (identified through multi–stage feature selection), 67 were found to intersect with the 2065 previously reported PE–linked genes from Genecard, Disgenet, and the published literature (Figure 68). This overlap suggests that these 67 UTS may play a significant role in the genetic basis of PE. The intersection of the UTS with the PE–linked gene list further supports the relevance of these features for future research and potential diagnostic or therapeutic applications related to PE.

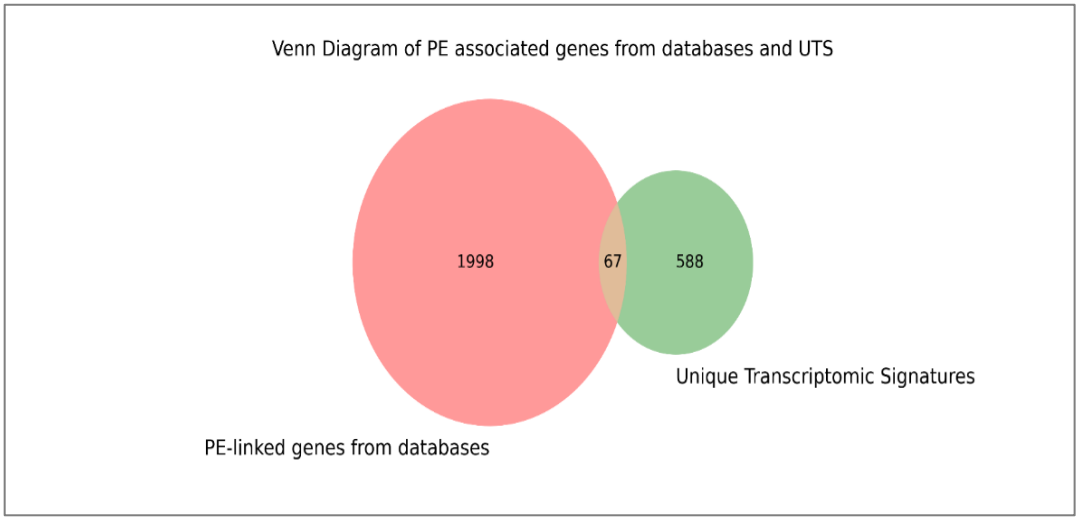


Figure 68: Overlap of PE-associated Genes from Databases and UTS

The analysis reveals that 67 genes are shared between the two groups, with 1998 genes uniquely identified from databases and 588 genes unique to the UTS dataset.

Among the 67 intersecting UTS, representing the overlap between PE–linked genes from publicly available databases and UTS obtained through the multi–stage feature selection process, 20 UTS were identified within the UDEGS category via overlap analysis (Table 24). Nine were also present within these 20 UDEGS, and nine were in the CDEGS sets, indicating their shared significance in differential and consistent gene expression profiles. The remaining 11 UDEGS were found to be exclusive to this category, suggesting their unique relevance in the context of unidentified differentially expressed genes. In contrast, the remaining 47 UTS were identified within the UCEGS group (Table 25), highlighting their distinct role in the unidentified, consistently expressed genes (Figure 69).

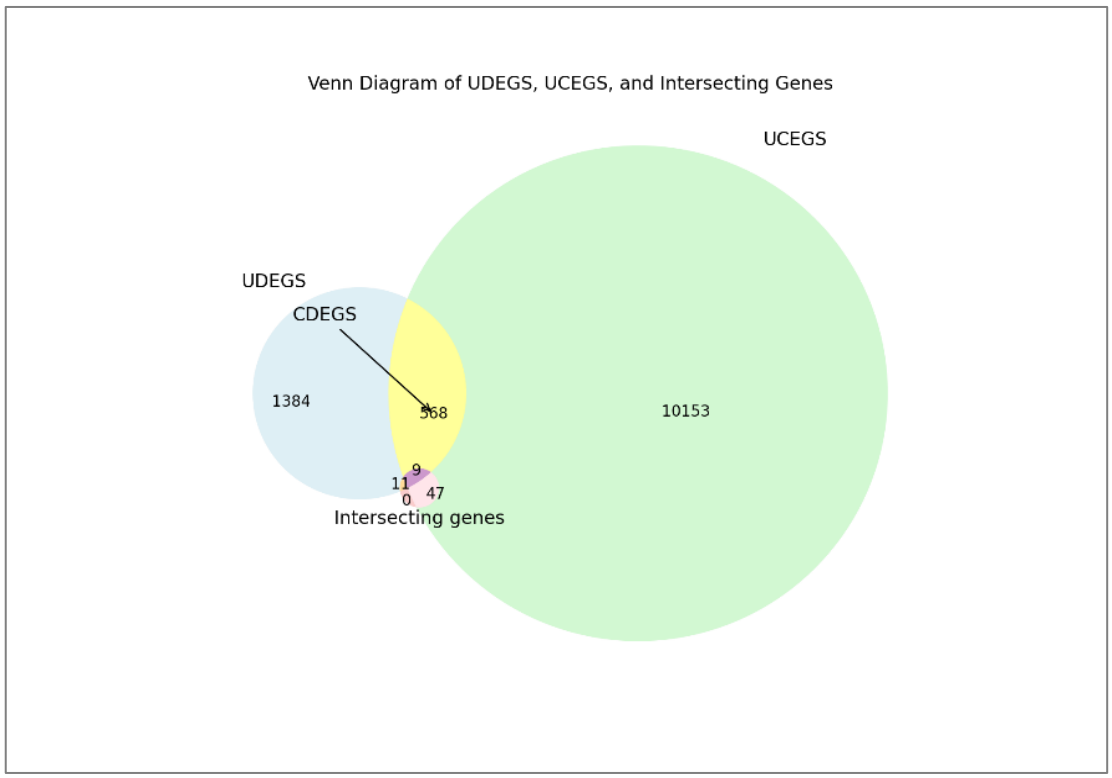


Figure 69: Distribution of intersecting UTS between PE–linked genes from databases and UTS identified in UDEGS, UCEGS, and CDEGS.

Table 24: Key statistical metrics of 20 intersecting genes between PE–linked genes from databases and UTS, including gene symbols, expression categories, base mean expression, log2 fold change, standard error (lfcSE), test statistic (stat), p–value, and adjusted p–value (padj).

No	Gene Symbol	Type	baseMean	log2FoldChange	lfcSE	stat	p–value	padj
1	SETD1A	CDEGS	10.65072	3.530349	1.020987	3.457781	0.000545	0.010168
2	IKBKB	CDEGS	11.70066	3.800923	1.332371	2.852751	0.004334	0.0361
3	PEPD	CDEGS	5.018012	5.360271	1.517639	3.53198	0.000412	0.008436
4	GDF15	CDEGS	1.926707	3.51187	1.301919	2.697456	0.006987	0.047747
5	TLR2	CDEGS	3.169479	4.469616	1.453868	3.074292	0.00211	0.023519
6	PDX1	CDEGS	30.27114	4.335096	1.423609	3.045146	0.002326	0.024805
7	PARP1	CDEGS	22.26868	3.245204	0.781433	4.15289	3.28E–05	0.001699
8	GPT2	CDEGS	2.828495	4.307902	1.583897	2.719811	0.006532	0.045891
9	SMAD3	CDEGS	16.41456	3.216609	0.877115	3.66726	0.000245	0.00604
10	HMGA1	UDEGS	13.14271	2.616767	0.928348	2.818735	0.004821	0.038738
11	ARHGDIA	UDEGS	18.73667	2.381136	0.822675	2.894381	0.003799	0.032868
12	EPHA1	UDEGS	51.88889	3.31037	1.063388	3.113042	0.001852	0.021425
13	RETREG1	UDEGS	4.245687	5.127342	1.813963	2.826597	0.004705	0.038029
14	XPC	UDEGS	9.811823	3.644554	0.972391	3.748035	0.000178	0.005003
15	APOA2	UDEGS	4.43132	5.188315	1.388097	3.737718	0.000186	0.005091
16	ITGB2	UDEGS	71.5974	3.403597	0.757271	4.494556	6.97E–06	0.000624
17	ABCA1	UDEGS	6.954211	4.018999	1.221494	3.290232	0.001001	0.014258
18	CYBB	UDEGS	18.19058	2.513443	0.857134	2.93238	0.003364	0.0309
19	STAT6	UDEGS	80.36779	2.795444	0.787116	3.551501	0.000383	0.008011
20	LGALS9	UDEGS	53.06131	1.707708	0.624876	2.732873	0.006278	0.044893

Table 25: Key statistical metrics of 47 intersecting genes between PE–linked genes from databases and UTS in UCEGS, including gene symbols, expression categories, base mean expression, log2 fold change, standard error (lfcSE), test statistic (stat), p–value, and adjusted p–value (padj).

No	GeneSymbol	RP value	Mean Expression	std_dev	CV	P value
1	TBXAS1	4512.454	12.97989	12.11125	0.933078	0.013986
2	ACSL4	2707.847	28.2636	25.83136	0.913944	0.000999
3	COL4A4	2.03E+50	0.79498	0.113013	0.142158	0.018981
4	IL5RA	1.38E+91	1.202169	0.987729	0.821622	0.000999
5	CD200	5.79E+39	0.663878	0.473238	0.712838	0.042957
6	BMP7	7.79E+49	0.793754	0.125916	0.158633	0.027972
7	RGCC	3958.826	17.7035	23.74542	1.341284	0.007992
8	KCNN4	2.36E+91	1.413509	1.303966	0.922502	0.001998
9	RETN	7.28E+49	0.935697	0.743397	0.794484	0.024975
10	HDAC5	3702.699	16.44961	14.92929	0.907577	0.005994
11	STRN	3028.662	31.98156	47.86977	1.496793	0.001998
12	RPS6KA1	9.71E+39	0.502982	0.248038	0.493135	0.025974
13	TEK	2.33E+89	1.226028	1.006653	0.821069	0.048951
14	TNFRSF8	5.46E+90	2.658046	2.592568	0.975366	0.037962
15	MTRR	3.17E+11	1.205665	0.504208	0.418199	0.01998
16	ID1	1.48E+89	1.464423	1.270257	0.867412	0.03996
17	H19	8.68E+39	0.548254	0.319692	0.58311	0.038961
18	COL5A1	4.05E+49	1.132428	1.062524	0.93827	0.041958
19	BPIFA3	5.05E+49	0.777457	0.221614	0.28505	0.046953
20	F12	4.01E+49	0.795562	0.325099	0.40864	0.035964
21	TRIM21	3.78E+47	1.112612	0.877277	0.788484	0.017982
22	PCNA	2.50E+40	0.529517	0.306334	0.578517	0.005994
23	NES	4.98E+39	0.60107	0.332979	0.553977	0.045954
24	SLC38A2	3557.201	21.13339	27.00372	1.277775	0.002997
25	PUM1	3.01E+11	1.221511	0.483815	0.396079	0.036963
26	MDM2	4522.447	12.58148	11.00989	0.875087	0.016983
27	FLNB	5177.016	7.316757	6.864858	0.938238	0.03996
28	HADHB	3081.548	22.51796	16.74507	0.743632	0.000999
29	ENPEP	1.04E+50	0.757977	0.323127	0.426302	0.01998
30	FTO	2.86E+11	1.31728	1.009775	0.766561	0.041958
31	EPHA2	1.51E+50	0.956447	0.631225	0.659969	0.023976
32	REN	2.46E+90	1.2072	0.995811	0.824894	0.008991
33	GSN	3180.425	22.18984	19.62659	0.884486	0
34	DGKE	3.12E+11	1.286205	0.625323	0.486177	0.040959
35	SLC7A7	2.97E+11	1.187924	0.386063	0.32499	0.02997
36	AIMP1	5038.317	13.21429	18.08366	1.368493	0.032967
37	ANXA5	4.08E+11	0.677784	0.070511	0.104031	0.00999
38	PRDX3	1.06E+40	0.53357	0.32588	0.610754	0.015984
39	FGA	6.13E+47	0.962905	0.725761	0.75372	0.005994
40	ABO	3063.93	21.47788	16.60208	0.772985	0
41	ERBB4	9.46E+48	0.849291	0.632993	0.745319	0.001998
42	SOCS3	5.81E+88	2.105569	2.080713	0.988195	0.04995
43	SDHD	9.69E+48	0.784626	0.470849	0.600094	0.004995
44	**FLT1P1**	5.12E+86	1.379828	1.095425	0.793885	0.013986
45	AGAP2–AS1	1.63E+48	1.134214	0.739475	0.651971	0.00999
46	LINC00261	2.86E+92	1.220193	0.996696	0.816835	0.000999
47	EGLN2	4.27E+50	1.004326	0.67527	0.672361	0.036963

7.3.1.3 Dimensionality Reduction

The combined dataset, consisting of 754 features, underwent dimensionality reduction using PCA to identify the optimal number of components for classification. A systematic evaluation was conducted by testing principal components ranging from 1 to 100 and performing cross–validation with fold values between 2 and 10. The results revealed that the optimal configuration consisted of 22 principal components with 9–fold cross–validation, achieving the highest cross–validation accuracy of 0.7814 (Figure 70). This selection of 22 components effectively captured the essential variance within the dataset while minimising noise and redundancy, leading to improved classification efficiency. The observed accuracy indicated that incorporating additional components beyond this threshold did not yield significant performance gains. By reducing the dimensionality, PCA ensured that the most relevant features contributed to the ML model while preserving critical information for classification. These findings confirm the effectiveness of PCA in enhancing model interpretability and optimising predictive performance.

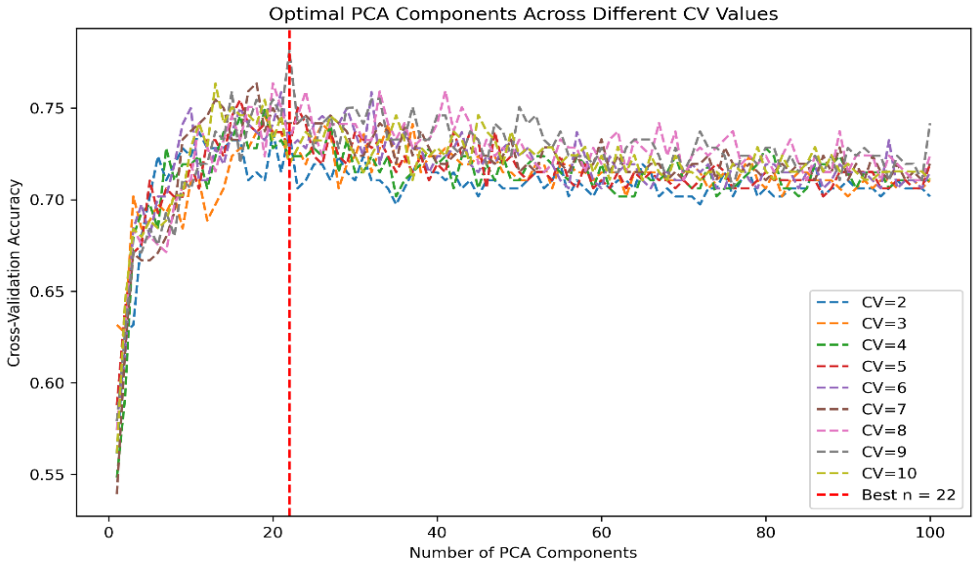


Figure 70: Determination of the optimal number of PCA components across different CV fold values.

The graph depicts the relationship between the number of PCA components and cross–validation accuracy for CV values ranging from 2 to 10. The red dashed vertical line indicates the optimal number of PCA components (n = 22), which achieved the highest cross–validation accuracy, ensuring a balance between dimensionality reduction and model performance.

7.3.2 ML–Model Development

Following dimensionality reduction using PCA with 22 components, several ML algorithms were utilised to develop predictive models for PE. The selected algorithms—SVM, DT, RF, LR, LDA, and GBM—were trained and evaluated to determine their effectiveness in classifying PE cases. The dataset was split into training and testing sets using a 70–30 ratio with stratified sampling to ensure a balanced representation of both classes. Each model's performance was assessed using various evaluation metrics, including accuracy, precision, recall, and F1–score. The performance metrics for each model are summarised in the Table 26 and visualised in (Figure 71).

Table 26: Performance evaluation of various ML models for disease prediction.
The table presents accuracy, precision, recall, and F1–score for each model, highlighting the SVM model as the top–performing classifier with the highest accuracy (92.75%) and a balanced trade–off between precision (89.47%) and recall (85%). Logistic Regression also achieved competitive performance, whereas the Decision Tree model demonstrated the lowest effectiveness across all metrics.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1–score (%)
SVM	92.75362	89.47368	85	87.17949
Decision Tree	66.66667	42.85714	45	43.90244
Random Forest	81.15942	70.58824	60	64.86486
Logistic Regression	89.85507	78.26087	90	83.72093
LDA	84.05797	84.61538	55	66.66667
Gradient Boosting	75.36232	60	45	51.42857

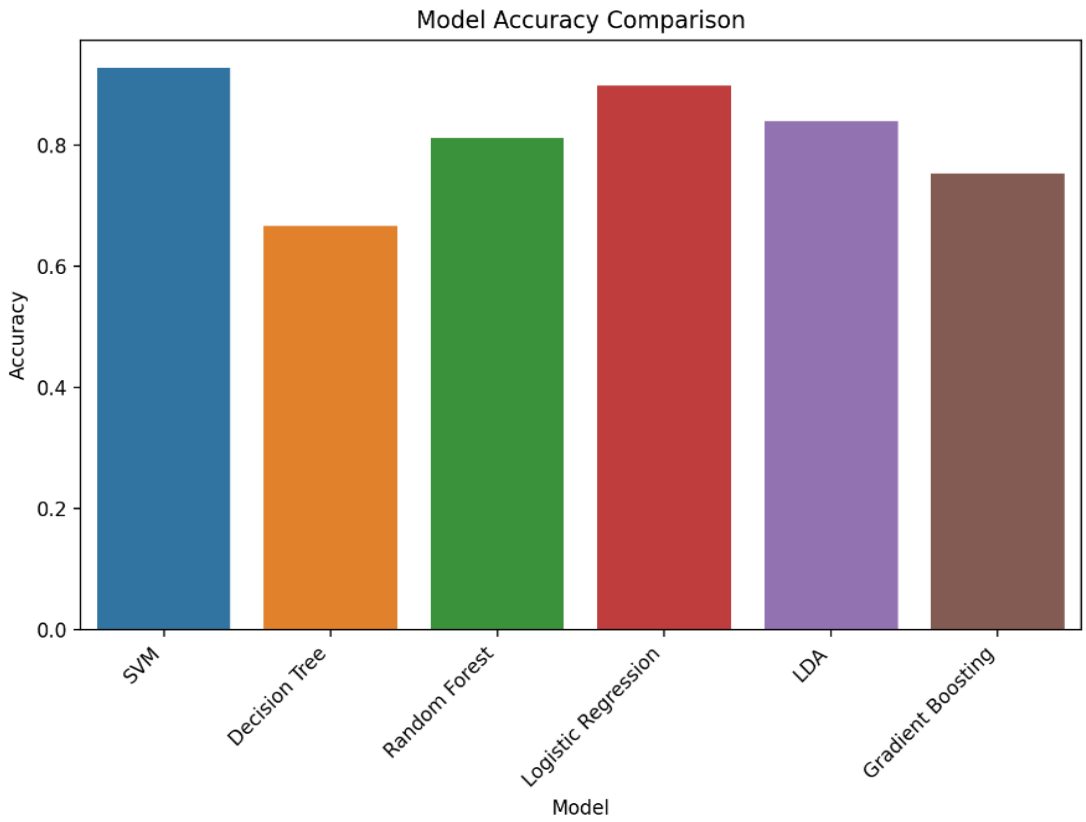


Figure 71: Model Accuracy Comparison.
The bar chart illustrates the accuracy of different ML models, including SVM, DT, RF, LR, LDA, and GBM. Among the models, SVM and LR achieved the highest accuracy, while DT exhibited the lowest performance.

The results indicate that the SVM model achieved the highest accuracy (92.75%) with intense precision (89.47%) and recall (85.00%), making it the most effective classifier for this dataset. LR followed closely with an accuracy of 89.86%, showing a higher recall (90.00%) than SVM but slightly lower precision (78.26%), suggesting it may better identify true positive cases. Random Forest also performed well, achieving an accuracy of 81.16% with a balanced precision–recall trade–off. In contrast, the Decision Tree model exhibited the lowest accuracy (66.67%), indicating a tendency to overfit or underperform in complex feature spaces. LDA performed moderately, with an accuracy of 84.06% but a relatively lower recall of 55.00%, suggesting its limitations in capturing complex relationships. Gradient Boosting, while often a strong performer in other contexts,

achieved the lowest recall (45.00%), indicating challenges in identifying positive cases effectively.

The ROC curves for all models were plotted to visualise their discriminatory power, with the SVM and LR models achieving the highest AUC values of 0.962 and 0.952, respectively, indicating superior performance in distinguishing between positive and negative cases (Figure 72). LDA also demonstrated strong discriminatory capability with an AUC of 0.955, reinforcing its potential as an effective predictive model. In contrast, ensemble models such as RF and Gradient Boosting achieved moderate AUC values of 0.814 and 0.864, respectively, suggesting their capability to handle complex data patterns but with slightly lower discrimination ability than SVM and LDA. The DT model exhibited the lowest performance across all metrics, with an AUC of 0.614 and an accuracy of 0.667, highlighting its limitations in this predictive task.

The evaluation results were compiled and saved for further comparative analysis, providing insights into the strengths and limitations of each approach. Overall, the findings suggest that SVM and LR models are particularly well–suited for developing robust predictive models for PE, achieving high accuracy (92.75% and 89.86%, respectively) and strong discriminatory power. Integrating dimensionality reduction techniques such as PCA enhanced model performance by reducing noise and improving feature interpretability, ultimately contributing to better prediction accuracy and reliability.

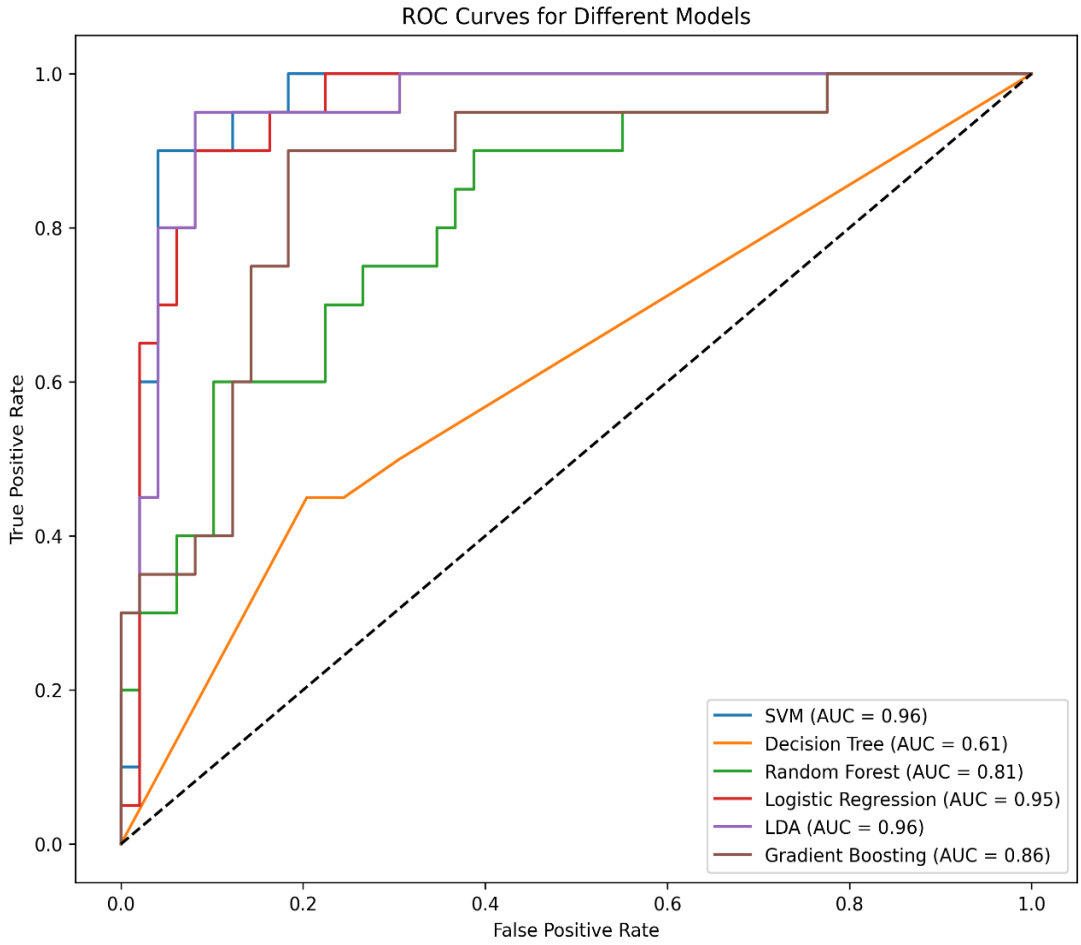


Figure 72: Receiver Operating Characteristic (ROC) Curves for Different ML Models. The ROC curves illustrate the true positive rate against the false positive rate for various classification models, including SVM, DT, RF, LR, LDA, and GBM. The Area Under the Curve (AUC) values indicate model performance, with SVM and LDA achieving the highest AUC (0.96), followed by LR (0.95), GBM (0.86), and RF (0.81). The DT model exhibited the lowest AUC (0.61), suggesting lower discriminative ability.

7.4 Discussion

The findings from this study underscore the dynamic nature of transcriptomic and variant biomarkers in early pregnancy and their potential application in ML–based predictive models for PE. The multi–stage feature selection process effectively identified GW–specific transcriptomic and variant signatures, demonstrating significant variability across weeks (Eidem *et al.*, 2015; Wong *et al.*, 2022). The increasing trend of unique signatures from GW 8 to GW 13, followed by a decline by GW 16, suggests a critical period of molecular changes during mid–gestation,

aligning with key developmental transitions. The peak number of transcriptomic signatures observed at GW 12 (1,738) and GW 13 (666) indicates heightened transcriptional activity (Uusküla et al., 2012; Tan et al., 2018; Ng et al., 2024). In contrast, the highest number of variant signatures identified at GW 11 (2,677) suggests a substantial contribution of genetic variations to pregnancy outcomes during this period.

The development of this automated prediction model, while demonstrating high accuracy, also brings to the fore the challenges outlined in the introduction. Firstly, the 'black-box' nature of the top-performing SVM model presents a barrier to clinical trust. Future work will employ explainable AI (XAI) techniques, such as SHAP analysis, to elucidate the contribution of key genomic features, thereby linking model predictions to biologically plausible pathways in PE pathogenesis (Chhatbar, Bird and Sanguinetti, 2025). Secondly, while this study focused on genomic biomarkers, a truly multi-dimensional approach for clinical deployment would involve integrating this model with established clinical risk factors into a unified risk score. Finally, the reliance on cfRNA sequencing, while powerful for discovery, presents a practicality hurdle. The transition to a targeted, cost-effective PCR panel measuring the top biomarkers identified here is a critical next step for feasible integration into routine prenatal screening pathways (Ronzoni *et al.*, 2025).

7.4.1 Significance of Feature Selection in Different Gestational Weeks

The feature selection results revealed that different ML algorithms performed variably across gestational weeks, with DT achieving relatively lower CV accuracy for variant signatures. In contrast, SVM, RF, and GBM showed higher performance in transcriptomic feature selection. The observed variations in feature selection effectiveness across weeks highlight the necessity of a tailored approach for feature selection, as certain weeks contain stronger predictive signatures than others (Schmidt et al., 2022a; Zheng et al., 2022; T. Li et al., 2024). Additionally, the final dataset refinement process, which excluded unannotated non–coding

genes and variants, further optimised the feature set, reducing noise and improving classification performance.

The effectiveness of GW–specific feature selection is significant, as it allows for developing precise predictive models that account for dynamic molecular changes over time. Early weeks (GW 8–10) showed fewer unique features, potentially due to lower biological variability in early pregnancy, whereas mid–gestation (GW 11–13) exhibited peak feature importance (GOETZINGER *et al.*, 2010; Yadama *et al.*, 2020; Slade *et al.*, 2025). This suggests that mid–gestational molecular patterns may be the most informative early PE prediction period. The decline in unique features beyond GW 13 indicates possible stabilisation of transcriptomic expression, which aligns with the transition from placental development to foetal growth (Prater *et al.*, 2021).

7.4.2 Dimensionality Reduction and Model Optimisation

Dimensionality reduction using PCA ensured that only the most relevant features were retained while minimising redundancy (Jolliffe and Cadima, 2016; Espinosa *et al.*, 2021). The selection of 22 principal components with a 9–fold cross–validation accuracy of 0.7814 reinforced the effectiveness of PCA in capturing essential variance within the dataset. PCA improved classification efficiency by reducing the feature space, preventing overfitting and enhancing model interpretability (Dai *et al.*, 2025).

These findings underscore the importance of dimensionality reduction techniques in high–dimensional biological datasets, where raw transcriptomic and variant data often contain excessive noise (Meng *et al.*, 2016; W. Liu *et al.*, 2022). Feature compression via PCA retained only the most influential gene expression signatures, eliminating non–informative variations (Bhadra *et al.*, 2022; Haji, 2024). The trade–off between feature retention and classification accuracy highlights the necessity of balancing data complexity with model performance, particularly in predictive modelling for PE.

7.4.3 Machine Learning Model Performance

The ML model evaluation demonstrated that SVM emerged as the top-performing classifier, achieving the highest accuracy (92.75%), precision (89.47%), and recall (85.00%) (Table 25). These results suggest that SVM effectively captured complex relationships within the dataset, making it the most reliable model for PE prediction. Logistic Regression (LR) also exhibited strong predictive capability, with an accuracy of 89.86% and the highest recall (90.00%), indicating its robustness in identifying true positive cases. RF performed moderately well, achieving an accuracy of 81.16%, demonstrating its effectiveness in handling non–linear relationships. However, DT exhibited the weakest performance, with an accuracy of 66.67% and an Area Under the Curve (AUC) of 0.61, suggesting overfitting or limited discriminative ability in this dataset. Despite being an ensemble model, Gradient Boosting (GBM) showed suboptimal performance compared to SVM and LR, indicating potential challenges in capturing subtle patterns in high–dimensional data.

The ROC analysis further validated these findings, with SVM and Linear Discriminant Analysis (LDA) achieving the highest AUC values (0.96), followed by LR (0.95), RF (0.81), and GBM (0.86). These results emphasise the superior classification ability of SVM and LDA, reinforcing their suitability for predictive modelling in early pregnancy. The low AUC of the Decision Tree (0.61) further corroborates its weaker predictive performance, suggesting its limited applicability for PE classification in complex datasets.

7.4.4 Implications for Early Pre-eclampsia Detection

These findings have significant implications for early PE detection. They underscore the importance of integrating transcriptomic and variant signatures in ML–based disease prediction models, particularly for conditions like PE, where early detection is crucial for maternal and foetal health (Marić *et al.*, 2022). The superior performance of SVM and LR suggests that linear models with well–defined decision boundaries are particularly effective in distinguishing PE cases,

possibly due to the structured nature of transcriptomic and variant data. Moreover, the observed gestational week–specific variations in feature selection emphasise the need for dynamic models that can adapt to temporal changes in molecular signatures. Traditional static models may overlook critical variations occurring at specific gestational weeks (Ranjbar *et al.*, 2024). This study advocates for adaptive ML frameworks that continuously refine predictive signatures based on evolving pregnancy dynamics.

7.4.5 Future Directions

For future research, incorporating additional biological datasets, such as proteomics, metabolomics, and epigenomic profiles, could further refine predictive accuracy and enhance model generalisability (Giesen, Verlohren and Blois, 2025). Additionally, exploring deep learning approaches, particularly convolutional and recurrent neural networks, may provide deeper insights into complex temporal patterns in gene expression. Multi–omics integration will likely improve PE prediction, providing a holistic understanding of the molecular mechanisms underlying disease onset (Wei et al., 2025).

Another crucial direction involves clinical validation of the proposed predictive models (Montgomery-Csobán *et al.*, 2024). External validation using independent patient cohorts and real–world clinical trials is necessary to assess the practical applicability of these models (Dathan-Stumpf *et al.*, 2022). Furthermore, developing user–friendly clinical decision support tools, such as risk stratification software based on ML models, could facilitate real–time diagnosis in obstetric settings. This underscores the need for continued research and validation in the field, as it is through these efforts that meaningful progress can be made in maternal-foetal health.

Identifying 577 CDEGS (Figure 67) provides a foundation for advancing transcriptomic biomarker discovery in PE. Future research should focus on validating these genes through independent datasets and functional assays, exploring their molecular pathways, and integrating multi–omics data to enhance

predictive accuracy. ML models incorporating these signatures could aid early diagnosis, while clinical trials are needed to assess their real–world applicability. Additionally, postdoctoral research for tracking gene expression changes in high–risk pregnancies will further refine their predictive value in the longitudinal studies. These efforts could lead to reliable, non–invasive diagnostic tools, ultimately improving maternal and foetal health outcomes.

A key limitation of the current model, and a primary avenue for future work, relates back to the extensive clinical risk factors detailed in the Chapter 2 literature review (e.g., maternal age, BMI, prior history of PE, and diabetes). This thesis focused on establishing the predictive power of novel genomic biomarkers (UDEGs, UCEGs, and UVS) in isolation. However, a truly robust clinical tool would undoubtedly integrate these genomic signatures with the established clinical risk factors (Chaemsaithong, Sahota and Poon, 2022). Future iterations of this AI model should be trained on a multi-dimensional dataset, combining both molecular and clinical data. This integrated approach is expected to significantly enhance predictive accuracy and generalisability, creating a more comprehensive and clinically applicable risk assessment tool.

Finally, ethical considerations surrounding data privacy and interpretability must be addressed (Coghlan, Gyngell and Vears, 2024). Given the complexity of genomic datasets, ensuring transparent decision–making and explainable AI models is crucial for clinical adoption. Future model deployment should prioritise regulatory compliance, patient consent, and robust data security mechanisms.

7.5 Conclusion

This study successfully demonstrated the feasibility of using ML models for predicting PE based on GW–specific transcriptomic and variant signatures. The results highlight SVM as the most effective classifier, followed by LR and RF, while DT exhibited the weakest performance. Applying feature selection and dimensionality reduction techniques significantly enhanced classification accuracy, emphasising the importance of careful data preprocessing in ML–based

disease prediction. These findings provide a strong foundation for developing personalised, data–driven approaches for early PE detection, ultimately improving maternal–foetal healthcare outcomes.

The transition of the developed AI-assisted prediction model from a research tool to a clinically applicable test requires a defined pathway for translation. The immediate next step involves prospective clinical validation in large, multi-centre cohorts to confirm the model's performance with new, prospectively collected samples. Concurrently, the complex RNA-seq-based assay must be simplified into a targeted, cost-effective diagnostic panel (e.g., qRT-PCR) for high-throughput clinical use. This panel would quantify the most robust genomic biomarkers (UDEGs, UCEGs, UVS) identified in this work.

Following successful validation, the test would require regulatory approval as an in vitro diagnostic. Finally, integration into routine prenatal care would involve establishing a standardised clinical workflow: maternal blood draw at 8-13 weeks, processing in a certified laboratory, and automated analysis by the AI model. The output would be a personalised risk report within the electronic health record, directly informing clinical decisions on prophylactic interventions like aspirin therapy, thereby enabling a shift from late diagnosis to early, targeted prevention. By leveraging ML–driven precision medicine, this study paves the way for early, non–invasive, and highly accurate PE prediction models, potentially revolutionising obstetric care and improving pregnancy management worldwide.

7.6 Summary of the Chapter

This study aimed to develop an AI-assisted prediction model for the early identification of PE in pregnant women by integrating UTS and UVS. GWs 8 to 16 data were used each week as a separate dataset to uncover temporal variations in gene expression. A multi–stage feature selection process, including RFECV, identified 655 UTS and 99 UVS, with 67 overlapping previously reported PE–linked genes.

Dimensionality reduction using PCA optimised the dataset to 22 components, ensuring efficient data analysis without compromising predictive accuracy. Six ML algorithms were evaluated: SVM, LR, RF, DT, GBM, and LDA. The SVM model demonstrated the highest accuracy (92.75%), followed by LR (89.86%) and RF (81.16%), while DT showed the lowest performance (66.67%).

The study highlighted mid–gestation (GWs 11–13) as a critical period for identifying robust predictive biomarkers, with significant molecular changes aligning with key developmental transitions. Integrating gestational week–specific transcriptomic and variant data enhanced the model's ability to predict PE, emphasising the importance of dynamic molecular profiling.

These findings underscore the potential of ML in precision medicine, offering a robust framework for non–invasive, early detection of PE. Future research should incorporate additional multi–omics datasets and conduct clinical validation to enhance model generalisability and practical application in obstetrics. This study paves the way for early interventions, improving maternal and foetal health outcomes.

Key Findings of the Chapter 7:

- Development of a novel AI-assisted prediction model for early identification of PE using unique transcriptomic signatures (UTS) and variant signatures (UVS).
- Implementation of multi–stage feature selection and recursive feature elimination with cross–validation (RFECV) to identify significant biomarkers across GWs 8 to 16.
- Discovery of 655 UTS and 99 UVS, refined to 67 intersecting genes with known PE–linked genes, highlighting their potential diagnostic relevance.
- Integration of dimensionality reduction using PCA, retaining 22 principal components to optimize model performance and classification accuracy.

- Superior performance of the Support Vector Machine (SVM) model, achieving the highest accuracy (92.75%) and precision (89.47%) among ML algorithms tested.
- The findings emphasize gestational week–specific variations in molecular signatures, with mid–gestation (GW 11–13) showing heightened predictive potential.

Chapter 8
Overall Discussion

8.1 Introduction

The overall discussion chapter serves to integrate and contextualise the major findings derived from Chapters 4 through 7 of this thesis, highlighting their collective significance in the development of an AI–driven model for the early detection of PE. Pre-eclampsia remains a significant global health concern, contributing extensively to maternal and perinatal morbidity and mortality (Dimitriadis, Rolnik, Zhou, Estrada-Gutierrez, Koga, Francisco, Whitehead, Hyett, Costa, *et al.*, 2023). Despite numerous clinical and molecular studies, early detection and prediction of PE remain suboptimal due to the heterogeneity of the disease and the lack of robust biomarkers (Than *et al.*, 2022). This study addresses that critical gap by employing a novel bioinformatics and AI framework that leverages multi-omic cfRNA data, aiming to improve diagnostic accuracy during the early gestational period.

This chapter critically evaluates the methodology, outcomes, and implications of the work conducted in previous chapters, placing the findings within the broader context of current literature. The discussion is structured to reflect the progression of the study – from the identification of unique transcriptomic signatures and consistently expressed genes to the discovery of unique genetic variants and the development of a machine learning model. It further examines the translational potential of these findings and outlines future directions for research and clinical implementation. Through this integrated approach, the study contributes a novel and scalable solution to one of obstetrics' most pressing challenges.

8.2 Overall Discussion

The research presented in this thesis marks a significant advancement in the early detection of PE, a complex and multifactorial pregnancy disorder that remains a

major cause of maternal and foetal morbidity and mortality. The core innovation of this study lies in its systematic exploration of cell-free RNA (cfRNA) transcriptomic and variant data across gestational weeks (GWs) 8 to 16, and the subsequent development of a machine learning (ML)–based predictive model. By identifying unique differentially expressed genes (UDEGs), unique consistent expression signatures (UCEGs), and unique variant signatures (UVS) at each gestational stage, the research constructs a highly granular and temporally sensitive foundation for an AI–driven diagnostic framework. The overall methodology for this study, depicted in Figure 73, details the comprehensive, multi-stage bioinformatics pipeline employed, tracking the workflow from the initial data pre-processing of raw sequencing data to the final identification and integration of unique biomarker signatures for developing the automated PE prediction model. The framework initiates with Data Pre-processing, where raw sequencing (SRA) data from cfRNA samples were acquired from public repositories. This data was converted to FastQ format using the SRAtoolkit and subjected to rigorous quality control with FastQC and MultiQC. High-quality reads were subsequently aligned to the human reference genome (hg38) using STAR to produce the Bam files that serve as the foundational dataset for all subsequent analyses.

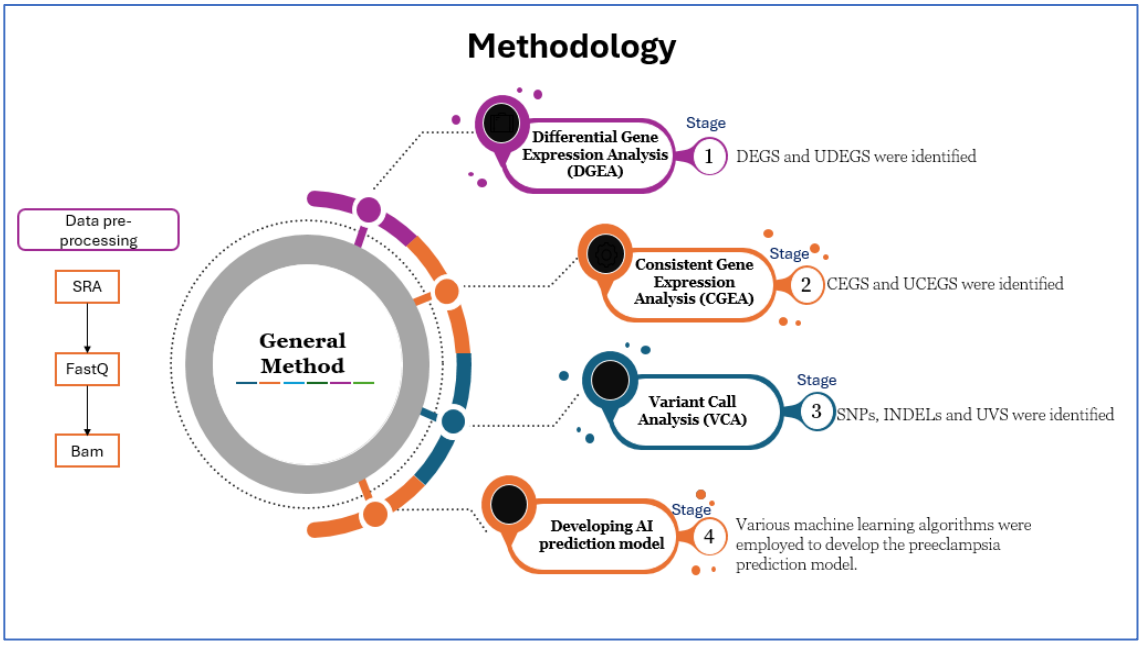


Figure 73: Overview of the general methodology.

This flowchart illustrates the comprehensive four-stage bioinformatics pipeline employed throughout the thesis to develop an automated PE prediction model. The process begins with Data pre-processing, where raw sequencing data (SRA) is converted to FastQ files for quality control and then aligned into Bam files. This processed data feeds into three parallel biomarker discovery pipelines (Stages 1-3): (1) Differential Gene Expression Analysis (DGEA) to identify DEGS and UDEGS, (2) Consistent Gene Expression Analysis (CGEA) to identify CEGS and UCEGS, and (3) Variant Call Analysis (VCA) to identify SNPs, INDELs, and UVS. Finally, the unique biomarker signatures from all three pipelines are integrated as features for (4) Developing the AI prediction model, as detailed in Chapter 7.

8.2.1 Alignment with Existing Literature

The findings of this study are framed within the broader context of existing literature, which has consistently highlighted the challenges of early PE detection due to its multifactorial aetiology and asymptomatic onset (Burton et al., 2019; Riano-Moreno et al., 2025; Dimitriadis et al., 2023). Prior studies have proposed several key mechanisms contributing to PE pathogenesis, including abnormal placentation, immune dysregulation, oxidative stress, and endothelial dysfunction (Mukherjee et al., 2021; Geldenhuys et al., 2018; Torres-Torres et al., 2024; Martini et al., 2025). The literature also emphasises the two-stage model of PE, where impaired trophoblast invasion and shallow spiral artery remodelling in early

pregnancy lead to placental hypoxia and the release of anti-angiogenic factors, ultimately triggering systemic maternal inflammation and endothelial injury (Ahmed and Ramma, 2015; Kornacki et al., 2023; Staff, 2019). These theoretical foundations strongly align with the molecular pathways uncovered in this thesis (Chen *et al.,* 2025).

Additionally, recent literature underscores the inadequacy of current diagnostic approaches, which primarily rely on maternal blood pressure, proteinuria, and clinical symptoms that manifest in the second or third trimester (Rath et al., 2014; Jena et al., 2020). Several biomarkers, including sFlt–1, PlGF, and PP13, have been explored, yet no single biomarker has proven sufficiently sensitive or specific for early prediction (Tal et al., 2010; Harapan and Yeni, 2015; Opichka et al., 2023). This gap in diagnostic capabilities further validates the need for an AI–driven, multi–omic model as developed in this study, which integrates genomics and machine learning to offer a data–rich alternative to conventional methods.

The molecular findings from this study align closely with the established hypotheses outlined in Chapter 2, including oxidative stress, endothelial dysfunction, and immune dysregulation. For instance, the identification of upregulated genes involved in oxidative stress response and angiogenic regulation (e.g., VEGFA, HIF1A, and IFNG) reinforces the theory that placental hypoxia triggers a cascade of inflammatory and vascular events leading to PE (Govender, Moodley and Naicker, 2018; Gerasimova *et al.*, 2019; Kim *et al.*, 2025). Similarly, the presence of variants in genes such as HLA family members supports the immunological dysregulation model of PE pathogenesis, suggesting impaired maternal–foetal immune tolerance. These results provide molecular validation for classical mechanistic models and offer a genomic basis for their integration into AI-driven prediction frameworks (Stepan *et al.*, 2023).

8.2.2 Chapter 4: Temporal Dynamics of UDEGs

Chapter 4 initiated this exploration by identifying UDEGs across each gestational week, providing a temporal map of dynamic molecular changes. Figure 74 visually

summarises the methodology from Chapter 4 that enabled this temporal analysis. The diagram details a two-stage process core to this approach: first, performing Differential Gene Expression Analysis (DGEA) independently on the cfRNA-seq data for each gestational week (GW 8 to 16) to generate weekly lists of all DEGs . Second, an overlap analysis was conducted across all these weekly sets to isolate the UDEGS.

The use of cfRNA provided a non–invasive method to capture placental and foetal gene expression patterns from maternal blood samples, offering a powerful tool for early diagnostics (Harris *et al.*, 2024; Q. Li *et al.*, 2024; Sarker *et al.*, 2025). A standout observation from this chapter was the high variability in UDEG expression across gestational ages, with gestational week 13 displaying a particularly high number of uniquely upregulated genes (Sufriyana *et al.*, 2022; Zheng *et al.*, 2024). This temporal specificity supports the hypothesis that mid–gestation may be a critical window for the development of molecular abnormalities in PE. Functional enrichment analyses showed that these genes were involved in key biological processes such as immune response, oxidative stress, inflammation, and angiogenesis – all of which are well–documented contributors to PE pathogenesis (G. Li *et al.*, 2022; Yu *et al.*, 2025). The overlap between UDEGs and known PE-associated genes further substantiated the biological relevance of the findings and highlighted the potential of UDEGs as predictive biomarkers.

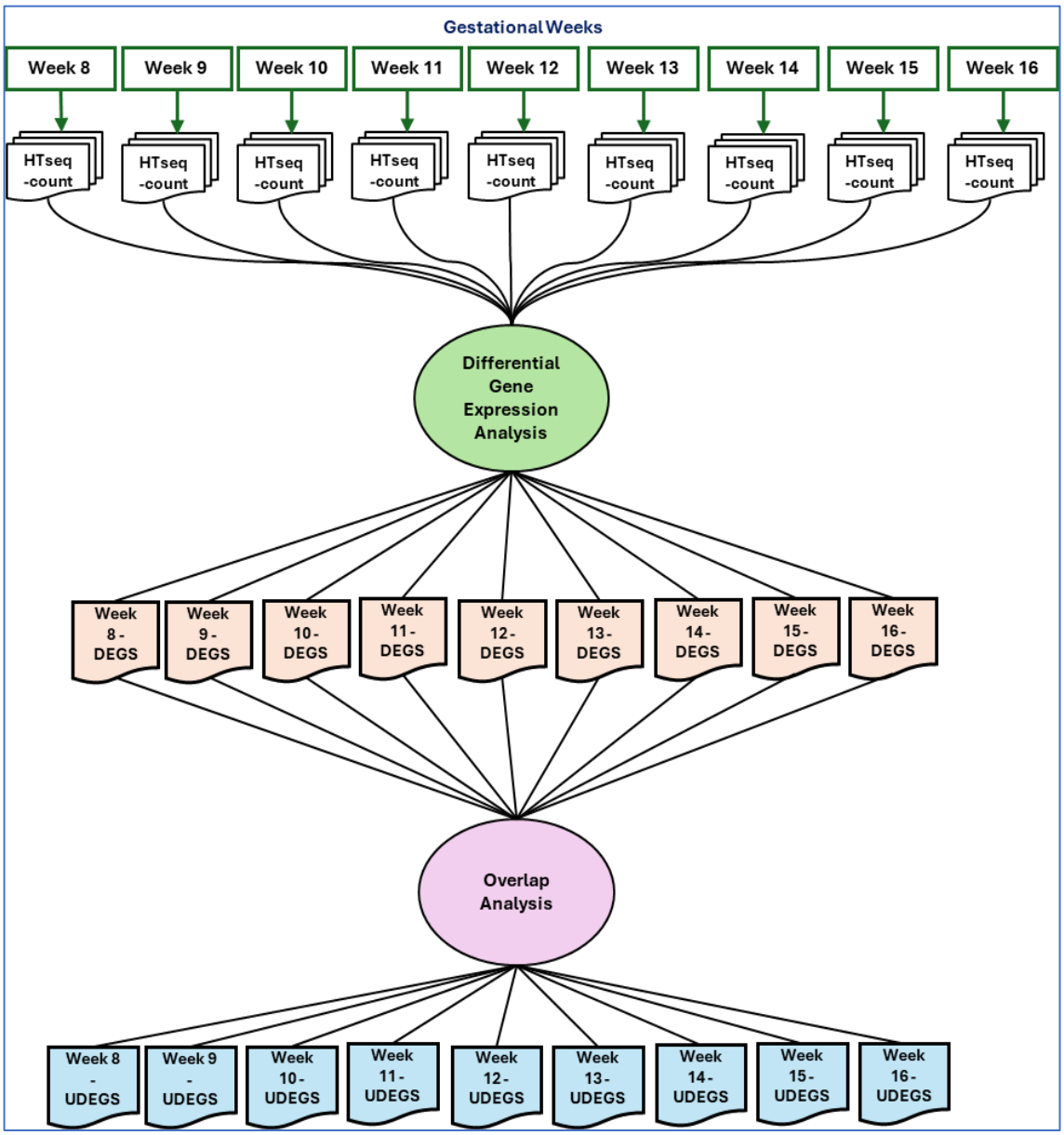


Figure 74: Summary of the UDEGS identification method.

This diagram details the methodology from Chapter 4 for identifying Unique Differentially Expressed Gene Signatures (UDEGS). The core of this approach is a temporal, week-by-week analysis. First, Differential Gene Expression Analysis (DGEA) was performed independently on the cfRNA-seq data for each gestational week (GW) from 8 to 16, comparing pre-eclamptic (PE) and normal pregnancy (NP) samples at each specific time point. This initial step generated distinct lists of all Differentially Expressed Genes (DEGs) for each gestational week (e.g., "Week 8 - DEGS", "Week 13 - DEGS"). Following this, an overlap analysis was conducted across all the weekly DEG sets for identifying UDEGS.

8.2.3 Chapter 5: Stability of UCEGs

Building on this, Chapter 5 focused on consistent gene expression patterns across PE and normal pregnancy (NP) cases. Figure 75 illustrates the novel analytical pipeline from Chapter 5 used to identify these UCEGS. This methodology differs from DGEA by first stratifying the cfRNA-seq data by both condition (NP and PE) and by each individual gestational week. The Consistent Gene Expression Analysis (CGEA) was then applied to each of these distinct groups to generate lists of CEGS. Finally, an overlap analysis was performed to isolate the UCEGS, identifying the stable expression patterns that persist in PE cases.

The identification of significant consistent genes (SCGs) and biologically relevant consistent genes (BRCGs) offered a contrasting view to the variability observed in UDEGs. Notably, the PE group exhibited a higher number of BRCGs, which were stable across samples but distinct from those in the NP group. This finding is significant as it indicates underlying transcriptomic disruptions that persist throughout early gestation in PE cases. The unique consistent expression signatures (UCEGs) derived from this analysis provided another layer of biomarkers that were temporally consistent yet specific to PE. Again, gestational weeks 12 and 13 were especially rich in UCEGs, further pointing to this period as a crucial developmental phase and critical window for PE development. Pathway analysis of UCEGs echoed earlier findings, showing enrichment in inflammatory, hypoxic, and vascular signalling pathways, with particular emphasis on placental ischemia, oxidative stress, and antiangiogenic factor production that compromise vascular function (Hao et al., 2020; Admati et al., 2023; Han et al., 2025).

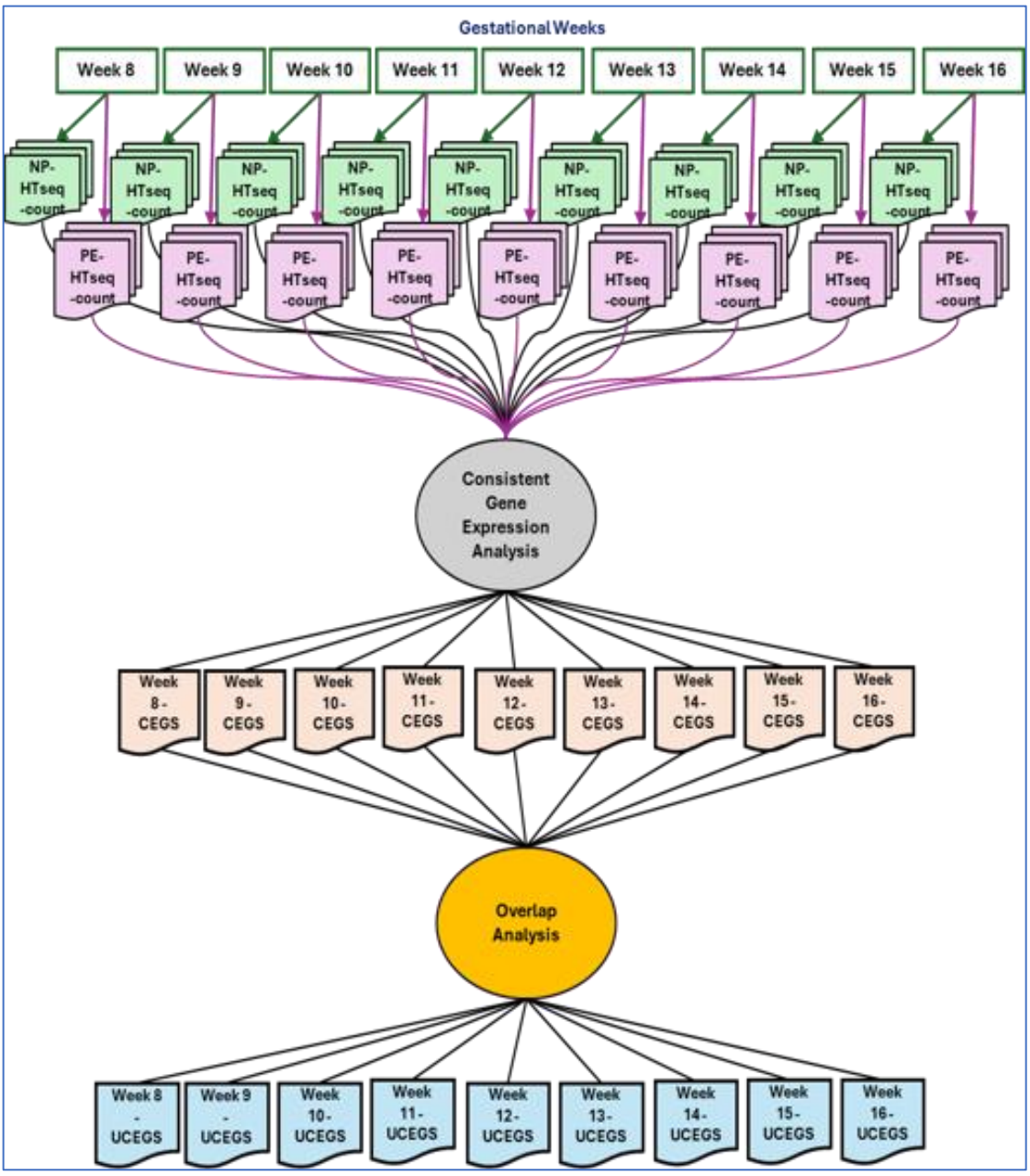


Figure 75: Summary of the methodology for identifying UCEGS.

This flowchart illustrates the novel analytical pipeline from Chapter 5 used to identify Unique Consistently Expressed Gene Signatures (UCEGS). The core of this approach is a temporal, week-by-week analysis stratified by condition. First, the cfRNA-seq count data (HTseq-count) was separated into distinct groups for both condition (Normal Pregnancy - NP and Pre-eclampsia - PE) and each gestational week (GW 8 to 16). Consistent Gene Expression Analysis (CGEA) was then applied independently to each group. This initial step generated distinct lists of all Consistently Expressed Gene Signatures (CEGS) for each condition at each week (e.g., "Week 8 - CEGS", "Week 13 - CEGS") . Following this, an overlap analysis was conducted across all the weekly CEGS sets (within each condition) for identifying UCEGS.

8.2.4 Chapter 6: Genomic Insights from UVS

Chapter 6 marked a pivotal shift from the transcriptomic focus of prior chapters to the genomic landscape, exploring unique variant signatures (UVS) within cfRNA sequencing data – a resource typically leveraged for expression profiling rather than variant analysis. This innovative repurposing employed a rigorous bioinformatics pipeline, utilizing tools like GATK and Funcotator to perform variant calling and functional annotation. The discussion highlighted the study's success in uncovering UVS with a temporal distribution mirroring that of UDEGs and UCEGs from Chapters 4 and 5, with gestational weeks 12 and 13 exhibiting the highest density of variants. These UVS were mapped to genes implicated in PE, such as IFNG, VEGFA, and HLA family members, which are intricately tied to immune signalling, coagulation, and angiogenesis – processes central to PE's pathogenesis (Tersigni *et al.*, 2020; Zheng *et al.*, 2020; Aisagbonhi and Morris, 2022; Luo *et al.*, 2022; Lintao *et al.*, 2024). This genetic lens not only validated the transcriptomic shifts observed earlier but also suggested that these variants might precede or co–occur with expression changes, offering a deeper mechanistic insight into the disease's progression. The chapter's findings underscore the potential of cfRNA as a dual–purpose biomarker source, capturing both expression dynamics and genetic alterations reflective of foetal–maternal interactions.

The novelty of this approach lies in its departure from conventional variant studies, which predominantly rely on DNA from placental tissue or maternal blood. While cfRNA's utility for variant calling remains underexplored in the literature, Chapter 6 demonstrated its capacity to reveal potentially functional mutations linked to PE-associated pathways. The discussion emphasized the significance of identifying variants in genes like VEGFA, critical for angiogenesis, and IFNG and HLA members, key players in immunological adaptation during pregnancy. This aligns with emerging research suggesting that cfRNA may serve as a dynamic archive of molecular changes at the maternal–foetal interface, extending beyond its traditional role in expression profiling (Tsang *et al.*, 2017). By integrating these genetic findings with the transcriptomic signatures from Chapters 4 and 5, the

study constructed a more comprehensive molecular profile of PE, bridging RNA–level changes with underlying genomic alterations. However, the discussion also cautioned that cfRNA–based variant calling introduces challenges, such as potential RNA editing artifacts or sequencing noise, necessitating validation against DNA–based datasets to confirm the functional relevance of these UVS (Piskol, Ramaswami and Li, 2013).

8.2.5 Chapter 7: AI–Driven Predictive Modelling

Chapter 7 brought together the outputs of Chapters 4 to 6 to construct and validate a predictive ML model. Figure 76 provides a high-level visual summary of this integrative methodology. It illustrates how the distinct biomarker categories—UDEGS (Chapter 4) and UCEGS (Chapter 5)—were first unified into a single transcriptomic feature set (UTS). This UTS dataset was then combined with the UVS (Chapter 6) to create the final, multi-omic dataset that was used as the input for the AI models. The integration of transcriptomic and variant signatures into a unified dataset enabled a comprehensive analysis of potential PE biomarkers (Sufriyana, Wu and Su, 2020; X. Li *et al.*, 2022). Principal Component Analysis (PCA) was employed for dimensionality reduction, preserving the most informative features while mitigating overfitting risks. Various ML algorithms were tested, including Decision Trees, Random Forest, Logistic Regression, Gradient Boosting, Linear Discriminant Analysis, and Support Vector Machines (SVM). Among these, the SVM model exhibited superior performance, achieving an accuracy of 92.75% and a precision of 89.47%. These metrics significantly outperform traditional risk assessment methods and underscore the utility of AI in clinical diagnostics (Jhee et al., 2019; Bennett et al., 2022). Importantly, the feature selection process revealed that the most predictive features were drawn from gestational weeks 11 to 13, reinforcing the centrality of this time frame in PE pathogenesis. The model's ability to stratify patients based on molecular risk profiles offers a promising avenue for early intervention and personalized obstetric care.

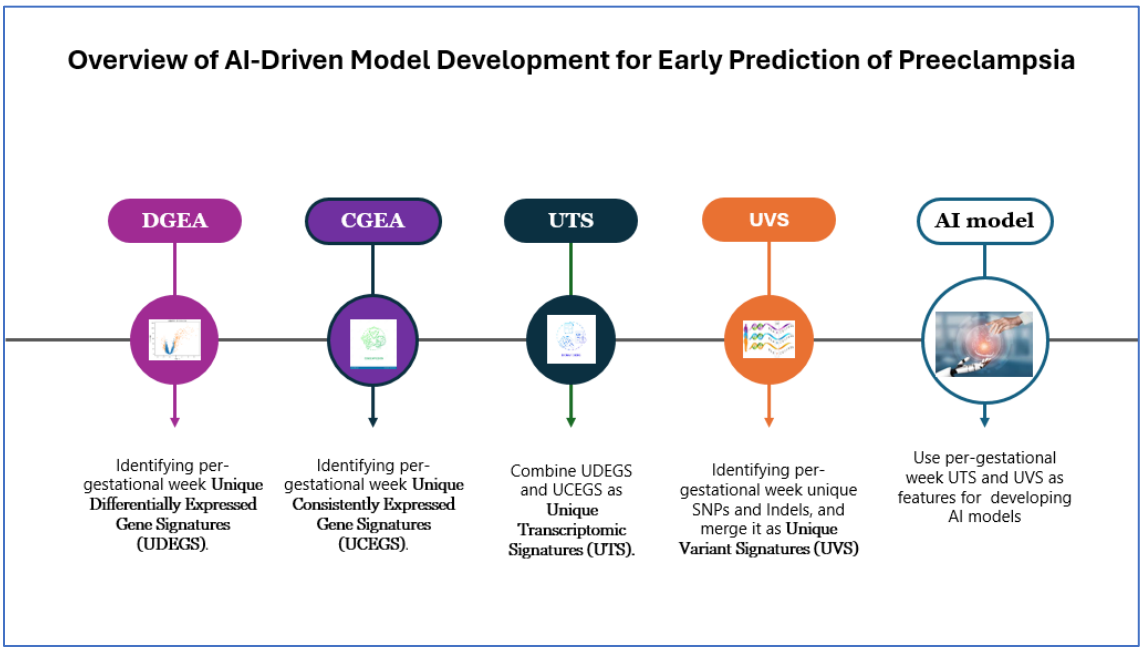


Figure 76: Overview of the AI-Driven model development.

This figure illustrates the data integration and feature engineering strategy from Chapter 7, which combines the outputs of all previous biomarker discovery chapters. The pipeline begins by identifying per-gestational week Unique Differentially Expressed Gene Signatures (UDEGS) from the DGEA (Chapter 4) and Unique Consistently Expressed Gene Signatures (UCEGS) from the CGEA (Chapter 5). These two sets of biomarkers are then combined to form the comprehensive Unique Transcriptomic Signatures (UTS). Concurrently, per-gestational week unique SNPs and INDELs from the Variant Call Analysis (Chapter 6) are identified and merged to create the Unique Variant Signatures (UVS) dataset. Finally, both the UTS and UVS feature sets are used as the integrated, multi-omic input for developing the AI models.

8.2.6 Synthesis and Clinical Implications

The integration of findings across these chapters reveals a coherent and compelling narrative. Each layer of analysis – from differential expression to consistent expression and variant profiling – converges on a critical insight: the mid–gestation period (particularly GWs 11 to 13) represents a pivotal phase in the onset of molecular events leading to PE. This insight is not only scientifically significant but also of great clinical utility, as it defines a window during which diagnostic screening and therapeutic intervention may be most effective (Bujold *et al.*, 2010; Poon and Nicolaides, 2014; O'Gorman *et al.*, 2017). The convergence of evidence from multiple data modalities enhances the robustness of the

predictive model and validates the multi–omic approach taken in this research (S. Li *et al.*, 2022; TARCA *et al.*, 2022).

The clinical implications of this research are profound. The development of a reliable, non–invasive, and interpretable AI-assisted model for early PE detection holds promise for transforming prenatal care. It offers clinicians a tool for risk stratification, enabling targeted surveillance and timely intervention. For researchers, the study opens avenues for further exploration into the molecular underpinnings of PE, potentially leading to the discovery of new therapeutic targets (Dimitriadis et al., 2023). For healthcare systems, the adoption of such predictive models could lead to better resource allocation, reduced complications, and improved maternal–foetal outcomes (Leslie, Thilaganathan and Papageorghiou, 2011; Bohr and Memarzadeh, 2020; Nirupama et al., 2021).

The biochemical biomarkers discussed in Chapter 2, such as PlGF, sFlt1, and PP13, have shown variable predictive value in conventional screening approaches (Kornacki and Wender-Ożegowska, 2020). The multi-omic AI model developed in this study complements these biomarkers by integrating transcriptomic and variant-level data, which enhances diagnostic specificity. Notably, genes associated with angiogenic balance, such as VEGFA and Flt1, were found to have significant expression and variant alterations, indicating that cfRNA-based genomic features could augment or even outperform traditional biomarkers when incorporated into AI-assisted models (Moufarrej *et al.*, 2022). This integrative framework represents a step toward precision-based biomarker discovery and risk stratification.

8.2.7 Interpretation and Implications

The collective findings suggest that PE has distinct molecular signatures that evolve dynamically through early gestation (Rasmussen *et al.*, 2022). The study confirms that cfRNA can serve as a reliable source of transcriptomic and variant data, facilitating non–invasive monitoring. The mid–gestational peak in molecular

disturbances corresponds well with the literature's emphasis on placental remodelling and immune adaptation during this time (Bogias *et al.*, 2022).

The implications for clinical practice are substantial. If validated in diverse populations, the AI-assisted model could allow early stratification of PE risk, enabling pre-emptive interventions such as aspirin therapy or close monitoring (Ansbacher-Feldman *et al.*, 2022). The use of interpretable and transparent models also supports clinical integration by providing clinicians with understandable decision–support tools (Schmidt *et al.*, 2022). Additionally, the ability to detect both transient and stable molecular changes may provide insights into disease trajectory, offering opportunities for both early intervention and monitoring.

Beyond PE, the model demonstrates the broader applicability of cfRNA–based diagnostics, but it also raises the critical question of diagnostic specificity—a point alluded to in the literature review. PE is often considered a heterogeneous syndrome (i.e., "more than one disease") and shares significant pathogenic origins with other placental insufficiency disorders, most notably Intrauterine Growth Restriction (IUGR). Both conditions can stem from the same initial insults of poor placentation and shallow spiral artery remodelling. Consequently, it is plausible that a subset of the genomic signatures (UDEGs, UCEGs, and UVS) identified in this study are not strictly PE-specific but rather reflect a broader 'placental stress' phenotype (Baker *et al.*, 2024). The current AI model was trained to distinguish PE from normal pregnancy and was not designed to differentiate PE from IUGR. This represents a key limitation and a clear direction for future work. Validating these biomarkers against an IUGR cohort is essential to isolate PE-unique signatures and enhance the model's diagnostic specificity, a necessary step for its eventual clinical application.

Overall, these findings substantiate the initial hypothesis outlined in Chapter 2 that integrating gene expression signatures and variant data with maternal characteristics can significantly enhance the early prediction of PE. The discovery

of temporally consistent molecular signatures across gestational weeks validates the conceptual framework established in the introduction. This convergence between hypothesis and results underscores the robustness of the study design and highlights the translational potential of cfRNA-based AI prediction models for clinical application (Ng et al., 2024).

Furthermore, this study's findings highlight a powerful opportunity for a synergistic AI model. The extensive discussion on clinical risk factors in Chapter 2 (such as obesity, chronic hypertension, and IVF) was not merely foundational but provides a clear roadmap for the model's next evolution. The genomic signatures identified in Chapters 4-6 represent the molecular dimension of PE risk, while the clinical factors represent the phenotypic dimension. By not following up on these factors in the current model, this thesis has established a baseline for genomic-only prediction. The clear next step is to create a multi-dimensional model. Fusing the 655 UTS and 99 UVS identified here with standard clinical data would allow the AI to learn complex interactions between a patient's baseline risk (e.g., from BMI) and their real-time molecular state (e.g., from cfRNA), thus creating a far more powerful and personalised predictive tool (Moufarrej *et al.*, 2022; T. Li *et al.*, 2024).

In summary, this thesis demonstrates that the integration of cfRNA–based transcriptomic and variant data into an AI–driven diagnostic framework is both feasible and effective for early PE detection. It lays the groundwork for future clinical translation and sets a precedent for the use of multi–omic data in personalized obstetric care.

8.2.8 Summary of the Chapter

This chapter has synthesized the principal findings of the thesis, underscoring their coherence with existing literature and their implications for future research and clinical practice. The study introduces a comprehensive, AI–driven framework for the early detection of PE, leveraging cfRNA–derived transcriptomic and genomic variant data. Through a multi–faceted analysis – encompassing differentially

expressed gene signatures (UDEGs), consistently expressed gene signatures (UCEGs), and unique variant signatures (UVS) – the research constructs a detailed molecular portrait of PE's development during early gestation, specifically from weeks 8 to 16. This layered approach elucidates the dynamic and stable molecular changes associated with PE, providing a robust foundation for predictive modelling.

Although the study acknowledges several limitations, such as reliance on retrospective data and challenges in cfRNA–based variant calling, the methodological strengths – including non–invasive sampling, gestational week–specific resolution, and multi–omic integration –demonstrate significant potential for clinical application. The robustness of the Support Vector Machine (SVM) model, achieving a 92% accuracy, further reinforces its promise as a reliable diagnostic tool. This work advances the field of maternal–foetal medicine by exemplifying the synergy of bioinformatics and artificial intelligence in enabling early disease detection, bridging molecular insights with practical utility. The chapter concludes by highlighting the transformative potential of this research to enhance prenatal care and promote personalized medicine in obstetrics, setting the stage for future validation and broader implementation.

Chapter 9
Conclusion

This thesis aimed to pioneer an AI–driven predictive model for the early recognition of PE women by harnessing multi–omics data derived from cfRNA, specifically targeting transcriptomic and genomic variant signatures across gestational weeks (GWs) 8 to 16. The specific objectives were threefold: to identify unique differentially expressed genes (UDEGs) and consistent expression signatures (UCEGs) reflective of PE's molecular dynamics, to uncover unique variant signatures (UVS) offering genetic insights, and to integrate these biomarkers into a machine learning (ML) framework for accurate early detection. By addressing these goals, the research sought to bridge the gap in prenatal diagnostics, where current methods often fail to detect PE before its symptomatic onset, thereby enhancing maternal and foetal health outcomes (Feng and Luo, 2024).

The major findings underscore the efficacy of this approach. Through systematic analyses in Chapters 4–6, the study pinpointed UDEGs, UCEGs, and UVS, revealing a critical molecular window in GW11–13 where PE-associated changes peak. These signatures, enriched in pathways like angiogenesis, immune dysregulation, and oxidative stress, aligned with PE's established pathogenesis (Yu *et al.*, 2025). Chapter 7's Support Vector Machine (SVM) model, integrating these features, achieved an impressive accuracy of 92.75% and precision of 89.47%, outperforming traditional risk assessments reliant on late–stage clinical markers. This high predictive performance directly answered the research question of whether multi–omics data from cfRNA could enable early PE detection, demonstrating that a combination of dynamic gene expression and stable genetic variants provides a robust basis for risk stratification.

A standout original contribution of this work lies in its pioneering integration of transcriptomic and genomic signatures from cfRNA for PE prediction. To the best of current knowledge, this is the first study to repurpose cfRNA, a typically expression-focused resource, for variant calling using tools like GATK and

Funcotator, and to combine these dual insights into an AI model with such high accuracy in early pregnancy. This multi–omic, temporally resolved approach not only advances beyond single–biomarker studies (e.g., sFlt–1, PlGF) but also sets a new benchmark for non–invasive diagnostics, leveraging the maternal plasma as a rich, accessible molecular archive of foetal–maternal interactions (Popp *et al.*, 2024; Bhosale *et al.*, 2025).

Nevertheless, critical reflection reveals limitations that temper these achievements. The study's reliance on retrospective, publicly available datasets introduced variability in sample quality and metadata, potentially limiting generalizability across diverse populations (Eberhard *et al.*, 2025). The small sample size and lack of detailed clinical data, such as blood pressure or proteinuria levels, restricted correlations with PE severity or outcomes (Nilsson *et al.*, 2024). Additionally, cfRNA's susceptibility to degradation and contamination posed technical challenges, particularly for variant analysis. These constraints highlight the need for cautious interpretation and further validation to ensure the model's broader applicability.

The implications of this research are profound across practical, theoretical, and clinical domains. Practically, the SVM model offers a decision–support tool for antenatal care, enabling clinicians to identify at–risk women early and tailor monitoring or interventions, potentially reducing PE–related complications like preterm delivery or maternal hypertension (Melinte-Popescu et al., 2023). Theoretically, the identification of GW11 to GW13 as a molecular tipping point refines the two-stage model of PE pathogenesis, enriching the understanding of its preclinical phase. Clinically, this framework could shift prenatal care from reactive to proactive, optimizing resource allocation and improving maternal–foetal outcomes, particularly in high–risk settings where PE mortality remains a pressing concern.

Looking ahead, future research should prioritize validation in larger, multi–ethnic prospective cohorts to address the current study's diversity and sample size

limitations. Collaborations with obstetric clinics could integrate clinical metrics – blood pressure, proteinuria, or ultrasound data –enhancing the model's real–world relevance. Expanding to additional omics layers, such as proteomics or metabolomics, could further boost predictive power and uncover PE subtypes, while developing point–of–care platforms would accelerate clinical translation. Ethical considerations, including equitable access and patient consent, also warrant exploration to ensure responsible implementation (Riano-Moreno *et al.*, 2025).

In conclusion, this research establishes a ground-breaking foundation for AI-assisted early diagnostics in obstetrics, offering a scalable, non–invasive solution to one of pregnancy's most formidable challenges. By blending bioinformatics, multi-omics, and machine learning, it enhances understanding of PE's molecular underpinnings and charts a promising course for personalised maternal healthcare, with the potential to save lives and redefine prenatal standards globally.

9.1 Strengths

This research is distinguished by a constellation of strengths that collectively elevate its scientific rigour, clinical relevance, and potential for future impact in the domain of PE detection. One of its foremost advantages lies in the utilization of non–invasive cfRNA data, derived from maternal plasma, as the cornerstone of its molecular analyses. Unlike invasive methods requiring placental biopsies or amniocentesis, cfRNA enables safe, repeatable sampling throughout pregnancy, facilitating longitudinal monitoring without risk to the mother or foetus (Abbasi, 2022; Zhou *et al.*, 2023). This non–invasive approach not only aligns with ethical standards in prenatal care but also enhances the feasibility of scaling the methodology into routine clinical practice, offering a practical alternative to traditional diagnostics that often depend on late–emerging symptoms or invasive procedures (Melinte-Popescu et al., 2023; Castillo-Marco et al., 2025).

A second key strength is the adoption of a gestational week–specific analysis framework, which introduces an unprecedented level of temporal resolution to the study of PE. By dissecting molecular signatures – spanning unique differentially expressed genes (UDEGs), unique consistently expressed genes (UCEGs), and unique variant signatures (UVS) – across gestational weeks 8 to 16, the research maps the dynamic evolution of PE-associated changes with precision (Tarca *et al.*, 2019; Than *et al.*, 2024). This granularity contrasts with much of the prior literature, which often aggregates data across trimesters or focuses solely on late–stage manifestations, overlooking the subtle, stage–specific shifts that precede clinical onset (Degnes *et al.*, 2024; Han *et al.*, 2025). For instance, the identification of peak molecular activity in GW12 and GW13, as observed in Chapters 4–6, underscores critical windows of vulnerability, enabling a more targeted approach to risk assessment and intervention timing that could pre–empt severe complications like IUGR or maternal hypertension.

Thirdly, the multi–omic integration of transcriptomic and variant data stands out as a robust methodological pillar, enriching the biological depth and plausibility of the findings (Gumusoglu *et al.*, 2025). By combining gene expression profiles (Chapters 4 and 5) with genetic variants (Chapter 6), the study captures both the functional consequences of RNA–level changes and the structural genomic alterations that may underpin them. This dual perspective – evidenced by the alignment of UVS with genes like VEGFA and IFNG, alongside transcriptomic shifts in angiogenesis and immune pathways – provides a more holistic understanding of PE's molecular aetiology than single–omic approaches. Such integration not only validates the consistency of findings across data types but also strengthens the biological narrative, linking placental dysfunction to systemic maternal responses in a manner that resonates with the established two-stage model of PE pathogenesis.

The application of a sophisticated machine learning (ML) framework represents another significant strength, ensuring high reliability, accuracy, and potential

generalizability of the predictive model developed in Chapter 7. The study employed a suite of ML algorithms – ranging from Support Vector Machines (SVM) to Random Forests and Gradient Boosting – coupled with rigorous techniques like Principal Component Analysis (PCA) for dimensionality reduction and Recursive Feature Elimination with Cross–Validation (RFE–CV) for feature selection (Özcan and Peker, 2025). The standout performance of the SVM model, achieving a 92% accuracy, was bolstered by cross–validation strategies that mitigated overfitting and confirmed robustness across internal datasets. This multi–algorithmic approach, paired with systematic optimisation, distinguishes the research from simpler predictive models in the literature, offering a level of statistical confidence that enhances its credibility for future external validation in diverse cohorts (Y. Xu *et al.*, 2025).

Moreover, the incorporation of established databases, functional annotations, and pathway analyses – such as Gene Ontology (GO), KEGG, and protein–protein interaction (PPI) networks – endows the model with a high degree of interpretability, a critical attribute for clinical adoption (Alharbi *et al.*, 2025; Zhou, Madala and Huang, 2025). By leveraging tools like GATK, Funcotator, and enrichR, the study annotated UDEGs, UCEGs, and UVS to biologically meaningful pathways (e.g., angiogenesis, immune regulation), translating complex omics data into actionable insights. This transparency not only bridges the gap between computational outputs and clinical relevance but also addresses a common critique of omics–based diagnostics: the “black box” nature of AI models (Hassija *et al.*, 2024; Xu and Shuttleworth, 2024; Abbas, Jeong and Lee, 2025). The detailed downstream analyses, such as the PPI clusters identified in GW13 (Chapter 4) or the KEGG pathways enriched in UVS (Chapter 6), provide clinicians with a mechanistic rationale for why certain biomarkers predict PE risk, fostering trust and facilitating integration into decision–making workflows.

Finally, the emphasis on methodological reproducibility enhances the study’s scientific integrity, addressing a persistent challenge in omics research where

replication is often hampered by proprietary datasets or opaque pipelines (Wilkinson *et al.*, 2016). The use of publicly available datasets from repositories like GEO and SRA, combined with well–documented bioinformatics workflows (e.g., STAR alignment, GATK variant calling), ensures that the analyses can be independently verified or adapted by other researchers (Clough *et al.*, 2024). The open–source nature of tools like R and Python, employed throughout the thesis, further supports this reproducibility, while the provision of detailed appendices (Chapter 9) with step–by–step analytical protocols reinforce transparency (Sandve *et al.*, 2013). This commitment to replicability not only strengthens the study's academic foundation but also positions it as a blueprint for future investigations into PE or other pregnancy–related disorders.

In aggregate, these strengths – non–invasive data collection, temporal specificity, multi–omic integration, advanced ML modelling, interpretability, and reproducibility – collectively position this thesis as a pioneering contribution to precision maternal health. They enable the research to transcend the limitations of traditional PE diagnostics, offering a comprehensive, reliable, and clinically translatable framework for early detection. By harnessing cutting-edge bioinformatics and AI within a biologically grounded approach, the study advances understanding of PE's molecular underpinnings and sets a high standard for interdisciplinary innovation in prenatal care, with the potential to influence both research and practice on a global scale.

9.2 Limitations

Despite the promising results achieved across this thesis, several critical limitations must be acknowledged to contextualize the findings and guide future research. A primary constraint stems from the study's reliance on retrospective, publicly available datasets, such as those sourced from repositories like the Gene Expression Omnibus (GEO) and Sequence Read Archive (SRA). While these datasets enabled a robust analysis of cfRNA across gestational weeks, their retrospective nature introduces inherent variability in sample quality, including

differences in RNA extraction protocols, storage conditions, and sequencing platforms (Lachmann *et al.*, 2018). This variability may exacerbate batch effects, where technical artifacts obscure biological signals, potentially skewing the identification of UDEGs, UCEGs, and UVS (Momeni *et al.*, 2023; Kovacheva *et al.*, 2024). Furthermore, the metadata accompanying these datasets is often incomplete, lacking detailed information on maternal demographics, lifestyle factors, or environmental exposures. This limitation hampers the generalizability of the findings, as subtle population–specific effects – such as genetic predispositions or socioeconomic influences on PE risk – may remain undetected, underscoring the need for prospective, standardized data collection to validate these molecular signatures across diverse cohorts.

Another significant limitation is the restricted access to comprehensive clinical phenotype data, which poses challenges in correlating the identified molecular signatures with the clinical severity or long–term outcomes of PE (Founds et al., 2018; Zhao et al., 2025). While the study successfully pinpointed transcriptomic and genetic markers linked to PE pathogenesis, the absence of detailed clinical annotations – such as blood pressure measurements, proteinuria levels, or neonatal outcomes like intrauterine growth restriction (IUGR) – limits the ability to assess how these biomarkers reflect disease progression or patient–specific risk profiles (Slobodchikova et al., 2025). This disconnects between molecular and clinical data weakens the translational potential of the findings, as it remains unclear whether the signatures distinguish mild from severe PE cases or predict complications like HELLP syndrome. Consequently, while the machine learning models, particularly the SVM with its 92% accuracy, demonstrated strong internal performance on the available data, their external validity remains untested. Prospective studies incorporating ethnically diverse populations – spanning variations in genetic ancestry and healthcare access – are essential to confirm the models' predictive reliability and clinical utility beyond the controlled confines of this research.

The use of cfRNA as the primary molecular substrate, while innovative and non–invasive, introduces additional technical challenges that warrant scrutiny. Although cfRNA offers a rich source of transcriptomic and variant information reflective of maternal–foetal dynamics, its susceptibility to degradation and contamination poses risks to data integrity. Unlike DNA, cfRNA exists in a fragmented state in maternal plasma, subject to rapid turnover and environmental stressors during sample handling (Lee and Schatz, 2012). These properties can introduce noise into sequencing data, potentially leading to false positives in differential expression analyses or inaccuracies in variant calling, despite the application of sophisticated bioinformatic corrections like those implemented with GATK and Funcotator. The fragmented nature of cfRNA further complicates variant detection, as short read lengths may hinder precise mapping to the reference genome, raising questions about the functional significance of some UVS (Marin *et al.*, 2022; Kosugi and Terao, 2024). Furthermore, a formal expression Quantitative Trait Loci (eQTL) analysis, which would directly link the identified variants to the observed gene expression changes, was not performed in this research (Gambardella, 2025). This analysis was considered beyond the scope of the current investigation due to these significant technical challenges, which complicate the reliable correlation of variant genotypes with transcript-level expression data. While the study mitigated these issues through quality control and preprocessing steps, the discussion across Chapters 4–6 consistently noted that validation against DNA–based datasets or paired placental tissue samples would strengthen confidence in these findings, highlighting an area ripe for methodological refinement.

Finally, the computational burden of integrating multi–omic data and deploying high–dimensional machine learning models presents a practical barrier to clinical implementation (Slobodchikova *et al.*, 2025). The analyses in this thesis – from differential gene expression and variant calling to the multi-step ML pipeline involving PCA, RFE–CV, and SVM training – required substantial computational resources, including high–performance computing infrastructure and expertise in bioinformatics. This complexity, while yielding robust results in a research setting,

poses significant challenges for real–time application in clinical environments, particularly in resource–limited settings where such infrastructure may be unavailable (S. Li *et al.*, 2022). The high–dimensional nature of the genomic data, encompassing thousands of features across UDEGs, UCEGs, and UVS, demands intensive preprocessing and optimization, as evidenced in Chapter 7's dimensionality reduction efforts. Moreover, the integration of additional omics layers (e.g., proteomics or metabolomics), as proposed for future work, would amplify this computational load, necessitating the development of streamlined algorithms or cloud–based solutions to ensure scalability. These technical and logistical hurdles underscore the gap between the study's proof–of–concept success and its deployment as a bedside tool, urging a focus on simplifying workflows without compromising predictive accuracy.

In sum, while the thesis delivers a compelling framework for early PE detection, these limitations highlight the multifaceted challenges of translating genomic and AI–driven insights into practice. The reliance on retrospective data, limited clinical context, cfRNA's technical constraints, and computational demands collectively temper the findings' immediate applicability, yet they also chart a clear path forward (Uzun *et al.*, 2016). Addressing these issues through prospective multi–centre studies, enhanced phenotyping, methodological advancements in cfRNA analysis, and optimized computational tools will be crucial to realizing the full potential of this research, ensuring it benefits diverse populations and integrates seamlessly into maternal healthcare systems.

9.3 Recommendations for Future Directions

The findings of this thesis lay a robust foundation for early PE detection, yet they also illuminate a spectrum of opportunities for future research to refine, validate, and translate this work into tangible clinical benefits. A primary direction involves validating the AI–driven predictive model in larger, prospective cohorts that reflect ethnic and geographic diversity. The current reliance on retrospective datasets, while effective for proof–of–concept, limits the model's generalizability across

populations with varying genetic ancestries, environmental exposures, and healthcare contexts (Urquia *et al.*, 2014). Collaborations with obstetric clinics, hospitals, and international consortia – such as the Global Pregnancy Collaboration (CoLab) –could enable longitudinal studies that track molecular profiles alongside clinical trajectories in real time. By linking cfRNA–derived signatures (UDEGs, UCEGs, UVS) to detailed phenotypic data – such as blood pressure trends, proteinuria onset, or neonatal outcomes like preterm birth –these studies could confirm the model's accuracy, sensitivity, and specificity in diverse settings, paving the way for its adoption as a universal screening tool (Nirupama et al., 2021; Reddy et al., 2021; Ranjbar et al., 2024).

Building on the multi–omic framework established here, integrating additional molecular layers – such as proteomics, metabolomics, and epigenomics – offers a promising avenue to enhance predictive power and unravel PE's heterogeneity (Wei *et al.*, 2025). While transcriptomic and variant data provide critical insights into gene expression and structural changes, proteomics could reveal downstream protein–level alterations (e.g., sFlt1 or PlGF imbalances), while metabolomics might capture metabolic shifts (e.g., lipid dysregulation) tied to PE's systemic effects (Tian *et al.*, 2024; Starodubtseva *et al.*, 2025). Epigenomic analyses, such as DNA methylation profiling, could further elucidate regulatory mechanisms influencing the identified biomarkers, particularly those in angiogenesis or immune pathways. This multi-omics approach, though computationally intensive, could be facilitated by emerging tools like multi–omics factor analysis (MOFA) or integrative AI models, potentially resolving subtypes of PE – such as early–onset versus late–onset – that remain poorly distinguished by current diagnostics (Han et al., 2025; Shyu et al., 2025). Such granularity would not only boost the SVM model's performance beyond its current 92% accuracy but also inform personalised risk stratification and tailored interventions (Kovacheva *et al.*, 2024).

The translation of this research into clinical practice hinges on developing point–of–care (POC) diagnostic platforms that embed the AI model within accessible

prenatal screening workflows (Butler *et al.*, 2024). Future efforts should focus on miniaturizing the cfRNA sequencing and analysis pipeline –potentially through innovations like nanopore sequencing or microfluidics – coupled with lightweight, cloud–based ML algorithms optimized for real–time risk assessment. These POC tools could be integrated into routine prenatal visits, delivering immediate PE risk scores based on a single blood draw, much like current glucose or haemoglobin tests. Partnerships with biomedical engineers and health technology firms could accelerate this development, ensuring the platforms are cost–effective and user–friendly for clinicians in both high–resource and low–resource settings. Additionally, adapting the model for use with wearable devices or mobile health applications could empower patients with continuous monitoring, aligning with the growing trend of digital health in maternal care (Liu *et al.*, 2024; Jaeger *et al.*, 2025).

Ethical considerations surrounding the use of AI in prenatal diagnostics represent another critical area for exploration (Coghlan, Gyngell and Vears, 2024). As this technology nears clinical deployment, research must address issues of informed consent, data privacy, and equity in access. For instance, how will expectant mothers be counselled about the implications of a high PE risk score, particularly in cases of false positives that might provoke unnecessary anxiety or interventions? The handling of sensitive genomic data, especially in cloud–based systems, requires robust encryption and compliance with regulations like GDPR or HIPAA. Moreover, ensuring equitable access – particularly in underserved regions where PE mortality rates are highest–demands strategies like subsidized testing or public health integration to prevent this tool from exacerbating existing disparities (Mapari *et al.*, 2024). Interdisciplinary studies involving bioethicists, policymakers, and patient advocates could establish guidelines that balance innovation with responsibility, fostering trust and inclusivity in its implementation.

Beyond validation and translation, future research should probe the biological roles of the newly identified biomarkers, particularly those with unknown or poorly

characterised functions, through wet–lab validation and preclinical models. Genes or variants highlighted in Chapters 4 to 6 – such as unannotated UDEGs or UVS in regulatory regions – warrant functional studies using techniques like CRISPR–Cas9 gene editing, RNA interference, or proteomics in cell lines mimicking placental trophoblasts. Animal models, such as PE–like rodent models induced by reduced uterine perfusion pressure (RUPP), could further test the causal roles of these biomarkers in driving placental dysfunction, immune dysregulation, or vascular pathology. These investigations might uncover novel therapeutic targets – e.g., small molecules to modulate VEGFA expression or biologics to correct HLA–mediated immune imbalances – offering preventive strategies that intervene before PE's systemic phase emerges (Mukherjee *et al.*, 2023). Such discoveries would extend the thesis's impact from diagnostics to therapeutics, addressing both detection and management of this complex disorder (Rasmussen *et al.*, 2022; Castillo-Marco *et al.*, 2025).

Additionally, the computational framework itself invites refinement to enhance scalability and real–world applicability (Ahmed *et al.*, 2023). Future work could explore streamlined algorithms—such as ensemble methods combining SVM with deep learning—to reduce the processing burden of high–dimensional multi–omic data, making the model viable on standard clinical hardware. Open–source development of these tools, accompanied by user–friendly interfaces, could democratize access for researchers and practitioners alike. Concurrently, expanding the gestational window beyond GW16 to later trimesters might reveal how early signatures evolve into late–stage PE, providing a more complete temporal map of the disease (Clare and Sakharkar, 2025). This could be paired with comparative studies against existing biomarkers (e.g., sFlt1/PlGF ratios) to benchmark the added value of the cfRNA–AI approach, solidifying its position in the diagnostic landscape.

A key consideration arising from this research is the specificity of the identified genomic signatures to PE. As noted, early pregnancy pathologies such as

Intrauterine Growth Restriction (IUGR) share analogous placental origins with PE, including impaired spiral artery remodelling and shallow trophoblast invasion (Lawless *et al.*, 2023; Azmi *et al.*, 2024). Consequently, it is plausible that a subset of the differentially expressed genes and variant signatures discovered in this study may not be exclusive to PE but could reflect a broader "placental stress" phenotype common to multiple obstetric complications (Gardner *et al.*, 2025; Zeng *et al.*, 2025). The analysis, which contrasted PE solely against normal pregnancy, was not designed to disentangle these shared pathways. Future studies would greatly benefit from a comparative analysis incorporating transcriptomic and genomic data from other placental disorders, such as IUGR and gestational hypertension. This approach would enable the discrimination of PE-unique biomarkers from those associated with generalised placental dysfunction, thereby enhancing the specificity and clinical predictive power of an AI-assisted model (Ng et al., 2024).

To effectively validate these biomarkers and enhance predictive power, future studies must progress from the current retrospective analysis to large-scale, prospective, multi-ethnic cohort studies. This approach is essential to confirm the model's generalisability across diverse genetic backgrounds and healthcare systems. Furthermore, to ensure biomarker specificity, comparative pathological studies are required to distinguish the identified PE signatures from those of other, similar placental-insufficiency disorders, such as Intrauterine Growth Restriction (IUGR) (Stepan, Hund and Andraczek, 2020; Hong and Kumar, 2023). For clinical translation, the model must be enhanced by integrating other critical factors. Future iterations should adopt a multi-dimensional approach, combining the genomic signatures (UDEGs, UCEGs, and UVS) with established clinical risk factors (e.g., maternal age, BMI, parity, chronic hypertension), biophysical markers (e.g., uterine artery Doppler), and additional multi-omics data, particularly proteomics (like the sFlt-1/PlGF ratio) and metabolomics. Finally, practical studies are needed to address model interpretability using explainable AI (XAI) and to determine the cost-

effectiveness of translating this RNA-seq discovery tool into a targeted, affordable clinical assay (Athanasopoulou *et al.*, 2025).

In essence, the future directions for this research span validation, integration, translation, ethics, biology, and computation, each building on the thesis's strengths to address its limitations. By pursuing these avenues—through collaborative, interdisciplinary efforts — the work has the potential to transform PE management from a reactive to a proactive paradigm, reducing its global burden on maternal and neonatal health. This roadmap not only extends the current findings but also positions them as a catalyst for broader advancements in precision prenatal medicine, with implications that could ripple across other pregnancy complications and beyond.

9.4 Novel Contributions to Knowledge

This research introduces several groundbreaking contributions to the field of maternal health and bioinformatics. First, it provides the first comprehensive identification of gestational week–specific molecular signatures for PE, enabling a more precise understanding of the disease's temporal progression. Unlike prior studies aggregating data across trimesters, this work delineates specific transcriptomic and variant signatures at each gestational week, offering unprecedented granularity in PE risk assessment.

Second, the study integrates multi–layered biomarkers – differential gene expression, consistent gene expression, and genetic variants – into an AI-assisted predictive model, a novel approach that enhances diagnostic accuracy. Previous studies have focused mainly on single–dimensional biomarkers, whereas this research demonstrates the synergistic value of combining genomic and clinical data for improved prediction.

Third, this study advances the field by leveraging recursive feature elimination and cross–validation techniques to refine biomarker selection (Han, Huang and Zhou, 2021; Abdelwahab *et al.*, 2022; Li, Ching and Liu, 2022). This approach enhances

the reliability of the identified biomarkers, ensuring that only the most predictive and biologically relevant features are incorporated into the AI model. Such rigorous feature selection methodologies help reduce overfitting and improve the model's generalisability across different datasets, strengthening its potential for clinical translation.

Fourth, the research introduces a novel application of gestational week–specific pathway analysis. The study sheds light on time–dependent biological mechanisms in PE progression by categorising biomarkers based on their activation during particular gestational windows (Estrela *et al.*, 2025). This insight not only refines the understanding of disease pathogenesis but also supports the development of gestation–specific therapeutic interventions, paving the way for personalised maternal–foetal medicine.

Fifth, the study employs a unique intersecting approach between identified biomarkers and known PE-associated genes from external databases such as GeneCards and DisGeNET. This validation step ensures that the novel biomarkers identified through cfRNA sequencing align with established PE–related genomic profiles. The intersectional analysis adds credibility to the identified biomarkers and strengthens their clinical relevance, enhancing confidence in their predictive potential (Huang et al., 2024; Q. Li et al., 2024).

This study marks a significant step forward in PE research, combining advanced bioinformatics with AI–driven predictive modelling to enhance early detection. The findings contribute to precision medicine in maternal health, offering a framework for non–invasive, early–stage PE diagnosis. By integrating multiple layers of genomic data with AI, this research provides a robust foundation for future advancements in predictive obstetrics and personalised maternal care.

9.5 Recommendations

This study provides a significant advancement in the early detection and prediction of PE using ML models based on gestational week–specific transcriptomic and

variant signatures. To enhance the practical applications of these findings in clinical and research settings, several recommendations are proposed. The integration of AI–driven predictive models into clinical practice is essential, as incorporating unique transcriptomic and variant biomarkers into electronic health record (EHR) systems can enable early risk stratification, allowing obstetricians to personalise prenatal care. Expanding multi–omics research by integrating proteomics, metabolomics, and epigenomics will refine PE prediction models, enhance the mechanistic understanding of PE pathophysiology, and improve their generalisability across diverse populations. Furthermore, validation across large–scale, ethnically diverse cohorts is crucial, given the genetic, environmental, and socioeconomic variations in PE, ensuring the robustness and reliability of predictive biomarkers. The development of non–invasive screening tools based on cfRNA should be prioritised, making these assays accessible, cost–effective, and suitable for routine prenatal screening in both high–resource and low–resource settings. Additionally, transitioning AI-based models from research to clinical practice requires rigorous clinical trials and regulatory approvals from health authorities to ensure safety and efficacy. Ethical and data privacy considerations must also be addressed, with stringent measures for data security, informed consent, and responsible use of genetic and clinical data. AI–powered risk assessment can further enable personalised medicine, allowing early interventions, such as aspirin therapy and close monitoring, to mitigate maternal and foetal complications associated with PE. Lastly, collaboration between researchers, clinicians, and policymakers is essential to drive innovation and ensure the successful implementation of AI in maternal health, fostering a multidisciplinary approach that bridges bioinformatics, obstetrics, healthcare administration, and policymaking.

Appendices

9.6 Appendix A

Differential Gene Expression Analysis

R Script for Differential Gene Expression Analysis

```
setwd("F:/DGE4myThesis2024")
```

9.6.1 Step 1: Load the libraries and per–gestational week data

```
# Load the required libraries
library(DESeq2)
library(edgeR)
library(limma)
library(ggplot2)
library(pheatmap)
library(enrichplot)
library(enrichR)
library(ggupset)

W8count = read.csv("NEW_HTseqCounts_Week8.csv", row.names = 1)
W9count = read.csv("NEW_HTseqCounts_Week9.csv", row.names = 1)
W10count = read.csv("NEW_HTseqCounts_Week10.csv", row.names = 1)
W11count = read.csv("NEW_HTseqCounts_Week11.csv", row.names = 1)
W12count = read.csv("NEW_HTseqCounts_Week12.csv", row.names = 1)
W13count = read.csv("NEW_HTseqCounts_Week13.csv", row.names = 1)
W14count = read.csv("NEW_HTseqCounts_Week14.csv", row.names = 1)
W15count = read.csv("NEW_HTseqCounts_Week15.csv", row.names = 1)
W16count = read.csv("NEW_HTseqCounts_Week16.csv", row.names = 1)

HTseqAll = read.csv("HTseqAll.csv", row.names = 1)
```

9.6.2 Step 2: Week 8

```
# Load the libraries
library(DESeq2)

condition <-  as.factor(c(rep("NP", 12), rep("PE", 4)))
colData <-  data.frame(row.names=colnames(W8count), condition=factor(cond
ition, levels=c('NP', 'PE')))

#DEseqDataset
```

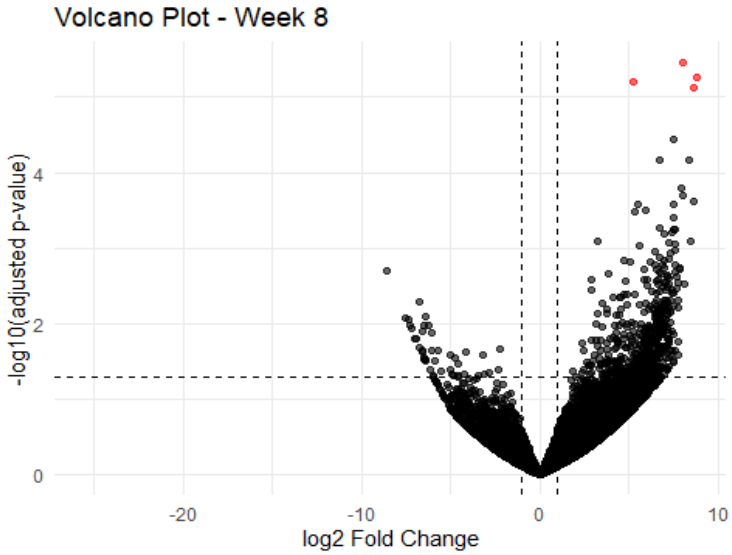
```
dataset <- DESeqDataSetFromMatrix(countData = W8count,
                                  colData = colData,
                                  design = ~condition)

# Pre–filtering: removing rows with low gene counts
# Keeping rows that have at least 10 reads total
keep <- rowSums(counts(dataset)) >= 10
dataset <- dataset[keep,]

# set the factor level
dataset$condition <- relevel(dataset$condition, ref = "NP")

#Then run the DESeq2 algorithm and extract results for the two-class compariso
n:

dds <- DESeq(dataset)

print(dds)

## class: DESeqDataSet
## dim: 19332 16
## metadata(1): version
## assays(6): counts mu ... replaceCounts replaceCooks
## rownames(19332): ENSG00000000419 ENSG00000000457 ... ENSG000002
91338
##   ENSG00000292285
## rowData names(23): baseMean baseVar ... maxCooks replace
## colnames(16): P3052NP P3090NP ... P3240SPE P3330PE
## colData names(3): condition sizeFactor replaceable

result8 <- results(dds, contrast=c('condition','NP','PE'), alpha = 0.05)
summary(result8)

##
## out of 18658 with nonzero total read count
## adjusted p–value < 0.05
## LFC > 0 (up)       : 4, 0.021%
## LFC < 0 (down)     : 0, 0%
## outliers [1]       : 4257, 23%
## low counts [2]     : 674, 3.6%
## (mean count < 0)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results

DEGup8 = result8[ which(result8$padj < 0.05 & result8$log2FoldChange > 1), ]
DEGdown8 = result8[ which(result8$padj < 0.05 & result8$log2FoldChange < 1),
```

```
]
W8_DEGs = rbind(DEGup8, DEGdown8)
#Extract the up and down regulated gene file
#write.csv(W8_DEGs, "W8_DEGS_p0.05.csv")
```

9.6.2.1 Volcano plot for gestational week 8

```
# Load the library
library(ggplot2)

# Create a volcano plot
volcano_plot8 <-  ggplot(data = result8, aes(x = log2FoldChange, y = –log10(pv
alue))) +
  geom_point(aes(color = ifelse(abs(log2FoldChange) > 1 & padj < 0.05, "red",
"black")), alpha = 0.6) +
  scale_color_identity() +
  geom_hline(yintercept = –log10(0.05), linetype = "dashed") +
  geom_vline(xintercept = c(–1, 1), linetype = "dashed") +
  labs(x = "log2 Fold Change", y = "–log10(adjusted p–value)", title = "Volcano Pl
ot –  Week 8") +
  theme_minimal()

# Print the volcano plot
print(volcano_plot8)
```

```
## Warning: Removed 4257 rows containing missing values or values outside the
scale range
## (`geom_point()`).
```



```
# Load the library
library(ggplot2)
```

9.6.3 Step 3: Week 9

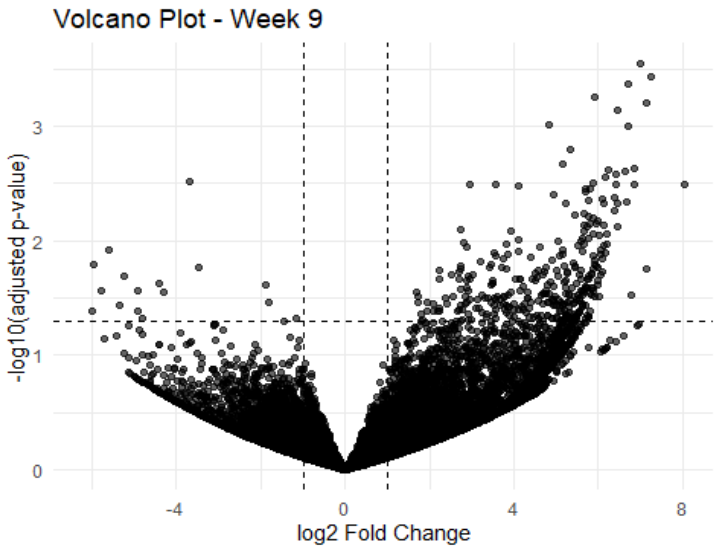
```
#Condition and Coldata
condition <- as.factor(c(rep("NP", 10), rep("PE", 3)))
colData <- data.frame(row.names=colnames(W9count), condition=factor(cond
ition, levels=c('NP', 'PE')))

#DEseqDataset
dataset <- DESeqDataSetFromMatrix(countData = W9count,
                              colData = colData,
                              design = ~condition)

# pre–filtering: removing rows with low gene counts
# keeping rows that have at least 10 reads total
keep <- rowSums(counts(dataset)) >= 10
dataset <- dataset[keep,]

# set the factor level
dataset$condition <- relevel(dataset$condition, ref = "NP")

#Then run the DESeq2 algorithm and extract results for the two-class compariso
n:

dds <- DESeq(dataset)

print(dds)

## class: DESeqDataSet
## dim: 15249 13
## metadata(1): version
## assays(6): counts mu ... replaceCounts replaceCooks
## rownames(15249): ENSG00000000419 ENSG00000000457 ... ENSG000002
91338
##   ENSG00000292277
## rowData names(23): baseMean baseVar ... maxCooks replace
## colnames(13): P3062NP P3102NP ... P3074SPE P3191PE
## colData names(3): condition sizeFactor replaceable

result9 <- results(dds, contrast=c('condition','NP','PE'), alpha = 0.05)
result9 <- results(dds, contrast=c('condition','NP','PE'), alpha = 0.05)
summary(result9)
```

```
##
## out of 15008 with nonzero total read count
## adjusted p–value < 0.05
## LFC > 0 (up)       : 0, 0%
## LFC < 0 (down)     : 0, 0%
## outliers [1]       : 647, 4.3%
## low counts [2]     : 241, 1.6%
## (mean count < 0)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results

DEGup9 = result9[ which(result9$padj < 0.05 & result9$log2FoldChange > 0), ]
DEGdown9 = result9[ which(result9$padj < 0.05 & result9$log2FoldChange < 0),
]
W9_DEGs = rbind(DEGup9, DEGdown9)

#Extract the up and down regulated gene file
#write.csv(W9_DEGs, "W9_DEGS_p0.05.csv")
```

9.6.3.1 Volcano plot for gestational week 9

```
# Load the library
library(ggplot2)

# Assuming 'result' contains the DESeq2 analysis results

# Create a volcano plot
volcano_plot9 <-  ggplot(data = result9, aes(x = log2FoldChange, y = –log10(pv
alue))) +
  geom_point(aes(color = ifelse(abs(log2FoldChange) > 1 & padj < 0.05, "red",
"black")), alpha = 0.6) +
  scale_color_identity() +
  geom_hline(yintercept = –log10(0.05), linetype = "dashed") +
  geom_vline(xintercept = c(–1, 1), linetype = "dashed") +
  labs(x = "log2 Fold Change", y = "–log10(adjusted p–value)", title = "Volcano Pl
ot –  Week 9") +
  theme_minimal()

# Print the volcano plot
print(volcano_plot9)
```

```
## Warning: Removed 647 rows containing missing values or values outside the
scale range
## (`geom_point()`).
```



9.6.4 Step 4: Week 10

```
#Condition and Coldata
condition <- as.factor(c(rep("NP", 15), rep("PE", 4)))
colData <- data.frame(row.names=colnames(W10count), condition=factor(con
dition, levels=c('NP', 'PE')))

#DEseqDataset
dataset <- DESeqDataSetFromMatrix(countData = W10count,
                          colData = colData,
                          design = ~condition)

# pre–filtering: removing rows with low gene counts
# keeping rows that have at least 10 reads total
keep <- rowSums(counts(dataset)) >= 10
dataset <- dataset[keep,]

# set the factor level
dataset$condition <- relevel(dataset$condition, ref = "NP")

#Then run the DESeq2 algorithm and extract results for the two-class compariso
n:

dds <- DESeq(dataset)

print(dds)

## class: DESeqDataSet
## dim: 17007 19
```

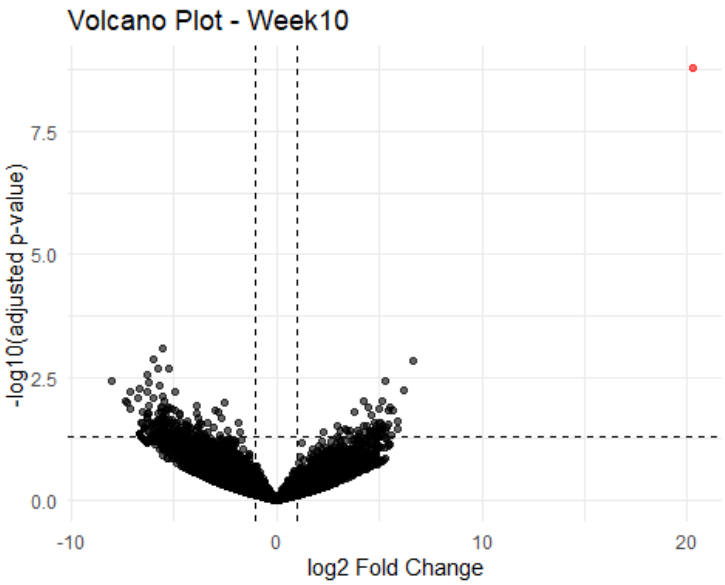
```
## metadata(1): version
## assays(6): counts mu ... replaceCounts replaceCooks
## rownames(17007): ENSG00000000419 ENSG00000000457 ... ENSG000002
92277
##   ENSG00000292285
## rowData names(23): baseMean baseVar ... maxCooks replace
## colnames(19): P3012NP P3059NP ... P3256PE P3302PE
## colData names(3): condition sizeFactor replaceable

result10 <– results(dds, contrast=c('condition','NP','PE'), alpha = 0.05)
result10 <– results(dds, contrast=c('condition','NP','PE'), alpha = 0.05)
summary(result10)

##
## out of 16614 with nonzero total read count
## adjusted p–value < 0.05
## LFC > 0 (up)       : 1, 0.006%
## LFC < 0 (down)     : 0, 0%
## outliers [1]       : 1207, 7.3%
## low counts [2]     : 393, 2.4%
## (mean count < 0)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results

DEGup10 = result10[ which(result10$padj < 0.05 & result10$log2FoldChange >
1), ]
DEGdown10 = result10[ which(result10$padj < 0.05 & result10$log2FoldChange
< 1), ]
W10_DEGs = rbind(DEGup10, DEGdown10)

#Extract the up and down regulated gene file
#write.csv(W10_DEGs, "W10_DEGS_p0.05.csv")
```

9.6.4.1 Volcano plot for gestational week 10

```
# Create a volcano plot
volcano_plot10 <– ggplot(data = result10, aes(x = log2FoldChange, y = –log10(
pvalue))) +
  geom_point(aes(color = ifelse(abs(log2FoldChange) > 1 & padj < 0.05, "red",
"black")), alpha = 0.6) +
  scale_color_identity() +
  geom_hline(yintercept = –log10(0.05), linetype = "dashed") +
  geom_vline(xintercept = c(–1, 1), linetype = "dashed") +
  labs(x = "log2 Fold Change", y = "–log10(adjusted p–value)", title = "Volcano Pl
ot –  Week10") +
  theme_minimal()
```

```
# Print the volcano plot
print(volcano_plot10)
```

```
## Warning: Removed 1207 rows containing missing values or values outside the
scale range
## (`geom_point()`).
```



9.6.5 Step 5: Week 11

```
#Remove single end data
W11count = subset(W11count, select = -S4611PE)

#Condition and Coldata
condition <- as.factor(c(rep("NP", 34), rep("PE", 12)))
colData <- data.frame(row.names=colnames(W11count), condition=factor(con
dition, levels=c('NP', 'PE')))

#DEseqDataset
dataset <- DESeqDataSetFromMatrix(countData = W11count,
                                  colData = colData,
                                  design = ~condition)

# pre-filtering: removing rows with low gene counts
# keeping rows that have at least 10 reads total
keep <- rowSums(counts(dataset)) >= 10
dataset <- dataset[keep,]
```

```r
# set the factor level
dataset$condition <- relevel(dataset$condition, ref = "NP")

#Then run the DESeq2 algorithm and extract results for the two-class compariso
n:
dds <- DESeq(dataset)

print(dds)
```

```
## class: DESeqDataSet
## dim: 23915 46
## metadata(1): version
## assays(6): counts mu ... replaceCounts replaceCooks
## rownames(23915): ENSG00000000419 ENSG00000000457 ... ENSG000002
92277
##   ENSG00000292285
## rowData names(23): baseMean baseVar ... maxCooks replace
## colnames(46): P2994NP P2995NP ... P3377SPE P3390SPE
## colData names(3): condition sizeFactor replaceable
```

```r
#result11 <- results(dds, contrast=c('condition','NP','PE'), alpha = 0.05)
result11 <- results(dds, contrast=c('condition','NP','PE'), alpha = 0.05)
summary(result11)
```

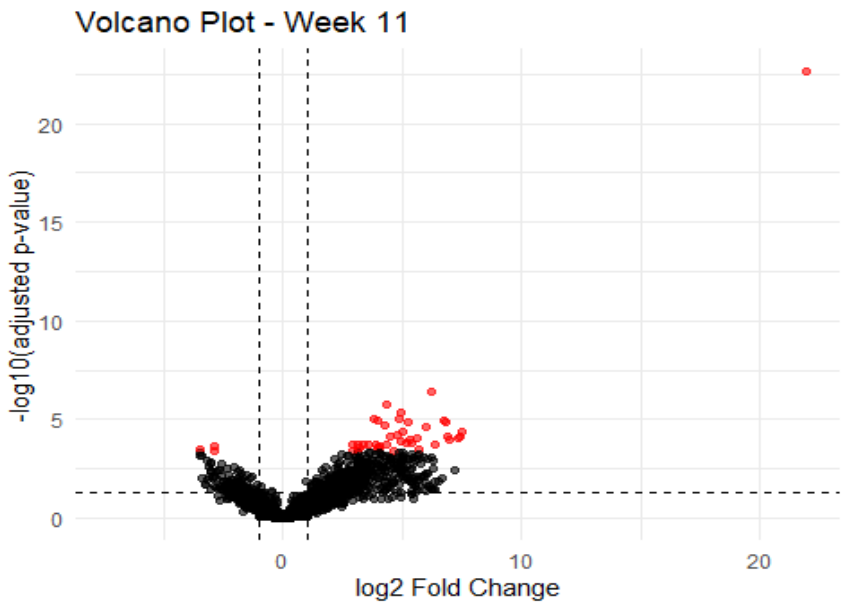
```
##
## out of 22174 with nonzero total read count
## adjusted p-value < 0.05
## LFC > 0 (up)       : 40, 0.18%
## LFC < 0 (down)     : 3, 0.014%
## outliers [1]       : 0, 0%
## low counts [2]     : 18438, 83%
## (mean count < 4)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

```r
DEGup11 = result11[ which(result11$padj < 0.05 & result11$log2FoldChange >
1), ]
DEGdown11 = result11[ which(result11$padj < 0.05 & result11$log2FoldChange
< 1), ]
W11_DEGs = rbind(DEGup11, DEGdown11)
#W11_DEGs
#Extract the up and down regulated gene file
#write.csv(W11_DEGs, "W11_DEGS_p0.05.csv")
```

9.6.5.1 Volcano plot for gestational week 11

```
# Create a volcano plot
volcano_plot11 <- ggplot(data = result11, aes(x = log2FoldChange, y = -log10(
pvalue))) +
  geom_point(aes(color = ifelse(abs(log2FoldChange) > 1 & padj < 0.05, "red",
"black")), alpha = 0.6) +
  scale_color_identity() +
  geom_hline(yintercept = -log10(0.05), linetype = "dashed") +
  geom_vline(xintercept = c(-1, 1), linetype = "dashed") +
  labs(x = "log2 Fold Change", y = "-log10(adjusted p-value)", title = "Volcano Pl
ot -  Week 11") +
  theme_minimal()

# Print the volcano plot
print(volcano_plot11)
```

```
## Warning: Removed 11528 rows containing missing values or values outside th
e scale range
## (`geom_point()`).
```

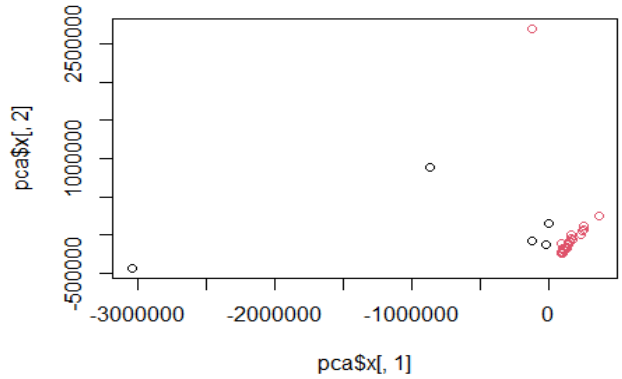


9.6.6 Step 6: Week 12

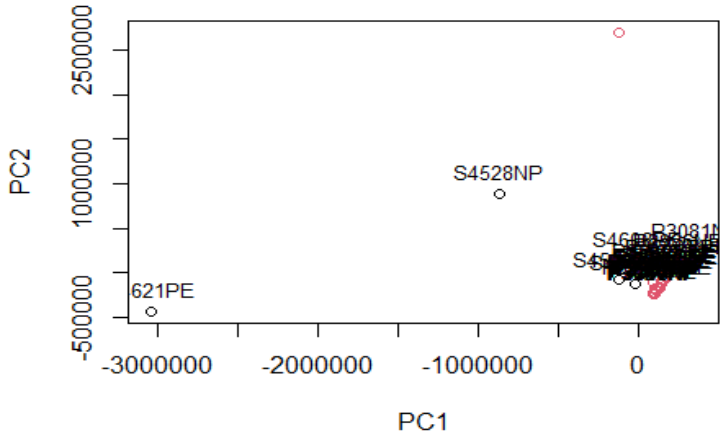
9.6.6.1 Batch effect removal

Batch identification

```
pca <- prcomp(t(W12count))
batch = c(rep(1, 1), rep(2,21), rep(1,4), rep(2, 8))
plot(pca$x[,1], pca$x[,2], col = batch)
```




```
pca <- prcomp(t(W12count))
batch <- c(rep(1, 1), rep(2,21), rep(1,4), rep(2, 8))
plot(pca$x[, 1], pca$x[, 2], col = batch, xlab = "PC1", ylab = "PC2")
# Add sample names as text labels
text(pca$x[, 1], pca$x[, 2], labels = colnames(W12count), pos = 3, cex = 0.8)
```



Normalised count – Batch effect

```
library(sva)
```

```r
# Convert count to matrix
countMatrix = as.matrix(W12count)

# Define batch
batch = c(rep(1, 1), rep(2,21), rep(1,4), rep(2, 8))

# include condition (group variable)
adjusted_counts <- ComBat_seq(countMatrix, batch=batch, full_mod=TRUE)

## Found 2 batches
## Using null model in ComBat–seq.
## Adjusting for 0 covariate(s) or covariate level(s)
## Estimating dispersions
## Fitting the GLM model
## Shrinkage off – using GLM estimates for parameters
## Adjusting the data

adjusted_countsdf12 = as.data.frame(adjusted_counts)

#Condition and Coldata
condition <- as.factor(c(rep("NP", 22), rep("PE", 12)))
colData <- data.frame(row.names=colnames(adjusted_countsdf12), condition=
factor(condition, levels=c('NP', 'PE')))

#DEseqDataset
dataset <- DESeqDataSetFromMatrix(countData = adjusted_countsdf12,
                                  colData = colData,
                                  design = ~condition)

## converting counts to integer mode

# pre–filtering: removing rows with low gene counts
# keeping rows that have at least 10 reads total
keep <- rowSums(counts(dataset)) >= 10
dataset <- dataset[keep,]

# set the factor level
dataset$condition <- relevel(dataset$condition, ref = "NP")

#Then run the DESeq2 algorithm and extract results for the two-class compariso
n:
dds <- DESeq(dataset)

print(dds)

## class: DESeqDataSet
## dim: 25794 34
```

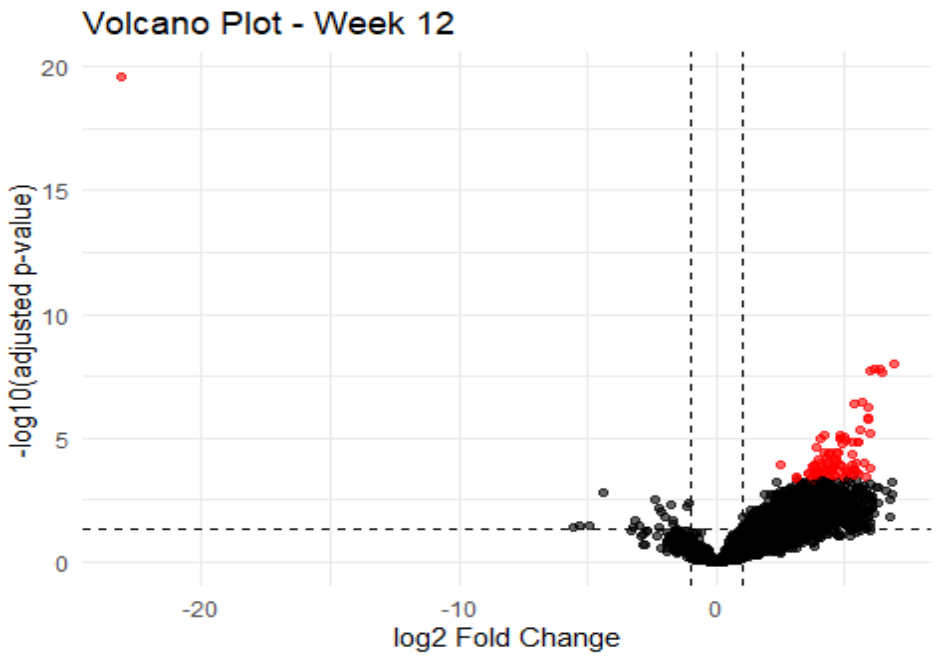
```
## metadata(1): version
## assays(6): counts mu ... replaceCounts replaceCooks
## rownames(25794): ENSG00000000003 ENSG00000000419 ... ENSG000002
92277
##   ENSG00000292285
## rowData names(23): baseMean baseVar ... maxCooks replace
## colnames(34): S4528NP P2996NP ... P3249SPE P3250SPE
## colData names(3): condition sizeFactor replaceable

result12 <– results(dds, contrast=c('condition','NP','PE'), alpha = 0.05)
result12 <– results(dds, contrast=c('condition','NP','PE'), alpha = 0.05)
summary(result12)

##
## out of 25392 with nonzero total read count
## adjusted p–value < 0.05
## LFC > 0 (up)       : 87, 0.34%
## LFC < 0 (down)     : 1, 0.0039%
## outliers [1]       : 0, 0%
## low counts [2]     : 16634, 66%
## (mean count < 3)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results

DEGup12 = result12[ which(result12$padj < 0.05 & result12$log2FoldChange >
1), ]
DEGdown12 = result12[ which(result12$padj < 0.05 & result12$log2FoldChange
< 1), ]
W12_DEGs = rbind(DEGup12, DEGdown12)

#Extract the up and down regulated gene file
#write.csv(W12_DEGs, "W12_DEGS_p0.05.csv")
```

9.6.6.1.1 Volcano plot for gestational week 12

```
# Create a volcano plot
volcano_plot12 <– ggplot(data = result12, aes(x = log2FoldChange, y = –log10(
pvalue))) +
  geom_point(aes(color = ifelse(abs(log2FoldChange) > 1 & padj < 0.05, "red",
"black")), alpha = 0.6) +
  scale_color_identity() +
  geom_hline(yintercept = –log10(0.05), linetype = "dashed") +
  geom_vline(xintercept = c(–1, 1), linetype = "dashed") +
  labs(x = "log2 Fold Change", y = "–log10(adjusted p–value)", title = "Volcano Pl
ot –  Week 12") +
  theme_minimal()
```

```
# Print the volcano plot
print(volcano_plot12)
```

```
## Warning: Removed 9259 rows containing missing values or values outside the
scale range
## (`geom_point()`).
```



9.6.7 Step 7: Week 13

```
# Batch effect removal
library(sva)
# Convert count to matrix
countMatrix = as.matrix(W13count)

# Define batch
batch = c(rep(1, 3), rep(2,14), rep(1,1), rep(2, 11))

# include condition (group variable)
adjusted_counts <-  ComBat_seq(countMatrix, batch=batch, full_mod=TRUE)
```

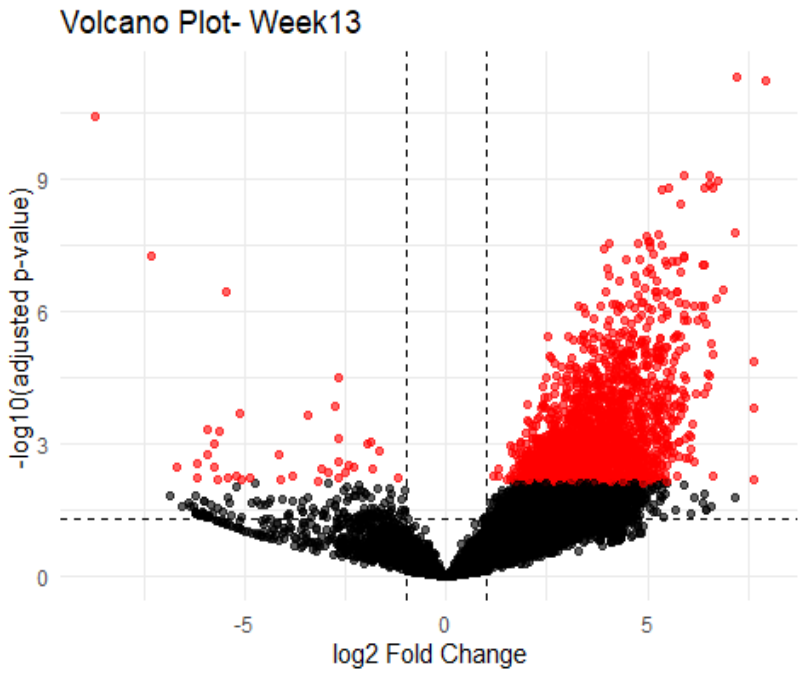
```
## Found 2 batches
## Using null model in ComBat–seq.
## Adjusting for 0 covariate(s) or covariate level(s)
## Estimating dispersions
## Fitting the GLM model
```

```
## Shrinkage off –  using GLM estimates for parameters
## Adjusting the data

adjusted_countsdf13 = as.data.frame(adjusted_counts)


#Condition and Coldata
condition <-  as.factor(c(rep("NP", 17), rep("PE", 12)))
colData <-  data.frame(row.names=colnames(adjusted_countsdf13), condition=
factor(condition, levels=c('NP', 'PE')))

#DEseqDataset
dataset <-  DESeqDataSetFromMatrix(countData = adjusted_countsdf13,
                           colData = colData,
                           design = ~condition)

## converting counts to integer mode

# pre–filtering: removing rows with low gene counts
# keeping rows that have at least 10 reads total
keep <-  rowSums(counts(dataset)) >= 10
dataset <-  dataset[keep,]

# set the factor level
dataset$condition <-  relevel(dataset$condition, ref = "NP")

#Then run the DESeq2 algorithm and extract results for the two-class compariso
n:
dds <-  DESeq(dataset)

print(dds)

## class: DESeqDataSet
## dim: 26061 29
## metadata(1): version
## assays(6): counts mu ... replaceCounts replaceCooks
## rownames(26061): ENSG00000000419 ENSG00000000457 ... ENSG000002
92277
##   ENSG00000292285
## rowData names(23): baseMean baseVar ... maxCooks replace
## colnames(29): S4523NP S4538NP ... P3284PE P3297SPE
## colData names(3): condition sizeFactor replaceable

result13 <-  results(dds, contrast=c('condition','NP','PE'), alpha = 0.05)
summary(result13)
```

```
##
## out of 24950 with nonzero total read count
## adjusted p–value < 0.05
## LFC > 0 (up)       : 1886, 7.6%
## LFC < 0 (down)     : 37, 0.15%
## outliers [1]       : 0, 0%
## low counts [2]     : 13176, 53%
## (mean count < 2)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results

DEGup13 = result13[ which(result13$padj < 0.05 & result13$log2FoldChange >
1), ]
DEGdown13 = result13[ which(result13$padj < 0.05 & result13$log2FoldChange
< 1), ]
W13_DEGs = rbind(DEGup13, DEGdown13)

#Extract the up and down regulated gene file
#write.csv(W13_DEGs, "W13_DEGS_p0.05.csv")
```

9.6.7.1 Volcano plot for gestational week 13

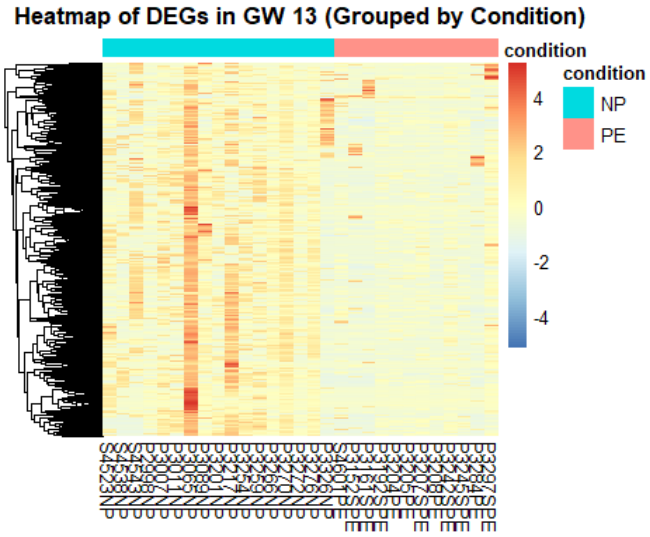
```
# Create a volcano plot
volcano_plot13 <–  ggplot(data = result13, aes(x = log2FoldChange, y = –log10(
pvalue))) +
  geom_point(aes(color = ifelse(abs(log2FoldChange) > 1 & padj < 0.05, "red",
"black")), alpha = 0.6) +
  scale_color_identity() +
  geom_hline(yintercept = –log10(0.05), linetype = "dashed") +
  geom_vline(xintercept = c(–1, 1), linetype = "dashed") +
  labs(x = "log2 Fold Change", y = "–log10(adjusted p–value)", title = "Volcano Pl
ot–  Week13") +
  theme_minimal()

# Print the volcano plot
print(volcano_plot13)

## Warning: Removed 7249 rows containing missing values or values outside the
scale range
## (`geom_point()`).
```



9.6.7.1.1 Heat map for gestational week 13

```
#install.packages("pheatmap")  # If not installed
library(pheatmap)

# Extract the gene symbols for all significant genes
all_geneID <- row.names(W13_DEGs)

# Subset the count data to include only the all significant genes
all_gene_expression <- W13count[all_geneID, ]


# Normalize the data using rlog
rlog_data <- rlog(dds, blind = TRUE)

# Subset the normalized data for the all significant genes
all_gene_rlog_expression <- assay(rlog_data)[all_geneID, ]

png("GW13_heatmap_large.png", width = 2000, height = 3500, res = 300) # widt
h/height in pixels
pheatmap(all_gene_rlog_expression,
         scale = "row",              # Scale by row (gene)
         clustering_distance_rows = "euclidean",
         clustering_distance_cols = "euclidean",
         clustering_method = "complete",
```

```
        cluster_cols = FALSE,         # Disable clustering for columns
        show_rownames = FALSE,        # Hide gene names if too many
        show_colnames = TRUE,         # Show sample names
        annotation_col = colData,  # Use sorted annotations
        main = "Heatmap of DEGs in GW 13 (Grouped by Condition)")
```




```
dev.off()

## png
##   3
```

9.6.8 Step 8: Week 14

```
# Batch effect removal
library(sva)
# Convert count to matrix
countMatrix = as.matrix(W14count)

# Define batch
batch = c(rep(1, 1), rep(2,5), rep(1,1), rep(2, 1))

# include condition (group variable)
adjusted_counts <-  ComBat_seq(countMatrix, batch=batch, full_mod=TRUE)

## Found 2 batches
## Using null model in ComBat-seq.
## Adjusting for 0 covariate(s) or covariate level(s)
## Estimating dispersions
## Fitting the GLM model
```

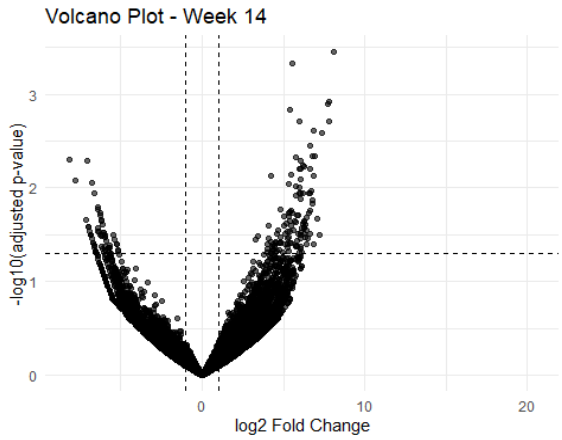
```
## Shrinkage off –  using GLM estimates for parameters
## Adjusting the data

adjusted_countsdf14 = as.data.frame(adjusted_counts)

#Condition and Coldata
condition <–  as.factor(c(rep("NP", 6), rep("PE", 2)))
colData <–  data.frame(row.names=colnames(adjusted_countsdf14), condition=
factor(condition, levels=c('NP', 'PE')))

#DEseqDataset
dataset <–  DESeqDataSetFromMatrix(countData = adjusted_countsdf14,
                           colData = colData,
                           design = ~condition)

## converting counts to integer mode

# pre–filtering: removing rows with low gene counts
# keeping rows that have at least 10 reads total
keep <–  rowSums(counts(dataset)) >= 10
dataset <–  dataset[keep,]

# set the factor level
dataset$condition <–  relevel(dataset$condition, ref = "NP")

#Then run the DESeq2 algorithm and extract results for the two-class compariso
n:
dds <–  DESeq(dataset)

print(dds)

## class: DESeqDataSet
## dim: 14578 8
## metadata(1): version
## assays(4): counts mu H cooks
## rownames(14578): ENSG00000000003 ENSG00000000419 ... ENSG000002
91336
##   ENSG00000291338
## rowData names(22): baseMean baseVar ... deviance maxCooks
## colnames(8): S4518NP P3005NP ... S4591PE P3181SPE
## colData names(2): condition sizeFactor

result14 <–  results(dds, contrast=c('condition','NP','PE'), alpha = 0.05)
result14 <–  results(dds, contrast=c('condition','NP','PE'), alpha = 0.05)
summary(result14)
```

```
##
## out of 14578 with nonzero total read count
## adjusted p–value < 0.05
## LFC > 0 (up)       : 0, 0%
## LFC < 0 (down)     : 0, 0%
## outliers [1]       : 1914, 13%
## low counts [2]     : 0, 0%
## (mean count < 0)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results

DEGup14 = result14[which(result14$padj < 0.05 & result14$log2FoldChange > 1), ]
DEGdown14 = result14[which(result14$padj < 0.05 & result14$log2FoldChange < 1), ]
W14_DEGs = rbind(DEGup14, DEGdown14)

#Extract the up and down regulated gene file
#write.csv(W14_DEGs, "W14_DEGS_p0.05.csv")
```

9.6.8.1 Volcano plot for gestational week 14

```
# Create a volcano plot
volcano_plot14 <–  ggplot(data = result14, aes(x = log2FoldChange, y = –log10(pvalue))) +
  geom_point(aes(color = ifelse(abs(log2FoldChange) > 1 & padj < 0.05, "red", "black")), alpha = 0.6) +
  scale_color_identity() +
  geom_hline(yintercept = –log10(0.05), linetype = "dashed") +
  geom_vline(xintercept = c(–1, 1), linetype = "dashed") +
  labs(x = "log2 Fold Change", y = "–log10(adjusted p–value)", title = "Volcano Plot –  Week 14") +
  theme_minimal()

# Print the volcano plot
print(volcano_plot14)
```

```
## Warning: Removed 1914 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



9.6.9 Step 9: Week 15

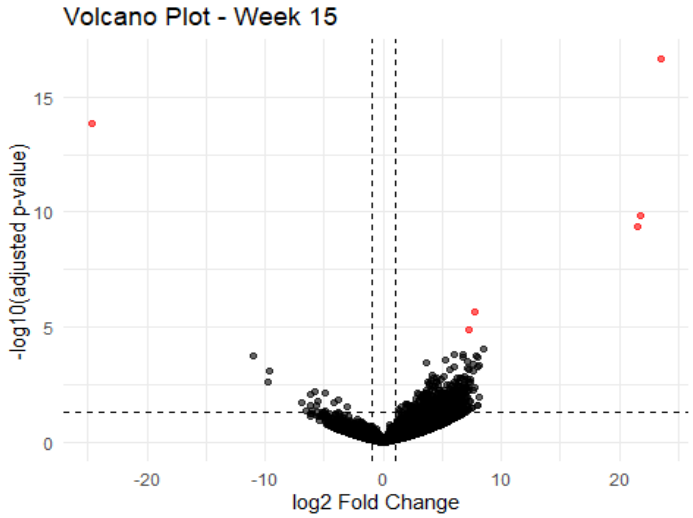
```
# Batch effect removal
library(sva)
# Convert count to matrix
countMatrix = as.matrix(W15count)

# Define batch
batch = c(rep(1, 1), rep(2,22), rep(1,1), rep(2, 6))

# include condition (group variable)
adjusted_counts <- ComBat_seq(countMatrix, batch=batch, full_mod=TRUE)

## Found 2 batches
## Using null model in ComBat–seq.
## Adjusting for 0 covariate(s) or covariate level(s)
## Estimating dispersions
## Fitting the GLM model
## Shrinkage off – using GLM estimates for parameters
## Adjusting the data

adjusted_countsdf15 = as.data.frame(adjusted_counts)

#Condition and Coldata
condition <- as.factor(c(rep("NP", 23), rep("PE", 7)))
colData <- data.frame(row.names=colnames(adjusted_countsdf15), condition=
factor(condition, levels=c('NP', 'PE')))

#DEseqDataset
dataset <- DESeqDataSetFromMatrix(countData = adjusted_countsdf15,
                                  colData = colData,
                                  design = ~condition)
```

```
## converting counts to integer mode

# pre–filtering: removing rows with low gene counts
# keeping rows that have at least 10 reads total
keep <– rowSums(counts(dataset)) >= 10
dataset <– dataset[keep,]

# set the factor level
dataset$condition <– relevel(dataset$condition, ref = "NP")

#Then run the DESeq2 algorithm and extract results for the two-class compariso
n:
dds <– DESeq(dataset)

print(dds)

## class: DESeqDataSet
## dim: 23978 30
## metadata(1): version
## assays(6): counts mu ... replaceCounts replaceCooks
## rownames(23978): ENSG00000000003 ENSG00000000419 ... ENSG000002
92277
##   ENSG00000292285
## rowData names(23): baseMean baseVar ... maxCooks replace
## colnames(30): S4513NP P3004NP ... P3322PE P3360SPE
## colData names(3): condition sizeFactor replaceable

result15 <– results(dds, contrast=c('condition','NP','PE'), alpha = 0.05)
result15 <– results(dds, contrast=c('condition','NP','PE'), alpha = 0.05)
summary(result15)

##
## out of 21422 with nonzero total read count
## adjusted p–value < 0.05
## LFC > 0 (up)       : 5, 0.023%
## LFC < 0 (down)     : 1, 0.0047%
## outliers [1]       : 0, 0%
## low counts [2]     : 2556, 12%
## (mean count < 0)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results

DEGup15 = result15[ which(result15$padj < 0.05 & result15$log2FoldChange >
1), ]
DEGdown15 = result15[ which(result15$padj < 0.05 & result15$log2FoldChange
< 1), ]
W15_DEGs = rbind(DEGup15, DEGdown15)
```

```
#Extract the up and down regulated gene file
#write.csv(W15_DEGs, "W15_DEGS_p0.05.csv")
```

9.6.9.1 Volcano plot for gestational week 15

```
# Create a volcano plot
volcano_plot15 <-  ggplot(data = result15, aes(x = log2FoldChange, y = –log10(
pvalue))) +
  geom_point(aes(color = ifelse(abs(log2FoldChange) > 1 & padj < 0.05, "red",
"black")), alpha = 0.6) +
  scale_color_identity() +
  geom_hline(yintercept = –log10(0.05), linetype = "dashed") +
  geom_vline(xintercept = c(–1, 1), linetype = "dashed") +
  labs(x = "log2 Fold Change", y = "–log10(adjusted p–value)", title = "Volcano Pl
ot –  Week 15") +
  theme_minimal()

# Print the volcano plot
print(volcano_plot15)
```



9.6.10 Step 10: Week 16

```
# Batch effect removal
library(sva)
# Convert count to matrix
countMatrix = as.matrix(W16count)

# Define batch
```

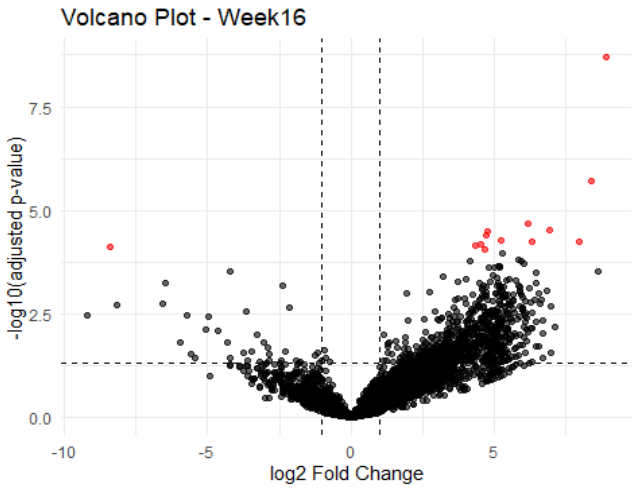
```
batch = c(rep(1, 2), rep(2,30))

# include condition (group variable)
adjusted_counts <- ComBat_seq(countMatrix, batch=batch, full_mod=TRUE)

## Found 2 batches
## Using null model in ComBat–seq.
## Adjusting for 0 covariate(s) or covariate level(s)
## Estimating dispersions
## Fitting the GLM model
## Shrinkage off – using GLM estimates for parameters
## Adjusting the data

adjusted_countsdf16 = as.data.frame(adjusted_counts)

#Condition and Coldata
condition <- as.factor(c(rep("NP", 22), rep("PE", 10)))
colData <- data.frame(row.names=colnames(adjusted_countsdf16), condition=
factor(condition, levels=c('NP', 'PE')))

#DEseqDataset
dataset <- DESeqDataSetFromMatrix(countData = adjusted_countsdf16,
                                  colData = colData,
                                  design = ~condition)

## converting counts to integer mode

# pre–filtering: removing rows with low gene counts
# keeping rows that have at least 10 reads total
keep <- rowSums(counts(dataset)) >= 10
dataset <- dataset[keep,]

# set the factor level
dataset$condition <- relevel(dataset$condition, ref = "NP")

#Then run the DESeq2 algorithm and extract results for the two-class compariso
n::
dds <- DESeq(dataset)

print(dds)

## class: DESeqDataSet
## dim: 23506 32
## metadata(1): version
## assays(6): counts mu ... replaceCounts replaceCooks
## rownames(23506): ENSG00000000003 ENSG00000000419 ... ENSG000002
92277
```

```
##   ENSG00000292285
## rowData names(23): baseMean baseVar ... maxCooks replace
## colnames(32): S4533NP S4548NP ... P3282PE P3314SPE
## colData names(3): condition sizeFactor replaceable

result16 <- results(dds, contrast=c('condition','NP','PE'), alpha = 0.05)
result16 <- results(dds, contrast=c('condition','NP','PE'), alpha = 0.05)
summary(result16)

##
## out of 22618 with nonzero total read count
## adjusted p-value < 0.05
## LFC > 0 (up)       : 12, 0.053%
## LFC < 0 (down)     : 1, 0.0044%
## outliers [1]       : 0, 0%
## low counts [2]     : 16204, 72%
## (mean count < 3)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results

DEGup16 = result16[ which(result16$padj < 0.05 & result16$log2FoldChange >
1), ]
DEGdown16 = result16[ which(result16$padj < 0.05 & result16$log2FoldChange
< 1), ]
W16_DEGs = rbind(DEGup16, DEGdown16)

#Extract the up and down regulated gene file
#write.csv(W16_DEGs, "W16_DEGS_p0.05.csv")
```

9.6.10.1 Volcano plot for gestational week 16

```
# Create a volcano plot
volcano_plot16 <- ggplot(data = result16, aes(x = log2FoldChange, y = -log10(
pvalue))) +
  geom_point(aes(color = ifelse(abs(log2FoldChange) > 1 & padj < 0.05, "red",
"black")), alpha = 0.6) +
  scale_color_identity() +
  geom_hline(yintercept = -log10(0.05), linetype = "dashed") +
  geom_vline(xintercept = c(-1, 1), linetype = "dashed") +
  labs(x = "log2 Fold Change", y = "-log10(adjusted p-value)", title = "Volcano Pl
ot -  Week16") +
  theme_minimal()

# Print the volcano plot
print(volcano_plot16)
```

```
## Warning: Removed 9272 rows containing missing values or values outside the
scale range
## (`geom_point()`).
```



9.6.11 Step 11: Merging volcano plots

```
# Load necessary libraries
library(cowplot)

# Assume you have volcano plots named volcano_plot8, volcano_plot9, ..., volca
no_plot16

# Merge volcano plots into one sheet
merged_plot <-  plot_grid(volcano_plot8, volcano_plot9, volcano_plot10, volcan
o_plot11,
                          volcano_plot12, volcano_plot13, volcano_plot14, volcano_plot15,
                          volcano_plot16, ncol = 3)

# Print the merged plot
print(merged_plot)
```

9.7 Appendix B

Consistent gene expression analysis

9.7.1 Step 1: Set the working directory

```
import os
os.chdir("F:/CGE_June2024/CGE_10_2024")
```

9.7.2 Step 2: Load the required Python libraries

```
import pandas as pd
import numpy as np
from scipy.stats import rankdata
from statsmodels.stats.multitest import multipletests
from itertools import permutations
import random
```

9.7.3 Step 3: Load the data and data pre–processing for consistent genes analysis process

9.7.3.1 Step 3.1: Load the data

```
gene_expression_data = pd.read_csv("NP13_TMM_normalized_counts.csv", ind
ex_col = 0)
# Rename the index column
gene_expression_data.index.name = 'GeneID'


gene_expression_data.shape
```

```
(23962, 17)
```

9.7.3.2 Step 3.2: Convert data frame into array data

```
array = np.array(gene_expression_data)
```

9.7.3.3 Step 3.3: Join the Gene ID column with array data

```
index_values = gene_expression_data.index.values
index_column_name = gene_expression_data.index.name
mapped_array = np.hstack((index_values.reshape(-1, 1), array))
```

9.7.4 Step 4: Calculate rank product

9.7.4.1 Step 4.1: Rank the gene expression levels

```
# Extract the expression columns from the mapped array
expression_columns = mapped_array[:, 1:].astype(float)

# Rank the expression columns
ranked_columns = np.apply_along_axis(lambda x: np.argsort(-x).argsort() + 1, a
xis=0, arr=expression_columns)

# Create a new array by combining the index column and ranked expression colu
mns
ranked_array = np.hstack((mapped_array[:, 0].reshape(-1, 1), ranked_columns))

# Get the number of columns in the array
num_columns = ranked_array.shape[1]
```

9.7.4.2 Step 4.2: Convert the ranked array data into data frame

```
ranked_df = pd.DataFrame(ranked_array)

# Rename the numeric column names
ranked_df = ranked_df.rename(columns={0: 'GeneID'})
#ranked_df.head(5)
```

9.7.4.3 Step 4.3: Calculate the rank product using RP formula by confirming the values of column 1(GeneID) as float value

```
ranked_array2 = ranked_array[:, 1:].astype(float)

# Number of conditions (number of columns in ranked_array2)
k = ranked_array2.shape[1]

rank_prod = np.prod(ranked_array2, axis=1)

# Calculate the Rank Product using the geometric mean formula
rank_prod = rank_prod ** (1 / k)
#rank_prod
```

9.7.4.4 Step 4.4: Convert the RP array data into data frame

```
rankProductdf = pd.DataFrame(rank_prod)
```

```
# Rename the numeric column names
rankProductdf = rankProductdf.rename(columns={0: 'RP value'})
#rankProductdf.head(5)
```

9.7.4.5 Step 4.5: Join the Gene ID column with RP values

```
# Select the column 1 using iloc
GeneID_column = ranked_df.iloc[:, 0]

# Select the column 1 using iloc
RP_column = rankProductdf.iloc[:, 0]

# Create a new DataFrame by concatenating the two columns
new_df = pd.concat([GeneID_column, RP_column], axis=1)
#new_df.head(5)
```

9.7.4.6 Step 5: RP Table Sorted in Ascending Order

```
# Sort the DataFrame by the 'RP value' column in ascending order
RP_Ascend_sort = new_df.sort_values('RP value', ascending=True)
#RP_Ascend_sort.head(5)
```

9.7.4.7 Step 6: Merged RP table with gene expression

```
# Merge the DataFrames based on the 'ID' column
merged_Ascend_df = pd.merge(RP_Ascend_sort, gene_expression_data, on='G
eneID')

# Set the first column as the index
merged_Ascend_df  = merged_Ascend_df.set_index(merged_Ascend_df.column
s[0])
#merged_Ascend_df.head(5)
```

9.7.4.8 Step 7: RP sorted gene expression data in Ascending order

```
Ascen_sorted_geneexpressionData = merged_Ascend_df.drop('RP value', axis=
1)

# Convert 'GeneID' to the index column
#Ascen_sorted_geneexpressionData.set_index('GeneID', inplace=True)
#Ascen_sorted_geneexpressionData.head(5)
```

9.7.5 Step 8: RP sorted table with Mean Expression and Standard Deviation

9.7.5.1 Step 8.1: Calculating mean expression

```
# Calculate the mean expression without considering the 'GeneID' column
merged_Ascend_df['Mean Expression'] = merged_Ascend_df.drop('RP value', ax
is=1).mean(axis=1)
#Ascend_mean_expression = merged_Ascend_df.drop('RP value', axis=0).mean
()
#merged_Ascend_df.head(5)
```

9.7.5.2 Step 8.2: Calculating standard deviation

```
# Drop columns "RP value', and 'Mean Expression' from the DataFrame
merged_Ascend_df_dropped = merged_Ascend_df.drop(['RP value', 'Mean Expr
ession'], axis=1)

# Calculate the standard deviation by gene (row–wise)
merged_Ascend_df_dropped['std_dev'] = merged_Ascend_df_dropped.std(axis=
1)
#merged_Ascend_df_dropped.head(5)
```

9.7.5.3 Step 8.3: Calculating MAD

```
# Step 1: Calculate the median for each row (gene) across all samples
gene_medians = Ascen_sorted_geneexpressionData.median(axis=1)

# Step 2: Calculate the absolute deviations from the median for each row (gene)
absolute_deviation = Ascen_sorted_geneexpressionData.sub(gene_medians, axi
s=0).abs()

# Step 3: Calculate the MAD for each row (gene)
MAD = absolute_deviation.median(axis=1)

# Step 4: Convert it to a DataFrame
df_MAD = pd.DataFrame(MAD, columns=['MAD'])  # Name the column 'MAD'
```

9.7.5.4 Step 8.4: RP table with mean of gene expression, standard deviation, and MAD

```
columnRP_Mean = merged_Ascend_df[[ 'RP value', 'Mean Expression']]
column_StdDev = merged_Ascend_df_dropped[['std_dev']]

# Create a new DataFrame by concatenating the two columns
```

```
Ascen_RP_SortedTable1 = pd.concat([columnRP_Mean, column_StdDev, df_MAD], axis=1)

# Set the first column as index
#Ascen_RP_SortedTable1.set_index(Ascen_RP_SortedTable1.columns[0], inplace=True)
```

9.7.5.5 Step 8.4: RP table with Coefficient of Variation (CV)

```
# Create a new DataFrame with the CV column, keeping the original intact
Ascend_CV = Ascen_RP_SortedTable1.copy()  # Make a copy of the original DataFrame

# Calculate the Coefficient of Variation (CV = std_dev / Mean Expression)
Ascend_CV['CV'] = Ascend_CV['std_dev'] / Ascend_CV['Mean Expression']
```

9.7.6 Step 9: Permutation for 'p' value calculation for RP table in Ascending order

9.7.6.1 Step 9.1: Select the observed RP value column and convert it into array

```
# Select the column 1 using iloc
rp = Ascen_RP_SortedTable1.iloc[:, 0]
rp_array = np.array(rp)
#rp_array
```

9.7.6.2 Step 9.2: Generate the permuted RP values for p value calculation

```
def calculate_rank_product(gene_expression_data):

    # Convert DataFrame to array
    array_data = gene_expression_data.values
    index_values = gene_expression_data.index.values
    index_column_name = gene_expression_data.index.name
    mapped_array = np.hstack((index_values.reshape(-1, 1), array_data))

    # Extract the expression columns from the mapped array
    expression_columns = mapped_array[:, 1:].astype(float)

    # Rank the expression columns
    ranked_columns = np.apply_along_axis(lambda x: np.argsort(-x).argsort() + 1, axis=0, arr=expression_columns)
```

```
    # Create a new array by combining the index column and ranked expression c
olumns
    ranked_array = np.hstack((mapped_array[:, 0].reshape(-1, 1), ranked_column
s))

    # Get the number of columns in the array
    num_columns = ranked_array.shape[1]

    # Create a copy of the original array
    sorted_array = ranked_array.copy()

    sorted_array = sorted_array[:, 1:].astype(float)
    #Rprod = np.prod(sorted_array, axis=1)
    Rank_Product = np.prod(sorted_array, axis=1)

    return Rank_Product


def permutation_test(gene_expression_data, Rank_Product, num_permutations,
seed=None):
    if seed is not None:
        np.random.seed(seed)
        random.seed(seed)

    permuted_rps = np.zeros((num_permutations, len(Rank_Product)))

    for i in range(num_permutations):
        permuted_data = gene_expression_data.copy()
        for column in gene_expression_data:
            permuted_data[column] = np.random.permutation(permuted_data[column
])

        permuted_rank_product = calculate_rank_product(permuted_data)
        permuted_rps[i] = permuted_rank_product

    return permuted_rps

Rank_Product = calculate_rank_product(Ascen_sorted_geneexpressionData)
#Rank_Product

# Set the random seed
seed_value = 42

# Call permutation_test with the seed
```

```
permutedRP = permutation_test(Ascen_sorted_geneexpressionData, Rank_Prod
uct, num_permutations=1000, seed=seed_value)
#permutedRP
```

9.7.6.3 Step 9.3: Generate the P values

```
def calculate_p_values(rank_product, permuted_values, num_permutations):
    num_genes = len(rank_product)
    p_values = np.zeros(num_genes)

    for i in range(num_genes):
        # Calculate the observed rank product for the current gene
        observed_rank_product = rank_product[i]

        # Calculate the number of permuted rank products that are greater than the
observed rank product
        num_greater = np.sum(permuted_values[:, i] > observed_rank_product)

        # Calculate the p–value for the current gene by dividing the number of perm
uted rank products by the total number of permutations
        p_value = (num_greater + 1) / (num_permutations + 1)

        # Invert the p–value to represent lower consistency levels as higher values
        inverted_p_value = 1 – p_value

        p_values[i] = inverted_p_value

    return p_values
num_permutations=1000
p_values = calculate_p_values(Rank_Product, permutedRP, num_permutations)
p_values

array([0.      , 0.      , 0.      , ..., 0.999001, 0.999001, 0.999001])
```

9.7.6.4 Step 9.4: Generate the P values array into data frame

```
Ascen_pval_df = pd.DataFrame(p_values)

# Rename the column
Ascen_pval_df = Ascen_pval_df.rename(columns={0: 'P value'})
```

9.7.6.5 Step 10: RP table in ascending order with Mean Expression, Standard Deviation, CV, and P values

```
# Assign the index from df1 to df2
Ascen_pval_df.index = Ascen_RP_SortedTable1.index

# Perform the merge based on the index column
Ascen_RP_Sorted_PvalTable = pd.merge(Ascen_RP_SortedTable1, Ascen_pval_df, left_index=True, right_index=True)

# Convert 'Column1' to the index column
#Ascen_RP_Sorted_PvalTable.set_index('GeneID', inplace=True)
#RP_Sorted_PvalTable.to_csv('GW8PE_RP_Sorted_PvalTable.csv')
#Ascen_RP_Sorted_PvalTable.head(5)

# Assign the index from df1 to df2
Ascend_CV.index = Ascend_CV.index

# Perform the merge based on the index column
Ascen_RP_Sorted_PvalTable = pd.merge(Ascend_CV, Ascen_pval_df, left_index=True, right_index=True)

# Convert 'Column1' to the index column
#Ascen_RP_Sorted_PvalTable.set_index('GeneID', inplace=True)
#RP_Sorted_PvalTable.to_csv('GW8PE_RP_Sorted_PvalTable.csv')
#Ascen_RP_Sorted_PvalTable.head(5)
```

Identification of Significant Consistent Genes

9.7.6.6 Step 11: Filtration 1 – Identifying Significant Genes (SCGs) based on p–values

```
# Set the P–value cutoff and calculate the 75th percentile of Mean Expression
p_value_cutoff = 0.05

# Filter genes based on P–value cutoff
SG_bypval_Ascend = Ascen_RP_Sorted_PvalTable[Ascen_RP_Sorted_PvalTable['P value'] <= p_value_cutoff]
SG_bypval_Ascend.to_csv('NP13_SG_bypval_Ascend.csv')
#SG_bypval_Ascend.head(5)

SG_bypval_Ascend.shape

(4558, 6)

# Merge the datasets based on the 'GeneID' column
SG_GeneExp = pd.merge(SG_bypval_Ascend, merged_Ascend_df, on="GeneID
```

```
", how="inner")

# Drop the duplicated columns ('RP value_x', 'Mean Expression_x', etc.)
SG_GeneExp = SG_GeneExp.loc[:, ~SG_GeneExp.columns.str.endswith('_x')]  # Drop columns ending with '_x'
SG_GeneExp = SG_GeneExp.rename(columns=lambda x: x.replace('_y', ''))  # Remove '_y' from column names
SG_GeneExp.to_csv('NP13_SG_bypval_Ascend_WithGeneExp_Metrics.csv')
#SG_GeneExp

# List of column names to remove
columns_to_remove = ['std_dev', 'MAD', 'CV', 'P value', 'Mean Expression', 'RP value']

# Remove columns by their names
SG_GeneExp_cleaned = SG_GeneExp.drop(columns=columns_to_remove)
#SG_GeneExp_cleaned
```

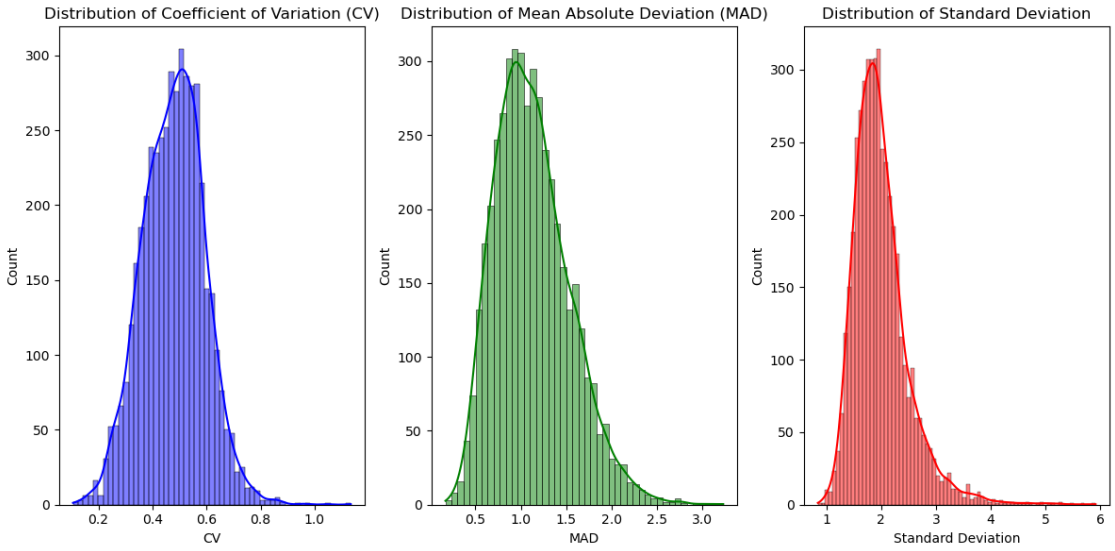
9.7.6.7 Step 12: Filtration 2 – CV

```
# Get the GeneID values from SG_bypval_Ascend
#selected_geneids = SG_bypval_Ascend.index

# Subset the gene_expression_data based on selected GeneIDs
#GeneExp = gene_expression_data[gene_expression_data.index.isin(selected_geneids)]
#GeneExp

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import zscore

# Load gene expression data
# Rows: Genes, Columns: Samples
data = SG_GeneExp_cleaned.copy()

# Normalize the data (log transformation to handle skewed data)
log_data = np.log2(data + 1)

# Function to calculate Coefficient of Variation (CV)
def calculate_cv(gene_expression):
    return np.std(gene_expression) / np.mean(gene_expression)
```

```
# Function to calculate Mean Absolute Deviation (MAD)
def calculate_mad(gene_expression):
    return np.median(np.abs(gene_expression - np.median(gene_expression)))

# Calculate CV, MAD, and Standard Deviation for each gene
cv_values = log_data.apply(calculate_cv, axis=1)
mad_values = log_data.apply(calculate_mad, axis=1)
std_values = log_data.std(axis=1)

# Create a DataFrame to store the consistency metrics
consistency_metrics = pd.DataFrame({
    'CV': cv_values,
    'MAD': mad_values,
    'Standard Deviation': std_values
})

# Rank genes by their consistency (lower values are more consistent)
ranked_genes = consistency_metrics.sort_values(by='CV')  # You can also rank by MAD or std_dev

# Display the ranked genes (showing the most consistent genes at the top)
#print("Genes ranked by Coefficient of Variation (CV):")
#print(ranked_genes)

# Optionally: Visualize the distribution of CV, MAD, and Standard Deviation
plt.figure(figsize=(12, 6))

# Plot Coefficient of Variation (CV)
plt.subplot(1, 3, 1)
sns.histplot(ranked_genes['CV'], kde=True, color='blue')
plt.title('Distribution of Coefficient of Variation (CV)')

# Plot Mean Absolute Deviation (MAD)
plt.subplot(1, 3, 2)
sns.histplot(ranked_genes['MAD'], kde=True, color='green')
plt.title('Distribution of Mean Absolute Deviation (MAD)')

# Plot Standard Deviation
plt.subplot(1, 3, 3)
sns.histplot(ranked_genes['Standard Deviation'], kde=True, color='red')
plt.title('Distribution of Standard Deviation')

plt.tight_layout()
plt.show()
```

```
# Optionally: Filter genes with CV < a threshold (for example, CV < 0.5)
threshold = 0.5
consistent_genes = ranked_genes[ranked_genes['CV'] < threshold]
#print("\nGenes with Coefficient of Variation (CV) < 0.5:")
#print(consistent_genes)
print(len(consistent_genes))
```

```
Genes ranked by Coefficient of Variation (CV):
                     CV      MAD  Standard Deviation
GeneID
ENSG00000276168  0.106097  0.577007            1.568888
ENSG00000105329  0.121926  0.661620            0.996773
ENSG00000087086  0.125883  0.337083            1.200554
ENSG00000067066  0.134387  0.648088            1.001176
ENSG00000184009  0.139542  0.655931            1.371408
...                   ...       ...                 ...
ENSG00000090376  0.889564  0.804524            2.444638
ENSG00000120063  0.946907  0.937357            2.510346
ENSG00000180999  0.972445  0.909585            2.961565
ENSG00000199994  1.046242  0.789543            3.253824
ENSG00000201822  1.134029  0.661383            3.186060

[4558 rows x 3 columns]
```



png

```
2558
```

```
# Merge the datasets based on the 'GeneID' column
SG_GeneExp2 = pd.merge(consistent_genes, SG_GeneExp_cleaned, on="Gen
```

```
eID", how="inner")

# Drop the duplicated columns ('RP value_x', 'Mean Expression_x', etc.)
SG_GeneExp2 = SG_GeneExp2.loc[:, ~SG_GeneExp2.columns.str.endswith('_x')]  # Drop columns ending with '_x'
SG_GeneExp2 = SG_GeneExp2.rename(columns=lambda x: x.replace('_y', ''))
# Remove '_y' from column names
SG_GeneExp2.to_csv('NP13_SCG_byCV_Ascend_WithGeneExp_Metrics.csv')
#SG_GeneExp2

# Extract gene names from df1
gene_list = SG_GeneExp2.index.tolist()

# Subset df2 based on the gene names from df1
filtered_df = SG_bypval_Ascend.loc[SG_bypval_Ascend.index.isin(gene_list)]
filtered_df.to_csv("NP13_SCG_byCV_Ascend_2025.csv")
```

9.8 Appendix C

ML model development with candidate gene expression signatures and variant signatures followed by a systematic feature selection process

9.8.1 Load the data

```
import os
import pandas as pd
os.chdir("F:/AUTOMATED_PEmolde_2024/PEmodel_2025/Data_4Model_2025")
```

9.8.2 Read the selected feature data

```
data = pd.read_csv("Final_GeneVariantData_19Jan2025.csv", index_col=0)
```

9.8.3 Dimensionality Reduction – PCA

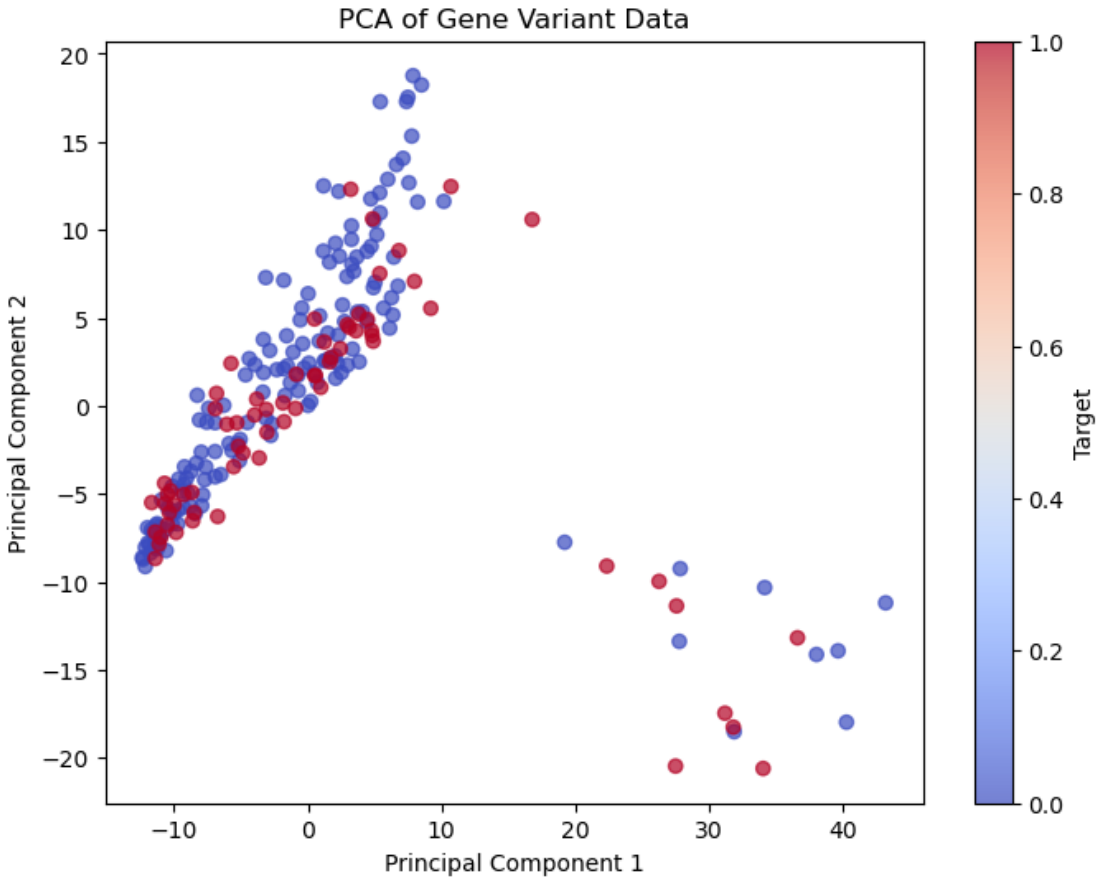
9.8.3.1 Standardize the Data

```
from sklearn.preprocessing import StandardScaler
import pandas as pd

# Standardize the data
scaler = StandardScaler()
scaled_data = scaler.fit_transform(data.iloc[:, 1:])  # Exclude non–numeric colum
ns (like IDs)
#scaled_data
```

9.8.3.2 Perform PCA

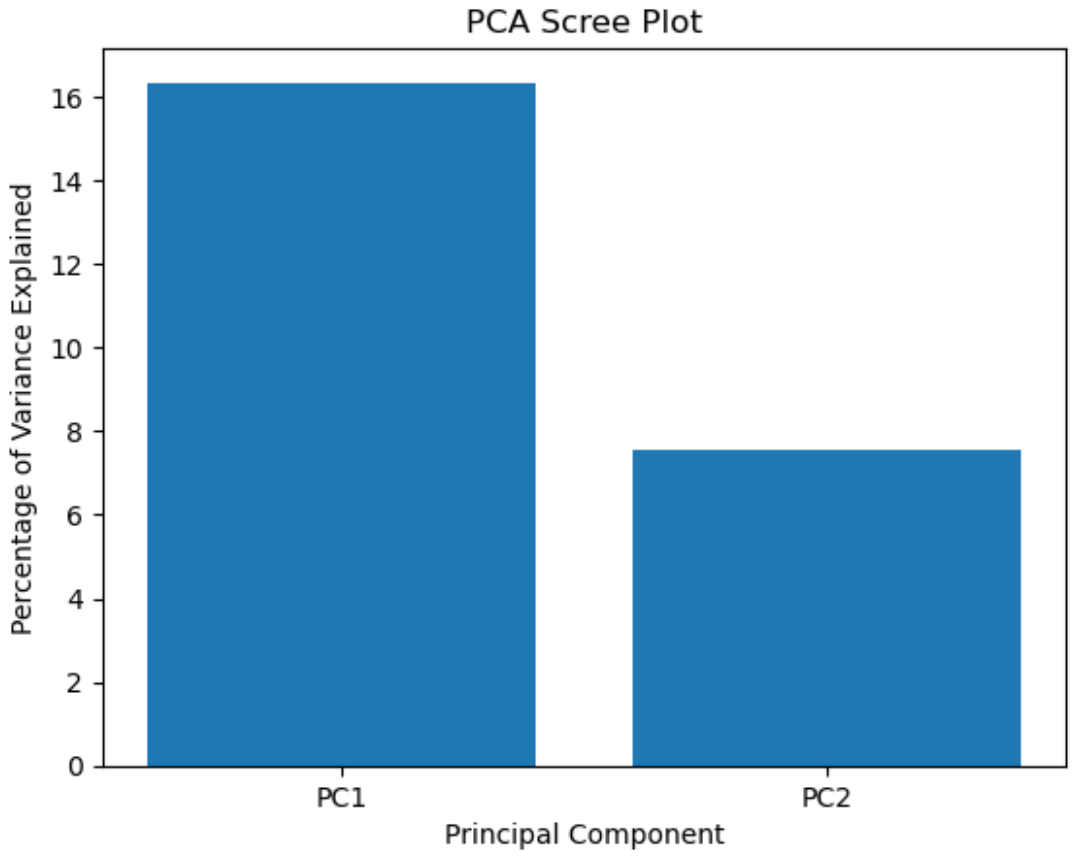
```
import pandas as pd
import numpy as np
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

# Load your variant data
data = pd.read_csv('Final_GeneVariantData_19Jan2025.csv', index_col=0)

# Create the target column
data['Target'] = [0] * 161 + [1] * 67

# Select features excluding the target column
features = data.columns.drop('Target')
x = data.loc[:, features].values
y = data['Target'].values

# Standardize the features
x = StandardScaler().fit_transform(x)

# Apply PCA to reduce to 2 components
pca = PCA(n_components=2)
principalComponents = pca.fit_transform(x)

# Create a DataFrame with principal components
principalDataframe = pd.DataFrame(data=principalComponents, columns=['PC1',
'PC2'])

# Add the target variable back for analysis
principalDataframe['Target'] = y
```

```
# Display the first few rows
#print(principalDataframe)

# Save the results to CSV if needed
principalDataframe.to_csv('PCA_Results.csv', index=False)

# Plot the PCA results
plt.figure(figsize=(8,6))
plt.scatter(principalDataframe['PC1'], principalDataframe['PC2'], c=y, cmap='cool
warm', alpha=0.7)
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.title('PCA of Gene Variant Data')
plt.colorbar(label='Target')
plt.show()
```




```
import numpy as np
import matplotlib.pyplot as plt
```

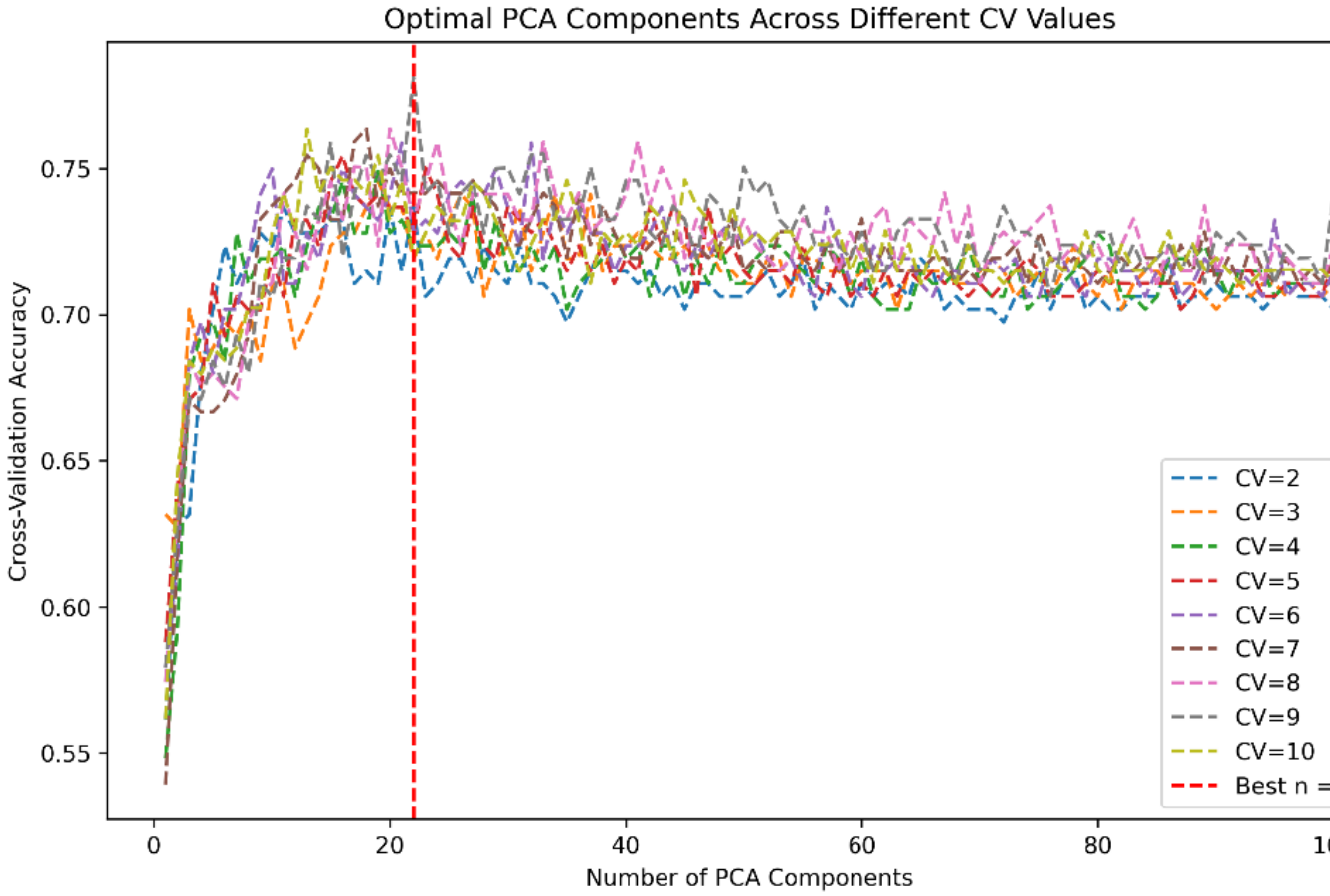
```
# Calculate percentage of variance explained
percent_variance = np.round(pca.explained_variance_ratio_ * 100, decimals=2)

# Adjust range and labels for the number of principal components
columns = ['PC1', 'PC2']
plt.bar(x=range(1, len(columns) + 1), height=percent_variance, tick_label=column
s)

# Adding labels and title
plt.ylabel('Percentage of Variance Explained')
plt.xlabel('Principal Component')
plt.title('PCA Scree Plot')

# Show the plot
plt.show()
```




```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.model_selection import cross_val_score
```

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.pipeline import Pipeline
from sklearn.decomposition import PCA

data2 = pd.read_csv('Final_GeneVariantData_19Jan2025.csv')
# Drop the SampleID column and separate features and target
X = data2.drop(columns=['SampleID', 'target'])
y = data2['target']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_stat
e=42)

# Try different numbers of components
scores = []
num_components = list(range(1, 101))  # Trying 1 to 100 components
cv_values = range(2, 11)  # Cross–validation values from 2 to 10

best_n = None
best_cv = None
best_score = –np.inf  # Initialize with a very low score

for cv in cv_values:
    for n in num_components:
        pca = PCA(n_components=n)
        model = Pipeline([("pca", pca), ("classifier", RandomForestClassifier(random
_state=42, n_jobs=–1))])
        score = np.mean(cross_val_score(model, X, y, cv=cv))  # Cross–validation
with varying folds
        scores.append((cv, n, score))

        # Track the best combination of n_components and cv
        if score > best_score:
            best_score = score
            best_n = n
            best_cv = cv

# Plot the results
plt.figure(figsize=(10, 6))
for cv in cv_values:
    plt.plot(num_components, [s[2] for s in scores if s[0] == cv], label=f"CV={cv}",
linestyle="— ")

plt.axvline(x=best_n, color='r', linestyle='— ', label=f"Best n = {best_n}")
plt.xlabel("Number of PCA Components")
```

```
plt.ylabel("Cross–Validation Accuracy")
plt.title("Optimal PCA Components Across Different CV Values")
plt.legend()
plt.show()
```



9.8.4 Modelling

```
import pandas as pd
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
import numpy as np

# Set random seed for reproducibility
np.random.seed(42)

# Load the data
data2 = pd.read_csv('Final_GeneVariantData_19Jan2025.csv')
```

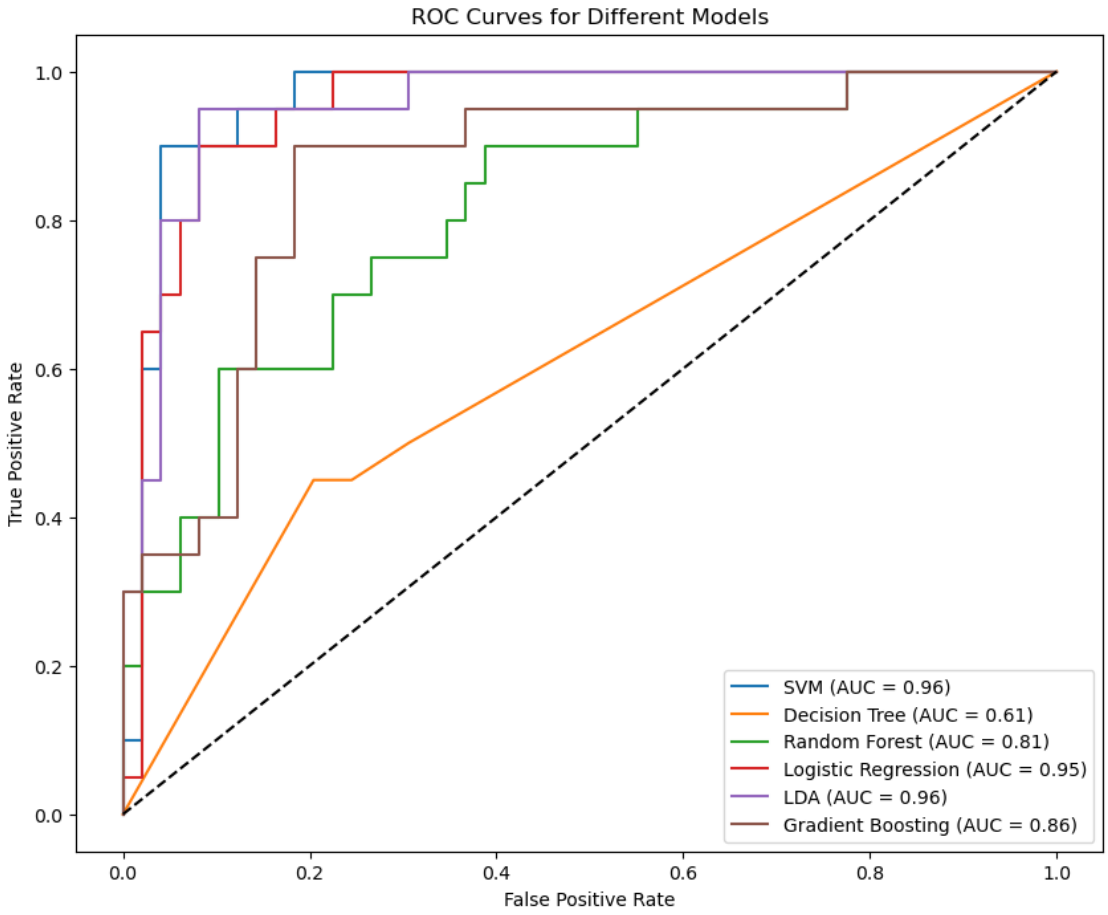
```
# Drop the SampleID column and separate features and target
X = data2.drop(columns=['SampleID', 'target'])
y = data2['target']

# Standardize data before PCA
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Apply PCA with the optimal number of components
pca = PCA(n_components=22, random_state=42)
X_pca = pca.fit_transform(X_scaled)

# Split the data into training and testing sets with fixed random state
X_train_pca, X_test_pca, y_train, y_test = train_test_split(X_pca, y, test_size=0.3
, random_state=42, stratify=y)

# Train SVM model with fixed random state
svm_model = SVC(C=0.1, gamma='scale', kernel='linear', random_state=42)
svm_model.fit(X_train_pca, y_train)

# Evaluate on test set
accuracy = svm_model.score(X_test_pca, y_test)
print(f"SVM Test Accuracy: {accuracy:.4f}")
```

```
SVM Test Accuracy: 0.9275
```

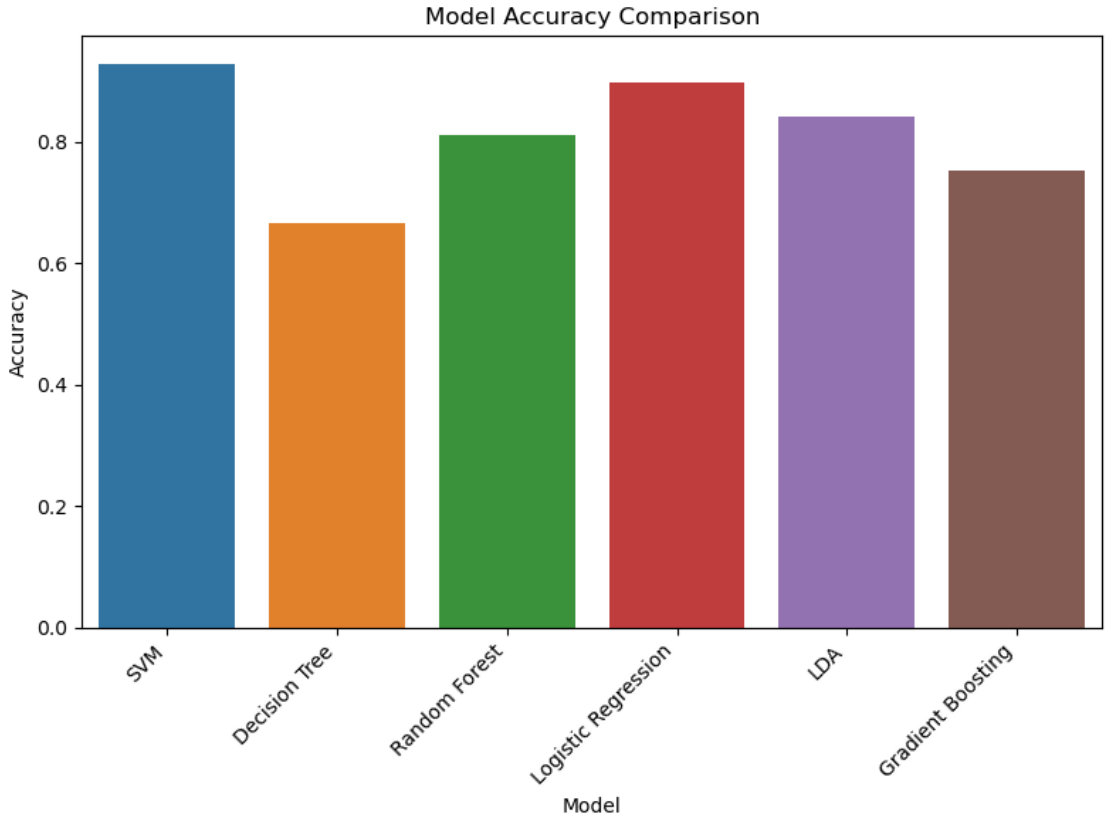
```
import pandas as pd
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifi
er
from sklearn.linear_model import LogisticRegression
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.metrics import accuracy_score, roc_auc_score, classification_report
, roc_curve, auc
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Set random seed for reproducibility
np.random.seed(42)
```

```
# Load the data
data2 = pd.read_csv('Final_GeneVariantData_19Jan2025.csv')

# Drop the SampleID column and separate features and target
X = data2.drop(columns=['SampleID', 'target'])
y = data2['target']

# Standardize data before PCA
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Apply PCA with the optimal number of components
pca = PCA(n_components=22, random_state=42)
X_pca = pca.fit_transform(X_scaled)

# Split the data into training and testing sets with fixed random state
X_train_pca, X_test_pca, y_train, y_test = train_test_split(X_pca, y, test_size=0.3
, random_state=42, stratify=y)

# Define models
models = {
    'SVM': SVC(C=0.1, gamma='scale', kernel='linear', probability=True, random_s
tate=42),
    'Decision Tree': DecisionTreeClassifier(criterion='gini', max_depth=20, min_sa
mples_split=5, random_state=42),
    'Random Forest': RandomForestClassifier(criterion='gini', max_depth=10, min_
samples_split=5, n_estimators=50, random_state=42),
    'Logistic Regression': LogisticRegression(C=10, penalty='l1', solver='saga', ma
x_iter=5000, random_state=42),
    'LDA': LinearDiscriminantAnalysis(),
    'Gradient Boosting': GradientBoostingClassifier(learning_rate=0.5, max_depth
=3, n_estimators=200, random_state=42)
}

# Evaluate models and collect metrics
results = []
plt.figure(figsize=(10, 8))

for name, model in models.items():
    model.fit(X_train_pca, y_train)
    y_pred = model.predict(X_test_pca)
    y_prob = model.predict_proba(X_test_pca)[:, 1]

    acc = accuracy_score(y_test, y_pred)
```

```
    auc_score = roc_auc_score(y_test, y_prob)

    # Generate ROC curve
    fpr, tpr, _ = roc_curve(y_test, y_prob)
    plt.plot(fpr, tpr, label=f'{name} (AUC = {auc_score:.2f})')

    # Store results
    results.append({'Model': name, 'Accuracy': acc, 'AUC': auc_score})

# Plot ROC curve
plt.plot([0, 1], [0, 1], 'k— ')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curves for Different Models')
plt.legend()
plt.savefig('roc_curve_comparison.png', dpi=200)
plt.show()

# Create a results DataFrame and save to CSV
results_df = pd.DataFrame(results)
results_df.to_csv('model_evaluation_results.csv', index=False)

# Print the evaluation table
print(results_df)

# Visualize the model performance
plt.figure(figsize=(8, 6))
sns.barplot(x='Model', y='Accuracy', data=results_df)
plt.title('Model Accuracy Comparison')

# Rotate x–axis labels and adjust layout
plt.xticks(rotation=45, ha='right')  # Rotate labels and align them to the right
plt.tight_layout()  # Automatically adjust the layout to avoid overlap

# Save the figure with adjusted layout
plt.savefig('model_accuracy_comparison.png', dpi=200)
plt.show()
```



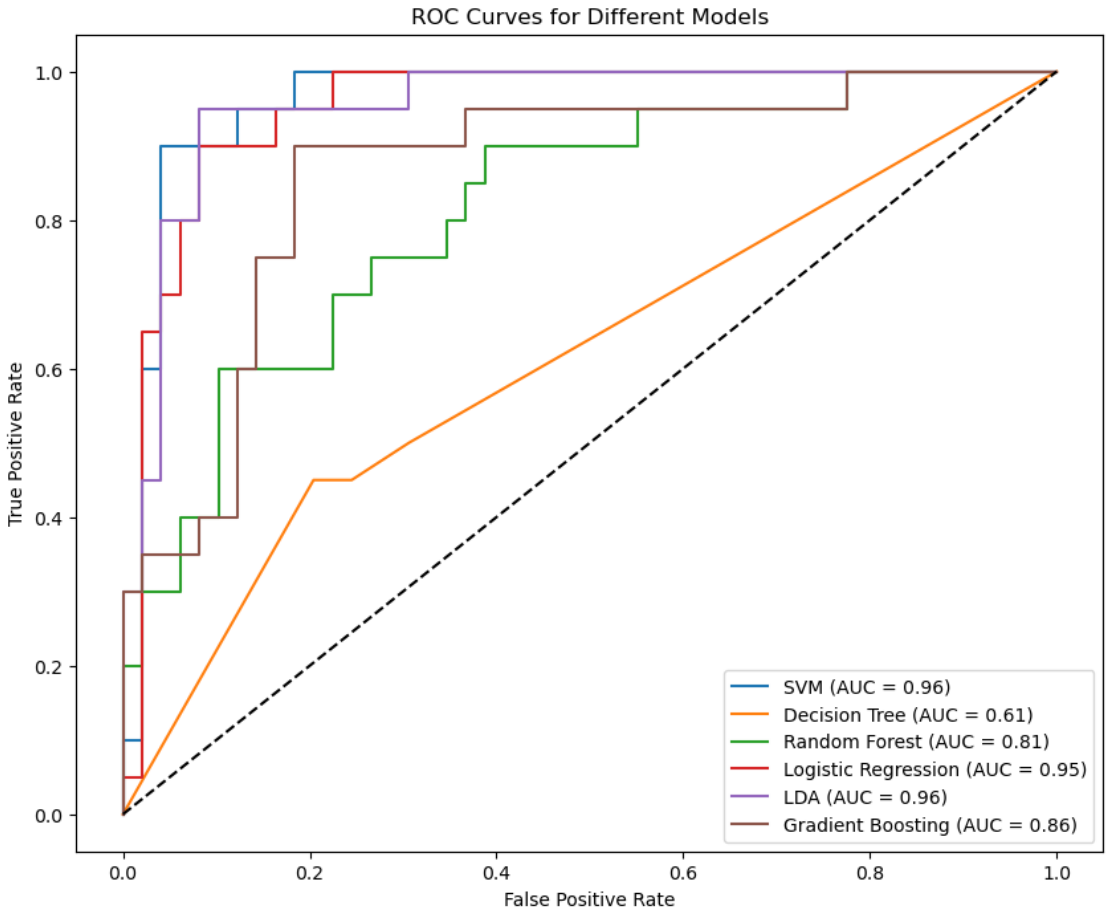
	Model	Accuracy	AUC
0	SVM	0.927536	0.962245
1	Decision Tree	0.666667	0.613776
2	Random Forest	0.811594	0.814286
3	Logistic Regression	0.898551	0.952041
4	LDA	0.840580	0.955102
5	Gradient Boosting	0.753623	0.864286




```
import pandas as pd
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifi
er
from sklearn.tree import DecisionTreeClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.metrics import accuracy_score, roc_auc_score, confusion_matrix, cl
assification_report, roc_curve, auc
import matplotlib.pyplot as plt
import numpy as np

# Set random seed for reproducibility
np.random.seed(42)

# Load the data
data2 = pd.read_csv('Final_GeneVariantData_19Jan2025.csv')
```

```
# Drop the SampleID column and separate features and target
X = data2.drop(columns=['SampleID', 'target'])
y = data2['target']

# Standardize data before PCA
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Apply PCA with the optimal number of components
pca = PCA(n_components=22, random_state=42)
X_pca = pca.fit_transform(X_scaled)

# Split the data into training and testing sets with fixed random state
X_train_pca, X_test_pca, y_train, y_test = train_test_split(X_pca, y, test_size=0.3
, random_state=42, stratify=y)

# Define models
models = {
    'SVM': SVC(C=0.1, gamma='scale', kernel='linear', probability=True, random_s
tate=42),
    'Decision Tree': DecisionTreeClassifier(criterion='gini', max_depth=20, min_sa
mples_split=5, random_state=42),
    'Random Forest': RandomForestClassifier(criterion='gini', max_depth=10, min_
samples_split=5, n_estimators=50, random_state=42),
    'Logistic Regression': LogisticRegression(C=10, penalty='l1', solver='saga', ran
dom_state=42, max_iter=10000),
    'LDA': LinearDiscriminantAnalysis(solver='svd'),
    'Gradient Boosting': GradientBoostingClassifier(learning_rate=0.5, max_depth
=3, n_estimators=200, random_state=42)
}

# Train and evaluate models
evaluation_results = []
plt.figure(figsize=(10, 8))
for name, model in models.items():
    model.fit(X_train_pca, y_train)
    y_pred = model.predict(X_test_pca)
    y_proba = model.predict_proba(X_test_pca)[:, 1]

    acc = accuracy_score(y_test, y_pred)
    auc_score = roc_auc_score(y_test, y_proba)
    conf_matrix = confusion_matrix(y_test, y_pred)
    class_report = classification_report(y_test, y_pred, output_dict=True)
```

```
    evaluation_results.append([name, acc, auc_score, conf_matrix])

    report_df = pd.DataFrame(class_report).transpose()
    report_df.to_csv(f'{name}_classification_report.csv')

    fpr, tpr, _ = roc_curve(y_test, y_proba)
    plt.plot(fpr, tpr, label=f'{name} (AUC = {auc_score:.2f})')

# Save evaluation metrics to CSV
columns = ['Model', 'Accuracy', 'AUC Score', 'Confusion Matrix']
eval_df = pd.DataFrame(evaluation_results, columns=columns)
eval_df.to_csv('model_evaluation_metrics.csv', index=False)

# Show ROC curve
plt.plot([0, 1], [0, 1], 'k— ')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curves for Different Models')
plt.legend()
plt.savefig('roc_curves.png', dpi=300, bbox_inches='tight')
plt.show()

# Print evaluation summary
print(eval_df[['Model', 'Accuracy', 'AUC Score']])
```

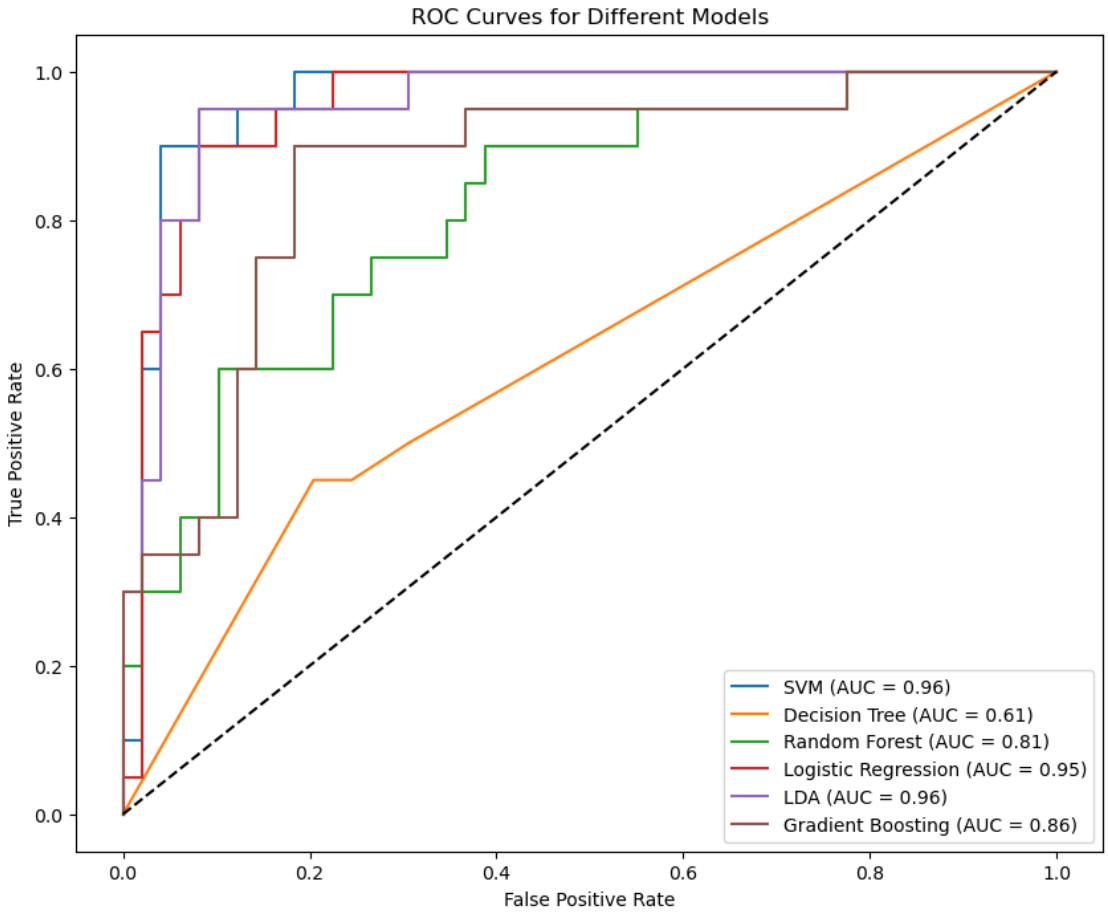


	Model	Accuracy	AUC Score
0	SVM	0.927536	0.962245
1	Decision Tree	0.666667	0.613776
2	Random Forest	0.811594	0.814286
3	Logistic Regression	0.898551	0.952041
4	LDA	0.840580	0.955102
5	Gradient Boosting	0.753623	0.864286

```
import pandas as pd
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifi
er
from sklearn.tree import DecisionTreeClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
```

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_sc
ore, roc_auc_score, confusion_matrix, classification_report, roc_curve
import matplotlib.pyplot as plt
import numpy as np

# Set random seed for reproducibility
np.random.seed(42)

# Load the data
data2 = pd.read_csv('Final_GeneVariantData_19Jan2025.csv')

# Drop the SampleID column and separate features and target
X = data2.drop(columns=['SampleID', 'target'])
y = data2['target']

# Standardize data before PCA
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Apply PCA with the optimal number of components
pca = PCA(n_components=22, random_state=42)
X_pca = pca.fit_transform(X_scaled)

# Split the data into training and testing sets with fixed random state
X_train_pca, X_test_pca, y_train, y_test = train_test_split(X_pca, y, test_size=0.3
, random_state=42, stratify=y)

# Define models
models = {
    'SVM': SVC(C=0.1, gamma='scale', kernel='linear', probability=True, random_s
tate=42),
    'Decision Tree': DecisionTreeClassifier(criterion='gini', max_depth=20, min_sa
mples_split=5, random_state=42),
    'Random Forest': RandomForestClassifier(criterion='gini', max_depth=10, min_
samples_split=5, n_estimators=50, random_state=42),
    'Logistic Regression': LogisticRegression(C=10, penalty='l1', solver='saga', ran
dom_state=42, max_iter=10000),
    'LDA': LinearDiscriminantAnalysis(solver='svd'),
    'Gradient Boosting': GradientBoostingClassifier(learning_rate=0.5, max_depth
=3, n_estimators=200, random_state=42)
}

# Train and evaluate models
evaluation_results = []
plt.figure(figsize=(10, 8))
```

```
for name, model in models.items():
    model.fit(X_train_pca, y_train)
    y_pred = model.predict(X_test_pca)
    y_proba = model.predict_proba(X_test_pca)[:, 1]

    acc = accuracy_score(y_test, y_pred) * 100
    precision = precision_score(y_test, y_pred) * 100
    recall = recall_score(y_test, y_pred) * 100
    f1 = f1_score(y_test, y_pred) * 100
    auc_score = roc_auc_score(y_test, y_proba)

    evaluation_results.append([name, acc, precision, recall, f1])

    fpr, tpr, _ = roc_curve(y_test, y_proba)
    plt.plot(fpr, tpr, label=f'{name} (AUC = {auc_score:.2f})')

# Save evaluation metrics to CSV
columns = ['Model', 'Accuracy (%)', 'Precision (%)', 'Recall (%)', 'F1–score (%)']
eval_df = pd.DataFrame(evaluation_results, columns=columns)
eval_df.to_csv('model_evaluation_metrics_detailed.csv', index=False)

# Show ROC curve
plt.plot([0, 1], [0, 1], 'k— ')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curves for Different Models')
plt.legend()
plt.savefig('roc_curves.png', dpi=300, bbox_inches='tight')
plt.show()

# Print evaluation summary
print(eval_df)
```



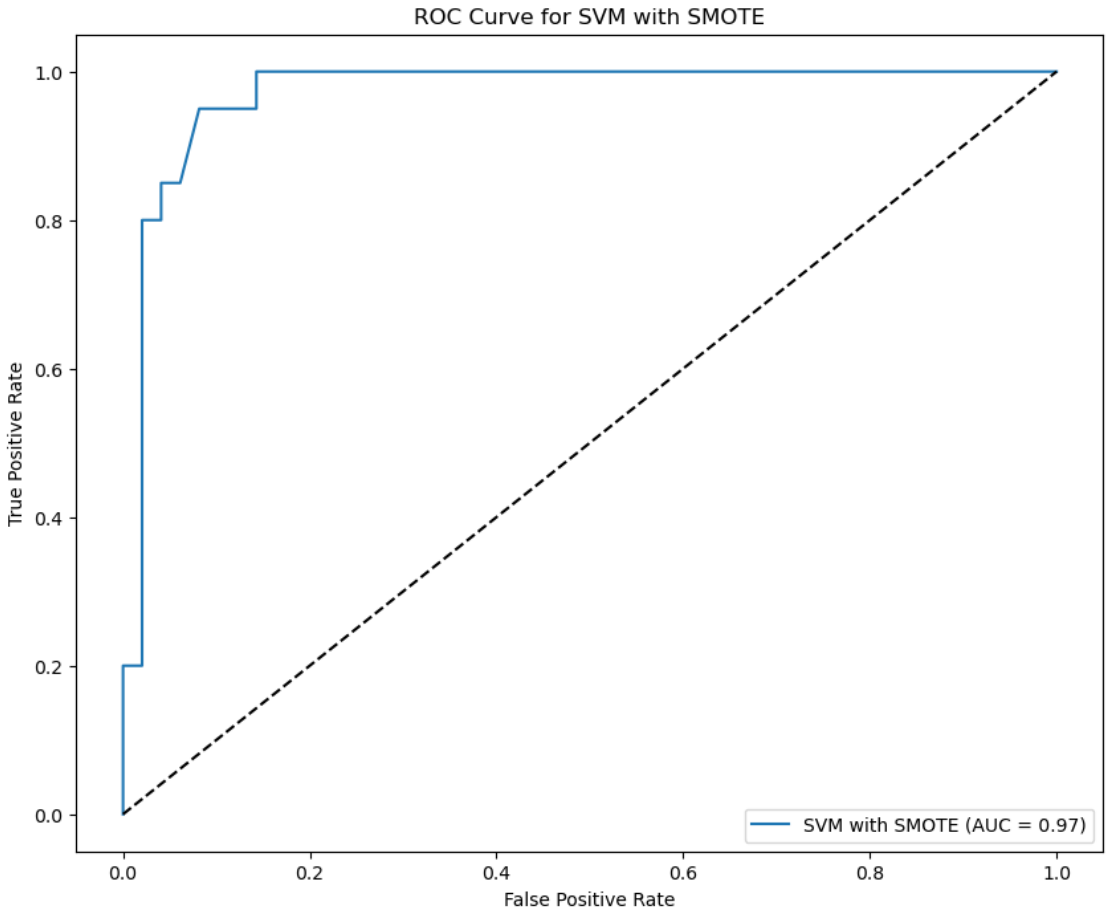
	Model	Accuracy (%)	Precision (%)	Recall (%)	F1–score (%)
0	SVM	92.753623	89.473684	85.0	87.179487
1	Decision Tree	66.666667	42.857143	45.0	43.902439
2	Random Forest	81.159420	70.588235	60.0	64.864865
3	Logistic Regression	89.855072	78.260870	90.0	83.720930
4	LDA	84.057971	84.615385	55.0	66.666667
5	Gradient Boosting	75.362319	60.000000	45.0	51.428571

```
import pandas as pd
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, roc_auc_score, classification_report
, roc_curve, auc
from imblearn.over_sampling import SMOTE
```

```
from imblearn.pipeline import Pipeline  # Use this to integrate SMOTE in the pipel
ine

# Set random seed for reproducibility
np.random.seed(42)

# Load the data
data2 = pd.read_csv('Final_GeneVariantData_19Jan2025.csv')

# Drop the SampleID column and separate features and target
X = data2.drop(columns=['SampleID', 'target'])
y = data2['target']

# Standardize data before PCA
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Apply PCA with the optimal number of components
pca = PCA(n_components=22, random_state=42)
X_pca = pca.fit_transform(X_scaled)

# Split the data into training and testing sets with fixed random state
X_train_pca, X_test_pca, y_train, y_test = train_test_split(X_pca, y, test_size=0.3
, random_state=42, stratify=y)

# Define SVM model
svm_model = SVC(C=0.1, gamma='scale', kernel='linear', probability=True, rand
om_state=42)

# Define SMOTE
smote = SMOTE(sampling_strategy='minority', random_state=42)

# Create a pipeline with SMOTE and SVM
pipeline = Pipeline([
    ('smote', smote),  # Apply SMOTE before SVM
    ('svm', svm_model)
])

# Train the model and make predictions
pipeline.fit(X_train_pca, y_train)
y_pred = pipeline.predict(X_test_pca)
y_prob = pipeline.predict_proba(X_test_pca)[:, 1]

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
```

```
auc_score = roc_auc_score(y_test, y_prob)
print(f"Accuracy: {accuracy:.4f}")
print(f"AUC Score: {auc_score:.4f}")

# Classification report
print("\nClassification Report:")
print(classification_report(y_test, y_pred))

# Generate ROC curve
fpr, tpr, _ = roc_curve(y_test, y_prob)
plt.figure(figsize=(10, 8))
plt.plot(fpr, tpr, label=f'SVM with SMOTE (AUC = {auc_score:.2f})')
plt.plot([0, 1], [0, 1], 'k— ')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curve for SVM with SMOTE')
plt.legend()
plt.show()

# Save evaluation results to CSV
results = {
    'Model': 'SVM with SMOTE',
    'Accuracy': accuracy,
    'AUC': auc_score
}

results_df = pd.DataFrame([results])
results_df.to_csv('svm_with_smote_evaluation.csv', index=False)

# Print results
print("\nEvaluation Results:")
print(results_df)
```

```
Accuracy: 0.9130
AUC Score: 0.9714

Classification Report:
              precision    recall  f1–score   support

           0       0.94      0.94      0.94        49
           1       0.85      0.85      0.85        20

    accuracy                           0.91        69
   macro avg       0.89      0.89      0.89        69
weighted avg       0.91      0.91      0.91        69
```



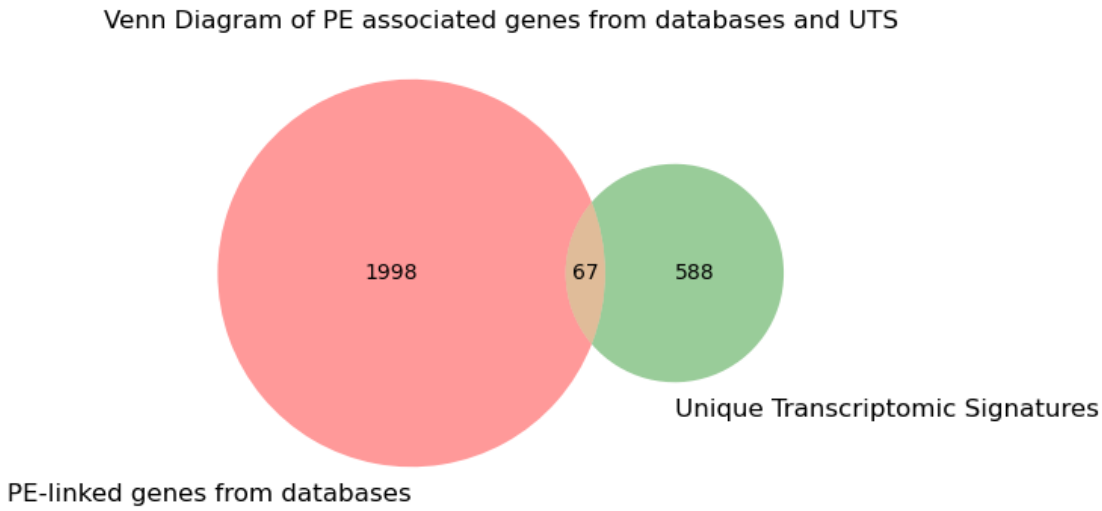
png

```
Evaluation Results:
            Model  Accuracy       AUC
0  SVM with SMOTE  0.913043  0.971429
```

9.8.5 Overlap – UDEGS with Known PE Genes

```
udegs = pd.read_csv("GeneCards–SearchResults.csv")
#udegs.head(5)

# Convert a specific column to a set (replace 'ColumnName' with your column na
me)
genes_db = set(udegs['Gene Symbol'].dropna())
uts = set(udegs['UTS'].dropna())

import matplotlib.pyplot as plt
from matplotlib_venn import venn2, venn3
```

```
# Create a Venn diagram for two sets
plt.figure(figsize=(10, 4))
venn2([genes_db, uts], ('PE–linked genes from databases', 'Unique Transcriptomic Signatures'))
plt.title('Venn Diagram of PE associated genes from databases and UTS')
plt.savefig('PElinkedGenesandUdegs_venn_diagram.png', dpi=400)
plt.show()
```



png

```
# Find intersecting genes
intersecting_genes = genes_db.intersection(uts)

# Convert to a DataFrame
df_intersect = pd.DataFrame(intersecting_genes, columns=['Intersecting_Genes'])

# Save to a CSV file
#df_intersect.to_csv('67_intersecting_genes.csv', index=False)
```

References

Abalos, E., Cuesta, C., Grosso, A.L., Chou, D. and Say, L. (2013) 'Global and regional estimates of preeclampsia and eclampsia: a systematic review', European Journal of Obstetrics, Gynecology, and Reproductive Biology, 170(1), pp. 1–7. Available at: https://doi.org/10.1016/j.ejogrb.2013.05.005.

Abbas, Q., Jeong, W. and Lee, S.W. (2025) 'Explainable AI in Clinical Decision Support Systems: A Meta-Analysis of Methods, Applications, and Usability Challenges', Healthcare, 13(17), p. 2154. Available at: https://doi.org/10.3390/healthcare13172154.

Abbas-Aghababazadeh, F., Li, Q. and Fridley, B.L. (2018) 'Comparison of normalization approaches for gene expression studies completed with high-throughput sequencing', PLoS ONE, 13(10), p. e0206312. Available at: https://doi.org/10.1371/journal.pone.0206312.

Abbasi, J. (2022) 'Cell-Free RNA in Maternal Plasma Predicts Preeclampsia', JAMA, 327(6), p. 517. Available at: https://doi.org/10.1001/jama.2022.0327.

Abdalla, R.A.H., Parveen, N., Iqbal, N., Mohamed, A.A.A., Shahid, S.M.A., Elhussein, G.E.M.O., Saleem, Mohd. and Khan, M.S. (2025) 'Adipokines in preeclampsia: disrupted signaling pathways and novel therapeutic strategies', European Journal of Medical Research, 30, p. 702. Available at: https://doi.org/10.1186/s40001-025-02972-y.

Abdelwahab, O., Awad, N., Elserafy, M. and Badr, E. (2022) 'A feature selection-based framework to identify biomarkers for cancer diagnosis: A focus on lung adenocarcinoma', PLoS ONE, 17(9), p. e0269126. Available at: https://doi.org/10.1371/journal.pone.0269126.

Abrams, Z.B., Johnson, T.S., Huang, K., Payne, P.R.O. and Coombes, K. (2019) 'A protocol to evaluate RNA sequencing normalization methods', BMC Bioinformatics, 20(24), p. 679. Available at: https://doi.org/10.1186/s12859-019-3247-x.

Admati, I., Skarbianskis, N., Hochgerner, H., Ophir, O., Weiner, Z., Yagel, S., Solt, I. and Zeisel, A. (2023) 'Two distinct molecular faces of preeclampsia revealed by single-cell transcriptomics', Med, 4(10), pp. 687-709.e7. Available at: https://doi.org/10.1016/j.medj.2023.07.005.

Agarwal, A., Aponte-Mellado, A., Premkumar, B.J., Shaman, A. and Gupta, S. (2012) 'The effects of oxidative stress on female reproduction: A review', Reproductive biology and endocrinology: RB&E, 10, pp. 49. Available at: https://doi.org/10.1186/1477-7827-10-49.

Aggarwal, R., Jain, A.K., Mittal, P., Kohli, M., Jawanjal, P. and Rath, G. (2019) 'Association of pro- and anti-inflammatory cytokines in preeclampsia', Journal of

Clinical Laboratory Analysis, 33(4), pp. e22834. Available at: https://doi.org/10.1002/jcla.22834.

Ahmed, M.I., Spooner, B., Isherwood, J., Lane, M., Orrock, E. and Dennison, A. (2023) ‘A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare’, Cureus, 15(10), p. e46454. Available at: https://doi.org/10.7759/cureus.46454.

Ahn, H., Park, J., Gilman-Sachs, A. and Kwak-Kim, J. (2011) 'Immunologic characteristics of preeclampsia, a comprehensive review', American journal of reproductive immunology, 65(4), pp. 377–394.

Ahrens, K.A., Nelson, H., Stidd, R.L., Moskosky, S. and Hutcheon, J.A. (2019) 'Short interpregnancy intervals and adverse perinatal outcomes in high-resource settings: An updated systematic review', Paediatric and perinatal epidemiology, 33(1), pp. O25–O47.

Airlangga, G. (2024) 'Enhancing medical diagnostics with ensemble machine learning: A comparative study of gradient boosting, XGBoost, LightGBM, and blended models', Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika), 9(2), pp. 1117–1132.

Aisagbonhi, O. and Morris, G.P. (2022) ‘Human Leukocyte Antigens in Pregnancy and Preeclampsia’, Frontiers in Genetics, 13, p. 884275. Available at: https://doi.org/10.3389/fgene.2022.884275.

Akyea, R.K., Leonardi-Bee, J., Asselbergs, F.W., Patel, R.S., Durrington, P., Wierzbicki, A.S., Ibiwoye, O.H., Kai, J., Qureshi, N. and Weng, S.F. (2020) 'Predicting major adverse cardiovascular events for secondary prevention: Protocol for a systematic review and meta-analysis of risk prediction models', BMJ Open, 10(7). Available at: https://doi.org/10.1136/bmjopen-2019-034564.

Alam, S.F. and Gonzalez Suarez, M.L. (2024) 'Transforming healthcare: The AI revolution in the comprehensive care of hypertension', Clinics and Practice, 14(4), pp. 1357–1374. Available at: https://doi.org/10.3390/clinpract14040109.

Albrecht, E.D. and Pepe, G.J. (2020) 'Regulation of uterine spiral artery remodeling: A review', Reproductive Sciences, 27(10), pp. 1932–1942.

Alese, M.O., Moodley, J. and Naicker, T. (2021) 'Preeclampsia and HELLP syndrome, the role of the liver', The Journal of Maternal-Fetal & Neonatal Medicine, 34(1), pp. 117–123.

Alharbi, F., Budhiraja, N., Vakanski, A., Zhang, B., Elbashir, M.K., Guduru, H. and Mohammed, M. (2025) ‘Interpretable graph Kolmogorov–Arnold networks for multi-cancer classification and biomarker identification using multi-omics data’, Scientific Reports, 15(1), p. 27607. Available at: https://doi.org/10.1038/s41598-025-13337-0.

Ali, S. and Ali, S. (2011) Differential gene expression and its possible therapeutic implications. IntechOpen.

Aljameel, S.S., Alzahrani, M., Almusharraf, R., Altukhais, M., Alshaia, S., Sahlouli, H., Aslam, N., Khan, I.U., Alabbad, D.A. and Alsumayt, A. (2023) ‘Prediction of Preeclampsia Using Machine Learning and Deep Learning Models: A Review’, Big Data and Cognitive Computing, 7(1), p. 32. Available at: https://doi.org/10.3390/bdcc7010032.

Alkhateeb, A. and Rueda, L. (2017) 'Zseq: An approach for preprocessing next-generation sequencing data', Journal of computational biology, 24(8), pp. 746–755. Available at: https://doi.org/10.1089/cmb.2017.0021.

Allen, R.E., Rogozinska, E., Cleverly, K., Aquilina, J. and Thangaratinam, S. (2014) 'Abnormal blood biomarkers in early pregnancy are associated with preeclampsia: A meta-analysis', European Journal of Obstetrics & Gynecology and Reproductive Biology, 182, pp. 194–201.

Alqaissi, E.Y., Alotaibi, F.S. and Ramzan, M.S. (2022) 'Modern machine-learning predictive models for diagnosing infectious diseases', Computational and Mathematical Methods in Medicine, 2022, pp. 6902321. Available at: https://doi.org/10.1155/2022/6902321.

Alser, M., Rotman, J., Deshpande, D., Taraszka, K., Shi, H., Baykal, P.I., Yang, H.T., Xue, V., Knyazev, S. and Singer, B.D. (2021) 'Technology dictates algorithms: Recent developments in read alignment', Genome biology, 22(1), pp. 249.

Alshareef, S.A., Eltom, A.M., Nasr, A.M., Hamdan, H.Z. and Adam, I. (2017) 'Rubella, herpes simplex virus type 2 and preeclampsia', Virology Journal, 14, pp. 1–4.

Altar, C.A. (2008) 'The biomarkers consortium: On the critical path of drug discovery', Clinical Pharmacology & Therapeutics, 83(2), pp. 361–364.

Amirian, A., Mahani, M.B. and Abdi, F. (2020) 'Role of interleukin-6 (IL-6) in predicting gestational diabetes mellitus', Obstetrics & Gynecology Science, 63(4), pp. 407–416. Available at: https://doi.org/10.5468/ogs.20020.

and prevention', International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics, 145(Suppl 1), pp. 1–33. Available at: https://doi.org/10.1002/ijgo.12802.

Anders, S., McCarthy, D.J., Chen, Y., Okoniewski, M., Smyth, G.K., Huber, W. and Robinson, M.D. (2013) 'Count-based differential expression analysis of RNA sequencing data using R and bioconductor', Nature protocols, 8(9), pp. 1765–1786.

Anders, S., Pyl, P.T. and Huber, W. (2015) 'HTSeq—a python framework to work with high-throughput sequencing data', bioinformatics, 31(2), pp. 166–169.

Andirkó, A., Moriano, J., Vitriolo, A., Kuhlwilm, M., Testa, G. and Boeckx, C. (2022) 'Temporal mapping of derived high-frequency gene variants supports the mosaic nature of the evolution of homo sapiens', Scientific Reports, 12(1), pp. 9937. Available at: https://doi.org/10.1038/s41598-022-13589-0.

Anjum, A., Jaggi, S., Varghese, E., Lall, S., Bhowmik, A. and Rai, A. (2016) 'Identification of differentially expressed genes in RNA-seq data of arabidopsis thaliana: A compound distribution approach', Journal of computational biology, 23(4), pp. 239–247. Available at: https://doi.org/10.1089/cmb.2015.0205.

Ansbacher-Feldman, Z., Syngelaki, A., Meiri, H., Cirkin, R., Nicolaides, K.H. and Louzoun, Y. (2022) ‘Machine-learning-based prediction of pre-eclampsia using first-trimester maternal characteristics and biomarkers’, Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology, 60(6), pp. 739–745. Available at: https://doi.org/10.1002/uog.26105.

Ansbacher-Feldman, Z., Syngelaki, A., Meiri, H., Cirkin, R., Nicolaides, K.H. and Louzoun, Y. (2022) 'Machine-learning-based prediction of pre-eclampsia using first-trimester maternal characteristics and biomarkers', Ultrasound in Obstetrics & Gynecology, 60(6), pp. 739–745.

Aouache, R., Biquard, L., Vaiman, D. and Miralles, F. (2018) ‘Oxidative Stress in Preeclampsia and Placental Diseases’, International Journal of Molecular Sciences, 19(5), p. 1496. Available at: https://doi.org/10.3390/ijms19051496.

Aouache, R., Biquard, L., Vaiman, D. and Miralles, F. (2018) 'Oxidative stress in preeclampsia and placental diseases', International Journal of Molecular Sciences, 19(5), pp. 1496. Available at: https://doi.org/10.3390/ijms19051496.

Armingol, E., Officer, A., Harismendy, O. and Lewis, N.E. (2021) 'Deciphering cell–cell interactions and communication from gene expression', Nature Reviews Genetics, 22(2), pp. 71–88.

Artemaki, P.I. and Kontos, C.K. (2022) 'Alternative splicing in human physiology and disease', Genes, 13(10), pp. 1820.

Artimovič, P., Špaková, I., Macejková, E., Pribulová, T., Rabajdová, M., Mareková, M. and Zavacká, M. (2024) ‘The ability of microRNAs to regulate the immune response in ischemia/reperfusion inflammatory pathways’, Genes & Immunity, 25(4), pp. 277–296. Available at: https://doi.org/10.1038/s41435-024-00283-6.

Arutyunyan, A., Roberts, K., Troulé, K., Wong, F.C., Sheridan, M.A., Kats, I., Garcia-Alonso, L., Velten, B., Hoo, R. and Ruiz-Morales, E.R. (2023) 'Spatial multiomics map of trophoblast development in early pregnancy', Nature, , pp. 1–9.

Assani, A.-D., Boldeanu, L., Siloși, I., Boldeanu, M.V., Dijmărescu, A.L., Assani, M.-Z., Manolea, M.-M. and Văduva, C.-C. (2025) ‘Pregnancy Under Pressure:

Oxidative Stress as a Common Thread in Maternal Disorders', Life, 15(9), p. 1348. Available at: https://doi.org/10.3390/life15091348.

Athanasopoulou, K., Michalopoulou, V.-I., Scorilas, A. and Adamopoulos, P.G. (2025) 'Integrating Artificial Intelligence in Next-Generation Sequencing: Advances, Challenges, and Future Directions', Current Issues in Molecular Biology, 47(6), p. 470. Available at: https://doi.org/10.3390/cimb47060470.

Aucamp, J., Bronkhorst, A.J., Badenhorst, C.P. and Pretorius, P.J. (2018) 'The diverse origins of circulating cell-free DNA in the human body: A critical re-evaluation of the literature', Biological Reviews, 93(3), pp. 1649–1683.

Auger, N., Amar, S., Paradis, G., Ayoub, A., Lewin, A., Maniraho, A. and Potter, B.J. (2025) 'Preeclampsia and Onset of Renal Disorders in the Long-Term Period following Pregnancy', American Journal of Nephrology [Preprint]. Available at: https://doi.org/10.1159/000547538.

Auwera, G.A.V.d., Carneiro, M.O., Hartl, C., Poplin, R., Angel, G.d., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S. and DePristo, M.A. (2013) 'From FastQ data to high confidence variant calls: The genome analysis toolkit best practices pipeline', Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.], 11(1110), pp. 11.10.1. Available at: https://doi.org/10.1002/0471250953.bi1110s43.

Azmi, M.B., Nasir, M.F., Asif, U., Kazi, M., Uddin, M.N. and Qureshi, S.A. (2024) 'Analyzing molecular signatures in preeclampsia and fetal growth restriction: Identifying key genes, pathways, and therapeutic targets for preterm birth', Frontiers in Molecular Biosciences, 11. Available at: https://doi.org/10.3389/fmolb.2024.1384214.

Baetens, M., Van Gaever, B., Deblaere, S., De Koker, A., Meuris, L., Callewaert, N., Janssens, S., Roelens, K., Roets, E., Van Dorpe, J., Dehaene, I. and Menten, B. (2024) 'Advancing diagnosis and early risk assessment of preeclampsia through noninvasive cell-free DNA methylation profiling', Clinical Epigenetics, 16, p. 182. Available at: https://doi.org/10.1186/s13148-024-01798-5.

Bai, L., Li, Y., Huang, H. and Liu, X. (2023) 'Machine learning and bioinformatics framework integration reveal potential characteristic genes related to immune cell infiltration in preeclampsia', Frontiers in Physiology, 14, pp. 1078166.

Baker, B.H., Freije, S., MacDonald, J.W., Bammler, T.K., Benson, C., Carroll, K.N., Enquobahrie, D.A., Karr, C.J., LeWinn, K.Z., Zhao, Q., Bush, N.R., Sathyanarayana, S. and Paquette, A.G. (2024) 'Placental transcriptomic signatures of prenatal and preconceptional maternal stress', Molecular Psychiatry, 29(4), pp. 1179–1191. Available at: https://doi.org/10.1038/s41380-023-02403-6.

Bakulski, K.M., Blostein, F. and London, S.J. (2023) 'Linking prenatal environmental exposures to lifetime health with epigenome-wide association studies: State-of-the-science review and future recommendations', Environmental Health Perspectives, 131(12), pp. 126001. Available at: https://doi.org/10.1289/EHP12956.

Bao, R., Huang, L., Andrade, J., Tan, W., Kibbe, W.A., Jiang, H. and Feng, G. (2014) 'Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing', Cancer informatics, 13, pp. CIN. S13779.

Baralle, F.E. and Giudice, J. (2017) 'Alternative splicing as a regulator of development and tissue identity', Nature reviews Molecular cell biology, 18(7), pp. 437–451.

Barbeito, C.G. and Miglino, M.A. (2023) ‘Editorial: Cellular processes in placental morphogenesis’, Frontiers in Cell and Developmental Biology, 11. Available at: https://doi.org/10.3389/fcell.2023.1298298.

Barman, M., Nilsson, S., Torinsson Naluai, Å, Sandin, A., Wold, A.E. and Sandberg, A. (2015) 'Single nucleotide polymorphisms in the FADS gene cluster but not the ELOVL2 gene are associated with serum polyunsaturated fatty acid composition and development of allergy (in a swedish birth cohort)', Nutrients, 7(12), pp. 10100–10115.

Baron, L.M. and Baron, B. (2023) 'Predictive biochemical factors for pre-eclampsia in early pregnancy', Current Women's Health Reviews, 19(1), pp. 19–34.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M. and Holko, M. (2012) 'NCBI GEO: Archive for functional genomics data sets—update', Nucleic acids research, 41(D1), pp. D991–D995.

Barton, J.R. and Sibai, B.M. (2008) 'Prediction and prevention of recurrent preeclampsia', Obstetrics & Gynecology, 112(2 Part 1), pp. 359–372.

Bartsch, E., Medcalf, K.E., Park, A.L. and Ray, J.G. (2016) 'Clinical risk factors for pre-eclampsia determined in early pregnancy: Systematic review and meta-analysis of large cohort studies', BMJ, 353, pp. i1753. Available at: https://doi.org/10.1136/bmj.i1753.

Bdolah, Y., Lam, C., Rajakumar, A., Shivalingappa, V., Mutter, W., Sachs, B.P., Lim, K.H., Bdolah-Abram, T., Epstein, F.H. and Karumanchi, S.A. (2008) 'Twin pregnancy and the risk of preeclampsia: Bigger placenta or relative ischemia?', American Journal of Obstetrics and Gynecology, 198(4), pp. 428.e1–6. Available at: https://doi.org/10.1016/j.ajog.2007.10.783.

Bengston, S.E., Dahan, R.A., Donaldson, Z., Phelps, S.M., Van Oers, K., Sih, A. and Bell, A.M. (2018) 'Genomic tools for behavioural ecologists to understand

repeatable individual differences in behaviour', Nature ecology & evolution, 2(6), pp. 944–955.

Ben-Israel, D., Jacobs, W.B., Casha, S., Lang, S., Ryu, W.H.A., de Lotbiniere-Bassett, M. and Cadotte, D.W. (2020) 'The impact of machine learning on patient care: A systematic review', Artificial Intelligence in Medicine, 103, pp. 101785.

Bennett, R., Mulla, Z.D., Parikh, P., Hauspurg, A. and Razzaghi, T. (2022) ‘An imbalance-aware deep neural network for early prediction of preeclampsia’, PLOS ONE, 17(4), p. e0266042. Available at: https://doi.org/10.1371/journal.pone.0266042.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L. and Bignell, H.R. (2008) 'Accurate whole human genome sequencing using reversible terminator chemistry', Nature, 456(7218), pp. 53–59.

Bergmann, F., Rotmensch, S., Rosenzweig, B., How, H. and Chediak, J. (1991) 'The role of von willebrand factor in pre-eclampsia', Thrombosis and Haemostasis, 66(5), pp. 525–528.

Bezeau, S. and Graves, R. (2001) 'Statistical power and effect sizes of clinical neuropsychology research', Journal of clinical and experimental neuropsychology, 23(3), pp. 399–406.

Bhadra, T., Mallik, S., Hasan, N. and Zhao, Z. (2022) ‘Comparison of five supervised feature selection algorithms leading to top features and gene signatures from multi-omics data in cancer’, BMC Bioinformatics, 23(3), p. 153. Available at: https://doi.org/10.1186/s12859-022-04678-y.

Bhattacharya, S., Campbell, D.M., Liston, W.A. and Bhattacharya, S. (2007) 'Effect of body mass index on pregnancy outcomes in nulliparous women delivering singleton babies', BMC public Health, 7, pp. 1–8.

Bhosale, Y.H., Nakkella, M., E, D.S.B. rani, Vamsikrishna, D.M., Garg, A. and Tiwari, R.G. (2025) ‘AI-Enabled Early Detection of Preeclampsia: A Predictive Model Based on Multivariate Biomarker Analysis.’, Vascular and Endovascular Review, 8(1s), pp. 296–303.

Bi, R. and Liu, P. (2016) 'Sample size calculation while controlling false discovery rate for differential expression analysis with RNA-sequencing experiments', BMC bioinformatics, 17(1), pp. 1–13.

Bianchi, D.W. (2004) 'Circulating fetal DNA: Its origin and diagnostic potential—a review', Placenta, 25, pp. S93–S101.

Bianchi, D.W., Parker, R.L., Wentworth, J., Madankumar, R., Saffer, C., Das, A.F., Craig, J.A., Chudova, D.I., Devers, P.L. and Jones, K.W. (2014) 'DNA sequencing versus standard prenatal aneuploidy screening', New England journal of medicine, 370(9), pp. 799–808.

Bischoff, F.Z., Lewis, D.E. and Simpson, J.L. (2005) 'Cell-free fetal DNA in maternal blood: Kinetics, source and structure', Human reproduction update, 11(1), pp. 59–67.

Bischoff, F.Z., Sinacori, M.K., Dang, D.D., Marquez-Do, D., Horne, C., Lewis, D.E. and Simpson, J.L. (2002) 'Cell-free fetal DNA and intact fetal cells in maternal blood circulation: Implications for first and second trimester non-invasive prenatal diagnosis', Human reproduction update, 8(6), pp. 493–500.

Blanchard, S.C., Kim, H.D., Gonzalez Jr, R.L., Puglisi, J.D. and Chu, S. (2004) 'tRNA dynamics on the ribosome during translation', Proceedings of the National Academy of Sciences, 101(35), pp. 12893–12898.

Bodnar, L.M., Catov, J.M. and Roberts, J.M. (2007) 'Racial/ethnic differences in the monthly variation of preeclampsia incidence', American Journal of Obstetrics and Gynecology, 196(4), pp. 324. e1–324. e5.

Bodnar, L.M., Catov, J.M., Klebanoff, M.A., Ness, R.B. and Roberts, J.M. (2007) 'Prepregnancy body mass index and the occurrence of severe hypertensive disorders of pregnancy', Epidemiology, , pp. 234–239.

Boeldt, D.S. and Bird, I.M. (2017) 'Vascular adaptation in pregnancy and endothelial dysfunction in preeclampsia', The Journal of endocrinology, 232(1), pp. R27.

Bogias, K.J., Pederson, S.M., Leemaqz, S., Smith, M.D., McAninch, D., Jankovic-Karasoulos, T., McCullough, D., Wan, Q., Bianco-Miotto, T., Breen, J. and Roberts, C.T. (2022) ‘Placental Transcription Profiling in 6–23 Weeks’ Gestation Reveals Differential Transcript Usage in Early Development’, International Journal of Molecular Sciences, 23(9), p. 4506. Available at: https://doi.org/10.3390/ijms23094506.

Bohr, A. and Memarzadeh, K. (2020) ‘The rise of artificial intelligence in healthcare applications’, Artificial Intelligence in Healthcare, pp. 25–60. Available at: https://doi.org/10.1016/B978-0-12-818438-7.00002-2.

Borzychowski, A.M., Sargent, I.L. and Redman, C.W.G. (2006) 'Inflammation and pre-eclampsia', Seminars in Fetal and Neonatal Medicine, 11(5), pp. 309–316. Available at: https://doi.org/10.1016/j.siny.2006.04.001.

Bosio, P.M., Cannon, S., McKenna, P.J., O'Herlihy, C., Conroy, R. and Brady, H. (2001) 'Plasma P-selectin is elevated in the first trimester in women who subsequently develop pre-eclampsia', British journal of obstetrics and gynaecology, 108(7), pp. 709–715.

Boss, A.L., Chamley, L.W. and James, J.L. (2018) 'Placental formation in early pregnancy: How is the centre of the placenta made?', Human Reproduction Update, 24(6), pp. 750–760. Available at: https://doi.org/10.1093/humupd/dmy030.

Bosschieter, T.M., Xu, Z., Lan, H., Lengerich, B.J., Nori, H., Sitcov, K., Souter, V. and Caruana, R. (2022) 'Using interpretable machine learning to predict maternal and fetal outcomes', arXiv preprint arXiv:2207.05322, .

Bottaci, L., Drew, P.J., Hartley, J.E., Hadfield, M.B., Farouk, R., Lee, P.W., Macintyre, I.M., Duthie, G.S. and Monson, J.R. (1997) 'Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions', The Lancet, 350(9076), pp. 469–472. Available at: https://doi.org/10.1016/S0140-6736(96)11196-X.

Bouter, A.R. and Duvekot, J.J. (2020) 'Evaluation of the clinical impact of the revised ISSHP and ACOG definitions on preeclampsia', Pregnancy Hypertension, 19, pp. 206–211. Available at: https://doi.org/10.1016/j.preghy.2019.11.011.

Boutin, A., Gasse, C., Demers, S., Giguère, Y., Tétu, A. and Bujold, E. (2018) 'Maternal characteristics for the prediction of preeclampsia in nulliparous women: The great obstetrical syndromes (GOS) study', Journal of obstetrics and gynaecology Canada: JOGC = Journal d'obstetrique et gynecologie du Canada: JOGC, 40(5), pp. 572–578. Available at: https://doi.org/10.1016/j.jogc.2017.07.025.

Bramham, K., Parnell, B., Nelson-Piercy, C., Seed, P.T., Poston, L. and Chappell, L.C. (2014) 'Chronic hypertension and pregnancy outcomes: Systematic review and meta-analysis', Bmj, 348.

Breitling, R. and Herzyk, P. (2005) 'Rank-based methods as a non-parametric alternative of the T-statistic for the analysis of biological microarray data', Journal of bioinformatics and computational biology, 3(05), pp. 1171–1189.

Breitling, R., Armengaud, P., Amtmann, A. and Herzyk, P. (2004) ‘Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments’, FEBS Letters, 573(1–3), pp. 83–92. Available at: https://doi.org/10.1016/j.febslet.2004.07.055.

Brew, O., Sullivan, M.H.F. and Woodman, A. (2016) ‘Comparison of Normal and Pre-Eclamptic Placental Gene Expression: A Systematic Review with Meta-Analysis’, PLOS ONE, 11(8), p. e0161504. Available at: https://doi.org/10.1371/journal.pone.0161504.

Briceño-Pérez, C., Briceño-Sanabria, L. and Vigil-De Gracia, P. (2009) 'Prediction and prevention of preeclampsia', Hypertension in pregnancy, 28(2), pp. 138–155.

Brito Querido, J., Sokabe, M., Díaz-López, I., Gordiyenko, Y., Fraser, C.S. and Ramakrishnan, V. (2024) 'The structure of a human translation initiation complex reveals two independent roles for the helicase eIF4A', Nature Structural & Molecular Biology, , pp. 1–10.

Bronson, S.L. and Bale, T.L. (2014) 'Prenatal stress-induced increases in placental inflammation and offspring hyperactivity are male-specific and ameliorated by maternal antiinflammatory treatment', Endocrinology, 155(7), pp. 2635–2646.

Brosens, I., Pijnenborg, R., Vercruysse, L. and Romero, R. (2011) 'The “great obstetrical syndromes” are associated with disorders of deep placentation', American Journal of Obstetrics and Gynecology, 204(3), pp. 193–201.

Brouard, J. and Bissonnette, N. (2012) 'Variant calling from RNA-seq data using the GATK joint genotyping workflow'Variant calling: Methods and protocols Springer, pp. 205–233. .

Brouard, J., Schenkel, F., Marete, A. and Bissonnette, N. (2019) 'The GATK joint genotyping workflow is appropriate for calling variants in RNA-seq experiments', Journal of animal science and biotechnology, 10, pp. 1–6.

Brown, M.A., Lindheimer, M.D., de Swiet, M., Van Assche, A. and Moutquin, J.M. (2001) 'The classification and diagnosis of the hypertensive disorders of pregnancy: Statement from the international society for the study of hypertension in pregnancy (ISSHP)', Hypertension in Pregnancy, 20(1), pp. IX–XIV. Available at: https://doi.org/10.1081/PRG-100104165.

Brown, M.A., Magee, L.A., Kenny, L.C., Karumanchi, S.A., McCarthy, F.P., Saito, S., Hall, D.R., Warren, C.E., Adoyi, G. and Ishaku, S. (2018) 'Hypertensive disorders of pregnancy', Hypertension, 72(1), pp. 24–43. Available at: https://doi.org/10.1161/HYPERTENSIONAHA.117.10803.

Brümmer, A. and Bergmann, S. (2024) ‘Disentangling genetic effects on transcriptional and post-transcriptional gene regulation through integrating exon and intron expression QTLs’, Nature Communications, 15, p. 3786. Available at: https://doi.org/10.1038/s41467-024-48244-x.

Bujold, E., Roberge, S., Lacasse, Y., Bureau, M., Audibert, F., Marcoux, S., Forest, J. and Giguere, Y. (2010) 'Prevention of preeclampsia and intrauterine growth restriction with aspirin started in early pregnancy: A meta-analysis', Obstetrics & Gynecology, 116(2 Part 1), pp. 402–414.

Bujold, E., Roberge, S., Lacasse, Y., Bureau, M., Audibert, F., Marcoux, S., Forest, J.-C. and Giguère, Y. (2010) ‘Prevention of Preeclampsia and Intrauterine Growth Restriction With Aspirin Started in Early Pregnancy: A Meta-Analysis’, Obstetrics & Gynecology, 116(2 Part 1), p. 402. Available at: https://doi.org/10.1097/AOG.0b013e3181e9322a.

Burel, J.G. and Peters, B. (2018) 'Discovering transcriptional signatures of disease for diagnosis versus mechanism', Nature reviews. Immunology, 18(5), pp. 289–290. Available at: https://doi.org/10.1038/nri.2018.26.

Burton, G.J. and Fowden, A.L. (2015) 'The placenta: A multifaceted, transient organ', Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1663), pp. 20140066.

Burton, G.J., Redman, C.W., Roberts, J.M. and Moffett, A. (2019) ‘Pre-eclampsia: pathophysiology and clinical implications’, Bmj, 366(Journal Article).

Burton, G.J., Redman, C.W., Roberts, J.M. and Moffett, A. (2019) 'Pre-eclampsia: Pathophysiology and clinical implications', Bmj, 366.

Bustin, S.A., Benes, V., Nolan, T. and Pfaffl, M.W. (2005) 'Quantitative real-time RT-PCR–a perspective', Journal of Molecular Endocrinology, 34(3), pp. 597–601.

Butler, L., Gunturkun, F., Chinthala, L., Karabayir, I., Tootooni, M.S., Bakir-Batu, B., Celik, T., Akbilgic, O. and Davis, R.L. (2024) ‘AI-based preeclampsia detection and prediction with electrocardiogram data’, Frontiers in Cardiovascular Medicine, 11. Available at: https://doi.org/10.3389/fcvm.2024.1360238.

Butler, L., Gunturkun, F., Chinthala, L., Karabayir, I., Tootooni, M.S., Bakir-Batu, B., Celik, T., Akbilgic, O. and Davis, R.L. (2024) 'AI-based preeclampsia detection and prediction with electrocardiogram data', Frontiers in Cardiovascular Medicine, 11. Available at: https://doi.org/10.3389/fcvm.2024.1360238.

Buyon, J.P., Kim, M.Y., Guerra, M.M., Laskin, C.A., Petri, M., Lockshin, M.D., Sammaritano, L., Branch, D.W., Porter, T.F. and Sawitzke, A. (2015) 'Predictors of pregnancy outcomes in patients with lupus: A cohort study', Annals of Internal Medicine, 163(3), pp. 153–163.

Cabús, L., Lagarde, J., Curado, J., Lizano, E. and Pérez-Boza, J. (2022) 'Current challenges and best practices for cell-free long RNA biomarker discovery', Biomarker Research, 10(1), pp. 62. Available at: https://doi.org/10.1186/s40364-022-00409-w.

Cao, C., Saxena, R. and Gray, K.J. (2024) ‘Placental Origins of Preeclampsia: Insights from Multi-Omic Studies’, International Journal of Molecular Sciences, 25(17), p. 9343. Available at: https://doi.org/10.3390/ijms25179343.

Carreño, J.F. and Qiu, P. (2020) Feature selection algorithms for predicting preeclampsia: A comparative approach. IEEE, pp. 2626.

Carter, A.M. and Enders, A.C. (2004) ‘Comparative aspects of trophoblast development and placentation’, Reproductive biology and endocrinology : RB&E, 2, p. 46. Available at: https://doi.org/10.1186/1477-7827-2-46.

Castellanos Gutierrez, A.S., Figueras, F., Morales-Prieto, D.M., Schleußner, E., Espinosa, G. and Baños, N. (2022) ‘Placental damage in pregnancies with systemic lupus erythematosus: A narrative review’, Frontiers in Immunology, 13. Available at: https://doi.org/10.3389/fimmu.2022.941586.

Castillo-Marco, N., Cordero, T., Igual, M., Muñoz-Blat, I., Gómez-Álvarez, C., Bernat-González, N., Gaspar-Doménech, Á., Ortiz-Domingo, É., Vives, A.,

Ortega-Sanchís, S., Monfort-Ortiz, R., Volkov, P., Delgado, J.L., Hernandez-Hernandez, L., Canovas, E., Gil, M. del M., Santacruz, B., Gonzalez-Gonzalez, N.L., Plasencia, W., Perales-Marín, A., Marcos-Puig, B., Palacios-Marqués, A.M., Melchor, Í., Martin-Martinez, A., Benitez-Delgado, T., Simón, C. and Garrido-Gómez, T. (2025) 'Maternal plasma cell-free RNA as a predictor of early and late-onset preeclampsia throughout pregnancy', Nature Communications, 16(1), p. 9208. Available at: https://doi.org/10.1038/s41467-025-64215-2.

Catalano, P. and Demouzon, S.H. (2015) 'Maternal obesity and metabolic risk to the offspring: Why lifestyle interventions may have not achieved the desired outcomes', International journal of obesity, 39(4), pp. 642–649.

Caulfield, T., McGuire, A.L., Cho, M., Buchanan, J.A., Burgess, M.M., Danilczyk, U., Diaz, C.M., Fryer-Edwards, K., Green, S.K., Hodosh, M.A., Juengst, E.T., Kaye, J., Kedes, L., Knoppers, B.M., Lemmens, T., Meslin, E.M., Murphy, J., Nussbaum, R.L., Otlowski, M., Pullman, D., Ray, P.N., Sugarman, J. and Timmons, M. (2008) 'Research ethics recommendations for whole-genome research: Consensus statement', PLoS Biology, 6(3). Available at: https://doi.org/10.1371/journal.pbio.0060073.

Cavazos-Rehg, P.A., Krauss, M.J., Spitznagel, E.L., Bommarito, K., Madden, T., Olsen, M.A., Subramaniam, H., Peipert, J.F. and Bierut, L.J. (2015) 'Maternal age and risk of labor and delivery complications', Maternal and child health journal, 19, pp. 1202–1211.

Cech, T.R. and Steitz, J.A. (2014) 'The noncoding RNA revolution—trashing old rules to forge new ones', Cell, 157(1), pp. 77–94.

Chaemsaithong, P., Pooh, R.K., Zheng, M., Ma, R., Chaiyasit, N., Tokunaka, M., Shaw, S.W., Seshadri, S., Choolani, M., Wataganara, T., Yeo, G.S.H., Wright, A., Leung, W.C., Sekizawa, A., Hu, Y., Naruse, K., Saito, S., Sahota, D., Leung, T.Y. and Poon, L.C. (2019) 'Prospective evaluation of screening performance of first-trimester prediction models for preterm preeclampsia in an Asian population', American Journal of Obstetrics and Gynecology, 221(6), p. 650.e1-650.e16. Available at: https://doi.org/10.1016/j.ajog.2019.09.041.

Chaemsaithong, P., Sahota, D.S. and Poon, L.C. (2022) 'First trimester preeclampsia screening and prediction', American Journal of Obstetrics and Gynecology, 226(2S), pp. S1071-S1097.e2. Available at: https://doi.org/10.1016/j.ajog.2020.07.020.

Chaemsaithong, P., Sahota, D.S. and Poon, L.C. (2022) 'First trimester preeclampsia screening and prediction', American Journal of Obstetrics and Gynecology, 226(2), pp. S1071–S1097. e2.

Chaiworapongsa, T., Chaemsaithong, P., Yeo, L. and Romero, R. (2014) 'Pre-eclampsia part 1: Current understanding of its pathophysiology', Nature Reviews Nephrology, 10(8), pp. 466–480.

Chaiworapongsa, T., Romero, R., Whitten, A., Tarca, A.L., Bhatti, G., Draghici, S., Chaemsaithong, P., Miranda, J., Kim, C.J. and Hassan, S.S. (2013) 'Differences and similarities in the transcriptional profile of peripheral whole blood in early and late-onset preeclampsia: Insights into the molecular basis of the phenotype of preeclampsia', Journal of perinatal medicine, 41(5), pp. 485–504. Available at: https://doi.org/10.1515/jpm-2013-0082.

Chan, K.A., Zhang, J., Hui, A.B., Wong, N., Lau, T.K., Leung, T.N., Lo, K., Huang, D.W. and Lo, Y.D. (2004) 'Size distributions of maternal and fetal DNA in maternal plasma', Clinical chemistry, 50(1), pp. 88–92.

Chandrasekhar, N. and Peddakrishna, S. (2023) 'Enhancing heart disease prediction accuracy through machine learning techniques and optimization', Processes, 11(4), pp. 1210. Available at: https://doi.org/10.3390/pr11041210.

Chang, C., Cheng, W., Chen, C., Shu, W., Tsai, M., Huang, C. and Hsu, I.C. (2011) 'Identification of human housekeeping genes and tissue-selective genes by microarray meta-analysis', PloS one, 6(7), pp. e22859.

Chang, K., Seow, K. and Chen, K. (2023) 'Preeclampsia: Recent advances in predicting, preventing, and managing the maternal and fetal life-threatening condition', International Journal of Environmental Research and Public Health, 20(4), pp. 2994.

Chao, Y., Jiang, Y., Zhong, M., Wei, K., Hu, C., Qin, Y., Zuo, Y., Yang, L., Shen, Z. and Zou, C. (2021) 'Regulatory roles and mechanisms of alternative RNA splicing in adipogenesis and human metabolic health', Cell & Bioscience, 11(1), pp. 1–16.

Chappell, L.C., Cluver, C.A., Kingdom, J. and Tong, S. (2021) ‘Pre-eclampsia’, The Lancet, 398(10297), pp. 341–354. Available at: https://doi.org/10.1016/S0140-6736(20)32335-7.

Chappell, L.C., Cluver, C.A., Kingdom, J. and Tong, S. (2021) 'Pre-eclampsia', The Lancet, 398(10297), pp. 341–354. Available at: https://doi.org/10.1016/S0140-6736(20)32335-7.

Charan, J. and Biswas, T. (2013) 'How to calculate sample size for different study designs in medical research?', Indian journal of psychological medicine, 35(2), pp. 121–126.

Chatterjee, S., Ouidir, M. and Tekola-Ayele, F. (2021) 'Genetic and in utero environmental contributions to DNA methylation variation in placenta', Human Molecular Genetics, 30(21), pp. 1968–1976. Available at: https://doi.org/10.1093/hmg/ddab161.

Chen, H.H., Jin, Y., Huang, Y. and Chen, Y. (2016) 'Detection of high variability in gene expression from single-cell RNA-seq profiling', BMC genomics, 17, pp. 119–128.

Chen, I., Jhangri, G.S. and Chandra, S. (2014) ‘Relationship between interpregnancy interval and congenital anomalies’, American Journal of Obstetrics and Gynecology, 210(6), p. 564.e1–8. Available at: https://doi.org/10.1016/j.ajog.2014.02.002.

Chen, I., Jhangri, G.S. and Chandra, S. (2014) 'Relationship between interpregnancy interval and congenital anomalies', American Journal of Obstetrics and Gynecology, 210(6), pp. 564. e1–564. e8.

Chen, J., Shao, J., Zhang, X., Wei, S., Cai, H. and Wang, G. (2022) 'Monitoring early dynamic changes of plasma cell-free DNA and pretreatment pre-albumin to predict chemotherapy effectiveness and survival outcomes in advanced non-small cell lung cancer', Annals of Translational Medicine, 10(5), pp. 253–253. Available at: https://doi.org/10.21037/atm-22-12.

Chen, J.H. and Asch, S.M. (2017) 'Machine learning and prediction in medicine — beyond the peak of inflated expectations', The New England journal of medicine, 376(26), pp. 2507–2509. Available at: https://doi.org/10.1056/NEJMp1702071.

Chen, R., Morgan, A.A., Dudley, J., Deshpande, T., Li, L., Kodama, K., Chiang, A.P. and Butte, A.J. (2008) 'FitSNPs: Highly differentially expressed genes are more likely to have variants associated with disease', Genome biology, 9(12), pp. 1–15.

Chen, S., Mukherjee, N., Janjanam, V.D., Arshad, S.H., Kurukulaaratchy, R.J., Holloway, J.W., Zhang, H. and Karmaus, W. (2017) 'Consistency and variability of DNA methylation in women during puberty, young adulthood, and pregnancy', Genetics & Epigenetics, 9, pp. 1179237X17721540.

Chen, X. and Lin, X. (2014) 'Big data deep learning: Challenges and perspectives', IEEE access, 2, pp. 514–525.

Chen, Y., Lun, A.T. and Smyth, G.K. (2014) 'Differential expression analysis of complex RNA-seq experiments using edgeR', Statistical analysis of next generation sequencing data, , pp. 51–74.

Chhatbar, K., Bird, A. and Sanguinetti, G. (2025) ‘Modelling transcription with explainable AI uncovers context-specific epigenetic gene regulation at promoters and gene bodies’, PLOS Genetics, 21(10), p. e1011908. Available at: https://doi.org/10.1371/journal.pgen.1011908.

Chiang, Y.-T., Seow, K.-M. and Chen, K.-H. (2024) ‘The Pathophysiological, Genetic, and Hormonal Changes in Preeclampsia: A Systematic Review of the Molecular Mechanisms’, International Journal of Molecular Sciences, 25(8), p. 4532. Available at: https://doi.org/10.3390/ijms25084532.

Chibon, F. (2013) 'Cancer gene expression signatures–the rise and fall?', European journal of cancer, 49(8), pp. 2000–2009.

Chim, S.S., Shing, T.K., Hung, E.C., Leung, T., Lau, T., Chiu, R.W. and Dennis Lo, Y.M. (2008) 'Detection and characterization of placental microRNAs in maternal plasma', Clinical chemistry, 54(3), pp. 482–490.

Chiu, R.W., Akolekar, R., Zheng, Y.W., Leung, T.Y., Sun, H., Chan, K.A., Lun, F.M., Go, A.T., Lau, E.T. and To, W.W. (2011) 'Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: Large scale validity study', Bmj, 342.

Chiu, R.W.K. and Lo, Y.M.D. (2021) 'Cell-free fetal DNA coming in all sizes and shapes', Prenatal Diagnosis, 41(10), pp. 1193–1201. Available at: https://doi.org/10.1002/pd.5952.

Choi, J. (2013) 'Guide: A desktop application for analysing gene expression data', BMC genomics, 14(1), pp. 1–7.

Chotaliya, J.R., Ladumor, D.G. and Dhorajiya, P.S. (2025) ‘Circulating Cell-Free Fetal DNA Methylation Profiles as Early Predictors of Preeclampsia: A Prospective Study’, Journal of Heart Valve Disease, 30, pp. 121–125.

Chowdhury, H.A., Bhattacharyya, D.K. and Kalita, J.K. (2020) 'Differential expression analysis of RNA-seq reads: Overview, taxonomy, and tools', IEEE/ACM Transactions on Computational Biology and Bioinformatics, 17(2), pp. 566–586. Available at: https://doi.org/10.1109/TCBB.2018.2873010.

Cirulli, E.T., Singh, A., Shianna, K.V., Ge, D., Smith, J.P., Maia, J.M., Heinzen, E.L., Goedert, J.J., Goldstein, D.B. and Center for HIV/AIDS Vaccine Immunology (CHAVI) (2010) 'Screening the human exome: A comparison of whole genome and whole transcriptome sequencing', Genome biology, 11, pp. 1–8.

Clare, A. and Sakharkar, V. (2025) ‘Role of second trimester uterine artery doppler in early prediction of preeclampsia and intrauterine growth restriction among high-risk Bahamian residents’, SAGE Open Medicine, 13, p. 20503121251378109. Available at: https://doi.org/10.1177/20503121251378109.

Clausen, T., Djurovic, S., Reseland, J.E., Berg, K., Drevon, C.A. and Henriksen, T. (2002) 'Altered plasma concentrations of leptin, transforming growth factor-β1 and plasminogen activator inhibitor type 2 at 18 weeks of gestation in women destined to develop pre-eclampsia. circulating markers of disturbed placentation?', Placenta, 23(5), pp. 380–385.

Clough, E., Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Lee, H., Zhang, N., Serova, N., Wagner, L., Zalunin, V., Kochergin, A. and Soboleva, A. (2024) ‘NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update’, Nucleic Acids Research, 52(D1), pp. D138–D144. Available at: https://doi.org/10.1093/nar/gkad965.

Cnattingius, S., Reilly, M., Pawitan, Y. and Lichtenstein, P. (2004) 'Maternal and fetal genetic factors account for most of familial aggregation of preeclampsia: A population-based swedish cohort study', American journal of medical genetics Part A, 130(4), pp. 365–371.

Cock, P.J.A., Fields, C.J., Goto, N., Heuer, M.L. and Rice, P.M. (2010) 'The sanger FASTQ file format for sequences with quality scores, and the solexa/illumina FASTQ variants', Nucleic Acids Research, 38(6), pp. 1767–1771. Available at: https://doi.org/10.1093/nar/gkp1137.

Coghlan, S., Gyngell, C. and Vears, D.F. (2024) 'Ethics of artificial intelligence in prenatal and pediatric genomic medicine', Journal of Community Genetics, 15(1), pp. 13–24. Available at: https://doi.org/10.1007/s12687-023-00678-4.

Cogswell, M.E. and Yip, R. (1995) 'The influence of fetal and maternal factors on the distribution of birthweight', Seminars in Perinatology, 19(3), pp. 222–240. Available at: https://doi.org/10.1016/s0146-0005(05)80028-x.

Cole, M.B., Risso, D., Wagner, A., DeTomaso, D., Ngai, J., Purdom, E., Dudoit, S. and Yosef, N. (2019) 'Performance assessment and selection of normalization procedures for single-cell RNA-seq', Cell systems, 8(4), pp. 315–328. e8.

Conde-Agudelo, A. and Belizán, J.M. (2000) 'Risk factors for pre-eclampsia in a large cohort of latin american and caribbean women', BJOG: An International Journal of Obstetrics & Gynaecology, 107(1), pp. 75–83.

Conde-Agudelo, A., Rosas-Bermúdez, A. and Kafury-Goeta, A.C. (2006) 'Birth spacing and risk of adverse perinatal outcomes: A meta-analysis', Jama, 295(15), pp. 1809–1823.

Conde-Agudelo, A., Villar, J. and Lindheimer, M. (2004) 'World health organization systematic review of screening tests for preeclampsia', Obstetrics & Gynecology, 104(6), pp. 1367–1391.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L. and Zhang, X. (2016) 'A survey of best practices for RNA-seq data analysis', Genome biology, 17(1), pp. 1–19.

Corchete, L.A., Rojas, E.A., Alonso-López, D., De Las Rivas, J., Gutiérrez, N.C. and Burguillo, F.J. (2020) 'Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis', Scientific Reports, 10(1), pp. 1–15. Available at: https://doi.org/10.1038/s41598-020-76881-x.

Cornwell, M., Vangala, M., Taing, L., Herbert, Z., Köster, J., Li, B., Sun, H., Li, T., Zhang, J. and Qiu, X. (2018) 'VIPER: Visualization pipeline for RNA-seq, a snakemake workflow for efficient and complete RNA-seq analysis', BMC bioinformatics, 19(1), pp. 1–14.

Costa-Silva, J., Domingues, D. and Lopes, F.M. (2017) 'RNA-seq differential expression analysis: An extended review and a software tool', PLOS

ONE, 12(12), pp. e0190152. Available at: https://doi.org/10.1371/journal.pone.0190152.

Cox, L.S. and Redman, C. (2017) 'The role of cellular senescence in ageing of the placenta', Placenta, 52, pp. 139–145.

Da Sacco, L., Baldassarre, A. and Masotti, A. (2011) 'Bioinformatics tools and novel challenges in long non-coding RNAs (lncRNAs) functional analysis', International journal of molecular sciences, 13(1), pp. 97–114.

Dacaj, R., Izetbegovic, S., Stojkanovic, G. and Dreshaj, S. (2016) 'Elevated liver enzymes in cases of preeclampsia and intrauterine growth restriction', Medical Archives, 70(1), pp. 44–47. Available at: https://doi.org/10.5455/medarh.2016.70.44-47.

Dahlstrom, J.E., Subramaniam, A. and Faye-Petersen, O.M. (2022) 'Maternal vascular malperfusion and associated maternal diseases', Benirschke's Pathology of the Human Placenta, , pp. 507–554.

Dai, K., Yu, R., Meng, Y., Yang, X., Jiang, Y., Luo, J., Fang, K., Wang, S. and Rong, Z. (2025) ‘Construction and Validation of a PCA-Based Prediction Model for Preterm Infant Stunting Risk: A Retrospective Study’, Children, 12(5), p. 583. Available at: https://doi.org/10.3390/children12050583.

Das, S., Mitra, K. and Mandal, M. (2016) 'Sample size calculation: Basic principles', Indian Journal of Anaesthesia, 60(9), pp. 652.

Dathan-Stumpf, A., Rieger, A., Verlohren, S., Wolf, C. and Stepan, H. (2022) ‘sFlt-1/PlGF ratio for prediction of preeclampsia in clinical routine: A pragmatic real-world analysis of healthcare resource utilisation’, PLoS ONE, 17(2), p. e0263443. Available at: https://doi.org/10.1371/journal.pone.0263443.

Davignon, J. and Ganz, P. (2004) 'Role of endothelial dysfunction in atherosclerosis', Circulation, 109(23 Suppl 1), pp. III27–32. Available at: https://doi.org/10.1161/01.CIR.0000131515.03336.f8.

de Jesus, G.R., Agmon-Levin, N., Andrade, C.A., Andreoli, L., Chighizola, C.B., Porter, T.F., Salmon, J., Silver, R.M., Tincani, A. and Branch, D.W. (2014) '14th international congress on antiphospholipid antibodies task force report on obstetric antiphospholipid syndrome', Autoimmunity reviews, 13(8), pp. 795–813.

de Jong, T.V., Moshkin, Y.M. and Guryev, V. (2019) 'Gene expression variability: The other dimension in transcriptome analysis', Physiological Genomics, 51(5), pp. 145–158. Available at: https://doi.org/10.1152/physiolgenomics.00128.2018.

De Kat, A.C., Hirst, J., Woodward, M., Kennedy, S. and Peters, S.A. (2019) ‘Prediction models for preeclampsia: A systematic review’, Pregnancy Hypertension, 16(Journal Article), pp. 48–66. Available at: https://doi.org/10.1016/j.preghy.2019.03.005.

De Kat, A.C., Hirst, J., Woodward, M., Kennedy, S. and Peters, S.A. (2019) 'Prediction models for preeclampsia: A systematic review', Pregnancy Hypertension, 16, pp. 48–66. Available at: https://doi.org/10.1016/j.preghy.2019.03.005.

De Maio, A., Yalamanchili, H.K., Adamski, C.J., Gennarino, V.A., Liu, Z., Qin, J., Jung, S.Y., Richman, R., Orr, H. and Zoghbi, H.Y. (2018) 'RBM17 Interacts with U2SURP and CHERP to Regulate Expression and Splicing of RNA-Processing Proteins', Cell Reports, 25(3), pp. 726-736.e7. Available at: https://doi.org/10.1016/j.celrep.2018.09.041.

De Marco, O., Ruotolo, C., Njanjo, L., Mariani, C., Mazé, B., Santagati, G., Fois, A., Chevé, M.T., Chatrenet, A., Salomone, M., Torreggiani, M. and Piccoli, G.B. (2025) 'Recurrence of preeclampsia is common, even during rigorously controlled multidisciplinary follow-up: a pilot experience', Clinical Kidney Journal, 18(4), p. sfaf070. Available at: https://doi.org/10.1093/ckj/sfaf070.

de Sena Brandine, G. and Smith, A.D. (2019) 'Falco: High-speed FastQC emulation for quality control of sequencing data', F1000Research, 8.

de Vries, J., Bull, S.J., Doumbo, O., Ibrahim, M., Mercereau-Puijalon, O., Kwiatkowski, D. and Parker, M. (2011) 'Ethical issues in human genomics research in developing countries', BMC Medical Ethics, 12, pp. 5. Available at: https://doi.org/10.1186/1472-6939-12-5.

Degnes, M.-H.L., Westerberg, A.C., Andresen, I.J., Henriksen, T., Roland, M.C.P., Zucknick, M. and Michelsen, T.M. (2024) 'Protein biomarker signatures of preeclampsia - a longitudinal 5000-multiplex proteomics study', Scientific Reports, 14, p. 23654. Available at: https://doi.org/10.1038/s41598-024-73796-9.

Dekker, G.A. and Sibai, B.M. (1998) 'Etiology and pathogenesis of preeclampsia: Current concepts', American Journal of Obstetrics and Gynecology, 179(5), pp. 1359–1375. Available at: https://doi.org/10.1016/S0002-9378(98)70160-7.

Del Fabbro, C., Scalabrin, S., Morgante, M. and Giorgi, F.M. (2013) 'An extensive evaluation of read trimming effects on illumina NGS data analysis', PLoS ONE, 8(12). Available at: https://doi.org/10.1371/journal.pone.0085024.

Del Vecchio, G., Li, Q., Li, W., Thamotharan, S., Tosevska, A., Morselli, M., Sung, K., Janzen, C., Zhou, X., Pellegrini, M. and Devaskar, S.U. (2021) 'Cell-free DNA methylation and transcriptomic signature prediction of pregnancies with adverse outcomes', Epigenetics, 16(6), pp. 642–661. Available at: https://doi.org/10.1080/15592294.2020.1816774.

DeLuca, D.S., Levin, J.Z., Sivachenko, A., Fennell, T., Nazaire, M., Williams, C., Reich, M., Winckler, W. and Getz, G. (2012) 'RNA-SeQC: RNA-seq metrics for quality control and process optimization', Bioinformatics, 28(11), pp. 1530–1532.

Deng, C. and Liu, S. (2022) 'Factors affecting the fetal fraction in noninvasive prenatal screening: A review', Frontiers in Pediatrics, 10, pp. 812781. Available at: https://doi.org/10.3389/fped.2022.812781.

Deng, Q., Ji, T., Chu, G. and Tan, H. (2025) ‘Pregnancy Outcomes Following Assisted Reproductive Technology in Advanced Maternal Age Primiparous Women: A Retrospective Cohort Study’, International Journal of Women’s Health, 17, pp. 2443–2449. Available at: https://doi.org/10.2147/IJWH.S538128.

Dever, T.E. and Green, R. (2012) 'The elongation, termination, and recycling phases of translation in eukaryotes', Cold Spring Harbor perspectives in biology, 4(7), pp. a013706.

Dhallan, R., Guo, X., Emche, S., Damewood, M., Bayliss, P., Cronin, M., Barry, J., Betz, J., Franz, K., Gold, K., Vallecillo, B. and Varney, J. (2007) 'A non-invasive test for prenatal diagnosis based on fetal DNA present in maternal blood: A preliminary study', Lancet (London, England), 369(9560), pp. 474–481. Available at: https://doi.org/10.1016/S0140-6736(07)60115-9.

Dimitriadis, E., Rolnik, D.L., Zhou, W., Estrada-Gutierrez, G., Koga, K., Francisco, R.P.V., Whitehead, C., Hyett, J., da Silva Costa, F., Nicolaides, K. and Menkhorst, E. (2023) ‘Pre-eclampsia’, Nature Reviews Disease Primers, 9(1), p. 8. Available at: https://doi.org/10.1038/s41572-023-00417-6.

Dixon, D., Sattar, H., Moros, N., Kesireddy, S.R., Ahsan, H., Lakkimsetti, M., Fatima, M., Doshi, D., Sadhu, K. and Hassan, M.J. (2024) 'Unveiling the influence of AI predictive analytics on patient outcomes: A comprehensive narrative review', Cureus, 16(5).

Dobin, A. and Gingeras, T.R. (2015) 'Mapping RNA-seq reads with STAR', Current Protocols in Bioinformatics, 51(1), pp. 11.14.1–11.14.19. Available at: https://doi.org/10.1002/0471250953.bi1114s51.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) 'STAR: Ultrafast universal RNA-seq aligner', Bioinformatics, 29(1), pp. 15–21. Available at: https://doi.org/10.1093/bioinformatics/bts635.

Dong, X., Han, M., Zeb, S., Tong, M., Li, X. and Chen, Q. (2021) ‘Active Management Reduces the Incidence of Recurrent Pre-eclampsia and Improves Maternal and Fetal Outcomes in Women With Recurrent Pre-eclampsia’, Frontiers in Medicine, 8. Available at: https://doi.org/10.3389/fmed.2021.658022.

Donno, V., Prats, P., Rodriguez, I. and Polyzos, N.P. (2025) ‘First-trimester uterine artery pulsatility index and preeclampsia risk in pregnancies after artificial frozen embryo transfer: analysis of over 27,000 pregnancies’, American Journal of Obstetrics and Gynecology, 232(5), p. 464.e1-464.e9. Available at: https://doi.org/10.1016/j.ajog.2024.10.033.

Drejza, M.A., Rylewicz, K., Majcherek, E., Gross-Tyrkin, K., Mizgier, M., Plagens-Rotman, K., Wójcik, M., Panecka-Mysza, K., Pisarska-Krawczyk, M., Kędzia, W. and Jarząbek-Bielecka, G. (2022) 'Markers of oxidative stress in obstetrics and Gynaecology—A systematic literature review', Antioxidants, 11(8), pp. 1477. Available at: https://doi.org/10.3390/antiox11081477.

Dube, L. and Verster, T. (2023) 'Enhancing classification performance in imbalanced datasets: A comparative analysis of machine learning models', Data Science in Finance and Economics, 3(4), pp. 354–379.

Duckitt, K. and Harrington, D. (2005) 'Risk factors for pre-eclampsia at antenatal booking: Systematic review of controlled studies', Bmj, 330(7491), pp. 565.

Duggal, P., Gillanders, E.M., Holmes, T.N. and Bailey-Wilson, J.E. (2008) 'Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies', BMC genomics, 9, pp. 1–8.

Duley, L. (2009) The global impact of pre-eclampsia and eclampsia. Elsevier, pp. 130.

Dvinge, H. and Bradley, R.K. (2015) 'Widespread intron retention diversifies most cancer transcriptomes', Genome medicine, 7, pp. 1–13.

Eberhard, B.W., Cohen, R.Y., Wheeler, N., Kleinlein, R., Rigoni, J., Bates, D.W., Gray, K.J. and Kovacheva, V.P. (2025) ‘Development and validation of an interpretable longitudinal preeclampsia risk prediction using machine learning’, PLOS One, 20(6), p. e0323873. Available at: https://doi.org/10.1371/journal.pone.0323873.

Eddy, A.C., Bidwell, G.L. and George, E.M. (2018) 'Pro-angiogenic therapeutics for preeclampsia', Biology of sex differences, 9(1), pp. 1–11.

Egbor, M., Ansari, T., Morris, N., Green, C.J. and Sibbons, P.D. (2006) 'Maternal medicine: Morphometric placental villous and vascular abnormalities in early- and late-onset pre-eclampsia with and without fetal growth restriction', BJOG: An International Journal of Obstetrics & Gynaecology, 113(5), pp. 580–589. Available at: https://doi.org/10.1111/j.1471-0528.2006.00882.x.

Eghbal-Fard, S., Yousefi, M., Heydarlou, H., Ahmadi, M., Taghavi, S., Movasaghpour, A., Jadidi-Niaragh, F., Yousefi, B., Dolati, S. and Hojjat-Farsangi, M. (2019) 'The imbalance of Th17/treg axis involved in the pathogenesis of preeclampsia', Journal of cellular physiology, 234(4), pp. 5106–5116.

Eidem, H.R., Ackerman, W.E., McGary, K.L., Abbot, P. and Rokas, A. (2015) ‘Gestational tissue transcriptomics in term and preterm human pregnancies: a systematic review and meta-analysis’, BMC Medical Genomics, 8, p. 27. Available at: https://doi.org/10.1186/s12920-015-0099-8.

Eisenberg, E. and Levanon, E.Y. (2013) 'Human housekeeping genes, revisited', TRENDS in Genetics, 29(10), pp. 569–574.

Eisinga, R., Breitling, R. and Heskes, T. (2013) ‘The exact probability distribution of the rank product statistics for replicated experiments’, FEBS Letters, 587(6), pp. 677–682. Available at: https://doi.org/10.1016/j.febslet.2013.01.037.

Elmore, S. (2007) 'Apoptosis: A review of programmed cell death', Toxicologic pathology, 35(4), pp. 495–516. Available at: https://doi.org/10.1080/01926230701320337.

El-Sayed, A.A.F. (2017) 'Preeclampsia: A review of the pathogenesis and possible management strategies based on its pathophysiological derangements', Taiwanese Journal of Obstetrics and Gynecology, 56(5), pp. 593–598. Available at: https://doi.org/10.1016/j.tjog.2017.08.004.

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B. and Gunnarsdottir, S. (2008) 'Genetics of gene expression and its effect on disease', Nature, 452(7186), pp. 423–428.

Erez, O., Romero, R., Jung, E., Chaemsaithong, P., Bosco, M., Suksai, M., Gallo, D.M. and Gotsch, F. (2022) 'Preeclampsia and eclampsia: The conceptual evolution of a syndrome', American Journal of Obstetrics and Gynecology, 226(2, Supplement), pp. S786–S803. Available at: https://doi.org/10.1016/j.ajog.2021.12.001.

Esfahani, M.S., Hamilton, E.G., Mehrmohamadi, M., Nabet, B.Y., Alig, S.K., King, D.A., Steen, C.B., Macaulay, C.W., Schultz, A. and Nesselbush, M.C. (2022) 'Inferring gene expression from cell-free DNA fragmentation profiles', Nature biotechnology, 40(4), pp. 585–597.

Espinosa, C., Becker, M., Marić, I., Wong, R.J., Shaw, G.M., Gaudilliere, B., Aghaeepour, N. and Stevenson, D.K. (2021) ‘Data-driven modeling of pregnancy-related complications’, Trends in molecular medicine, 27(8), pp. 762–776. Available at: https://doi.org/10.1016/j.molmed.2021.01.007.

Espín-Pérez, A., Portier, C., Chadeau-Hyam, M., van Veldhoven, K., Kleinjans, J.C. and de Kok, T.M. (2018) 'Comparison of statistical methods and the use of quality control samples for batch effect correction in human transcriptome data', PloS one, 13(8), pp. e0202947.

Esplin, M.S., Fausett, M.B., Fraser, A., Kerber, R., Mineau, G., Carrillo, J. and Varner, M.W. (2001) 'Paternal and maternal components of the predisposition to preeclampsia', New England Journal of Medicine, 344(12), pp. 867–872.

Estrela, D., Santos, R.F., Masserdotti, A., Silini, A., Parolini, O., Pinto, I.M. and Cruz, A. (2025) ‘Molecular Biomarkers for Timely and Personalized Prediction of Maternal-Fetal Health Risk’, Biomolecules, 15(3), p. 312. Available at: https://doi.org/10.3390/biom15030312.

Etesami, E., Nikukar, H., Ramezanali, F., Ghotbizadeh Vahdani, F., Shahhoseini, M., Favaedi, R., Ghaheri, A., Seydabadi, S. and Zamanian, M. (2019) 'Gene expression analysis of MMPs in women with preeclampsia using cell-free fetal

RNA in maternal plasma', Pregnancy Hypertension, 17, pp. 261–268. Available at: https://doi.org/10.1016/j.preghy.2019.06.008.

Evans, C., Hardin, J. and Stoebel, D.M. (2018) ‘Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions’, Briefings in Bioinformatics, 19(5), pp. 776–792. Available at: https://doi.org/10.1093/bib/bbx008.

Everett, T.R. and Chitty, L.S. (2015) 'Cell-free fetal DNA: The new tool in fetal medicine', Ultrasound in Obstetrics & Gynecology, 45(5), pp. 499.

Ewels, P., Magnusson, M., Lundin, S. and Käller, M. (2016) 'MultiQC: Summarize analysis results for multiple tools and samples in a single report', Bioinformatics, 32(19), pp. 3047–3048.

Fabbro, C.D., Scalabrin, S., Morgante, M. and Giorgi, F.M. (2013) 'An extensive evaluation of read trimming effects on illumina NGS data analysis', PLOS ONE, 8(12), pp. e85024. Available at: https://doi.org/10.1371/journal.pone.0085024.

Fakhr, Y., Koshti, S., Habibyan, Y.B., Webster, K. and Hemmings, D.G. (2022) 'Tumor necrosis factor-α induces a preeclamptic-like phenotype in placental villi via sphingosine kinase 1 activation', International Journal of Molecular Sciences, 23(7), pp. 3750. Available at: https://doi.org/10.3390/ijms23073750.

Fan, H.C., Blumenfeld, Y.J., Chitkara, U., Hudgins, L. and Quake, S.R. (2010) 'Analysis of the size distributions of fetal and maternal cell-free DNA by paired-end sequencing', Clinical chemistry, 56(8), pp. 1279–1286.

Fan, M., Xu, Y., Hong, F., Gao, X., Xin, G., Hong, H., Dong, L. and Zhao, X. (2016) ‘Rac1/β-Catenin Signalling Pathway Contributes to Trophoblast Cell Invasion by Targeting Snail and MMP9’, Cellular Physiology and Biochemistry: International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology, 38(4), pp. 1319–1332. Available at: https://doi.org/10.1159/000443076.

Farladansky-Gershnabel, S., Heusler, I., Biron-Shental, T., Shechter-Maor, G., Amiel, A., Kidron, D., Weisz, A., Einbinder, Y., Cohen-Hagai, K., Benchetrit, S. and Zitman-Gal, T. (2021) ‘Elevated expression of galectin-3, thioredoxin and thioredoxin interacting protein in preeclampsia’, Pregnancy Hypertension, 26, pp. 95–101. Available at: https://doi.org/10.1016/j.preghy.2021.10.003.

Feig, D.S., Shah, B.R., Lipscombe, L.L., Wu, C.F., Ray, J.G., Lowe, J., Hwee, J. and Booth, G.L. (2013) 'Preeclampsia as a risk factor for diabetes: A population-based cohort study', PLoS medicine, 10(4), pp. e1001425.

Feng, W. and Luo, Y. (2024) ‘Preeclampsia and its prediction: traditional versus contemporary predictive methods’, The Journal of Maternal-Fetal & Neonatal Medicine, 37(1), p. 2388171. Available at: https://doi.org/10.1080/14767058.2024.2388171.

Feng, W., Qiao, J., Tan, Y., Liu, Q., Wang, Q., Yang, B., Yang, S. and Cui, L. (2024) ‘Interaction of antiphospholipid antibodies with endothelial cells in antiphospholipid syndrome’, Frontiers in Immunology, 15. Available at: https://doi.org/10.3389/fimmu.2024.1361519.

Feng, Y. and Li, L.M. (2021) 'MUREN: A robust and multi-reference approach of RNA-seq transcript normalization', BMC bioinformatics, 22(1), pp. 1–18.

Finotello, F. and Di Camillo, B. (2015) 'Measuring differential gene expression with RNA-seq: Challenges and strategies for data analysis', Briefings in Functional Genomics, 14(2), pp. 130–142. Available at: https://doi.org/10.1093/bfgp/elu035.

Flint, E.J., Cerdeira, A.S., Redman, C.W. and Vatish, M. (2019) 'The role of angiogenic factors in the management of preeclampsia', Acta Obstetricia et Gynecologica Scandinavica, 98(6), pp. 700–707.

Fonseca, N.A., Marioni, J. and Brazma, A. (2014) 'RNA-seq gene profiling-a systematic empirical comparison', Plos one, 9(9), pp. e107026.

Founds, S.A., Tsigas, E., Ren, D. and Barmada, M.M. (2018) ‘Associating Symptom Phenotype and Genotype in Preeclampsia’, Biological Research for Nursing, 20(2), pp. 126–136. Available at: https://doi.org/10.1177/1099800417754140.

Freedman, B.I., Kopp, J.B., Winkler, C.A., Nelson, G.W., Rao, D.C., Eckfeldt, J.H., Leppert, M.F., Hicks, P.J., Divers, J., Langefeld, C.D. and Hunt, S.C. (2009) ‘Polymorphisms in the Nonmuscle Myosin Heavy Chain 9 Gene (MYH9) Are Associated with Albuminuria in Hypertensive African Americans: The HyperGEN Study’, American Journal of Nephrology, 29(6), pp. 626–632. Available at: https://doi.org/10.1159/000194791.

Fu, B., Zhou, Y., Ni, X., Tong, X., Xu, X., Dong, Z., Sun, R., Tian, Z. and Wei, H. (2017) 'Natural killer cells promote fetal development through the secretion of growth-promoting factors', Immunity, 47(6), pp. 1100–1113. e6.

Gagnon-Bartsch, J.A. and Speed, T.P. (2012) 'Using control genes to correct for unwanted variation in microarray data', Biostatistics, 13(3), pp. 539–552. Available at: https://doi.org/10.1093/biostatistics/kxr034.

Gambardella, G. (2025) ‘Joint processing of long- and short-read sequencing data with deep learning improves variant calling’, Cell Reports Methods, 5(7), p. 101107. Available at: https://doi.org/10.1016/j.crmeth.2025.101107.

Gammill, H.S., Guthrie, K.A., Aydelotte, T.M., Waldorf, K.M.A. and Nelson, J.L. (2010) 'Effect of parity on fetal and maternal microchimerism: Interaction of grafts within a host?', Blood, The Journal of the American Society of Hematology, 116(15), pp. 2706–2712.

Gao, D., Kim, J., Kim, H., Phang, T.L., Selby, H., Tan, A.C. and Tong, T. (2010) 'A survey of statistical software for analysing RNA-seq data', Human genomics, 5(1), pp. 1–5.

Gao, Y., Wu, Z., Liu, S., Chen, Y., Zhao, G. and Lin, H. (2023) 'Identification of key genes in the pathogenesis of preeclampsia via bioinformatic analysis and experimental verification', Frontiers in Endocrinology, 14, pp. 1190012. Available at: https://doi.org/10.3389/fendo.2023.1190012.

Gao, Y., Wu, Z., Liu, S., Chen, Y., Zhao, G. and Lin, H.-P. (2023) 'Identification of key genes in the pathogenesis of preeclampsia via bioinformatic analysis and experimental verification', Frontiers in Endocrinology, 14, p. 1190012. Available at: https://doi.org/10.3389/fendo.2023.1190012.

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T.F. and Conesa, A. (2012) 'Qualimap: Evaluating next-generation sequencing alignment data', Bioinformatics, 28(20), pp. 2678–2679.

Gardner, J.J., de Nazaré Oliveira da Silva, R., Bradshaw, J.L., Mabry, S., Wilson, E.N., Hula, N., Tucker, S.M., Gorham, I.K., Escalera, D., Lopez, L., Phillips, N.R., Cunningham, R.L. and Goulopoulou, S. (2025) 'Gestational Chronic Intermittent Hypoxia Triggers Maternal Inflammation and Disrupts Placental Stress Responses', American journal of physiology. Cell physiology, 329(2), pp. C630–C645. Available at: https://doi.org/10.1152/ajpcell.00446.2025.

Gathiram, P. and Moodley, J. (2016) 'Pre-eclampsia: Its pathogenesis and pathophysiolgy', Cardiovascular Journal of Africa, 27(2), pp. 71–78. Available at: https://doi.org/10.5830/CVJA-2016-009.

Gekas, J., Boomer, T.H., Rodrigue, M., Jinnett, K.N. and Bhatt, S. (2023) 'Use of cell-free signals as biomarkers for early and easy prediction of preeclampsia', Frontiers in Medicine, 10, pp. 1191163.

Geldenhuys, J., Rossouw, T.M., Lombaard, H.A., Ehlers, M.M. and Kock, M.M. (2018) 'Disruption in the Regulation of Immune Responses in the Placental Subtype of Preeclampsia', Frontiers in Immunology, 9, p. 1659. Available at: https://doi.org/10.3389/fimmu.2018.01659.

Geldenhuys, J., Rossouw, T.M., Lombaard, H.A., Ehlers, M.M. and Kock, M.M. (2018) 'Disruption in the regulation of immune responses in the placental subtype of preeclampsia', Frontiers in Immunology, 9, pp. 1659.

George, E.M. and Granger, J.P. (2011) 'Endothelin: Key mediator of hypertension in preeclampsia', American Journal of Hypertension, 24(9), pp. 964–969. Available at: https://doi.org/10.1038/ajh.2011.99.

Gerasimova, E.M., Fedotov, S.A., Kachkin, D.V., Vashukova, E.S., Glotov, A.S., Chernoff, Y.O. and Rubel, A.A. (2019) 'Protein Misfolding during Pregnancy: New

Approaches to Preeclampsia Diagnostics', International Journal of Molecular Sciences, 20(24), p. 6183. Available at: https://doi.org/10.3390/ijms20246183.

Giannakopoulos, B. and Krilis, S.A. (2013) 'The pathogenesis of the antiphospholipid syndrome', New England Journal of Medicine, 368(11), pp. 1033–1044.

Giannubilo, S.R., Landi, B. and Ciavattini, A. (2014) 'Preeclampsia: What could happen in a subsequent pregnancy?', Obstetrical & gynecological survey, 69(12), pp. 747–762.

Giesen, E.M., Verlohren, S. and Blois, S.M. (2025) 'Multi-omic insights of preeclampsia and cardiovascular health outcomes', Communications Medicine, 5, p. 446. Available at: https://doi.org/10.1038/s43856-025-01165-2.

Gill, S.S., Xu, M., Ottaviani, C., Patros, P., Bahsoon, R., Shaghaghi, A., Golec, M., Stankovski, V., Wu, H. and Abraham, A. (2022) 'AI for next generation computing: Emerging trends and future directions', Internet of Things, 19, pp. 100514.

Gliddon, H.D., Herberg, J.A., Levin, M. and Kaforou, M. (2018) 'Genome-wide host RNA signatures of infectious diseases: Discovery and clinical translation', Immunology, 153(2), pp. 171–178.

Go, A.T., van Vugt, J.M. and Oudejans, C.B. (2011) 'Non-invasive aneuploidy detection using free fetal DNA and RNA in maternal plasma: Recent progress and future possibilities', Human reproduction update, 17(3), pp. 372–382.

Go, A.T., Visser, A., Mulders, M.A., Blankenstein, M.A., Van Vugt, J.M. and Oudejans, C.B. (2004) 'Detection of placental transcription factor mRNA in maternal plasma', Clinical chemistry, 50(8), pp. 1413–1414.

GOETZINGER, K.R., SINGLA, A., GERKOWICZ, S., DICKE, J.M., GRAY, D.L. and ODIBO, A.O. (2010) 'Predicting the risk of pre-eclampsia between 11 and 13 weeks gestation by combining maternal characteristics and serum analytes, PAPP-A and free β-hCG', Prenatal diagnosis, 30(12–13), pp. 1138–1142. Available at: https://doi.org/10.1002/pd.2627.

Golby, P., Hatch, K.A., Bacon, J., Cooney, R., Riley, P., Allnutt, J., Hinds, J., Nunez, J., Marsh, P.D. and Hewinson, R.G. (2007) 'Comparative transcriptomics reveals key gene expression differences between the human and bovine pathogens of the mycobacterium tuberculosis complex', Microbiology, 153(10), pp. 3323–3336.

Gong, S., Gaccioli, F., Dopierala, J., Sovio, U., Cook, E., Volders, P., Martens, L., Kirk, P.D., Richardson, S. and Smith, G.C. (2021) 'The RNA landscape of the human placenta in health and disease', Nature Communications, 12(1), pp. 2639.

Gong, X., Li, J., Jiang, Y., Yuan, P., Chen, L., Yang, Y., Li, Y., Sun, M., Zhao, Y., Shi, H. and Wei, Y. (2022) 'Risk of preeclampsia by gestational weight gain in

women with varied prepregnancy BMI: A retrospective cohort study', Frontiers in Endocrinology, 13. Available at: https://doi.org/10.3389/fendo.2022.967102.

González-Vanegas, O. and Martinez-Perez, O. (2023) 'SARS-CoV-2 infection and Preeclampsia—How an infection can help us to know more about an obstetric condition', Viruses, 15(7), pp. 1564. Available at: https://doi.org/10.3390/v15071564.

Goswami, D., Tannetta, D.S., Magee, L.A., Fuchisawa, A., Redman, C., Sargent, I.L. and Von Dadelszen, P. (2006) 'Excess syncytiotrophoblast microparticle shedding is a feature of early-onset pre-eclampsia, but not normotensive intrauterine growth restriction', Placenta, 27(1), pp. 56–61.

Goulopoulou, S. and Davidge, S.T. (2015) 'Molecular mechanisms of maternal vascular dysfunction in preeclampsia', Trends in molecular medicine, 21(2), pp. 88–97.

Govender, N., Moodley, J. and Naicker, T. (2018) 'The Use of Soluble FMS-like Tyrosine Kinase 1/Placental Growth Factor Ratio in the Clinical Management of Pre-eclampsia', African Journal of Reproductive Health, 22(4). Available at: https://www.ajrh.info/index.php/ajrh/article/view/1622 (Accessed: 13 November 2025).

Grant, G.R., Farkas, M.H., Pizarro, A.D., Lahens, N.F., Schug, J., Brunk, B.P., Stoeckert, C.J., Hogenesch, J.B. and Pierce, E.A. (2011) 'Comparative analysis of RNA-seq alignment algorithms and the RNA-seq unified mapper (RUM)', Bioinformatics, 27(18), pp. 2518–2528. Available at: https://doi.org/10.1093/bioinformatics/btr427.

Gray, K.J., Hemberg, M. and Karumanchi, S.A. (2022) 'Cell-Free RNA Transcriptome and Prediction of Adverse Pregnancy Outcomes', Clinical chemistry, 68(11), pp. 1358–1360. Available at: https://doi.org/10.1093/clinchem/hvac109.

Graybill, P.M. and Davalos, R.V. (2020) 'Cytoskeletal Disruption after Electroporation and Its Significance to Pulsed Electric Field Therapies', Cancers, 12(5), p. 1132. Available at: https://doi.org/10.3390/cancers12051132.

Grotegut, C.A. (2016) 'Prevention of preeclampsia', The Journal of clinical investigation, 126(12), pp. 4396–4398. Available at: https://doi.org/10.1172/JCI91300.

Grover, P., Goel, P.N. and Greene, M.I. (2021) 'Regulatory T cells: Regulation of identity and function', Frontiers in Immunology, 12. Available at: https://doi.org/10.3389/fimmu.2021.750542.

Gu, J., Dai, J., Lu, H. and Zhao, H. (2023) 'Comprehensive analysis of ubiquitously expressed genes in humans from a data-driven perspective', Genomics, Proteomics & Bioinformatics, 21(1), pp. 164–176.

Gu, Q., Tian, J., Li, X. and Jiang, S. (2022) 'A novel random forest integrated model for imbalanced data classification problem', Knowledge-Based Systems, 250, pp. 109050.

Gumusoglu, S., Blaine, B., Bertolli, A., Weber, M.A., Kamal, M., Hazzard, H., Schickling, B., Lauffer, M., Zhang, Y., Taylor, R., Kirkpatrick, K., Santillan, D., Aldridge, G. and Santillan, M. (2025) 'Integrated multi-omics analysis reveals immunovascular mechanisms of the placenta-maternal brain axis and lifespan neurobehavior changes in a mouse model of preeclampsia', Neuroscience, 579, pp. 69–84. Available at: https://doi.org/10.1016/j.neuroscience.2025.05.301.

Gunderson, E.P., Greenberg, M., Najem, M., Sun, B., Alexeeff, S.E., Alexander, J., Nguyen-Huynh, M.N. and Roberts, J.M. (2025) 'Severe Maternal Morbidity Associated With Chronic Hypertension, Preeclampsia, and Gestational Hypertension', JAMA Network Open, 8(1), p. e2451406. Available at: https://doi.org/10.1001/jamanetworkopen.2024.51406.

Guo, J. (2014) 'Transcription: The epicenter of gene expression', Journal of Zhejiang University SCIENCE B, 15, pp. 409–411.

Guo, Y., Ye, F., Sheng, Q., Clark, T. and Samuels, D.C. (2014) 'Three-stage quality control strategies for DNA re-sequencing data', Briefings in Bioinformatics, 15(6), pp. 879–889. Available at: https://doi.org/10.1093/bib/bbt069.

Guo, Y., Zhao, S., Ye, F., Sheng, Q. and Shyr, Y. (2014) 'MultiRankSeq: Multiperspective approach for RNAseq differential expression analysis and quality control', BioMed research international, 2014.

Guo, Z., Yang, F., Zhang, J., Zhang, Z., Li, K., Tian, Q., Hou, H., Xu, C., Lu, Q. and Ren, Z. (2020) 'Whole-genome promoter profiling of plasma DNA exhibits diagnostic value for placenta-origin pregnancy complications', Advanced Science, 7(7), pp. 1901819.

Gupta, J.K., Care, A., Goodfellow, L., Alfirevic, Z., Müller-Myhsok, B. and Alfirevic, A. (2022) 'Genome and transcriptome profiling of spontaneous preterm birth phenotypes', Scientific Reports, 12, pp. 1003. Available at: https://doi.org/10.1038/s41598-022-04881-0.

Gurnadi, J.I., Mose, J., Handono, B., Satari, M.H., Anwar, A.D., Fauziah, P.N., Yogi Pramatirta, A. and Rihibiha, D.D. (2015) 'Difference of concentration of placental soluble fms-like tyrosine kinase-1 (sFlt-1), placental growth factor (PlGF), and sFlt-1/PlGF ratio in severe preeclampsia and normal pregnancy', BMC research notes, 8, pp. 1–5.

Hackelöer, M., Schmidt, L. and Verlohren, S. (2022) 'New advances in prediction and surveillance of preeclampsia: Role of machine learning approaches and remote monitoring', Archives of Gynecology and Obstetrics, , pp. 1–15.

Hahn, S., Rusterholz, C., Hoesli, I. and Lapaire, O. (2011) 'Cell-free nucleic acids as potential markers for preeclampsia', Placenta, 32 Suppl, pp. 17. Available at: https://doi.org/10.1016/j.placenta.2010.06.018.

Haji, A. (2024) Comparative analysis of autoencoder and PCA for dimensionality reduction in gene expression data. Available at: https://urn.kb.se/resolve?urn=urn:nbn:se:his:diva-24281 (Accessed: 17 November 2025).

Hajian-Tilaki, K. (2014) 'Sample size estimation in diagnostic test studies of biomedical informatics', Journal of Biomedical Informatics, 48, pp. 193–204.

Hall, M., de Marvao, A., Schweitzer, R., Cromb, D., Colford, K., Jandu, P., O'Regan, D.P., Ho, A., Price, A., Chappell, L.C., Rutherford, M.A., Story, L., Lamata, P. and Hutter, J. (2024) 'Preeclampsia Associated Differences in the Placenta, Fetal Brain, and Maternal Heart Can Be Demonstrated Antenatally: An Observational Cohort Study Using MRI', Hypertension, 81(4), pp. 836–847. Available at: https://doi.org/10.1161/HYPERTENSIONAHA.123.22442.

Hamada, M.M. and Saadeldin, I.M. (2022) 'Physiology of the reproductive system', Advances in Assisted Reproduction Technologies, 5, pp. 1.

Hammoud, G.M. and Ibdah, J.A. (2014) 'Preeclampsia-induced liver dysfunction, HELLP syndrome, and acute fatty liver of pregnancy', Clinical Liver Disease, 4(3), pp. 69–73. Available at: https://doi.org/10.1002/cld.409.

Han, L., da Silva Costa, F., Perkins, A. and Holland, O. (2025) 'Molecular signatures of preeclampsia subtypes determined through integrated weighted gene co-expression network analysis and differential gene expression analysis of placental transcriptomics', Frontiers in Cell and Developmental Biology, 13. Available at: https://doi.org/10.3389/fcell.2025.1635878.

Han, L., Holland, O.J., Da Silva Costa, F. and Perkins, A.V. (2023) 'Potential biomarkers for late-onset and term preeclampsia: A scoping review', Frontiers in Physiology, 14, pp. 338.

Han, Q., Zheng, W., Guo, X., Zhang, D., Liu, H., Yu, L. and Yan, J. (2020) 'A new predicting model of preeclampsia based on peripheral blood test values.', European Review for Medical & Pharmacological Sciences, 24(13).

Han, Y., Gao, S., Muegge, K., Zhang, W. and Zhou, B. (2015a) 'Advanced applications of RNA sequencing and challenges', Bioinformatics and biology insights, 9, pp. BBI. S28991.

Han, Y., Gao, S., Muegge, K., Zhang, W. and Zhou, B. (2015b) 'Advanced applications of RNA sequencing and challenges', Bioinformatics and Biology Insights, 9s1, pp. BBI.S28991. Available at: https://doi.org/10.4137/BBI.S28991.

Han, Y., Huang, L. and Zhou, F. (2021) 'A dynamic recursive feature elimination framework (dRFE) to further refine a set of OMIC biomarkers', Bioinformatics,

37(15), pp. 2183–2189. Available at: https://doi.org/10.1093/bioinformatics/btab055.

Hansen, K.D., Brenner, S.E. and Dudoit, S. (2010) 'Biases in illumina transcriptome sequencing caused by random hexamer priming', Nucleic acids research, 38(12), pp. e131.

Hao, S., You, J., Chen, L., Zhao, H., Huang, Y., Zheng, L., Tian, L., Maric, I., Liu, X., Li, T., Bianco, Y.K., Winn, V.D., Aghaeepour, N., Gaudilliere, B., Angst, M.S., Zhou, X., Li, Y.-M., Mo, L., Wong, R.J., Shaw, G.M., Stevenson, D.K., Cohen, H.J., Mcelhinney, D.B., Sylvester, K.G. and Ling, X.B. (2020) 'Changes in pregnancy-related serum biomarkers early in gestation are associated with later development of preeclampsia', PLOS ONE, 15(3), p. e0230000. Available at: https://doi.org/10.1371/journal.pone.0230000.

Haram, K., Mortensen, J.H., Myking, O., Roald, B., Magann, E.F. and Morrison, J.C. (2020) 'Early development of the human placenta and pregnancy complications', The Journal of Maternal-Fetal & Neonatal Medicine, 33(20), pp. 3538–3545.

Harmon, A.C., Cornelius, D.C., Amaral, L.M., Faulkner, J.L., Cunningham, M.W., Wallace, K. and LaMarca, B. (2016) 'The role of inflammation in the pathology of preeclampsia', Clinical science (London, England : 1979), 130(6), pp. 409–419. Available at: https://doi.org/10.1042/CS20150702.

Harris, K., Xu, L., Woodward, M., De Kat, A., Zhou, X., Shang, J., Hirst, J.E. and Henry, A. (2024) 'Early pregnancy maternal blood pressure and risk of preeclampsia: Does the association differ by parity? Evidence from 14,086 women across 7 countries', Pregnancy Hypertension, 37, p. 101136. Available at: https://doi.org/10.1016/j.preghy.2024.101136.

Hartmann, S., Botha, S.M., Gray, C.M., Valdes, D.S., Tong, S., Kaitu'u-Lino, T.J., Herse, F., Bergman, L., Cluver, C.A., Dechend, R. and Nonn, O. (2023) 'Can single-cell and spatial omics unravel the pathophysiology of pre-eclampsia?', Journal of Reproductive Immunology, 159, p. 104136. Available at: https://doi.org/10.1016/j.jri.2023.104136.

Hassan, H.E., Azzam, H., Othman, M., Hassan, M. and Selim, T. (2019) 'Soluble E-selectin, platelet count and mean platelet volume as biomarkers for pre-eclampsia', Pregnancy Hypertension, 17, pp. 1–4. Available at: https://doi.org/10.1016/j.preghy.2019.04.008.

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M. and Hussain, A. (2024) 'Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence', Cognitive Computation, 16(1), pp. 45–74. Available at: https://doi.org/10.1007/s12559-023-10179-8.

Hatem, A., Bozdağ, D., Toland, A.E. and Çatalyürek, ÜV. (2013) 'Benchmarking short sequence mapping tools', BMC bioinformatics, 14(1), pp. 1–25.

Haudiquet, M., D'Incau, M., Letouzey, V. and Moranne, O. (2025) 'A systematic review on the determinants of long-term kidney sequelae after hypertensive diseases of pregnancy', Acta Obstetricia et Gynecologica Scandinavica, 104(7), pp. 1254–1273. Available at: https://doi.org/10.1111/aogs.15127.

Havers-Borgersen, E., Fosbøl, E., Johansen, M., Køber, L., Morris, J.M. and Seeho, S.K.M. (2023) 'Pre-eclampsia in a first pregnancy and subsequent pregnancy outcomes: a nationwide cohort study', J Epidemiol Community Health, 77(11), pp. 694–703. Available at: https://doi.org/10.1136/jech-2023-220829.

He, B., Zhu, R., Yang, H., Lu, Q., Wang, W., Song, L., Sun, X., Zhang, G., Li, S., Yang, J., Tian, G., Bing, P. and Lang, J. (2020) 'Assessing the impact of data preprocessing on analyzing next generation sequencing data', Frontiers in Bioengineering and Biotechnology, 8, pp. 817. Available at: https://doi.org/10.3389/fbioe.2020.00817.

Hedley, P.L., Hagen, C.M., Wilstrup, C. and Christiansen, M. (2023) 'The use of artificial intelligence and machine learning methods in early pregnancy pre-eclampsia screening: A systematic review protocol', PloS One, 18(4), p. e0272465. Available at: https://doi.org/10.1371/journal.pone.0272465.

Hellen, C.U. (2018) 'Translation termination and ribosome recycling in eukaryotes', Cold Spring Harbor Perspectives in Biology, 10(10), pp. a032656.

Hernández-Díaz, S., Toh, S. and Cnattingius, S. (2009) 'Risk of pre-eclampsia in first and subsequent pregnancies: prospective cohort study', BMJ, 338, p. b2255. Available at: https://doi.org/10.1136/bmj.b2255.

Hernández-Díaz, S., Toh, S. and Cnattingius, S. (2009) 'Risk of pre-eclampsia in first and subsequent pregnancies: Prospective cohort study', Bmj, 338.

Herraiz, I., Simón, E., Gómez-Arriaga, P.I., Martínez-Moratalla, J.M., García-Burguillo, A., Jiménez, E.A.L. and Galindo, A. (2015) 'Angiogenesis-related biomarkers (sFlt-1/PLGF) in the prediction and diagnosis of placental dysfunction: An approach for clinical integration', International Journal of Molecular Sciences, 16(8), pp. 19009–19026. Available at: https://doi.org/10.3390/ijms160819009.

Herrock, O., Deer, E. and LaMarca, B. (2023) 'Setting a stage: Inflammation during preeclampsia and postpartum', Frontiers in Physiology, 14. Available at: https://doi.org/10.3389/fphys.2023.1130116.

Hershey, J.W., Sonenberg, N. and Mathews, M.B. (2012) 'Principles of translational control: An overview', Cold Spring Harbor perspectives in biology, 4(12), pp. a011528.

Heskes, T., Eisinga, R. and Breitling, R. (2014) 'A fast algorithm for determining bounds and accurate approximate p-values of the rank product statistic for replicate experiments', BMC bioinformatics, 15, pp. 1–11.

Hestand, M.S., Bessem, M., van Rijn, P., de Menezes, R.X., Sie, D., Bakker, I., Boon, E.M., Sistermans, E.A. and Weiss, M.M. (2019) 'Fetal fraction evaluation in non-invasive prenatal screening (NIPS)', European Journal of Human Genetics, 27(2), pp. 198–202.

Hezareh, F., Moghaddam-Banaem, L. and Shahali, S. (2021) 'The relationship between maternal serum interleukin-6 and CRP levels at first trimester of pregnancy and gestational diabetes occurrence', The Iranian Journal of Obstetrics, Gynecology and Infertility, 24(5), pp. 59–68. Available at: https://doi.org/10.22038/ijogi.2021.18572.

Hicks, S.A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M.A., Halvorsen, P. and Parasa, S. (2022) 'On evaluation metrics for medical applications of artificial intelligence', Scientific Reports, 12, pp. 5979. Available at: https://doi.org/10.1038/s41598-022-09954-8.

Hinnebusch, A.G. and Lorsch, J.R. (2012) 'The mechanism of eukaryotic translation initiation: New insights and challenges', Cold Spring Harbor perspectives in biology, 4(10), pp. a011544.

Hogg, K., Blair, J.D., McFadden, D.E., von Dadelszen, P. and Robinson, W.P. (2013) 'Early onset pre-eclampsia is associated with altered DNA methylation of cortisol-signalling and steroidogenic genes in the placenta', PloS one, 8(5), pp. e62969.

Hombach, S. and Kretz, M. (2016) 'Non-coding RNAs: Classification, biology and functioning', Non-coding RNAs in colorectal cancer, , pp. 3–17.

Hong, F., Breitling, R., McEntee, C.W., Wittner, B.S., Nemhauser, J.L. and Chory, J. (2006) ‘RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis’, Bioinformatics, 22(22), pp. 2825–2827. Available at: https://doi.org/10.1093/bioinformatics/btl476.

Hong, J. and Kumar, S. (2023) ‘Circulating biomarkers associated with placental dysfunction and their utility for predicting fetal growth restriction’, Clinical Science, 137(8), pp. 579–595. Available at: https://doi.org/10.1042/CS20220300.

Hong, K., Kim, S.H., Cha, D.H. and Park, H.J. (2021) ‘Defective Uteroplacental Vascular Remodeling in Preeclampsia: Key Molecular Factors Leading to Long Term Cardiovascular Disease’, International Journal of Molecular Sciences, 22(20), p. 11202. Available at: https://doi.org/10.3390/ijms222011202.

Hong, K., Park, H.J. and H Cha, D. (2021) 'Clinical implications of placenta-derived angiogenic/anti-angiogenic biomarkers in pre-eclampsia', Biomarkers in Medicine, 15(7), pp. 523–536.

Hong, M., Tao, S., Zhang, L., Diao, L., Huang, X., Huang, S., Xie, S., Xiao, Z. and Zhang, H. (2020) 'RNA sequencing: New technologies and applications in cancer research', Journal of hematology & oncology, 13(1), pp. 1–16.

Hoque, M.N., Sarkar, M.M.H., Khan, M.A., Hossain, M.A., Hasan, M.I., Rahman, M.H., Habib, M.A., Akter, S., Banu, T.A., Goswami, B., Jahan, I., Nafisa, T., Molla, M.M.A., Soliman, M.E., Araf, Y., Khan, M.S., Zheng, C. and Islam, T. (2022) 'Differential gene expression profiling reveals potential biomarkers and pharmacological compounds against SARS-CoV-2: Insights from machine learning and bioinformatics approaches', Frontiers in Immunology, 13, pp. 918692. Available at: https://doi.org/10.3389/fimmu.2022.918692.

Horii, M., Boyd, T.K. and Parast, M.M. (2018) 'Placental development and complications of previable pregnancy'Diagnostic gynecologic and obstetric pathology Elsevier, pp. 1070–1102. .

Hosseini, P., Tremblay, A., Matthews, B.F. and Alkharouf, N.W. (2010) 'An efficient annotation and gene-expression derivation tool for illumina solexa datasets', BMC research notes, 3(1), pp. 183. Available at: https://doi.org/10.1186/1756-0500-3-183.

Hrdlickova, R., Toloue, M. and Tian, B. (2017) 'RNA-seq methods for transcriptome analysis', Wiley interdisciplinary reviews. RNA, 8(1). Available at: https://doi.org/10.1002/wrna.1364.

Hu, X. and Zhang, L. (2021) ‘Uteroplacental Circulation in Normal Pregnancy and Preeclampsia: Functional Adaptation and Maladaptation’, International Journal of Molecular Sciences, 22(16), p. 8622. Available at: https://doi.org/10.3390/ijms22168622.

Huang, J., Yeung, A.M., Armstrong, D.G., Battarbee, A.N., Cuadros, J., Espinoza, J.C., Kleinberg, S., Mathioudakis, N., Swerdlow, M.A. and Klonoff, D.C. (2022) ‘Artificial Intelligence for Predicting and Diagnosing Complications of Diabetes’, Journal of Diabetes Science and Technology, 17(1), pp. 224–238. Available at: https://doi.org/10.1177/19322968221124583.

Huang, P., Song, Y., Yang, Y., Bai, F., Li, N., Liu, D., Li, C., Li, X., Gou, W. and Zong, L. (2024) ‘Identification and verification of diagnostic biomarkers based on mitochondria-related genes related to immune microenvironment for preeclampsia using machine learning algorithms’, Frontiers in Immunology, 14. Available at: https://doi.org/10.3389/fimmu.2023.1304165.

Huang, Q., Hao, S., You, J., Yao, X., Li, Z., Schilling, J., Thyparambil, S., Liao, W.-L., Zhou, X., Mo, L., Ladella, S., Davies-Balch, S.R., Zhao, H., Fan, D., Whitin, J.C., Cohen, H.J., McElhinney, D.B., Wong, R.J., Shaw, G.M., Stevenson, D.K., Sylvester, K.G. and Ling, X.B. (2021) ‘Early-pregnancy prediction of risk for pre-eclampsia using maternal blood leptin/ceramide ratio: discovery and confirmation’, BMJ Open, 11(11), p. e050963. Available at: https://doi.org/10.1136/bmjopen-2021-050963.

Huang, S., Cai, S., Li, H., Zhang, W., Xiao, H., Yu, D., Zhong, X., Tao, P. and Luo, Y. (2022) ‘Prediction of Differentially Expressed Genes and a Diagnostic Signature of Preeclampsia via Integrated Bioinformatics Analysis’, Disease Markers, 2022, p. 5782637. Available at: https://doi.org/10.1155/2022/5782637.

Huang, S., Cai, S., Li, H., Zhang, W., Xiao, H., Yu, D., Zhong, X., Tao, P. and Luo, Y. (2022) 'Prediction of differentially expressed genes and a diagnostic signature of preeclampsia via integrated bioinformatics analysis', Disease Markers, 2022, pp. 5782637. Available at: https://doi.org/10.1155/2022/5782637.

Hubel, C.A. (1999) 'Oxidative stress in the pathogenesis of preeclampsia', Proceedings of the Society for Experimental Biology and Medicine, 222(3), pp. 222–235. Available at: https://doi.org/10.1046/j.1525-1373.1999.d01-139.x.

Huggett, J., Dheda, K., Bustin, S. and Zumla, A. (2005) 'Real-time RT-PCR normalisation; strategies and considerations', Genes & Immunity, 6(4), pp. 279–284.

Hui, D. and Hladunewich, M.A. (2019) 'Chronic kidney disease and pregnancy', Obstetrics & Gynecology, 133(6), pp. 1182–1194.

Hui, L. and Bianchi, D.W. (2011) 'Cell-free fetal nucleic acids in amniotic fluid', Human reproduction update, 17(3), pp. 362–371.

Huppertz, B. (2006) 'Placental villous trophoblast: The altered balance between proliferation and apoptosis triggers pre-eclampsia', J Reproduktionsmed Endokrinol, 3(2), pp. 103–108.

Huppertz, B. (2008) ‘Placental Origins of Preeclampsia’, Hypertension, 51(4), pp. 970–975. Available at: https://doi.org/10.1161/HYPERTENSIONAHA.107.107607.

Huppertz, B. (2010) 'IFPA award in placentology lecture: Biology of the placental syncytiotrophoblast–myths and facts', Placenta, 31, pp. S75–S81.

Huppertz, B. (2023) 'Placental development with histological aspects'The placenta: Basics and clinical significance Springer, pp. 1–27. .

Huppertz, B. and Herrler, A. (2005) 'Regulation of proliferation and apoptosis during development of the preimplantation embryo and the placenta', Birth Defects Research Part C: Embryo Today: Reviews, 75(4), pp. 249–261.

Hwang, S., Kim, E., Lee, I. and Marcotte, E.M. (2015) 'Systematic comparison of variant calling pipelines using gold standard personal exome variants', Scientific Reports, 5(1), pp. 17875. Available at: https://doi.org/10.1038/srep17875.

Ingman, K., Cookson, V., Jones, C. and Aplin, J.D. (2010) 'Characterisation of hofbauer cells in first and second trimester placenta: Incidence, phenotype, survival in vitro and motility', Placenta, 31(6), pp. 535–544.

initiative on preeclampsia (PE): A pragmatic guide for first trimester screening

Itadani, H., Mizuarai, S. and Kotani, H. (2008) 'Can systems biology understand pathway activation? gene expression signatures as surrogate markers for understanding the complexity of pathway activation', Current Genomics, 9(5), pp. 349–360.

Ives, C.W., Sinkey, R., Rajapreyar, I., Tita, A.T. and Oparil, S. (2020) 'Preeclampsia—pathophysiology and clinical presentations: JACC state-of-the-art review', Journal of the American College of Cardiology, 76(14), pp. 1690–1702.

Iyer, G., Wang, A.R., Brennan, S.R., Bourgeois, S., Armstrong, E., Shah, P. and Harari, P.M. (2017) 'Identification of stable housekeeping genes in response to ionizing radiation in cancer research', Scientific reports, 7(1), pp. 43763.

Jaeger, K.M., Nissen, M., Leutheuser, H., Danzberger, N., Titzmann, A., Pontones, C.A., Goossens, C., Ziegler, P., Uhrig, S., Haeberle, L., Bleher, H., Kast, K., Kornhuber, J., Schoeffski, O., Braun, M., Fasching, P.A., Beckmann, M.W., Eskofier, B.M. and Huebner, H. (2025) 'Adherence to digital pregnancy care – lessons learned from the SMART start feasibility study', NPJ Digital Medicine, 8, p. 561. Available at: https://doi.org/10.1038/s41746-025-01966-8.

James, J.L., Lissaman, A., Nursalim, Y.N. and Chamley, L.W. (2022) 'Modelling human placental villous development: Designing cultures that reflect anatomy', Cellular and Molecular Life Sciences, 79(7), pp. 384.

Jamjute, P., Ahmad, A., Ghosh, T. and Banfield, P. (2009) 'Liver function test and pregnancy', The Journal of Maternal-Fetal & Neonatal Medicine, 22(3), pp. 274–283.

Jang, J.H., Jung, Y.W., Shim, S.H., Sin, Y.J., Lee, K.J., Shim, S.S., Ahn, E.H. and Cha, D.H. (2017) 'Global gene expression changes of amniotic fluid cell free RNA according to fetal development', European Journal of Obstetrics & Gynecology and Reproductive Biology, 216, pp. 104–110.

Jansen, C.H.J.R., Kastelein, A.W., Kleinrouweler, C.E., Leeuwen, E.V., Jong, K.H.D., Pajkrt, E. and Noorden, C.J.F.V. (2020) 'Development of placental abnormalities in location and anatomy', Acta Obstetricia et Gynecologica Scandinavica, 99(8), pp. 983–993. Available at: https://doi.org/10.1111/aogs.13834.

Jasper, E.A., Hellwege, J.N., Breeyear, J.H., Xiao, B., Jarvik, G.P., Stanaway, I.B., Leppig, K.A., Chittoor, G., Hayes, M.G., Dikilitas, O., Kullo, I.J., Holm, I.A., Verma, S.S., Edwards, T.L. and Velez Edwards, D.R. (2024) 'Genetic predictors of blood pressure traits are associated with preeclampsia', Scientific Reports, 14(1), p. 17613. Available at: https://doi.org/10.1038/s41598-024-68469-6.

Jayamol, A., Dey, M. and Sahoo, I. (2019) 'A correlation of first trimester serum levels of PAPP-A and PlGF with preeclampsia in western indian population', Medicine, 8(3), pp. 597–600.

Jena, M.K., Sharma, N.R., Petitt, M., Maulik, D. and Nayak, N.R. (2020) ‘Pathogenesis of Preeclampsia and Therapeutic Approaches Targeting the Placenta’, Biomolecules, 10(6), p. 953. Available at: https://doi.org/10.3390/biom10060953.

Jeyabalan, A. (2013) 'Epidemiology of preeclampsia: Impact of obesity', Nutrition reviews, 71(suppl_1), pp. S18–S25.

Jhee, J.H., Lee, S., Park, Y., Lee, S.E., Kim, Y.A., Kang, S.-W., Kwon, J.-Y. and Park, J.T. (2019) ‘Prediction model development of late-onset preeclampsia using machine learning-based methods’, PLoS ONE, 14(8). Available at: https://doi.org/10.1371/journal.pone.0221202.

Jiang, P., Chan, C.W., Chan, K.A., Cheng, S.H., Wong, J., Wong, V.W., Wong, G.L., Chan, S.L., Mok, T.S. and Chan, H.L. (2015) 'Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients', Proceedings of the National Academy of Sciences, 112(11), pp. E1317–E1325.

Jiang, Y., Zhang, Y., Yang, Q., Zeng, D., Zhao, K., Ma, X. and Yin, W. (2022) 'The association between fetal fraction and pregnancy-related complications among chinese population', PLoS ONE, 17(7), pp. e0271219. Available at: https://doi.org/10.1371/journal.pone.0271219.

Jolliffe, I.T. and Cadima, J. (2016) ‘Principal component analysis: a review and recent developments’, Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 374(2065), p. 20150202. Available at: https://doi.org/10.1098/rsta.2015.0202.

Jong, V.L., Novianti, P.W., Roes, K.C.B. and Eijkemans, M.J.C. (2016) 'Selecting a classification function for class prediction with gene expression data', Bioinformatics, 32(12), pp. 1814–1822. Available at: https://doi.org/10.1093/bioinformatics/btw034.

Joshi, C.J., Ke, W., Drangowska-Way, A., O’Rourke, E.J. and Lewis, N.E. (2022) 'What are housekeeping genes?', PLoS computational biology, 18(7), pp. e1010295.

Junaid, M.A.L. (2025) ‘Oxidative Stress and Nutrient Imbalance in Pre-eclampsia: The Roles of Homocysteine and Vitamin B12’, Journal of Pioneering Medical Sciences, 14, pp. 108–114. Available at: https://doi.org/10.47310/jpms202514S0217.

Jung, E., Romero, R., Yeo, L., Gomez-Lopez, N., Chaemsaithong, P., Jaovisidha, A., Gotsch, F. and Erez, O. (2022) 'The etiology of preeclampsia', American Journal of Obstetrics and Gynecology, 226(2), pp. S844–S866.

Kaartokallio, T., Cervera, A., Kyllönen, A., Laivuori, K., Kere, J. and Laivuori, H. (2015) 'Gene expression profiling of pre-eclamptic placentae by RNA

sequencing', Scientific reports, 5(1), pp. 14107. Available at: https://doi.org/10.1038/srep14107.

Kaitu'u-Lino, T.J., MacDonald, T.M., Keenan, E., Hastie, R., Cluver, C.A., Susic, D., Henry, A., Myers, J.E., McCowan, L.M., Taylor, R.S., Bergman, L., Marques, F.Z., Kaye, D.M., Bartho, L.A., Hannan, N.J., Cannon, P., Nguyen, T.-V., Kandel, M., Murphy, C., Wong, G.P., Masci, J., Pritchard, N., Walker, S.P. and Tong, S. (2025) 'The association between circulating SIGLEC6 and preeclampsia: observational studies of seven cohorts', eBioMedicine, 118, p. 105870. Available at: https://doi.org/10.1016/j.ebiom.2025.105870.

Kalari, K.R., Nair, A.A., Bhavsar, J.D., O'Brien, D.R., Davila, J.I., Bockol, M.A., Nie, J., Tang, X., Baheti, S., Doughty, J.B., Middha, S., Sicotte, H., Thompson, A.E., Asmann, Y.W. and Kocher, J.A. (2014) 'MAP-RSeq: Mayo analysis pipeline for RNA sequencing', BMC bioinformatics, 15(1), pp. 224. Available at: https://doi.org/10.1186/1471-2105-15-224.

Kallela, J., Jääskeläinen, T., Kortelainen, E., Heinonen, S., Kajantie, E., Kere, J., Kivinen, K., Pouta, A. and Laivuori, H. (2016) 'The diagnosis of pre-eclampsia using two revised classifications in the finnish pre-eclampsia consortium (FINNPEC) cohort', BMC Pregnancy and Childbirth, 16. Available at: https://doi.org/10.1186/s12884-016-1010-0.

Karp, P.D., Midford, P.E., Caspi, R. and Khodursky, A. (2021) 'Pathway size matters: The influence of pathway granularity on over-representation (enrichment analysis) statistics', BMC genomics, 22(1), pp. 1–11.

Karrar, S.A., Martingano, D.J. and Hong, P.L. (2025) 'Preeclampsia', in StatPearls. Treasure Island (FL): StatPearls Publishing. Available at: http://www.ncbi.nlm.nih.gov/books/NBK570611/ (Accessed: 3 November 2025).

Kavitha, C., Mani, V., Srividhya, S.R., Khalaf, O.I. and Tavera Romero, C.A. (2022) 'Early-stage alzheimer's disease prediction using machine learning models', Frontiers in public health, 10, pp. 853294.

Kaye, J., Boddington, P., de Vries, J., Hawkins, N. and Melham, K. (2010) 'Ethical implications of the use of whole genome methods in medical research', European Journal of Human Genetics, 18(4), pp. 398–403. Available at: https://doi.org/10.1038/ejhg.2009.191.

Kemp, M.W., Newnham, J.P., Challis, J.G., Jobe, A.H. and Stock, S.J. (2016) 'The clinical use of corticosteroids in pregnancy', Human reproduction update, 22(2), pp. 240–259.

Kenny, L.C., Thomas, G., Poston, L., Myers, J.E., Simpson, N.A., McCarthy, F.P., Brown, L.W., Bond, A.E., Tuytten, R. and Baker, P.N. (2020) 'Prediction of preeclampsia risk in first time pregnant women: Metabolite biomarkers for a clinical test', PLoS One, 15(12), pp. e0244369.

Khalil, A., Syngelaki, A., Maiz, N., Zinevich, Y. and Nicolaides, K.H. (2013) 'Maternal age and adverse pregnancy outcome: A cohort study', Ultrasound in Obstetrics & Gynecology, 42(6), pp. 634–643.

Khan, N., Andrade, W., Castro, H.D., Wright, A., Wright, D. and Nicolaides, K.H. (2020) 'Impact of new definitions of pre-eclampsia on incidence and performance of first-trimester screening', Ultrasound in Obstetrics & Gynecology, 55(1), pp. 50–57. Available at: https://doi.org/10.1002/uog.21867.

Khong, T.Y., De Wolf, F., Robertson, W.B. and Brosens, I. (1986) 'Inadequate maternal vascular response to placentation in pregnancies complicated by pre-eclampsia and by small-for-gestational age infants', BJOG: An International Journal of Obstetrics & Gynaecology, 93(10), pp. 1049–1059.

Kim, H.J., Jeong, D.E., Choi, E.J., Kim, N.Y., Lee, M.J., Choi, B.Y. and Park, J.Y. (2025) 'The sFlt-1/PlGF ratio of high-risk pregnant women for preeclampsia compared in singleton, twin, and triplet pregnancies', Taiwanese Journal of Obstetrics and Gynecology, 64(6), pp. 996–1003. Available at: https://doi.org/10.1016/j.tjog.2025.07.023.

Kim, S. and Kim, J. (2017) 'A review of mechanisms of implantation', Development & Reproduction, 21(4), pp. 351–359. Available at: https://doi.org/10.12717/DR.2017.21.4.351.

Kim, S., Ryu, H., Yang, J., Kim, M., Han, J., Kim, J., Chung, J., Park, S., Lee, M. and Kim, D. (2007) 'Increased sFlt-1 to PlGF ratio in women who subsequently develop preeclampsia', Journal of Korean Medical Science, 22(5), pp. 873–877. Available at: https://doi.org/10.3346/jkms.2007.22.5.873.

Knöfler, M., Haider, S., Saleh, L., Pollheimer, J., Gamage, T.K.J.B. and James, J. (2019) 'Human placenta and trophoblast development: Key molecular mechanisms and model systems', Cellular and Molecular Life Sciences, 76(18), pp. 3479–3496. Available at: https://doi.org/10.1007/s00018-019-03104-6.

Kokkinidis, M., Glykos, N.M. and Fadouloglou, V.E. (2020) 'Catalytic activity regulation through post-translational modification: The expanding universe of protein diversity', Advances in Protein Chemistry and Structural Biology, 122, pp. 97–125.

Kongwattanakul, K., Saksiriwuttho, P., Chaiyarach, S. and Thepsuthammarat, K. (2018) 'Incidence, characteristics, maternal complications, and perinatal outcomes associated with preeclampsia with severe features and HELLP syndrome', International Journal of Women's Health, 10, pp. 371–377. Available at: https://doi.org/10.2147/IJWH.S168569.

Konidena, B.K., Malaiyappan, J.N.A. and Tadimarri, A. (2024) 'Ethical considerations in the development and deployment of AI systems', European Journal of Technology, 8(2), pp. 41–53.

Kornacki, J. and Wender-Ożegowska, E. (2020) ‘Utility of biochemical tests in prediction, diagnostics and clinical management of preeclampsia: a review’, Archives of Medical Science : AMS, 16(6), pp. 1370–1375. Available at: https://doi.org/10.5114/aoms.2020.97762.

Kornacki, J., Gutaj, P., Kalantarova, A., Sibiak, R., Jankowski, M. and Wender-Ozegowska, E. (2021) ‘Endothelial Dysfunction in Pregnancy Complications’, Biomedicines, 9(12), p. 1756. Available at: https://doi.org/10.3390/biomedicines9121756.

Kornacki, J., Olejniczak, O., Sibiak, R., Gutaj, P. and Wender-Ożegowska, E. (2024) ‘Pathophysiology of Pre-Eclampsia—Two Theories of the Development of the Disease’, International Journal of Molecular Sciences, 25(1), p. 307. Available at: https://doi.org/10.3390/ijms25010307.

Kosugi, S. and Terao, C. (2024) ‘Comparative evaluation of SNVs, indels, and structural variations detected with short- and long-read sequencing data’, Human Genome Variation, 11(1), p. 18. Available at: https://doi.org/10.1038/s41439-024-00276-x.

Kovacheva, V.P., Eberhard, B.W., Cohen, R.Y., Maher, M., Saxena, R. and Gray, K.J. (2024) ‘Preeclampsia Prediction Using Machine Learning and Polygenic Risk Scores From Clinical and Genetic Risk Factors in Early and Late Pregnancies’, Hypertension, 81(2), pp. 264–272. Available at: https://doi.org/10.1161/HYPERTENSIONAHA.123.21053.

Koziol, J.A. (2010) 'The rank product method with two samples', FEBS letters, 584(21), pp. 4481–4484. Available at: https://doi.org/10.1016/j.febslet.2010.10.012.

Kristensen, J.H., Basit, S., Wohlfahrt, J., Damholt, M.B. and Boyd, H.A. (2019) 'Pre-eclampsia and risk of later kidney disease: Nationwide cohort study', The BMJ, 365, pp. l1516. Available at: https://doi.org/10.1136/bmj.l1516.

Krüger-Genge, A., Blocki, A., Franke, R. and Jung, F. (2019) 'Vascular endothelial cell biology: An update', International journal of molecular sciences, 20(18), pp. 4411.

Kukurba, K.R. and Montgomery, S.B. (2015) 'RNA sequencing and analysis', Cold Spring Harbor Protocols, 2015(11), pp. pdb. top084970.

Kumar, A. and Jain, J. (2024) ‘Serum Oxidative Stress Markers and Antioxidant Levels in Pre-Eclampsia: Evaluating Malondialdehyde, Total Antioxidant Capacity, and Superoxide Dismutase as Potential Biomarkers’, CME Journal Geriatric Medicine, 16, pp. 116–123. Available at: https://doi.org/10.61336/cmejgm.

Kumar, V. and Stewart, J.H. (2025) ‘The complement system in human pregnancy and preeclampsia’, Frontiers in Immunology, 16. Available at: https://doi.org/10.3389/fimmu.2025.1617140.

Kung, J.T., Colognori, D. and Lee, J.T. (2013) 'Long noncoding RNAs: Past, present, and future', Genetics, 193(3), pp. 651–669.

Łabaj, P.P., Leparc, G.G., Linggi, B.E., Markillie, L.M., Wiley, H.S. and Kreil, D.P. (2011) 'Characterization and improvement of RNA-seq precision in quantitative transcript expression profiling', Bioinformatics, 27(13), pp. i383–i391.

Lachmann, A., Torre, D., Keenan, A.B., Jagodnik, K.M., Lee, H.J., Wang, L., Silverstein, M.C. and Ma'ayan, A. (2018) 'Massive mining of publicly available RNA-seq data from human and mouse', Nature Communications, 9, p. 1366. Available at: https://doi.org/10.1038/s41467-018-03751-6.

Lai, C.-C., Zhang, L.-L., Sun, M.-Y., Sun, J.-F. and Jiang, H. (2023) '[Association between HLA-DRB1 gene polymorphisms and genetic susceptibility of early-onset severe preeclampsia]', Zhongguo Dang Dai Er Ke Za Zhi = Chinese Journal of Contemporary Pediatrics, 25(10), pp. 1022–1027. Available at: https://doi.org/10.7499/j.issn.1008-8830.2303100.

Lai, J., Syngelaki, A., Nicolaides, K.H., von Dadelszen, P. and Magee, L.A. (2021) 'Impact of new definitions of preeclampsia at term on identification of adverse maternal and perinatal outcomes', American journal of obstetrics and gynecology, 224(5), pp. 518.e1–518.e11. Available at: https://doi.org/10.1016/j.ajog.2020.11.004.

Lakens, D. (2013) 'Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs', Frontiers in psychology, 4, pp. 863.

Lala, V., Zubair, M. and Minter, D.A. (2023) 'Liver function tests'StatPearls StatPearls Publishing Available at: http://www.ncbi.nlm.nih.gov/books/NBK482489/ Aug 25, 2023.

Lam, C., Lim, K. and Karumanchi, S.A. (2005) 'Circulating angiogenic factors in the pathogenesis and prediction of preeclampsia', Hypertension, 46(5), pp. 1077–1085.

LaMarca, B. (2012) 'Endothelial dysfunction; an important mediator in the Pathophysiology of Hypertension during Preeclampsia', Minerva ginecologica, 64(4), pp. 309–320.

Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J., Subramanian, A. and Ross, K.N. (2006) 'The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease', Science, 313(5795), pp. 1929–1935.

Lamminpää, R., Vehviläinen-Julkunen, K., Gissler, M. and Heinonen, S. (2012) 'Preeclampsia complicated by advanced maternal age: A registry-based study on primiparous women in finland 1997–2008', BMC pregnancy and childbirth, 12, pp. 1–5.

Lavezzo, E., Barzon, L., Toppo, S. and Palù, G. (2016) 'Third generation sequencing technologies applied to diagnostic microbiology: Benefits and challenges in applications and data analysis', Expert Review of Molecular Diagnostics, 16(9), pp. 1011–1023. Available at: https://doi.org/10.1080/14737159.2016.1217158.

Lawless, L., Qin, Y., Xie, L. and Zhang, K. (2023) 'Trophoblast Differentiation: Mechanisms and Implications for Pregnancy Complications', Nutrients, 15(16), p. 3564. Available at: https://doi.org/10.3390/nu15163564.

Le Ray, C., Scherier, S., Anselem, O., Marszalek, A., Tsatsaris, V., Cabrol, D. and Goffinet, F. (2012) 'Association between oocyte donation and maternal and perinatal outcomes in women aged 43 years or older', Human reproduction, 27(3), pp. 896–901.

Leavey, K., Benton, S.J., Grynspan, D., Kingdom, J.C., Bainbridge, S.A. and Cox, B.J. (2016) 'Unsupervised Placental Gene Expression Profiling Identifies Clinically Relevant Subclasses of Human Preeclampsia', Hypertension, 68(1), pp. 137–147. Available at: https://doi.org/10.1161/HYPERTENSIONAHA.116.07293.

Lee, H. and Schatz, M.C. (2012) 'Genomic dark matter: the reliability of short read mapping illustrated by the genome mappability score', Bioinformatics, 28(16), pp. 2097–2105. Available at: https://doi.org/10.1093/bioinformatics/bts330.

Lee, S., Zhang, A.Y., Su, S., Ng, A.P., Holik, A.Z., Asselin-Labat, M.-L., Ritchie, M.E. and Law, C.W. (2020) 'Covering all your bases: incorporating intron signal from RNA-seq data', NAR Genomics and Bioinformatics, 2(3), p. lqaa073. Available at: https://doi.org/10.1093/nargab/lqaa073.

Lei, J., Zhao, M., Li, L., Ji, B., Xu, T., Sun, M., Chen, J., Qiu, J. and Gao, Q. (2022) 'Research progress of placental vascular pathophysiological changes in pregnancy-induced hypertension and gestational diabetes mellitus', Frontiers in Physiology, 13, p. 954636. Available at: https://doi.org/10.3389/fphys.2022.954636.

Leslie, K., Thilaganathan, B. and Papageorghiou, A. (2011) 'Early prediction and prevention of pre-eclampsia', Best Practice & Research Clinical Obstetrics & Gynaecology, 25(3), pp. 343–354. Available at: https://doi.org/10.1016/j.bpobgyn.2011.01.002.

Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N., Gnirke, A. and Regev, A. (2010) 'Comprehensive comparative analysis of strand-specific RNA sequencing methods', Nature methods, 7(9), pp. 709–715.

Levine, R.J., Maynard, S.E., Qian, C., Lim, K., England, L.J., Yu, K.F., Schisterman, E.F., Thadhani, R., Sachs, B.P. and Epstein, F.H. (2004) 'Circulating angiogenic factors and the risk of preeclampsia', New England journal of medicine, 350(7), pp. 672–683.

Levine, R.J., Thadhani, R., Qian, C., Lam, C., Lim, K., Kai, F.Y., Blink, A.L., Sachs, B.P., Epstein, F.H. and Sibai, B.M. (2005) 'Urinary placental growth factor and risk of preeclampsia', Jama, 293(1), pp. 77–85.

Li, C., Meng, Y., Zhou, B., Zhang, Y., Xia, Q., Huang, Y., Meng, L., Shan, C., Xia, J., Zhang, X., Wang, Q., Lv, M. and Long, W. (2024) 'ITGB3 is reduced in pregnancies with preeclampsia and its influence on biological behavior of trophoblast cells', Molecular Medicine, 30(1), p. 275. Available at: https://doi.org/10.1186/s10020-024-01050-z.

Li, C., Samuels, D.C., Zhao, Y., Shyr, Y. and Guo, Y. (2018) 'Power and sample size calculations for high-throughput sequencing-based experiments', Briefings in bioinformatics, 19(6), pp. 1247–1255.

Li, D., Zand, M.S., Dye, T.D., Goniewicz, M.L., Rahman, I. and Xie, Z. (2022) 'An evaluation of RNA-seq differential analysis methods', PLoS One, 17(9), pp. e0264246.

Li, G., Huang, S., Liu, X. and Du, Q. (2022) 'Potential biomarkers and molecular mechanisms in preeclampsia progression', Open Life Sciences, 17(1), pp. 529–543. Available at: https://doi.org/10.1515/biol-2022-0053.

Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows–Wheeler transform', bioinformatics, 25(14), pp. 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009) 'The sequence alignment/map format and SAMtools', bioinformatics, 25(16), pp. 2078–2079.

Li, H., Ruan, J. and Durbin, R. (2008) 'Mapping short DNA sequencing reads and calling variants using mapping quality scores', Genome Research, 18(11), pp. 1851–1858. Available at: https://doi.org/10.1101/gr.078212.108.

Li, J., Jiang, L., Kai, H., Zhou, Y., Cao, J. and Tang, W. (2025) 'Identifying preeclampsia-associated key module and hub genes via weighted gene co-expression network analysis', Scientific Reports, 15(1), pp. 1364. Available at: https://doi.org/10.1038/s41598-025-85599-7.

Li, J., Witten, D.M., Johnstone, I.M. and Tibshirani, R. (2012) 'Normalization, testing, and false discovery rate estimation for RNA-sequencing data', Biostatistics, 13(3), pp. 523–538.

Li, L., Ching, W.-K. and Liu, Z.-P. (2022) 'Robust biomarker screening from gene expression data by stable machine learning-recursive feature elimination methods', Computational Biology and Chemistry, 100, p. 107747. Available at: https://doi.org/10.1016/j.compbiolchem.2022.107747.

Li, L., Lu, X. and Dean, J. (2013) 'The maternal to zygotic transition in mammals', Molecular aspects of medicine, 34(5), pp. 919–938. Available at: https://doi.org/10.1016/j.mam.2013.01.003.

Li, L., Zhang, Z., Li, H., Zhou, M., Li, F., Chu, C., Zhang, Y., Zhu, X., Ju, H. and Li, X. (2024) ‘Research progress on the STAT signaling pathway in pregnancy and pregnancy-associated disorders’, Frontiers in Immunology, 14. Available at: https://doi.org/10.3389/fimmu.2023.1331964.

Li, Q., Wei, X., Wu, F., Qin, C., Dong, J., Chen, C. and Lin, Y. (2024) ‘Development and validation of preeclampsia predictive models using key genes from bioinformatics and machine learning approaches’, Frontiers in Immunology, 15, p. 1416297. Available at: https://doi.org/10.3389/fimmu.2024.1416297.

Li, S., Wang, Z., Vieira, L.A., Zheutlin, A.B., Ru, B., Schadt, E., Wang, P., Copperman, A.B., Stone, J.L., Gross, S.J., Kao, Y.-H., Lau, Y.K., Dolan, S.M., Schadt, E.E. and Li, L. (2022) ‘Improving preeclampsia risk prediction by modeling pregnancy trajectories from routinely collected electronic medical record data’, npj Digital Medicine, 5(1), p. 68. Available at: https://doi.org/10.1038/s41746-022-00612-x.

Li, T., Xu, M., Wang, Yuan, Wang, Ya, Tang, H., Duan, H., Zhao, G., Zheng, M. and Hu, Y. (2024) ‘Prediction model of preeclampsia using machine learning based methods: a population based cohort study in China’, Frontiers in Endocrinology, 15. Available at: https://doi.org/10.3389/fendo.2024.1345573.

Li, X., Liu, L., Whitehead, C., Li, J., Thierry, B., Le, T.D. and Winter, M. (2022) ‘Identifying preeclampsia-associated genes using a control theory method’, Briefings in Functional Genomics, 21(4), pp. 296–309. Available at: https://doi.org/10.1093/bfgp/elac006.

Li, X., Liu, L., Whitehead, C., Li, J., Thierry, B., Le, T.D. and Winter, M. (2022) 'Identifying preeclampsia-associated genes using a control theory method', Briefings in Functional Genomics, 21(4), pp. 296–309.

Li, X., Wu, Z., Fu, X. and Han, W. (2013) 'Long noncoding RNAs: Insights from biological features and functions to diseases', Medicinal research reviews, 33(3), pp. 517–553.

Li, X.L., Guo, P.L., Xue, Y., Gou, W.L., Tong, M. and Chen, Q. (2016) ‘An analysis of the differences between early and late preeclampsia with severe hypertension’, Pregnancy Hypertension: An International Journal of Women’s Cardiovascular Health, 6(1), pp. 47–52. Available at: https://doi.org/10.1016/j.preghy.2015.12.003.

Li, Y., Shen, X., Yang, C., Cao, Z., Du, R., Wang, J. and Wang, M. (2021) 'Novel electronic health records applied for prediction of pre-eclampsia: Machine-learning algorithms', Pregnancy Hypertension, 26, pp. 102–109.

Li, Y., Zimmermann, B., Rusterholz, C., Kang, A., Holzgreve, W. and Hahn, S. (2004) 'Size separation of circulatory DNA in maternal plasma permits ready detection of fetal DNA polymorphisms', Clinical chemistry, 50(6), pp. 1002–1011.

Lian, I.A., Langaas, M., Moses, E. and Johansson, Å (2013) 'Differential gene expression at the maternal-fetal interface in preeclampsia is influenced by gestational age', PloS one, 8(7), pp. e69848.

Liang, Y. (2024) 'Plasma free cell RNA profiling for the prediction of preeclampsia', American Journal of Clinical and Experimental Immunology, 13(3), pp. 140.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P.A., Clarke, M., Devereaux, P.J., Kleijnen, J. and Moher, D. (2009) 'The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration', PLOS Medicine, 6(7), pp. e1000100. Available at: https://doi.org/10.1371/journal.pmed.1000100.

Lim, K., Zhou, Y., Janatpour, M., McMaster, M., Bass, K., Chun, S. and Fisher, S.J. (1997) 'Human cytotrophoblast differentiation/invasion is abnormal in pre-eclampsia.', The American journal of pathology, 151(6), pp. 1809.

Lim, W.K. and Mathuru, A.S. (2020) 'Design, challenges, and the potential of transcriptomics to understand social behavior', Current Zoology, 66(3), pp. 321–330. Available at: https://doi.org/10.1093/cz/zoaa007.

Lin, L., Chen, Q., Hirsch, J.P., Yoo, S., Yeung, K., Bumgarner, R.E., Tu, Z., Schadt, E.E. and Zhu, J. (2018) 'Temporal genetic association and temporal genetic causality methods for dissecting complex networks', Nature Communications, 9(1), pp. 3980. Available at: https://doi.org/10.1038/s41467-018-06203-3.

Lin, Y., Chang, P., Hsu, C., Hung, M., Chien, Y., Hwu, W., Lai, F. and Lee, N. (2022) 'Comparison of GATK and DeepVariant by trio sequencing', Scientific Reports, 12, pp. 1809. Available at: https://doi.org/10.1038/s41598-022-05833-4.

Lin, Y.C., Mallia, D., Clark-Sevilla, A.O., Catto, A., Leshchenko, A., Haas, D.M., Wapner, R., Pe'er, I., Raja, A. and Salleb-Aouissi, A. (2022) 'Preeclampsia predictor with machine learning: A comprehensive and bias-free machine learning pipeline', medRxiv, , pp. 2022.06. 08.22276107.

Lintao, R.C.V., Richardson, L.S., Kammala, A.K., Chapa, J., Yunque-Yap, D.A., Khanipov, K., Golovko, G., Dalmacio, L.M.M. and Menon, R. (2024) 'PGRMC2 and HLA-G regulate immune homeostasis in a microphysiological model of human maternal-fetal membrane interface', Communications Biology, 7(1), p. 1041. Available at: https://doi.org/10.1038/s42003-024-06740-2.

Lisonkova, S. and Joseph, K.S. (2013) 'Incidence of preeclampsia: Risk factors and outcomes associated with early-versus late-onset disease', American Journal of Obstetrics and Gynecology, 209(6), pp. 544. e1–544. e12.

Liu, J., Campen, A., Huang, S., Peng, S., Ye, X., Palakal, M., Dunker, A.K., Xia, Y. and Li, S. (2008) 'Identification of a gene signature in cell cycle pathway for

breast cancer prognosis using gene expression profiling data', BMC medical genomics, 1(1), pp. 1–12.

Liu, L., Pu, Y., Fan, J., Yan, Y., Liu, W., Luo, K., Wang, Y., Zhao, G., Chen, T., Puiu, P.D. and Huang, H. (2024) ‘Wearable Sensors, Data Processing, and Artificial Intelligence in Pregnancy Monitoring: A Review’, Sensors, 24(19), p. 6426. Available at: https://doi.org/10.3390/s24196426.

Liu, M., Yang, X., Chen, G., Ding, Y., Shi, M., Sun, L., Huang, Z., Liu, J., Liu, T. and Yan, R. (2022) 'Development of a prediction model on preeclampsia using machine learning-based method: A retrospective cohort study in china', Frontiers in Physiology, 13, pp. 896969.

Liu, N., Guo, Y., Gong, L. and Wang, B. (2021) 'Advances in biomarker development and potential application for preeclampsia based on pathogenesis', European Journal of Obstetrics & Gynecology and Reproductive Biology: X, 9, pp. 100119. Available at: https://doi.org/10.1016/j.eurox.2020.100119.

Liu, Q., Fang, L. and Wu, C. (2022) ‘Alternative Splicing and Isoforms: From Mechanisms to Diseases’, Genes, 13(3), p. 401. Available at: https://doi.org/10.3390/genes13030401.

Liu, S., Xu, C., Zhang, Y., Liu, J., Yu, B., Liu, X. and Dehmer, M. (2018) 'Feature selection of gene expression data for cancer classification using double RBF-kernels', BMC bioinformatics, 19(1), pp. 1–14.

Liu, W. and Wang, J. (2022) 'Recursive elimination current algorithms and a distributed computing scheme to accelerate wrapper feature selection', Information Sciences, 589, pp. 636–654. Available at: https://doi.org/10.1016/j.ins.2021.12.086.

Liu, W., Liao, X., Yang, Y., Lin, H., Yeong, J., Zhou, X., Shi, X. and Liu, J. (2022) ‘Joint dimension reduction and clustering analysis of single-cell RNA-seq and spatial transcriptomics data’, Nucleic Acids Research, 50(12), pp. e72–e72. Available at: https://doi.org/10.1093/nar/gkac219.

Liu, X., Li, N., Liu, S., Wang, J., Zhang, N., Zheng, X., Leung, K. and Cheng, L. (2019) 'Normalization methods for the analysis of unbalanced transcriptome data: A review', Frontiers in Bioengineering and Biotechnology, 0. Available at: https://doi.org/10.3389/fbioe.2019.00358.

Liu, Z., Wang, T., Yang, X., Zhou, Q., Zhu, S., Zeng, J., Chen, H., Sun, J., Li, L. and Xu, J. (2022) 'Polyadenylation ligation-mediated sequencing (PALM-Seq) characterizes cell-free coding and non-coding RNAs in human biofluids', Clinical and Translational Medicine, 12(7), pp. e987.

Lo, Y.D., Chan, K.A., Sun, H., Chen, E.Z., Jiang, P., Lun, F.M., Zheng, Y.W., Leung, T.Y., Lau, T.K. and Cantor, C.R. (2010) 'Maternal plasma DNA

sequencing reveals the genome-wide genetic and mutational profile of the fetus', Science translational medicine, 2(61), pp. 61ra91.

Lo, Y.D., Corbetta, N., Chamberlain, P.F., Rai, V., Sargent, I.L., Redman, C.W. and Wainscoat, J.S. (1997) 'Presence of fetal DNA in maternal plasma and serum', The lancet, 350(9076), pp. 485–487.

Lo, Y.D., Zhang, J., Leung, T.N., Lau, T.K., Chang, A.M. and Hjelm, N.M. (1999) 'Rapid clearance of fetal DNA from maternal plasma', The American Journal of Human Genetics, 64(1), pp. 218–224.

Lo, Y.M., Corbetta, N., Chamberlain, P.F., Rai, V., Sargent, I.L., Redman, C.W. and Wainscoat, J.S. (1997) 'Presence of fetal DNA in maternal plasma and serum', Lancet (London, England), 350(9076), pp. 485–487. Available at: https://doi.org/10.1016/S0140-6736(97)02174-0.

Lokeswara, A.W., Hiksas, R., Irwinda, R. and Wibowo, N. (2021) 'Preeclampsia: From Cellular Wellness to Inappropriate Cell Death, and the Roles of Nutrition', Frontiers in Cell and Developmental Biology, 9. Available at: https://doi.org/10.3389/fcell.2021.726513.

Lokki, A.I., Heikkinen-Eloranta, J.K. and Laivuori, H. (2018) 'The immunogenetic conundrum of preeclampsia', Frontiers in Immunology, 9, pp. 2630. Available at: https://doi.org/10.3389/fimmu.2018.02630.

Lopes van Balen, V.A., Spaan, J.J., Cornelis, T. and Spaanderman, M.E.A. (2017) 'Prevalence of chronic kidney disease after preeclampsia', Journal of Nephrology, 30(3), pp. 403–409. Available at: https://doi.org/10.1007/s40620-016-0342-1.

Love, M.I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', Genome biology, 15(12), pp. 1–21.

Lovén, J., Orlando, D.A., Sigova, A.A., Lin, C.Y., Rahl, P.B., Burge, C.B., Levens, D.L., Lee, T.I. and Young, R.A. (2012) 'Revisiting global gene expression analysis', Cell, 151(3), pp. 476–482.

Lun, F.M., Chiu, R.W., Allen Chan, K.C., Yeung Leung, T., Kin Lau, T. and Dennis Lo, Y.M. (2008) 'Microfluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma', Clinical chemistry, 54(10), pp. 1664–1672.

Luo, F., Yue, J., Li, L., Mei, J., Liu, X. and Huang, Y. (2022) 'Narrative review of the relationship between the maternal-fetal interface immune tolerance and the onset of preeclampsia', Annals of Translational Medicine, 10(12), pp. 713–713. Available at: https://doi.org/10.21037/atm-22-2287.

Lykke, J.A., Langhoff-Roos, J., Lockwood, C.J., Triche, E.W. and Paidas, M.J. (2010) 'Mortality of mothers from cardiovascular and non-cardiovascular causes

following pregnancy complications in first delivery', Paediatric and perinatal epidemiology, 24(4), pp. 323–330.

Lytle, J.R., Yario, T.A. and Steitz, J.A. (2007) 'Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR', Proceedings of the National Academy of Sciences, 104(23), pp. 9667–9672.

MacDonald, T.M., Walker, S.P., Hannan, N.J., Tong, S. and Kaitu'u-Lino, T.J. (2022) 'Clinical tools and biomarkers to predict preeclampsia', eBioMedicine, 75. Available at: https://doi.org/10.1016/j.ebiom.2021.103780.

MacDonald, T.M., Walker, S.P., Hannan, N.J., Tong, S. and Tu'uhevaha, J. (2022) 'Clinical tools and biomarkers to predict preeclampsia', EBioMedicine, 75.

Magee, L.A., Singer, J., Lee, T., Rey, E., Asztalos, E., Hutton, E., Helewa, M., Logan, A.G., Ganzevoort, W., Welch, R., Thornton, J.G., Kinshella, M.W., Green, M., Tsigas, E. and Dadelszen, P.v. (2021) 'The impact of pre-eclampsia definitions on the identification of adverse outcome risk in hypertensive pregnancy – analyses from the CHIPS trial (control of hypertension in pregnancy study)', BJOG: An International Journal of Obstetrics & Gynaecology, 128(8), pp. 1373–1382. Available at: https://doi.org/10.1111/1471-0528.16602.

Malik, A., Jee, B. and Gupta, S.K. (2019) 'Preeclampsia: Disease biology and burden, its management strategies with reference to india', Pregnancy Hypertension, 15, pp. 23–31. Available at: https://doi.org/10.1016/j.preghy.2018.10.011.

Mallik, S. and Zhao, Z. (2018) 'Identification of gene signatures from RNA-seq data using pareto-optimal cluster algorithm', BMC Systems Biology, 12(8), pp. 21–29.

Mamsen, L.S., Ernst, E.H., Borup, R., Larsen, A., Olesen, R.H., Ernst, E., Anderson, R.A., Kristensen, S.G. and Andersen, C.Y. (2017) 'Temporal expression pattern of genes during the period of sex differentiation in human embryonic gonads', Scientific Reports, 7, pp. 15961. Available at: https://doi.org/10.1038/s41598-017-15931-3.

Manokhina, I., Wilson, S.L. and Robinson, W.P. (2015) 'Noninvasive nucleic acid–based approaches to monitor placental health and predict pregnancy-related complications', American Journal of Obstetrics and Gynecology, 213(4), pp. S197–S206.

Manoochehri, Z., Manoochehri, S., Soltani, F., Tapak, L. and Sadeghifar, M. (2021) 'Predicting preeclampsia and related risk factors using data mining approaches: A cross-sectional study', International Journal of Reproductive BioMedicine, 19(11), pp. 959.

Manuel, J.M., Guilloy, N., Khatir, I., Roucou, X. and Laurent, B. (2023) 'Re-evaluating the impact of alternative RNA splicing on proteomic diversity', Frontiers in Genetics, 14, pp. 1089053.

Mao, Y. and Yang, Y. (2019) 'A wrapper feature subset selection method based on randomized search and multilayer structure', BioMed Research International, 2019, pp. 9864213. Available at: https://doi.org/10.1155/2019/9864213.

Mapari, S.A., Shrivastava, D., Dave, A., Bedi, G.N., Gupta, A., Sachani, P., Kasat, P.R. and Pradeep, U. (2024) ‘Revolutionizing Maternal Health: The Role of Artificial Intelligence in Enhancing Care and Accessibility’, Cureus, 16(9), p. e69555. Available at: https://doi.org/10.7759/cureus.69555.

Mar, J.C., Kimura, Y., Schroder, K., Irvine, K.M., Hayashizaki, Y., Suzuki, H., Hume, D. and Quackenbush, J. (2009) 'Data-driven normalization strategies for high-throughput quantitative RT-PCR', BMC Bioinformatics, 10, pp. 110. Available at: https://doi.org/10.1186/1471-2105-10-110.

Marasco, L.E. and Kornblihtt, A.R. (2023) 'The physiology of alternative splicing', Nature Reviews Molecular Cell Biology, 24(4), pp. 242–254.

Marić, I., Contrepois, K., Moufarrej, M.N., Stelzer, I.A., Feyaerts, D., Han, X., Tang, A., Stanley, N., Wong, R.J., Traber, G.M., Ellenberger, M., Chang, A.L., Fallahzadeh, R., Nassar, H., Becker, M., Xenochristou, M., Espinosa, C., De Francesco, D., Ghaemi, M.S., Costello, E.K., Culos, A., Ling, X.B., Sylvester, K.G., Darmstadt, G.L., Winn, V.D., Shaw, G.M., Relman, D.A., Quake, S.R., Angst, M.S., Snyder, M.P., Stevenson, D.K., Gaudilliere, B. and Aghaeepour, N. (2022) ‘Early prediction and longitudinal modeling of preeclampsia from multiomics’, Patterns, 3(12), p. 100655. Available at: https://doi.org/10.1016/j.patter.2022.100655.

Marić, I., Tsur, A., Aghaeepour, N., Montanari, A., Stevenson, D.K., Shaw, G.M. and Winn, V.D. (2020) 'Early prediction of preeclampsia via machine learning', American Journal of Obstetrics & Gynecology MFM, 2(2), pp. 100100.

Marin, I., Pavaloiu, B., Marian, C., Racovita, V. and Goga, N. (2019) Early detection of preeclampsia based on a machine learning approach. IEEE, pp. 1.

Marin, M., Vargas, R., Harris, M., Jeffrey, B., Epperson, L.E., Durbin, D., Strong, M., Salfinger, M., Iqbal, Z., Akhundova, I., Vashakidze, S., Crudu, V., Rosenthal, A. and Farhat, M.R. (2022) ‘Benchmarking the empirical accuracy of short-read sequencing across the M. tuberculosis genome’, Bioinformatics, 38(7), pp. 1781–1787. Available at: https://doi.org/10.1093/bioinformatics/btac023.

Marín, R., Chiarello, D.I., Abad, C., Rojas, D., Toledo, F. and Sobrevia, L. (2020) 'Oxidative stress and mitochondrial dysfunction in early-onset and late-onset preeclampsia', Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1866(12), pp. 165961.

Maron, J.L., Johnson, K.L., Slonim, D., Lai, C., Ramoni, M., Alterovitz, G., Jarrah, Z., Yang, Z. and Bianchi, D.W. (2007) 'Gene expression analysis in pregnant women and their infants identifies unique fetal biomarkers that circulate in maternal blood', The Journal of Clinical Investigation, 117(10), pp. 3007. Available at: https://doi.org/10.1172/JCI29959.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', EMBnet.journal, 17(1), pp. 10–12.

Martínez-Velasco, A., Martinez-Villaseñor, L. and Miralles-Pechuán, L. (2018) Machine learning approach for pre-eclampsia risk factors association. pp. 232.

Martini, C., Saeed, Z., Simeone, P., Palma, S., Ricci, M., Arata, A., Sorella, A., Liani, R., Ricci, F., D'Antonio, F., Mattioli, A.V., Gallina, S., Santilli, F. and Renda, G. (2025) 'Preeclampsia: Insights into pathophysiological mechanisms and preventive strategies', American Journal of Preventive Cardiology, 23, p. 101054. Available at: https://doi.org/10.1016/j.ajpc.2025.101054.

Martins-Taylor, K., Hsiao, J.S., Chen, P., Glatt-Deeley, H., De Smith, A.J., Blakemore, A.I., Lalande, M. and Chamberlain, S.J. (2014) 'Imprinted expression of UBE3A in non-neuronal cells from a Prader–Willi syndrome patient with an atypical deletion', Human molecular genetics, 23(9), pp. 2364–2373.

Marzioni, D., Piani, F., Di Simone, N., Giannubilo, S.R., Ciavattini, A. and Tossetta, G. (2025) 'Importance of STAT3 signaling in preeclampsia (Review)', International Journal of Molecular Medicine, 55(4), p. 58. Available at: https://doi.org/10.3892/ijmm.2025.5499.

Matheny, M.E., Ohno-Machado, L. and Resnic, F.S. (2005) 'Discrimination and calibration of mortality risk prediction models in interventional cardiology', Journal of Biomedical Informatics, 38(5), pp. 367–375. Available at: https://doi.org/10.1016/j.jbi.2005.02.007.

Mathur, M., Kim, C.M., Munro, S.A., Rudina, S.S., Sawyer, E.M. and Smolke, C.D. (2019) 'Programmable mutually exclusive alternative splicing for generating RNA and protein diversity', Nature communications, 10(1), pp. 2673.

Matsumoto, H. (2017) 'Molecular and cellular events during blastocyst implantation in the receptive uterus: Clues from mouse models', Journal of Reproduction and Development, 63(5), pp. 445–454.

Mattick, J.S., Amaral, P.P., Carninci, P., Carpenter, S., Chang, H.Y., Chen, L., Chen, R., Dean, C., Dinger, M.E. and Fitzgerald, K.A. (2023) 'Long non-coding RNAs: Definitions, functions, challenges and recommendations', Nature Reviews Molecular Cell Biology, 24(6), pp. 430–447.

Mayer-Pickel, K., Stern, C., Cervar-Zivkovic, M., Schöll, W. and Moertl, M. (2020) 'Preeclampsia before fetal viability in women with primary antiphospholipid syndrome- materno-fetal outcomes in a series of 7 cases', Journal of

Reproductive Immunology, 138, p. 103101. Available at: https://doi.org/10.1016/j.jri.2020.103101.

Maynard, S.E., Min, J., Merchan, J., Lim, K., Li, J., Mondal, S., Libermann, T.A., Morgan, J.P., Sellke, F.W. and Stillman, I.E. (2003) 'Excess placental soluble fms-like tyrosine kinase 1 (sFlt1) may contribute to endothelial dysfunction, hypertension, and proteinuria in preeclampsia', The Journal of clinical investigation, 111(5), pp. 649–658.

Mayya, V.K. and Duchaine, T.F. (2019) 'Ciphers and executioners: How 3′-untranslated regions determine the fate of messenger RNAs', Frontiers in Genetics, 10, pp. 6.

Maza, E. (2016) 'In papyro comparison of TMM (edgeR), RLE (DESeq2), and MRN normalization methods for a simple two-conditions-without-replicates RNA-seq experimental design', Frontiers in Genetics, 7, pp. 164. Available at: https://doi.org/10.3389/fgene.2016.00164.

McCarthy, F.P., Ryan, R.M. and Chappell, L.C. (2018) 'Prospective biomarkers in preterm preeclampsia: A review', Pregnancy Hypertension, 14, pp. 72–78. Available at: https://doi.org/10.1016/j.preghy.2018.03.010.

McDermaid, A., Monier, B., Zhao, J., Liu, B. and Ma, Q. (2019) 'Interpretation of differential gene expression results of RNA-seq data: Review and integration', Briefings in bioinformatics, 20(6), pp. 2044–2054. Available at: https://doi.org/10.1093/bib/bby067.

McElwain, C.J., Tuboly, E., McCarthy, F.P. and McCarthy, C.M. (2020) ‘Mechanisms of Endothelial Dysfunction in Pre-eclampsia and Gestational Diabetes Mellitus: Windows Into Future Cardiometabolic Health?’, Frontiers in Endocrinology, 11, p. 655. Available at: https://doi.org/10.3389/fendo.2020.00655.

McGinnis, R., Steinthorsdottir, V., Williams, N.O., Thorleifsson, G., Shooter, S., Hjartardottir, S., Bumpstead, S., Stefansdottir, L., Hildyard, L. and Sigurdsson, J.K. (2017) 'Variants in the fetal genome near FLT1 are associated with risk of preeclampsia', Nature genetics, 49(8), pp. 1255–1260.

Meazaw, M.W., Chojenta, C., Muluneh, M.D. and Loxton, D. (2020) 'Systematic and meta-analysis of factors associated with preeclampsia and eclampsia in sub-saharan africa', PloS one, 15(8), pp. e0237600.

Mekinian, A., Lachassinne, E., Nicaise-Roland, P., Carbillon, L., Motta, M., Vicaut, E., Boinot, C., Avcin, T., Letoumelin, P. and De Carolis, S. (2013) 'European registry of babies born to mothers with antiphospholipid syndrome', Annals of the Rheumatic Diseases, 72(2), pp. 217–222.

Melchiorre, K., Sutherland, G.R., Baltabaeva, A., Liberati, M. and Thilaganathan, B. (2011) 'Maternal cardiac dysfunction and remodeling in women with preeclampsia at term', Hypertension, 57(1), pp. 85–93.

Melesse, M.F., Aynalem, G.L., Badi, M.B. and Aynalem, B.Y. (2025) ‘Maternal outcomes of severe preeclampsia and eclampsia and associated factors among women admitted at referral hospitals of amhara regional state, institutional-based cross-sectional study, North-West Ethiopia’, Frontiers in Global Women’s Health, 6. Available at: https://doi.org/10.3389/fgwh.2025.1555778.

Melinte-Popescu, A., Vasilache, I., Socolov, D. and Melinte-Popescu, M. (2023) 'Predictive performance of machine learning-based methods for the prediction of Preeclampsia—A prospective study', Journal of Clinical Medicine, 12(2), pp. 418.

Melinte-Popescu, A.-S., Vasilache, I.-A., Socolov, D. and Melinte-Popescu, M. (2023) ‘Predictive Performance of Machine Learning-Based Methods for the Prediction of Preeclampsia—A Prospective Study’, Journal of Clinical Medicine, 12(2), p. 418. Available at: https://doi.org/10.3390/jcm12020418.

Mendes, S., Timóteo-Ferreira, F., Almeida, H. and Silva, E. (2019) 'New insights into the process of placentation and the role of oxidative uterine microenvironment', Oxidative medicine and cellular longevity, 2019.

Meng, C., Zeleznik, O.A., Thallinger, G.G., Kuster, B., Gholami, A.M. and Culhane, A.C. (2016) ‘Dimension reduction techniques for the integrative analysis of multi-omics data’, Briefings in Bioinformatics, 17(4), pp. 628–641. Available at: https://doi.org/10.1093/bib/bbv108.

Meng, Y., Yan, Y., Yue, X., Li, Y. and Mo, Y. (2025) ‘Advances in cfDNA research for pregnancy-related diseases’, Frontiers in Cell and Developmental Biology, 13. Available at: https://doi.org/10.3389/fcell.2025.1600532.

Metzker, M.L. (2010) 'Sequencing technologies—the next generation', Nature reviews genetics, 11(1), pp. 31–46.

Mikat, B., Gellhaus, A., Wagner, N., Birdir, C., Kimmig, R. and Köninger, A. (2012) 'Early detection of maternal risk for preeclampsia', ISRN Obstetrics and Gynecology, 2012, pp. 172808. Available at: https://doi.org/10.5402/2012/172808.

Miller, J. and Ulrich, R. (2019) 'The quest for an optimal alpha', PLoS One, 14(1), pp. e0208631.

Miranda, M.L., Edwards, S.E. and Myers, E.R. (2011) 'Adverse birth outcomes among nulliparous vs. multiparous women', Public health reports, 126(6), pp. 797–805.

Mishra, P., Pandey, C.M., Singh, U., Keshri, A. and Sabaretnam, M. (2019) 'Selection of appropriate statistical methods for data analysis', Annals of cardiac anaesthesia, 22(3), pp. 297.

Mittal, V.K. and McDonald, J.F. (2012) 'R-SAP: A multi-threading computational pipeline for the characterization of high-throughput RNA-sequencing data', Nucleic acids research, 40(9), pp. e67.

Modak, R., Pal, Amitrajit, Pal, Amitava and Ghosh, M.K. (2020) ‘Prediction of preeclampsia by a combination of maternal spot urinary protein-creatinine ratio and uterine artery doppler.’, International Journal of Reproduction, Contraception, Obstetrics and Gynecology, 9(2), pp. 635–640.

Mohamad, M.A., Mohd Manzor, N.F., Zulkifli, N.F., Zainal, N., Hayati, A.R. and Ahmad Asnawi, A.W. (2020) ‘A Review of Candidate Genes and Pathways in Preeclampsia–An Integrated Bioinformatical Analysis’, Biology, 9(4), p. 62. Available at: https://doi.org/10.3390/biology9040062.

Molania, R., Foroutan, M., Gagnon-Bartsch, J.A., Gandolfo, L.C., Jain, A., Sinha, A., Olshansky, G., Dobrovic, A., Papenfuss, A.T. and Speed, T.P. (2023) 'Removing unwanted variation from large-scale RNA sequencing data with PRPS', Nature Biotechnology, 41(1), pp. 82–95. Available at: https://doi.org/10.1038/s41587-022-01440-w.

Momeni, K., Ghorbian, S., Ahmadpour, E. and Sharifi, R. (2023) ‘Unraveling the complexity: understanding the deconvolutions of RNA-seq data’, Translational Medicine Communications, 8(1), p. 21. Available at: https://doi.org/10.1186/s41231-023-00154-8.

Montagnana, M., Danese, E., Lippi, G. and Fava, C. (2017) ‘Blood laboratory testing for early prediction of preeclampsia: chasing the finish line or at the starting blocks?’, Annals of Medicine, 49(3), pp. 240–253. Available at: https://doi.org/10.1080/07853890.2016.1255350.

Montgomery-Csobán, T., Kavanagh, K., Murray, P., Robertson, C., Barry, S.J.E., Ukah, U.V., Payne, B.A., Nicolaides, K.H., Syngelaki, A., Ionescu, O., Akolekar, R., Hutcheon, J.A., Magee, L.A., Dadelszen, P.v., Brown, M.A., Davis, G.K., Parker, C., Walters, B.N., Sass, N., Ansermino, J.M., Cao, V., Cundiff, G.W., Dadelszen, E.C.M.v., Douglas, M.J., Dumont, G.A., Dunsmuir, D.T., Hutcheon, J.A., Joseph, K.S., Lalji, S., Lee, T., Li, J., Lim, K.I., Lisonkova, S., Lott, P., Menzies, J.M., Millman, A.L., Palmer, L., Payne, B.A., Qu, Z., Russell, J.A., Sawchuck, D., Shaw, D., Still, D.K., Ukah, U.V., Wagner, B., Walley, K.R., Hugo, D., Gruslin, T.I.A., Tawagi, G., Smith, G.N., Côté, A., Moutquin, J., Ouellet, A.B., Lee, S.K., Duan, T., Zhou, J., Haniff, T.I.F., Mahajan, S., Noovao, A., Karjalainend, H., Kortelainen, A., Laivuori, H., Ganzevoort, J.W., Groen, H., Kyle, P.M., Moore, M.P., Pullar, B., Bhutta, Z.A., Qureshi, R.N., Sikandar, R., Bhutta, T.I.S.Z., Cloete, G., Hall, D.R., Papendorp, T.I.E.v., Steyn, D.W., Biryabarema, C., Mirembe, F., Nakimuli, A., Allotey, J., Thangaratinam, S., Nicolaides, K.H., Ionescu, O., Syngelaki, A., Swiet, M.d., Magee, L.A., Dadelszen, P.v., Akolekar, R., Walker, J.J., Robson, S.C., Broughton-Pipkin, F., Loughna, P., Vatish, M., Redman, C.W.G., Barry, S.J.E., Kavanagh, K., Montgomery-Csobán, T., Murray, P., Robertson, C., Tsigas, E.Z., Woelkers, D.A., Lindheimer, M.D., Grobman, W.A., Sibai, B.M., Merialdi, M. and Widmer, M. (2024) 'Machine learning-enabled maternal risk assessment for women with pre-eclampsia (the PIERS-ML model):

A modelling study', The Lancet Digital Health, 6(4), pp. e238–e250. Available at: https://doi.org/10.1016/S2589-7500(23)00267-4.

Montgomery-Csobán, Tünde, Kavanagh, K., Murray, P., Robertson, C., Barry, S.J.E., Ukah, U.V., Payne, B.A., Nicolaides, K.H., Syngelaki, A., Ionescu, O., Akolekar, R., Hutcheon, J.A., Magee, L.A., Dadelszen, P. von, Brown, M.A., Davis, G.K., Parker, C., Walters, B.N., Sass, N., Ansermino, J.M., Cao, V., Cundiff, G.W., Dadelszen, E.C.M. von, Douglas, M.J., Dumont, G.A., Dunsmuir, D.T., Hutcheon, J.A., Joseph, K.S., Lalji, S., Lee, T., Li, J., Lim, K.I., Lisonkova, S., Lott, P., Menzies, J.M., Millman, A.L., Palmer, L., Payne, B.A., Qu, Z., Russell, J.A., Sawchuck, D., Shaw, D., Still, D.K., Ukah, U.V., Wagner, B., Walley, K.R., Hugo, D., Gruslin, T. late A., Tawagi, G., Smith, G.N., Côté, A.-M., Moutquin, J.-M., Ouellet, A.B., Lee, S.K., Duan, T., Zhou, J., Haniff, T. late F., Mahajan, S., Noovao, A., Karjalainend, H., Kortelainen, A., Laivuori, H., Ganzevoort, J.W., Groen, H., Kyle, P.M., Moore, M.P., Pullar, B., Bhutta, Z.A., Qureshi, R.N., Sikandar, R., Bhutta, T. late S.Z., Cloete, G., Hall, D.R., Papendorp, T. late E. van, Steyn, D.W., Biryabarema, C., Mirembe, F., Nakimuli, A., Allotey, J., Thangaratinam, S., Nicolaides, K.H., Ionescu, O., Syngelaki, A., Swiet, M. de, Magee, L.A., Dadelszen, P. von, Akolekar, R., Walker, J.J., Robson, S.C., Broughton-Pipkin, F., Loughna, P., Vatish, M., Redman, C.W.G., Barry, S.J.E., Kavanagh, K., Montgomery-Csobán, Tunde, Murray, P., Robertson, C., Tsigas, E.Z., Woelkers, D.A., Lindheimer, M.D., Grobman, W.A., Sibai, B.M., Merialdi, M. and Widmer, M. (2024) ‘Machine learning-enabled maternal risk assessment for women with pre-eclampsia (the PIERS-ML model): a modelling study’, The Lancet Digital Health, 6(4), pp. e238–e250. Available at: https://doi.org/10.1016/S2589-7500(23)00267-4.

Moore, M.J. and Proudfoot, N.J. (2009) 'Pre-mRNA processing reaches back totranscription and ahead to translation', Cell, 136(4), pp. 688–700.

Morano, D., Rossi, S., Lapucci, C., Pittalis, M.C. and Farina, A. (2018) ‘Cell-Free DNA (cfDNA) Fetal Fraction in Early- and Late-Onset Fetal Growth Restriction’, Molecular Diagnosis & Therapy, 22(5), pp. 613–619. Available at: https://doi.org/10.1007/s40291-018-0353-9.

Moreira, M.W.L., Rodrigues, J.J.P.C., Al-Muhtadi, J., Korotaev, V.V. and de Albuquerque, V.H.C. (2021) ‘Neuro-fuzzy model for HELLP syndrome prediction in mobile cloud computing environments’, Concurrency and Computation: Practice and Experience, 33(7), p. e4651. Available at: https://doi.org/10.1002/cpe.4651.

Mortazavipour, M.M. and Shahbazi, S. (2022) 'The current applications of cell-free fetal DNA in prenatal diagnosis of single-gene diseases: A review', International Journal of Reproductive BioMedicine, 20(8), pp. 613.

Motta-Mejia, C., Kandzija, N., Zhang, W., Mhlomi, V., Cerdeira, A.S., Burdujan, A., Tannetta, D., Dragovic, R., Sargent, I.L., Redman, C.W., Kishore, U. and Vatish, M. (2017) 'Placental vesicles carry active endothelial nitric oxide synthase

and their activity is reduced in preeclampsia', Hypertension (Dallas, Tex. : 1979), 70(2), pp. 372–381. Available at: https://doi.org/10.1161/HYPERTENSIONAHA.117.09321.

Mou, A.D., Barman, Z., Hasan, M., Miah, R., Hafsa, J.M., Das Trisha, A. and Ali, N. (2021) 'Prevalence of preeclampsia and the associated risk factors among pregnant women in bangladesh', Scientific Reports, 11(1), pp. 1–9. Available at: https://doi.org/10.1038/s41598-021-00839-w.

Moufarrej, M.N., Bianchi, D.W., Shaw, G.M., Stevenson, D.K. and Quake, S.R. (2023) 'Noninvasive Prenatal Testing Using Circulating DNA and RNA: Advances, Challenges, and Possibilities', Annual review of biomedical data science, 6, pp. 397–418. Available at: https://doi.org/10.1146/annurev-biodatasci-020722-094144.

Moufarrej, M.N., Vorperian, S.K., Wong, R.J., Campos, A.A., Quaintance, C.C., Sit, R.V., Tan, M., Detweiler, A.M., Mekonen, H., Neff, N.F., Baruch-Gravett, C., Litch, J.A., Druzin, M.L., Winn, V.D., Shaw, G.M., Stevenson, D.K. and Quake, S.R. (2022) 'Early prediction of preeclampsia in pregnancy with cell-free RNA', Nature, 602(7898), pp. 689–694. Available at: https://doi.org/10.1038/s41586-022-04410-z.

Moufarrej, M.N., Wong, R.J., Campos, A.A., Quaintance, C.C., Sit, R.V., Tan, M., Neff, N.F., Druzin, M.L., Winn, V.D. and Shaw, G.M. (2021) 'Early prediction of preeclampsia in pregnancy with circulating, cell-free RNA', medRxiv, .

Moufarrej, M.N., Wong, R.J., Shaw, G.M., Stevenson, D.K. and Quake, S.R. (2020) 'Investigating Pregnancy and Its Complications Using Circulating Cell-Free RNA in Women's Blood During Gestation', Frontiers in Pediatrics, 8, p. 605219. Available at: https://doi.org/10.3389/fped.2020.605219.

Mouillet, J., Ouyang, Y., Coyne, C.B. and Sadovsky, Y. (2015) 'MicroRNAs in placental health and disease', American Journal of Obstetrics and Gynecology, 213(4), pp. S163–S172.

Moutsopoulos, I., Maischak, L., Lauzikaite, E., Vasquez Urbina, S.A., Williams, E.C., Drost, H. and Mohorianu, I.I. (2021) 'noisyR: Enhancing biological signal in sequencing datasets by characterizing random technical noise', Nucleic acids research, 49(14), pp. e83.

Mubarik, S., Ghulam, N., Suqing, W. and Cuifang, F. (2020) 'Preeclampsia mediates the association between advanced maternal age and adverse pregnancy outcomes: A structural equation modeling approach', Iranian Journal of Public Health, 49(9), pp. 1727.

Mukherjee, I., Singh, S., Karmakar, A., Kashyap, N., Mridha, A.R., Sharma, J.B., Luthra, K., Sharma, R.S., Biswas, S., Dhar, R. and Karmakar, S. (2023) 'New immune horizons in therapeutics and diagnostic approaches to Preeclampsia',

American Journal of Reproductive Immunology, 89(2), p. e13670. Available at: https://doi.org/10.1111/aji.13670.

Murthi, P., Pinar, A.A., Dimitriadis, E. and Samuel, C.S. (2020) 'Inflammasomes—A Molecular Link for Altered Immunoregulation and Inflammation Mediated Vascular Dysfunction in Preeclampsia', International Journal of Molecular Sciences, 21(4), p. 1406. Available at: https://doi.org/10.3390/ijms21041406.

Musich, R., Cadle-Davidson, L. and Osier, M.V. (2021) 'Comparison of short-read sequence aligners indicates strengths and weaknesses for biologists to consider', Frontiers in Plant Science, 0. Available at: https://doi.org/10.3389/fpls.2021.657240.

Myatt, L. and Webster, R.P. (2009) 'Vascular biology of preeclampsia', Journal of Thrombosis and Haemostasis, 7(3), pp. 375–384.

Myhrer, D.-M.M., Frøystad, M., Paasche Roland, M.C., Ueland, T. and Lekva, T. (2024) 'The long non-coding RPPH1 is decreased in leukocytes and increased in plasma from women developing pre-eclampsia†', Biology of Reproduction, 111(2), pp. 427–435. Available at: https://doi.org/10.1093/biolre/ioae069.

Nahm, F.S. (2022) 'Receiver operating characteristic curve: Overview and practical use for clinicians', Korean Journal of Anesthesiology, 75(1), pp. 25–36. Available at: https://doi.org/10.4097/kja.21209.

Nakagawa, S. and Cuthill, I.C. (2007) 'Effect size, confidence interval and statistical significance: A practical guide for biologists', Biological reviews, 82(4), pp. 591–605.

Nakimuli, A., Chazara, O., Byamugisha, J., Elliott, A.M., Kaleebu, P., Mirembe, F. and Moffett, A. (2014) 'Pregnancy, parturition and preeclampsia in women of african ancestry', American Journal of Obstetrics and Gynecology, 210(6), pp. 510–520. e1.

Nath, M.C., Cubro, H., McCormick, D.J., Milic, N.M. and Garovic, V.D. (2020) 'Preeclamptic women have decreased circulating IL-10 values at the time of preeclampsia diagnosis: Systematic review and meta-analysis', Hypertension (Dallas, Tex. : 1979), 76(6), pp. 1817–1827. Available at: https://doi.org/10.1161/HYPERTENSIONAHA.120.15870.

Nayak, J.N., RajilaRajendran, H., Mahesh, S., Srinivasan, S., Nayak, J.N., RajilaRajendran, H., Mahesh, S. and Srinivasan, S. (2024) 'Comparison of morphometric and histological features of placenta of in vitro fertilization and naturally conceived pregnancies', Gulhane Medical Journal [Preprint]. Available at: https://doi.org/10.4274/gulhane.galenos.2024.92259.

Ness, R.B. and Roberts, J.M. (1996) 'Heterogeneous causes constituting the single syndrome of preeclampsia: A hypothesis and its implications', American Journal of Obstetrics and Gynecology, 175(5), pp. 1365–1370.

Nevis, I.F., Reitsma, A., Dominic, A., McDonald, S., Thabane, L., Akl, E.A., Hladunewich, M., Akbari, A., Joseph, G. and Sia, W. (2011) 'Pregnancy outcomes in women with chronic kidney disease: A systematic review', Clinical journal of the American Society of Nephrology: CJASN, 6(11), pp. 2587.

Newman, T.B., Browner, W.S., Cummings, S.R. and Hulley, S.B. (2013) 'Designing studies of medical tests', Designing clinical research, , pp. 183–205.

Ng, E.K., Tsui, N.B., Lau, T.K., Leung, T.N., Chiu, R.W., Panesar, N.S., Lit, L.C., Chan, K. and Lo, Y.D. (2003) 'mRNA of placental origin is readily detectable in maternal plasma', Proceedings of the National Academy of Sciences, 100(8), pp. 4748–4753.

Ng, K.W., Chaturvedi, N., Coté, G.L., Fisher, S.A. and Mabbott, S. (2024) 'Biomarkers and point of care screening approaches for the management of preeclampsia', Communications Medicine, 4, pp. 208. Available at: https://doi.org/10.1038/s43856-024-00642-4.

Ngo, T.T., Moufarrej, M.N., Rasmussen, M.H., Camunas-Soler, J., Pan, W., Okamoto, J., Neff, N.F., Liu, K., Wong, R.J. and Downes, K. (2018) 'Noninvasive blood tests for fetal development predict gestational age and preterm delivery', Science, 360(6393), pp. 1133–1136.

Ni, W., Gao, X., Su, X., Cai, J., Zhang, S., Zheng, L., Liu, J., Feng, Y., Chen, S., Ma, J., Cao, W. and Zeng, F. (2023) ‘Birth spacing and risk of adverse pregnancy and birth outcomes: A systematic review and dose–response meta-analysis’, Acta Obstetricia et Gynecologica Scandinavica, 102(12), pp. 1618–1633. Available at: https://doi.org/10.1111/aogs.14648.

Nie, Q., Zhou, B., Wang, Y., Ye, M., Chen, D. and He, F. (2024) ‘Evaluation of outcomes and risk factors for recurrent preeclampsia in a subsequent pregnancy’, Archives of Gynecology and Obstetrics, 310(5), pp. 2487–2495. Available at: https://doi.org/10.1007/s00404-024-07751-4.

Nieves, C., Victoria da Costa Ghignatti, P., Aji, N. and Bertagnolli, M. (2024) 'Immune cells and infectious diseases in preeclampsia susceptibility', Canadian Journal of Cardiology, 40(12), pp. 2340–2355. Available at: https://doi.org/10.1016/j.cjca.2024.09.012.

Nilsson, E.E., Winchester, P., Proctor, C., Beck, D. and Skinner, M.K. (2024) ‘Epigenetic biomarker for preeclampsia-associated preterm birth and potential preventative medicine’, Environmental Epigenetics, 10(1), p. dvae022. Available at: https://doi.org/10.1093/eep/dvae022.

Nirupama, R., Divyashree, S., Janhavi, P., Muthukumar, S.P. and Ravindra, P.V. (2021) 'Preeclampsia: Pathophysiology and management', Journal of Gynecology Obstetrics and Human Reproduction, 50(2), pp. 101975. Available at: https://doi.org/10.1016/j.jogoh.2020.101975.

Nishizawa, H., Pryor-Koishi, K., Kato, T., Kowa, H., Kurahashi, H. and Udagawa, Y. (2007) 'Microarray analysis of differentially expressed fetal genes in placental tissue derived from early and late onset severe pre-eclampsia', Placenta, 28(5-6), pp. 487–497.

Nissen, P., Hansen, J., Ban, N., Moore, P.B. and Steitz, T.A. (2000) 'The structural basis of ribosome activity in peptide bond synthesis', Science, 289(5481), pp. 920–930.

Njoroge, S., Kuriloff, M., Mueller, A., Lopes Perdigao, J., Dhir, R. and Rana, S. (2021) 'The interval between births and the risk of recurrent preeclampsia among predominantly high risk women in urban tertiary care center', Pregnancy Hypertension, 25, pp. 7–11. Available at: https://doi.org/10.1016/j.preghy.2021.05.009.

North, R.A., McCowan, L.M., Dekker, G.A., Poston, L., Chan, E.H., Stewart, A.W., Black, M.A., Taylor, R.S., Walker, J.J. and Baker, P.N. (2011) 'Clinical risk prediction for pre-eclampsia in nulliparous women: Development of model in international prospective cohort', Bmj, 342.

Norton, M.E., Jacobsson, B., Swamy, G.K., Laurent, L.C., Ranzini, A.C., Brar, H., Tomlinson, M.W., Pereira, L., Spitz, J.L. and Hollemon, D. (2015) 'Cell-free DNA analysis for noninvasive examination of trisomy', New England Journal of Medicine, 372(17), pp. 1589–1597.

Nwaimo, C.S., Adegbola, A.E. and Adegbola, M.D. (2024) 'Transforming healthcare with data analytics: Predictive models for patient outcomes', GSC Biological and Pharmaceutical Sciences, 27(3), pp. 25.

O'Gorman, N., Wright, D., Poon, L.C., Rolnik, D.L., Syngelaki, A., de Alvarado, M., Carbone, I.F., Dutemeyer, V., Fiolna, M., Frick, A., Karagiotis, N., Mastrodima, S., de Paco Matallana, C., Papaioannou, G., Pazos, A., Plasencia, W. and Nicolaides, K.H. (2017) 'Multicenter screening for pre-eclampsia by maternal factors and biomarkers at 11–13 weeks' gestation: comparison with NICE guidelines and ACOG recommendations', Ultrasound in Obstetrics & Gynecology, 49(6), pp. 756–760. Available at: https://doi.org/10.1002/uog.17455.

O'rourke, M.B., Town, S.E., Dalla, P.V., Bicknell, F., Koh Belic, N., Violi, J.P., Steele, J.R. and Padula, M.P. (2019) 'What is normalization? the strategies employed in top-down and bottom-up proteome analysis workflows', Proteomes, 7(3), pp. 29.

O'Brien, J., Hayder, H., Zayed, Y. and Peng, C. (2018) 'Overview of microRNA biogenesis, mechanisms of actions, and circulation', Frontiers in endocrinology, 9, pp. 402.

Ogawa, K., Urayama, K.Y., Tanigaki, S., Sago, H., Sato, S., Saito, S. and Morisaki, N. (2017) 'Association between very advanced maternal age and

adverse pregnancy outcomes: A cross sectional japanese study', BMC pregnancy and childbirth, 17, pp. 1–10.

O'Gorman, N., Wright, D., Poon, L.C., Rolnik, D.L., Syngelaki, A., de ALVARADO, M., Carbone, I.F., Dutemeyer, V., Fiolna, M. and Frick, A. (2017) 'Multicenter screening for pre-eclampsia by maternal factors and biomarkers at 11–13 weeks' gestation: Comparison with NICE guidelines and ACOG recommendations', Ultrasound in Obstetrics & Gynecology, 49(6), pp. 756–760.

Ogoyama, M., Takahashi, H., Suzuki, H., Ohkuchi, A., Fujiwara, H. and Takizawa, T. (2022) 'Non-coding RNAs and prediction of preeclampsia in the first trimester of pregnancy', Cells, 11(15), pp. 2428.

Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A.M. and Qasem, S.N. (2024) 'Machine learning-based predictive models for detection of cardiovascular diseases', Diagnostics, 14(2), pp. 144.

Okonechnikov, K., Conesa, A. and García-Alcalde, F. (2016) 'Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data', Bioinformatics, 32(2), pp. 292–294.

Ong, C., Lioo, A.W., Munim, S., Spencer, K. and Nicolaides, K.H. (2004) 'First-trimester maternal serum activin A in pre-eclampsia and fetal growth restriction', The Journal of Maternal-Fetal & Neonatal Medicine, 15(3), pp. 176–180.

Opichka, M.A., Rappelt, M.W., Gutterman, D.D., Grobe, J.L. and McIntosh, J.J. (2021) ‘Vascular Dysfunction in Preeclampsia’, Cells, 10(11), p. 3055. Available at: https://doi.org/10.3390/cells10113055.

Ortega, N.M., Winblad, N., Plaza Reyes, A. and Lanner, F. (2018) 'Functional genetics of early human development', Current Opinion in Genetics & Development, 52, pp. 1–6. Available at: https://doi.org/10.1016/j.gde.2018.04.005.

Oshlack, A. and Wakefield, M.J. (2009) 'Transcript length bias in RNA-seq data confounds systems biology', Biology direct, 4, pp. 1–10.

Özcan, M. and Peker, S. (2025) ‘Preeclampsia prediction via machine learning: a systematic literature review’, Health Systems, 14(3), pp. 208–222. Available at: https://doi.org/10.1080/20476965.2024.2435845.

Palei, A.C. (2021) 'Cell-free DNA as a potential biomarker for preeclampsia', Expert review of molecular diagnostics, 21(12), pp. 1253–1256. Available at: https://doi.org/10.1080/14737159.2021.1997593.

Palei, A.C., de O Cruz, J., Chaguri, J.L., Peraçoli, J.C., Romão-Veiga, M., Ribeiro-Vasques, V.R., Cavalli, R.C., Nunes, P.R., Luizon, M.R. and Sandrim, V.C. (2023) ‘Circulating levels of tissue inhibitor of metalloproteinase (TIMP)-3, a protein with inhibitory effects on angiogenesis, are increased in pre-eclampsia’, International journal of gynaecology and obstetrics: the official organ of the

International Federation of Gynaecology and Obstetrics, 161(2), pp. 544–551. Available at: https://doi.org/10.1002/ijgo.14552.

Palmer, K.R., Kaitu'u-Lino, T.J., Hastie, R., Hannan, N.J., Ye, L., Binder, N., Cannon, P., Tuohey, L., Johns, T.G. and Shub, A. (2015) 'Placental-specific sFLT-1 e15a protein is increased in preeclampsia, antagonizes vascular endothelial growth factor signaling, and has antiangiogenic activity', Hypertension, 66(6), pp. 1251–1259.

Parab, S. and Bhalerao, S. (2010) 'Choosing statistical test', International journal of Ayurveda research, 1(3), pp. 187.

Paria, B.C., Huet-Hudson, Y.M. and Dey, S.K. (1993) 'Blastocyst's state of activity determines the" window" of implantation in the receptive mouse uterus', Proceedings of the National Academy of Sciences, 90(21), pp. 10159–10162.

Parra-Cordero, M., Bosco, C., González, J., Gutiérrez, R., Barja, P. and Rodrigo, R. (2011) 'Immunohistochemical expression of von willebrand factor in the preeclamptic placenta', Journal of Molecular Histology, 42(5), pp. 459–465. Available at: https://doi.org/10.1007/s10735-011-9351-5.

Patel, K. (2024) 'Ethical reflections on data-centric AI: Balancing benefits and risks', International Journal of Artificial Intelligence Research and Development, 2(1), pp. 1–17.

Patel, R.K. and Jain, M. (2012) 'NGS QC toolkit: A toolkit for quality control of next generation sequencing data', PloS one, 7(2), pp. e30619. Available at: https://doi.org/10.1371/journal.pone.0030619.

Patino, C.M. and Ferreira, J.C. (2018) 'Meeting the assumptions of statistical tests: An important and often forgotten step to reporting valid results', Jornal Brasileiro de Pneumologia, 44, pp. 353.

Pei, C.-Z., Song, X.-X., Xu, H. and Baek, K.-H. (2025) 'Role of ubiquitin-proteasome system in preeclampsia (Review)', Molecular Medicine Reports, 33(1), p. 25. Available at: https://doi.org/10.3892/mmr.2025.13735.

Pélabon, C., Hilde, C.H., Einum, S. and Gamelon, M. (2020) 'On the use of the coefficient of variation to quantify and compare trait variation', Evolution letters, 4(3), pp. 180–188.

Peñailillo, R., Monteiro, L.J., Acuña-Gallardo, S., García, F., Velásquez, V., Correa, P., Díaz, P., Valdebenito, P.P., Navarro, C., Romero, R., Sánchez, M., Illanes, S.E. and Nardocci, G. (2022) 'Identification of LOC101927355 as a Novel Biomarker for Preeclampsia', Biomedicines, 10(6), p. 1253. Available at: https://doi.org/10.3390/biomedicines10061253.

Pérez-Rubio, P., Lottaz, C. and Engelmann, J.C. (2019) 'FastqPuri: High-performance preprocessing of RNA-seq data', BMC bioinformatics, 20(1), pp. 1–11.

Petca, A., Sinescu, R.D., Sandru, F., Petca, R., Dumitrascu, M.C., Mehedintu, C. and Zvanca, M.E. (2022) 'New approaches in predicting and diagnosing preeclampsia: Congo red dot paper test (review)', Experimental and therapeutic medicine, 23(4), pp. 270. Available at: https://doi.org/10.3892/etm.2022.11196.

Petkova-Parlapanska, K., Kostadinova-Slavova, D., Angelova, M., Sadi J. Al-Dahwi, R., Georgieva, E., Goycheva, P., Karamalakova, Y. and Nikolova, G. (2025) 'Oxidative Stress and Antioxidant Status in Pregnant Women with Gestational Diabetes Mellitus and Late-Onset Complication of Pre-Eclampsia', International Journal of Molecular Sciences, 26(8), p. 3605. Available at: https://doi.org/10.3390/ijms26083605.

Phillips, S.P., Spithoff, S. and Simpson, A. (2022) 'Artificial intelligence and predictive algorithms in medicine: Promise and problems', Canadian Family Physician, 68(8), pp. 570.

Phipps, E., Prasanna, D., Brima, W. and Jim, B. (2016) 'Preeclampsia: Updates in pathogenesis, definitions, and guidelines', Clinical Journal of the American Society of Nephrology : CJASN, 11(6), pp. 1102–1113. Available at: https://doi.org/10.2215/CJN.12081115.

Phipps, E.A., Thadhani, R., Benzing, T. and Karumanchi, S.A. (2019) 'Pre-eclampsia: pathogenesis, novel diagnostics and therapies', Nature reviews. Nephrology, 15(5), pp. 275–289. Available at: https://doi.org/10.1038/s41581-019-0119-6.

Phumsiripaiboon, P., Suksai, M., Suntharasaj, T. and Geater, A. (2020) 'Screening for pre-eclampsia: Performance of national institute for health and care excellence guidelines versus american college of obstetricians and gynecologists recommendations', Journal of Obstetrics and Gynaecology Research, 46(11), pp. 2323–2331. Available at: https://doi.org/10.1111/jog.14425.

Pinto-Souza, C.C., Kaihara, J.N.S., Rossini, B.C., Cavalli, R. de C., dos Santos, L.D. and Sandrim, V.C. (2025) 'Pathophysiological mechanisms in severe preeclampsia: role of upregulated proteins in blood pressure, extracellular matrix and immunity', Revista Brasileira de Ginecologia e Obstetrícia, 47, p. e-rbgo54. Available at: https://doi.org/10.61622/rbgo/2025rbgo54.

Piovesan, A., Pelleri, M.C., Antonaros, F., Strippoli, P., Caracausi, M. and Vitale, L. (2019) 'On the length, weight and GC content of the human genome', BMC Research Notes, 12(1), p. 106. Available at: https://doi.org/10.1186/s13104-019-4137-z.

Piskol, R., Ramaswami, G. and Li, J.B. (2013) 'Reliable Identification of Genomic Variants from RNA-Seq Data', The American Journal of Human Genetics, 93(4), pp. 641–651. Available at: https://doi.org/10.1016/j.ajhg.2013.08.008.

Podymow, T. and August, P. (2010) 'Postpartum course of gestational hypertension and preeclampsia', Hypertension in pregnancy, 29(3), pp. 294–300.

Poon, L.C. and Nicolaides, K.H. (2014) ‘Early Prediction of Preeclampsia’, Obstetrics and Gynecology International, 2014(1), p. 297397. Available at: https://doi.org/10.1155/2014/297397.

Poon, L.C., Shennan, A., Hyett, J.A., Kapur, A., Hadar, E., Divakar, H., McAuliffe, F., da Silva Costa, F., von Dadelszen, P., McIntyre, H.D., Kihara, A.B., Di Renzo, G.C., Romero, R., D’Alton, M., Berghella, V., Nicolaides, K.H. and Hod, M. (2019) 'The international federation of gynecology and obstetrics (FIGO)

Poon, L.C.Y., Kametas, N.A., Maiz, N., Akolekar, R. and Nicolaides, K.H. (2009) 'First-trimester prediction of hypertensive disorders in pregnancy', Hypertension, 53(5), pp. 812–818. Available at: https://doi.org/10.1161/HYPERTENSIONAHA.108.127977.

Popp, C., Carson, J.M., Drysdale, A.B., Arora, H., Johnstone, E.D., Myers, J.E. and van Loon, R. (2024) ‘Development of non-invasive biomarkers for pre-eclampsia through data-driven cardiovascular network models’, Scientific Reports, 14(1), p. 23144. Available at: https://doi.org/10.1038/s41598-024-72832-y.

Porcu, E., Sadler, M.C., Lepik, K., Auwerx, C., Wood, A.R., Weihs, A., Sleiman, M.S.B., Ribeiro, D.M., Bandinelli, S. and Tanaka, T. (2021) 'Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome', Nature communications, 12(1), pp. 1–9.

Porrua, O. and Libri, D. (2015) 'Transcription termination and the control of the transcriptome: Why, where and how to stop', Nature reviews Molecular cell biology, 16(3), pp. 190–202.

Powe, C.E., Levine, R.J. and Karumanchi, S.A. (2011) 'Preeclampsia, a disease of the maternal endothelium: The role of anti-angiogenic factors and implications for later cardiovascular disease', Circulation, 123(24), pp. 10.1161/CIRCULATIONAHA.109.853127. Available at: https://doi.org/10.1161/CIRCULATIONAHA.109.853127.

Prados, F.J., Debrock, S., Lemmen, J.G. and Agerholm, I. (2012) 'The cleavage stage embryo', Human Reproduction, 27(suppl_1), pp. i50–i71. Available at: https://doi.org/10.1093/humrep/des224.

Prater, M., Hamilton, R.S., Wa Yung, H., Sharkey, A.M., Robson, P., Abd Hamid, N.E., Jauniaux, E., Charnock-Jones, D.S., Burton, G.J. and Cindrova-Davies, T. (2021) ‘RNA-Seq reveals changes in human placental metabolism, transport and endocrinology across the first–second trimester transition’, Biology Open, 10(6), p. bio058222. Available at: https://doi.org/10.1242/bio.058222.

Pregnancy, World Health Organization International Collaborative Study of Hypertensive Disorders of (1988) 'Geographic variation in the incidence of

hypertension in pregnancy', American Journal of Obstetrics & Gynecology, 158(1), pp. 80–83. Available at: https://doi.org/10.5555/uri:pii:000293788890782X.

Prins, J.R., Boelens, H.M., Heimweg, J., Van der Heide, S., Dubois, A.E., Van Oosterhout, A.J. and Erwich, J.J.H. (2009) 'Preeclampsia is associated with lower percentages of regulatory T cells in maternal blood', Hypertension in pregnancy, 28(3), pp. 300–311.

Proudfoot, N.J. (2016) 'Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut', Science, 352(6291), pp. aad9926.

Qi, J., Wu, B., Chen, X., Wei, W. and Yao, X. (2022) 'Diagnostic biomolecules and combination therapy for pre-eclampsia', Reproductive Biology and Endocrinology : RB&E, 20, p. 136. Available at: https://doi.org/10.1186/s12958-022-01003-3.

Qiu, C., Luthy, D.A., Zhang, C., Walsh, S.W., Leisenring, W.M. and Williams, M.A. (2004) 'A prospective study of maternal serum C-reactive protein concentrations and risk of preeclampsia', American journal of hypertension, 17(2), pp. 154–160.

Quinn, J.J. and Chang, H.Y. (2016) 'Unique features of long non-coding RNA biogenesis and function', Nature reviews genetics, 17(1), pp. 47–62.

Rahat, B., Ali, T., Sapehia, D., Mahajan, A. and Kaur, J. (2020) 'Circulating cell-free nucleic acids as epigenetic biomarkers in precision medicine', Frontiers in genetics, 11, pp. 844.

Rai, M.F., Tycksen, E.D., Sandell, L.J. and Brophy, R.H. (2018) 'Advantages of RNA-seq compared to RNA microarrays for transcriptome profiling of anterior cruciate ligament tears', Journal of Orthopaedic Research®, 36(1), pp. 484–497.

Rainio, O., Teuho, J. and Klén, R. (2024) 'Evaluation metrics and statistical tests for machine learning', Scientific Reports, 14(1), pp. 1–14. Available at: https://doi.org/10.1038/s41598-024-56706-x.

Raja Xavier, J.P., Rianna, C., Hellwich, E., Nikolou, I., Lankapalli, A.K., Brucker, S.Y., Singh, Y., Lang, F., Schäffer, T.E. and Salker, M.S. (2024) 'Excessive endometrial PlGF- Rac1 signalling underlies endometrial cell stiffness linked to pre-eclampsia', Communications Biology, 7(1), p. 530. Available at: https://doi.org/10.1038/s42003-024-06220-7.

Rajkomar, A., Dean, J. and Kohane, I. (2019) 'Machine learning in medicine', New England Journal of Medicine, 380(14), pp. 1347–1358.

Ramazi, S. and Zahiri, J. (2021) 'Post-translational modifications in proteins: Resources, tools and prediction methods', Database, 2021, pp. baab012.

Ramazi, S., Allahverdi, A. and Zahiri, J. (2020) 'Evaluation of post-translational modifications in histone proteins: A review on histone modification defects in

developmental and neurological disorders', Journal of Biosciences, 45(1), pp. 135.

Ramma, W. and Ahmed, A. (2011) 'Is inflammation the cause of pre-eclampsia?', Biochemical Society transactions, 39(6), pp. 1619–1627.

Ramsköld, D., Wang, E.T., Burge, C.B. and Sandberg, R. (2009) 'An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data', PLoS computational biology, 5(12), pp. e1000598.

Rana, S., Burke, S.D. and Karumanchi, S.A. (2022) 'Imbalances in circulating angiogenic factors in the pathophysiology of preeclampsia and related disorders', American Journal of Obstetrics and Gynecology, 226(2), pp. S1019–S1034.

Rana, S., Lemoine, E., Granger, J.P. and Karumanchi, S.A. (2019) 'Preeclampsia', Circulation Research, 124(7), pp. 1094–1112. Available at: https://doi.org/10.1161/CIRCRESAHA.118.313276.

Ranjbar, A., Montazeri, F., Ghamsari, S.R., Mehrnoush, V., Roozbeh, N. and Darsareh, F. (2024) 'Machine learning models for predicting preeclampsia: A systematic review', BMC pregnancy and childbirth, 24(1), pp. 6. Available at: https://doi.org/10.1186/s12884-023-06220-1.

Rao, D., Chaudhari, N.K., Moore, R.M. and Jim, B. (2016) 'HELLP syndrome: A diagnostic conundrum with severe complications', BMJ Case Reports, 2016, pp. bcr2016216802. Available at: https://doi.org/10.1136/bcr-2016-216802.

Rao, M.S., Van Vleet, T.R., Ciurlionis, R., Buck, W.R., Mittelstadt, S.W., Blomme, E.A. and Liguori, M.J. (2019) 'Comparison of RNA-seq and microarray gene expression platforms for the toxicogenomic evaluation of liver from short-term rat toxicity studies', Frontiers in genetics, 9, pp. 636.

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C.E., Socci, N.D. and Betel, D. (2013) 'Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data', Genome biology, 14(9), pp. 1–13.

Rasmussen, M., Reddy, M., Nolan, R., Camunas-Soler, J., Khodursky, A., Scheller, N.M., Cantonwine, D.E., Engelbrechtsen, L., Mi, J.D., Dutta, A., Brundage, T., Siddiqui, F., Thao, M., Gee, E.P.S., La, J., Baruch-Gravett, C., Santillan, M.K., Deb, S., Ame, S.M., Ali, S.M., Adkins, M., DePristo, M.A., Lee, M., Namsaraev, E., Gybel-Brask, D.J., Skibsted, L., Litch, J.A., Santillan, D.A., Sazawal, S., Tribe, R.M., Roberts, J.M., Jain, M., Høgdall, E., Holzman, C., Quake, S.R., Elovitz, M.A. and McElrath, T.F. (2022) ‘RNA profiles reveal signatures of future health and disease in pregnancy’, Nature, 601(7893), pp. 422–427. Available at: https://doi.org/10.1038/s41586-021-04249-w.

Raymond, Y., Fernando, S., Menezes, M., MCLENNAN, A., COSTA, F.D.S., HARDY, T. and ROLNIK, D.L. (2023) 'Placental, maternal, fetal and technical

origins of false-positive cell-free DNA screening results', American Journal of Obstetrics and Gynecology, .

Reddy, M., Fenn, S., Rolnik, D.L., Mol, B.W., da Silva Costa, F., Wallace, E.M. and Palmer, K.R. (2021) ‘The impact of the definition of preeclampsia on disease diagnosis and outcomes: a retrospective cohort study’, American Journal of Obstetrics and Gynecology, 224(2), p. 217.e1-217.e11. Available at: https://doi.org/10.1016/j.ajog.2020.08.019.

Redman, C. and Sargent, I.L. (2000) 'Placental debris, oxidative stress and pre-eclampsia.', Placenta, 21(7).

Redman, C. and Sargent, I.L. (2009) 'Placental stress and pre-eclampsia: A revised view', Placenta, 30, pp. 38–42.

Redman, C., Tannetta, D.S., Dragovic, R.A., Gardiner, C., Southcombe, J.H., Collett, G.P. and Sargent, I.L. (2012) 'Does size matter? placental debris and the pathophysiology of pre-eclampsia', Placenta, 33, pp. S48–S54.

Redman, C.W. (1990) 'Platelets and the beginnings of preeclampsia', New England Journal of Medicine, 323(7), pp. 478–480.

Redman, C.W. and Sargent, I.L. (2005) 'Latest advances in understanding preeclampsia', Science (New York, N.Y.), 308(5728), pp. 1592–1594. Available at: https://doi.org/10.1126/science.1111726.

Redman, C.W., Sacks, G.P. and Sargent, I.L. (1999) 'Preeclampsia: An excessive maternal inflammatory response to pregnancy', American Journal of Obstetrics and Gynecology, 180(2), pp. 499–506.

Redman, C.W., Sargent, I.L. and Staff, A.C. (2014) 'IFPA senior award lecture: Making sense of pre-eclampsia–two placental causes of preeclampsia?', Placenta, 35, pp. S20–S25.

Redman, C.W.G. (1991) 'Pre-eclampsia and the placenta', Placenta, 12(4), pp. 301–308. Available at: https://doi.org/10.1016/0143-4004(91)90339-H.

Redman, C.W.G. and Sargent, I.L. (2010) ‘REVIEW ARTICLE: Immunology of Pre-Eclampsia’, American Journal of Reproductive Immunology, 63(6), pp. 534–543. Available at: https://doi.org/10.1111/j.1600-0897.2010.00831.x.

Reilly, K., McKenna, C., McCullough, S., McKee, S. and Mone, F. (2023) 'Prenatal genomic testing for ultrasound-detected fetal structural anomalies', The Obstetrician and Gynaecologist, 25(2), pp. 121–130.

Reines, D., Conaway, R.C. and Conaway, J.W. (1999) 'Mechanism and regulation of transcriptional elongation by RNA polymerase II', Current opinion in cell biology, 11(3), pp. 342–346.

Riano-Moreno, J.C., Vargas-Castellanos, E., Pedraza, A., Díaz-Quiñonez, L.S. and Rangel-Ramos, V.S. (2025) ‘Preeclampsia prediction and diagnosis: a comprehensive historical review from clinical insights to omics perspectives’,

Frontiers in Medicine, 12. Available at: https://doi.org/10.3389/fmed.2025.1689745.

Richards, O., Jenkins, C., Griffiths, H., Paczkowska, E., Dunstan, P.R., Jones, S., Morgan, M., Thomas, T., Bowden, J. and Nakimuli, A. (2021) 'Vibrational spectroscopy: A valuable screening and diagnostic tool for obstetric disorders?', Frontiers in Global Women's Health, 1, pp. 610582.

Rigau, M., Juan, D., Valencia, A. and Rico, D. (2019) ‘Intronic CNVs and gene expression variation in human populations’, PLoS Genetics, 15(1), p. e1007902. Available at: https://doi.org/10.1371/journal.pgen.1007902.

Rigó, J., Alasztics, B., Árokszállási, A., Dobó, N., Prosszer, M., Joó, J.G. and Molvarec, A. (2015) 'PL59. the impact of classification of hypertensive disorders of pregnancy based on the ACOG 2013 and ISSHP 2014 criteria', Pregnancy Hypertension: An International Journal of Women's Cardiovascular Health, 5(3), pp. 222–223. Available at: https://doi.org/10.1016/j.preghy.2015.07.035.

Rijnders, R.J., van der Schoot, C.E., Bossers, B., de Vroede, M.A. and Christiaens, G.C. (2001) 'Fetal sex determination from maternal plasma in pregnancies at risk for congenital adrenal hyperplasia', Obstetrics & Gynecology, 98(3), pp. 374–378.

Risso, D., Ngai, J., Speed, T.P. and Dudoit, S. (2014) 'Normalization of RNA-seq data using factor analysis of control genes or samples', Nature biotechnology, 32(9), pp. 896–902. Available at: https://doi.org/10.1038/nbt.2931.

Risso, D., Schwartz, K., Sherlock, G. and Dudoit, S. (2011) 'GC-content normalization for RNA-seq data', BMC bioinformatics, 12(1), pp. 1–17.

Ritchie, M.E., Phipson, B., Wu, D.I., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. (2015) 'Limma powers differential expression analyses for RNA-sequencing and microarray studies', Nucleic acids research, 43(7), pp. e47.

Ritis, K., Doumas, M., Mastellos, D., Micheli, A., Giaglis, S., Magotti, P., Rafail, S., Kartalis, G., Sideras, P. and Lambris, J.D. (2006) 'A novel C5a receptor-tissue factor cross-talk in neutrophils links innate immunity to coagulation pathways', The Journal of Immunology, 177(7), pp. 4794–4802.

Roberge, S., Giguère, Y., Villa, P., Nicolaides, K., Vainio, M., Forest, J., von Dadelzen, P., Vaiman, D., Tapp, S. and Bujold, E. (2012) 'Early administration of low-dose aspirin for the prevention of severe and mild preeclampsia: A systematic review and meta-analysis', American Journal of Perinatology, , pp. 551–556.

Roberge, S., Villa, P., Nicolaides, K., Giguère, Y., Vainio, M., Bakthi, A., Ebrashy, A. and Bujold, E. (2012) 'Early administration of low-dose aspirin for the prevention of preterm and term preeclampsia: A systematic review and meta-analysis', Fetal diagnosis and therapy, 31(3), pp. 141–146.

Roberts, C.L., Ford, J.B., Algert, C.S., Antonsen, S., Chalmers, J., Cnattingius, S., Gokhale, M., Kotelchuck, M., Melve, K.K. and Langridge, A. (2011) 'Population-based trends in pregnancy hypertension and pre-eclampsia: An international comparative study', BMJ open, 1(1), pp. e000101.

Roberts, J.M. and Cooper, D.W. (2001) 'Pathogenesis and genetics of pre-eclampsia', The Lancet, 357(9249), pp. 53–56.

Roberts, J.M. and Escudero, C. (2012) 'The placenta in preeclampsia', Pregnancy Hypertension: An International Journal of Women's Cardiovascular Health, 2(2), pp. 72–83. Available at: https://doi.org/10.1016/j.preghy.2012.01.001.

Roberts, J.M. and Hubel, C.A. (2004) 'Oxidative stress in preeclampsia', American Journal of Obstetrics & Gynecology, 190(5), pp. 1177–1178.

Roberts, J.M. and Hubel, C.A. (2009) 'The two stage model of preeclampsia: Variations on the theme', Placenta, 30, pp. 32–37.

Robinson, M.D. and Oshlack, A. (2010) 'A scaling normalization method for differential expression analysis of RNA-seq data', Genome Biology, 11(3), pp. R25. Available at: https://doi.org/10.1186/gb-2010-11-3-r25.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) 'edgeR: A bioconductor package for differential expression analysis of digital gene expression data', bioinformatics, 26(1), pp. 139–140.

Rodriguez-Esteban, R. and Jiang, X. (2017) 'Differential gene expression in disease: A comparison between high-throughput studies and the literature', BMC medical genomics, 10, pp. 1–10.

Rokotyanskaya, E.A., Panova, I.A., Malyshkina, A.I., Fetisova, I.N., Fetisov, N.S., Kharlamova, N.V. and Kuligina, M.V. (2020) ‘Technologies for Prediction of Preeclampsia’, Modern Technologies in Medicine, 12(5), pp. 78–84. Available at: https://doi.org/10.17691/stm2020.12.5.09.

Rolfo, A., Attini, R., Nuzzo, A.M., Piazzese, A., Parisi, S., Ferraresi, M., Todros, T. and Piccoli, G.B. (2013) 'Chronic kidney disease may be differentially diagnosed from preeclampsia by serum biomarkers', Kidney international, 83(1), pp. 177–181.

Romão, R.M., Levi, J.E., Carvalho, H.B.d., Francisco, R.P.V., Amorim Filho, A.G.d. and Zugaib, M. (2012) 'Use of cell-free fetal nucleic acids in maternal blood for prenatal diagnosis: The reality of this scenario in brazil', Revista da Associação Médica Brasileira, 58, pp. 615–619. Available at: https://doi.org/10.1590/S0104-42302012000500021.

Romero, R. and Chaiworapongsa, T. (2013) 'Preeclampsia: A link between trophoblast dysregulation and an antiangiogenic state', The Journal of clinical

investigation, 123(7), pp. 2775–2777. Available at: https://doi.org/10.1172/JCI70431.

Romundstad, L.B., Romundstad, P.R., Sunde, A., von Düring, V., Skjærven, R. and Vatten, L.J. (2006) 'Increased risk of placenta previa in pregnancies following IVF/ICSI; a comparison of ART and non-ART pregnancies in the same mother', Human reproduction, 21(9), pp. 2353–2358.

Ronzoni, S., Rashid, S., Santoro, A., Mei-Dan, E., Barrett, J., Okun, N. and Huang, T. (2025) 'Preterm preeclampsia screening and prevention: a comprehensive approach to implementation in a real-world setting', BMC Pregnancy and Childbirth, 25, p. 32. Available at: https://doi.org/10.1186/s12884-025-07154-6.

Rosati, D., Palmieri, M., Brunelli, G., Morrione, A., Iannelli, F., Frullanti, E. and Giordano, A. (2024) 'Differential gene expression analysis pipelines and bioinformatic tools for the identification of specific biomarkers: A review', Computational and Structural Biotechnology Journal, 23, pp. 1154–1168. Available at: https://doi.org/10.1016/j.csbj.2024.02.018.

Rosati, D., Palmieri, M., Brunelli, G., Morrione, A., Iannelli, F., Frullanti, E. and Giordano, A. (2024) 'Differential gene expression analysis pipelines and bioinformatic tools for the identification of specific biomarkers: A review', Computational and structural biotechnology journal, .

Rustici, G., Kolesnikov, N., Brandizi, M., Burdett, T., Dylag, M., Emam, I., Farne, A., Hastings, E., Ison, J. and Keays, M. (2012) 'ArrayExpress update—trends in database growth and links to data analysis tools', Nucleic acids research, 41(D1), pp. D987–D990.

Rustveld, L.O., Kelsey, S.F. and Sharma, R. (2008) 'Association between maternal infections and preeclampsia: A systematic review of epidemiologic studies', Maternal and Child Health Journal, 12(2), pp. 223–242. Available at: https://doi.org/10.1007/s10995-007-0224-1.

Rybak-Krzyszkowska, M., Staniczek, J., Kondracka, A., Bogusławska, J., Kwiatkowski, S., Góra, T., Strus, M. and Górczewski, W. (2023) 'From Biomarkers to the Molecular Mechanism of Preeclampsia-A Comprehensive Literature Review', International Journal of Molecular Sciences, 24(17), p. 13252. Available at: https://doi.org/10.3390/ijms241713252.

Saeidian, A.H., Youssefian, L., Vahidnezhad, H. and Uitto, J. (2020) 'Research techniques made simple: Whole-transcriptome sequencing by RNA-seq for diagnosis of monogenic disorders', Journal of Investigative Dermatology, 140(6), pp. 1117–1126. e1.

Sakinah, N., Tahir, M., Badriyah, T. and Syarif, I. (2019) LSTM with adam optimization-powered high accuracy preeclampsia classification. IEEE, pp. 314.

Salmon, J.E., Heuser, C., Triebwasser, M., Liszewski, M.K., Kavanagh, D., Roumenina, L., Branch, D.W., Goodship, T., Fremeaux-Bacchi, V. and Atkinson, J.P. (2011) 'Mutations in complement regulatory proteins predispose to preeclampsia: A genetic analysis of the PROMISSE cohort', PLoS medicine, 8(3), pp. e1001013.

Salomon, L.J., Benattar, C., Audibert, F., Fernandez, H., Duyme, M., Taieb, J. and Frydman, R. (2003) 'Severe preeclampsia is associated with high inhibin A levels and normal leptin levels at 7 to 13 weeks into pregnancy', American Journal of Obstetrics and Gynecology, 189(6), pp. 1517–1522.

San Juan-Reyes, S., Gómez-Oliván, L.M., Islas-Flores, H. and Dublán-García, O. (2020) 'Oxidative stress in pregnancy complicated by preeclampsia', Archives of Biochemistry and Biophysics, 681, p. 108255. Available at: https://doi.org/10.1016/j.abb.2020.108255.

Sandström, A., Snowden, J.M., Höijer, J., Bottai, M. and Wikström, A.-K. (2019) 'Clinical risk assessment in early pregnancy for preeclampsia in nulliparous women: A population based cohort study', PLOS ONE, 14(11), p. e0225716. Available at: https://doi.org/10.1371/journal.pone.0225716.

Sandve, G.K., Nekrutenko, A., Taylor, J. and Hovig, E. (2013) 'Ten Simple Rules for Reproducible Computational Research', PLOS Computational Biology, 9(10), p. e1003285. Available at: https://doi.org/10.1371/journal.pcbi.1003285.

Santoro, A., Lampinen, A., Mathewson, K., Lillicrap, T. and Raposo, D. (2021) 'Symbolic behaviour in artificial intelligence', arXiv preprint arXiv:2102.03406, .

Sarker, S., Mahmud, S.M.H., Hosen, M.F., Michael Goh, K.O. and Shoombuatong, W. (2025) 'AttBiomarker: unveiling preeclampsia biomarkers and molecular pathways through two-stage gene selection techniques and attention-based CNN with gene regulatory network analysis', Briefings in Bioinformatics, 26(5), p. bbaf473. Available at: https://doi.org/10.1093/bib/bbaf473.

Sasaki, Y., Darmochwal-Kolarz, D., Suzuki, D., Sakai, M., Ito, M., Shima, T., Shiozaki, A., Rolinski, J. and Saito, S. (2007) 'Proportion of peripheral blood and decidual CD4 CD25bright regulatory T cells in pre-eclampsia', Clinical & Experimental Immunology, 149(1), pp. 139–145.

Sasaki, Y., Sakai, M., Miyazaki, S., Higuma, S., Shiozaki, A. and Saito, S. (2004) 'Decidual and peripheral blood CD4 CD25 regulatory T cells in early pregnancy subjects and spontaneous abortion cases', MHR: Basic science of reproductive medicine, 10(5), pp. 347–353.

Sastry, A.V., Hu, A., Heckmann, D., Poudel, S., Kavvas, E. and Palsson, B.O. (2021) 'Independent component analysis recovers consistent regulatory signals from disparate datasets', PLoS Computational Biology, 17(2), pp. e1008647.

Savvidou, M.D., Hingorani, A.D., Tsikas, D., Frölich, J.C., Vallance, P. and Nicolaides, K.H. (2003) 'Endothelial dysfunction and raised plasma

concentrations of asymmetric dimethylarginine in pregnant women who subsequently develop pre-eclampsia', Lancet (London, England), 361(9368), pp. 1511–1517. Available at: https://doi.org/10.1016/S0140-6736(03)13177-7.

Schapkaitz, E., Libhaber, E., Jacobson, B.F., Meiring, M. and Büller, H.R. (2021) 'Von willebrand factor propeptide-to-antigen ratio in HIV-infected pregnancy: Evidence of endothelial activation', Journal of Thrombosis and Haemostasis, 19(12), pp. 3168–3176.

Schiffer, V., Evers, L., de Haas, S., Ghossein-Doha, C., Al-Nasiry, S. and Spaanderman, M. (2020) 'Spiral artery blood flow during pregnancy: A systematic review and meta-analysis', BMC Pregnancy and Childbirth, 20, pp. 680. Available at: https://doi.org/10.1186/s12884-020-03150-0.

Schlembach, D., Hund, M., Wolf, C. and Vatish, M. (2019) 'Diagnostic utility of angiogenic biomarkers in pregnant women with suspected preeclampsia: A health economics review', Pregnancy Hypertension, 17, pp. 28–35. Available at: https://doi.org/10.1016/j.preghy.2019.03.002.

Schmidt, L.J., Rieger, O., Neznansky, M., Hackelöer, M., Dröge, L.A., Henrich, W., Higgins, D. and Verlohren, S. (2022) ‘A machine-learning–based algorithm improves prediction of preeclampsia-associated adverse outcomes’, American Journal of Obstetrics & Gynecology, 227(1), p. 77.e1-77.e30. Available at: https://doi.org/10.1016/j.ajog.2022.01.026.

Schneider, H. (2017) 'Placental dysfunction as a key element in the pathogenesis of preeclampsia', Journal of Mother and Child, 21(4), pp. 309–316.

Schoots, M.H., Gordijn, S.J., Scherjon, S.A., van Goor, H. and Hillebrands, J. (2018) 'Oxidative stress in placental pathology', Placenta, 69, pp. 153–161.

Schurch, N.J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G.G. and Owen-Hughes, T. (2016) 'How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?', Rna, 22(6), pp. 839–851.

Schurch, N.J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G.G., Owen-Hughes, T., Blaxter, M. and Barton, G.J. (2016) ‘How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?’, RNA, 22(6), pp. 839–851. Available at: https://doi.org/10.1261/rna.053959.115.

Selwood, L. and Johnson, M.H. (2006) 'Trophoblast and hypoblast in the monotreme, marsupial and eutherian mammal: Evolution and origins', Bioessays, 28(2), pp. 128–145.

Serdar, C.C., Cihan, M., Yücel, D. and Serdar, M.A. (2021) 'Sample size, power and effect size revisited: Simplified and practical approaches in pre-clinical, clinical and laboratory studies', Biochemia medica, 31(1), pp. 27–53.

Serra, B., Mendoza, M., Scazzocchio, E., Meler, E., Nolla, M., Sabrià, E., Rodríguez, I. and Carreras, E. (2020) 'A new model for screening for early-onset preeclampsia', American Journal of Obstetrics and Gynecology, 222(6), pp. 608. e1–608. e18.

Serrano, N.C., Quintero-Lesmes, D.C., Dudbridge, F., Leon, L.J., Hingorani, A.D., Williams, D.J. and Casas, J.P. (2020) 'Family history of pre-eclampsia and cardiovascular disease as risk factors for pre-eclampsia: the GenPE case-control study', Hypertension in Pregnancy, 39(1), pp. 56–63. Available at: https://doi.org/10.1080/10641955.2019.1704003.

Shafabakhsh, R., Mirhosseini, N., Chaichian, S., Moazzami, B., Mahdizadeh, Z. and Asemi, Z. (2019) 'Could circRNA be a new biomarker for pre-eclampsia?', Molecular Reproduction and Development, 86(12), pp. 1773–1780. Available at: https://doi.org/10.1002/mrd.23262.

Shahryary, Y., Hazarika, R.R. and Johannes, F. (2020) 'MethylStar: A fast and robust pre-processing pipeline for bulk or single-cell whole-genome bisulfite sequencing data', BMC genomics, 21, pp. 1–8.

Shahshahan, Z., Hashemi, L. and Rasouli, O. (2014) 'Maternal serum interleukin 6 and 8 and C-reactive protein in predicting the tocolytic therapy in preterm labor', Journal of Research in Medical Sciences: The Official Journal of Isfahan University of Medical Sciences, 19(6), pp. 537–541.

Sharma, B.S., Prabhakaran, V., Desai, A.P., Bajpai, J., Verma, R.J. and Swain, P.K. (2019) 'Post-translational modifications (PTMs), from a cancer perspective: An overview', Oncogen, 2(3), pp. 12.

Sharp, A.N., Heazell, A.E.P., Crocker, I.P. and Mor, G. (2010) 'Placental apoptosis in health and disease', American journal of reproductive immunology (New York, N.Y. : 1989), 64(3), pp. 159–169. Available at: https://doi.org/10.1111/j.1600-0897.2010.00837.x.

Shaw, G.T.W., Shih, E.S.C., Chen, C. and Hwang, M. (2011) 'Preservation of ranking order in the expression of human housekeeping genes', PLoS ONE, 6(12), pp. e29314. Available at: https://doi.org/10.1371/journal.pone.0029314.

Shaw, L.J., Patel, K., Lala-Trindade, A., Feltovich, H., Vieira, L., Kontorovich, A., Ananth, C., Taqueti, V.R., Mitrani, L., Stern, T., DeBolt, C., Kase, N., Smith, R.T., Narula, J., Mehran, R., Bianco, A., Bhatt, D.L. and Stone, J.L. (2024) 'Pathophysiology of Preeclampsia-Induced Vascular Dysfunction and Implications for Subclinical Myocardial Damage and Heart Failure', JACC: Advances, 3(6), p. 100980. Available at: https://doi.org/10.1016/j.jacadv.2024.100980.

Shelansky, R. and Boeger, H. (2020) 'Nucleosomal proofreading of activator–promoter interactions', Proceedings of the National Academy of Sciences, 117(5), pp. 2456–2461.

Shennan, A.H. and Hurrell, A. (2021) 'The evolving definition of pre-eclampsia', BJOG: An International Journal of Obstetrics & Gynaecology, 128(8), pp. 1383–1383. Available at: https://doi.org/10.1111/1471-0528.16632.

Shi, H., Zhou, Y., Jia, E., Pan, M., Bai, Y. and Ge, Q. (2021) 'Bias in RNA-seq library preparation: Current challenges and solutions', BioMed Research International, 2021, pp. e6647597. Available at: https://doi.org/10.1155/2021/6647597.

Shrivastava, S., Naik, R., Suryawanshi, H. and Gupta, N. (2019) 'Microchimerism: A new concept', Journal of Oral and Maxillofacial Pathology : JOMFP, 23(2), pp. 311. Available at: https://doi.org/10.4103/jomfp.JOMFP_85_17.

Shyu, I., Liu, C.-F., Tsai, Y.-C., Ma, Y.-S., Kuo, T.-N. and Yow-Ling, S. (2025) 'Machine learning predictive system to predict the risk of developing pre-eclampsia', BMJ Health & Care Informatics, 32(1), p. e101151. Available at: https://doi.org/10.1136/bmjhci-2024-101151.

Sibai, B.M. (2005) 'Diagnosis, prevention, and management of eclampsia', Obstetrics and Gynecology, 105(2), pp. 402–410. Available at: https://doi.org/10.1097/01.AOG.0000152351.13671.99.

Sibai, B.M. and Stella, C.L. (2009) 'Diagnosis and management of atypical preeclampsia-eclampsia', American Journal of Obstetrics and Gynecology, 200(5), pp. 481.e1–481.e7. Available at: https://doi.org/10.1016/j.ajog.2008.07.048.

Silva, W.M.C., Hurtado, F.A., Simi, K., Barros, P.H.A., Sokolowskei, D., Silva-Pereira, I., Walter, M.E. and Brigido, M. (2022) 'Transcriptome analysis using RNA-seq and scRNA-seq'Transcriptomics in health and disease Springer, pp. 73–107. .

Simas, T.A.M., Crawford, S.L., Solitro, M.J., Frost, S.C., Meyer, B.A. and Maynard, S.E. (2007) 'Angiogenic factors for the prediction of preeclampsia in high-risk women', American Journal of Obstetrics and Gynecology, 197(3), pp. 244. e1–244. e8.

Simillion, C., Liechti, R., Lischer, H.E., Ioannidis, V. and Bruggmann, R. (2017) 'Avoiding the pitfalls of gene set enrichment analysis with SetRank', BMC bioinformatics, 18(1), pp. 1–14.

Sims, D., Sudbery, I., Ilott, N.E., Heger, A. and Ponting, C.P. (2014) 'Sequencing depth and coverage: Key considerations in genomic analyses', Nature Reviews Genetics, 15(2), pp. 121–132.

Singh, V., Kirtipal, N., Song, B. and Lee, S. (2024) 'Normalization of RNA-seq data using adaptive trimmed mean with multi-reference', Briefings in Bioinformatics, 25(3), pp. bbae241. Available at: https://doi.org/10.1093/bib/bbae241.

Sisti, G. and Colombi, I. (2019) 'New blood pressure cut off for preeclampsia definition: 130/80 mmHg', European Journal of Obstetrics & Gynecology and Reproductive Biology, 240, pp. 322–324. Available at: https://doi.org/10.1016/j.ejogrb.2019.07.019.

Sitras, V., Paulssen, R.H., Grønaas, H., Leirvik, J., Hanssen, T.A., Vårtun, Å and Acharya, G. (2009) 'Differential placental gene expression in severe preeclampsia', Placenta, 30(5), pp. 424–433.

Skoura, R., Andronikidi, P.-E., Anestakis, D., Petanidis, S., Orovou, E., Tzitiridou, M. and Eskitzis, P. (2022) ‘Antiphospholipid Syndrome and Preeclampsia in Pregnancy: A Case Report’, Cureus, 14(8), p. e28458. Available at: https://doi.org/10.7759/cureus.28458.

Slade, L.J., Syngelaki, A., Wilson, M., Mistry, H.D., Akolekar, R., von Dadelszen, P., Nicolaides, K.H. and Magee, L.A. (2025) ‘Blood pressure cutoffs at 11-13 weeks of gestation and risk of preeclampsia’, American Journal of Obstetrics and Gynecology, 232(2), p. 214.e1-214.e10. Available at: https://doi.org/10.1016/j.ajog.2024.04.032.

Slobodchikova, T., Tayzhanova, D., Amirbekova, Z., Vazenmiller, D., Mustafin, R. and Izmailovich, M. (2025) ‘Genetic Variants in Preeclampsia During Pregnancy: A Hospital-Based Case–Control Study’, Journal of Clinical Medicine, 14(11), p. 3850. Available at: https://doi.org/10.3390/jcm14113850.

Small, H.Y., Akehurst, C., Sharafetdinova, L., McBride, M.W., McClure, J.D., Robinson, S.W., Carty, D.M., Freeman, D.J. and Delles, C. (2017) ‘HLA gene expression is altered in whole blood and placenta from women who later developed preeclampsia’, Physiological Genomics, 49(3), pp. 193–200. Available at: https://doi.org/10.1152/physiolgenomics.00106.2016.

Smets, E.M.L., Visser, A., Go, A.T.J.I., van Vugt, J.M.G. and Oudejans, C.B.M. (2006) 'Novel biomarkers in preeclampsia', Clinica Chimica Acta, 364(1), pp. 22–32. Available at: https://doi.org/10.1016/j.cca.2005.06.011.

Smid, M., Coebergh van den Braak, R.R., van de Werken, H.J., van Riet, J., van Galen, A., de Weerd, V., van der Vlugt-Daane, M., Bril, S.I., Lalmahomed, Z.S. and Kloosterman, W.P. (2018) 'Gene length corrected trimmed mean of M-values (GeTMM) processing of RNA-seq data performs similarly in intersample analyses while improving intrasample comparisons', BMC bioinformatics, 19, pp. 1–13.

Smith, E.N., Jepsen, K., Khosroheidari, M., Rassenti, L.Z., D’Antonio, M., Ghia, E.M., Carson, D.A., Jamieson, C.H., Kipps, T.J. and Frazer, K.A. (2014) 'Biased estimates of clonal evolution and subclonal heterogeneity can arise from PCR

duplicates in deep sequencing experiments', Genome Biology, 15(7), pp. 420. Available at: https://doi.org/10.1186/s13059-014-0420-4.

Smith, G.C.S., Stenhouse, E.J., Crossley, J.A., Aitken, D.A., Cameron, A.D. and Connor, J.M. (2002) 'Early pregnancy levels of pregnancy-associated plasma protein a and the risk of intrauterine growth restriction, premature birth, preeclampsia, and stillbirth', The Journal of Clinical Endocrinology and Metabolism, 87(4), pp. 1762–1767. Available at: https://doi.org/10.1210/jcem.87.4.8430.

Smith, M.H. (2004) 'A sample/population size activity: Is it the sample size of the sample as a fraction of the population that matters?', Journal of Statistics Education, 12(2).

Snydal, S. (2014) 'Major changes in diagnosis and management of preeclampsia', Journal of Midwifery & Women's Health, 59(6), pp. 596–605. Available at: https://doi.org/10.1111/jmwh.12260.

Socha, M.W., Stankiewicz, M., Żołnieżewicz, K., Puk, O. and Wartęga, M. (2022) ‘Decrease in Nitric Oxide Production as a Key Mediator in the Pathogenesis of Preeclampsia and a Potential Therapeutic Target: A Case-Control Study’, Biomedicines, 10(10), p. 2653. Available at: https://doi.org/10.3390/biomedicines10102653.

Soneson, C. and Delorenzi, M. (2013) 'A comparison of methods for differential expression analysis of RNA-seq data', BMC bioinformatics, 14, pp. 91. Available at: https://doi.org/10.1186/1471-2105-14-91.

Soongsatitanon, A. and Phupong, V. (2022) 'Prediction of preeclampsia using first trimester placental protein 13 and uterine artery doppler', The Journal of Maternal-Fetal & Neonatal Medicine, 35(22), pp. 4412–4417.

Sotiriou, C. and Pusztai, L. (2009) 'Gene-expression signatures in breast cancer', New England Journal of Medicine, 360(8), pp. 790–800.

Sovio, U., McBride, N., Wood, A.M., Masconi, K.L., Cook, E., Gaccioli, F., Charnock-Jones, D.S., Lawlor, D.A. and Smith, G.C.S. (2020) ‘4-Hydroxyglutamate is a novel predictor of pre-eclampsia’, International Journal of Epidemiology, 49(1), pp. 301–311. Available at: https://doi.org/10.1093/ije/dyz098.

Spencer, C.A. and Groudine, M. (1990) 'Transcription elongation and eukaryotic gene regulation.', Oncogene, 5(6), pp. 777–785.

Spradley, F.T., Palei, A.C. and Granger, J.P. (2015) 'Increased risk for the development of preeclampsia in obese pregnancies: Weighing in on the mechanisms', American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 309(11), pp. R1326–R1343.

Sprang, M., Andrade-Navarro, M.A. and Fontaine, J. (2022) 'Batch effect detection and correction in RNA-seq data using machine-learning-based automated assessment of quality', BMC bioinformatics, 23(6), pp. 1–15.

Srivastava, A., Malik, L., Sarkar, H., Zakeri, M., Almodaresi, F., Soneson, C., Love, M.I., Kingsford, C. and Patro, R. (2020) 'Alignment and mapping methodology influence transcript abundance estimation', Genome biology, 21(1), pp. 1–29.

Staff, A.C. (2019) 'The two-stage placental model of preeclampsia: An update', Journal of reproductive immunology, 134, pp. 1–10.

Staff, A.C., Fjeldstad, H.E., Fosheim, I.K., Moe, K., Turowski, G., Johnsen, G.M., Alnaes-Katjavivi, P. and Sugulle, M. (2022) 'Failure of physiological transformation and spiral artery atherosis: Their roles in preeclampsia', American Journal of Obstetrics and Gynecology, 226(2), pp. S895–S906.

Stanley, K.E., Jatsenko, T., Tuveri, S., Sudhakaran, D., Lannoo, L., Van Calsteren, K., de Borre, M., Van Parijs, I., Van Coillie, L. and Van Den Bogaert, K. (2024) 'Cell type signatures in cell-free DNA fragmentation profiles reveal disease biology', Nature Communications, 15(1), pp. 2220.

Starodubtseva, N., Poluektova, A., Tokareva, A., Kukaev, E., Avdeeva, A., Rimskaya, E. and Khodzayeva, Z. (2025) ‘Proteome-Based Maternal Plasma and Serum Biomarkers for Preeclampsia: A Systematic Review and Meta-Analysis’, Life, 15(5), p. 776. Available at: https://doi.org/10.3390/life15050776.

Steinberg, G., Khankin, E.V. and Karumanchi, S.A. (2009) 'Angiogenic factors and preeclampsia', Thrombosis research, 123, pp. S93–S99.

Stepan, H., Galindo, A., Hund, M., Schlembach, D., Sillman, J., Surbek, D. and Vatish, M. (2023) ‘Clinical utility of sFlt-1 and PlGF in screening, prediction, diagnosis and monitoring of pre-eclampsia and fetal growth restriction’, Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology, 61(2), pp. 168–180. Available at: https://doi.org/10.1002/uog.26032.

Stepan, H., Hund, M. and Andraczek, T. (2020) ‘Combining Biomarkers to Predict Pregnancy Complications and Redefine Preeclampsia’, Hypertension, 75(4), pp. 918–926. Available at: https://doi.org/10.1161/HYPERTENSIONAHA.119.13763.

Steyerberg, E., Vickers, A., Cook, N., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. and Kattan, M. (2010) 'Assessing the performance of prediction models A framework for traditional and novel measures', Epidemiology (Cambridge, Mass.), 21, pp. 128–38. Available at: https://doi.org/10.1097/EDE.0b013e3181c30fb2.

Stuart, R. and Peter, N. (2016) 'Artificial intelligence-a modern approach 3rd ed', .

Stupnikov, A., McInerney, C.E., Savage, K.I., McIntosh, S.A., Emmert-Streib, F., Kennedy, R., Salto-Tellez, M., Prise, K.M. and McArt, D.G. (2021) 'Robustness of

differential gene expression analysis of RNA-seq', Computational and structural biotechnology journal, 19, pp. 3470–3481.

Sufriyana, H., Salim, H.M., Muhammad, A.R., Wu, Y.-W. and Su, E.C.-Y. (2022) ‘Blood biomarkers representing maternal-fetal interface tissues used to predict early-and late-onset preeclampsia but not COVID-19 infection’, Computational and Structural Biotechnology Journal, 20, pp. 4206–4224. Available at: https://doi.org/10.1016/j.csbj.2022.08.011.

Sufriyana, H., Wu, Y. and Su, E.C. (2020b) 'Prediction of preeclampsia and intrauterine growth restriction: Development of machine learning models on a prospective cohort', JMIR medical informatics, 8(5), pp. e15411.

Sufriyana, H., Wu, Y.-W. and Su, E.C.-Y. (2020a) ‘Artificial intelligence-assisted prediction of preeclampsia: Development and external validation of a nationwide health insurance dataset of the BPJS Kesehatan in Indonesia’, EBioMedicine, 54, p. 102710. Available at: https://doi.org/10.1016/j.ebiom.2020.102710.

Sullivan, G.M. and Feinn, R. (2012) 'Using effect size—or why the P value is not enough', Journal of graduate medical education, 4(3), pp. 279–282.

Sun, S., Li, W., Zhang, X., Aziz, A. ur R. and Zhang, N. (2025) ‘Trends in global and regional incidence and prevalence of hypertensive disordersin pregnancy (1990–2021): an age-period-cohort analysis’, Scientific Reports, 15(1), p. 1513. Available at: https://doi.org/10.1038/s41598-025-85819-0.

Suresh, K.P. and Chandrashekara, S. (2012) 'Sample size estimation and power analysis for clinical research studies', Journal of human reproductive sciences, 5(1), pp. 7.

Suzumori, N., Ebara, T., Yamada, T., Samura, O., Yotsumoto, J., Nishiyama, M., Miura, K., Sawai, H., Murotsuki, J. and Kitagawa, M. (2016) 'Fetal cell-free DNA fraction in maternal plasma is affected by fetal trisomy', Journal of human genetics, 61(7), pp. 647–652.

Suzumori, N., Sekizawa, A., Ebara, T., Samura, O., Sasaki, A., Akaishi, R., Wada, S., Hamanoue, H., Hirahara, F. and Izumi, H. (2018) 'Fetal cell-free DNA fraction in maternal plasma for the prediction of hypertensive disorders of pregnancy', European Journal of Obstetrics & Gynecology and Reproductive Biology, 224, pp. 165–169.

Szczepanski, J., Griffin, A., Novotny, S. and Wallace, K. (2020) 'Acute kidney injury in pregnancies complicated with preeclampsia or HELLP syndrome', Frontiers in medicine, 7, pp. 22.

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A. and Bork, P. (2016) 'The STRING database in 2017: Quality-controlled protein–protein association networks, made broadly accessible', Nucleic acids research, , pp. gkw937.

Taglauer, E.S., Wilkins-Haug, L. and Bianchi, D.W. (2014) 'Review: Cell-free fetal DNA in the maternal circulation as an indication of placental health and disease', Placenta, 35(Suppl), pp. S64–S68. Available at: https://doi.org/10.1016/j.placenta.2013.11.014.

Taher, F., Ali, M. and Otman, G. (2022) 'Knowledge about Preeclampsia among Women of Reproductive Age Attending Al-Beyda Medical Center', AlQalam Journal of Medical and Applied Sciences, pp. 242–249.

Tahir, M., Badriyah, T. and Syarif, I. (2018a) 'Classification Algorithms of Maternal Risk Detection For Preeclampsia With Hypertension During Pregnancy Using Particle Swarm Optimization', EMITTER International Journal of Engineering Technology, 6(2), pp. 236–253. Available at: https://doi.org/10.24003/emitter.v6i2.287.

Tahir, M., Badriyah, T. and Syarif, I. (2018b) 'Neural Networks Algorithm to Inquire Previous Preeclampsia Factors in Women with Chronic Hypertension During Pregnancy in Childbirth Process', in 2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC). 2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC), pp. 51–55. Available at: https://doi.org/10.1109/KCIC.2018.8628588.

Takabayashi, H., Kuwabara, S., Ukita, T., Ikawa, K., Yamafuji, K. and Igarashi, T. (1995) 'Development of non-invasive fetal DNA diagnosis from maternal blood', Prenatal Diagnosis, 15(1), pp. 74–77. Available at: https://doi.org/10.1002/pd.1970150116.

Tan, M.Y., Syngelaki, A., Poon, L.C., Rolnik, D.L., O'Gorman, N., Delgado, J.L., Akolekar, R., Konstantinidou, L., Tsavdaridou, M., Galeva, S., Ajdacka, U., Molina, F.S., Persico, N., Jani, J.C., Plasencia, W., Greco, E., Papaioannou, G., Wright, A., Wright, D. and Nicolaides, K.H. (2018) 'Screening for pre-eclampsia by maternal factors and biomarkers at 11-13 weeks' gestation', Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology, 52(2), pp. 186–195. Available at: https://doi.org/10.1002/uog.19112.

Tanaka, M., Tsujimoto, Y., Goto, K., Kumahara, K., Onishi, S., Iwanari, S., Fumihara, D., Miki, S., Ikeda, M., Sato, K., Sato, H., Hirose, M. and Takeoka, H. (2014) 'Preeclampsia before 20 weeks of gestation: a case report and review of the literature', CEN Case Reports, 4(1), pp. 55–60. Available at: https://doi.org/10.1007/s13730-014-0140-3.

Tang, Z., Wang, S., Li, X., Hu, C., Zhai, Q., Wang, J., Ye, Q., Liu, J., Zhang, G., Guo, Y., Su, F., Liu, H., Guan, L., Jiang, C., Chen, J., Li, M., Ren, F., Zhang, Y., Huang, M., Li, L., Zhang, H., Hou, G., Jin, X., Chen, F., Zhu, H., Li, L., Zeng, J., Xiao, H., Zhou, A., Feng, L., Gao, Y. and Liu, G. (2024) 'Longitudinal integrative cell-free DNA analysis in gestational diabetes mellitus', Cell Reports

Medicine, 5(8), pp. 101660. Available at: https://doi.org/10.1016/j.xcrm.2024.101660.

Tannetta, D. and Sargent, I. (2013) 'Placental disease and the maternal syndrome of preeclampsia: Missing links?', Current hypertension reports, 15, pp. 590–599.

Tarca, A.L., Romero, R., Benshalom-Tirosh, N., Than, N.G., Gudicha, D.W., Done, B., Pacora, P., Chaiworapongsa, T., Panaitescu, B., Tirosh, D., Gomez-Lopez, N., Draghici, S., Hassan, S.S. and Erez, O. (2019) 'The prediction of early preeclampsia: Results from a longitudinal proteomics study', PloS One, 14(6), p. e0217273. Available at: https://doi.org/10.1371/journal.pone.0217273.

Tarca, A.L., Romero, R., Erez, O., Gudicha, D.W., Than, N.G., Benshalom-Tirosh, N., Pacora, P., Hsu, C., Chaiworapongsa, T., Hassan, S.S. and Gomez-Lopez, N. (2021) 'Maternal whole blood mRNA signatures identify women at risk of early preeclampsia: A longitudinal study', The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstetricians, 34(21), pp. 3463–3474. Available at: https://doi.org/10.1080/14767058.2019.1685964.

Tarca, A.L., Romero, R., Xu, Z., Gomez-Lopez, N., Erez, O., Hsu, C., Hassan, S.S. and Carey, V.J. (2019) 'Targeted expression profiling by RNA-seq improves detection of cellular dynamics during pregnancy and identifies a role for T cells in term parturition', Scientific Reports, 9(1), pp. 848. Available at: https://doi.org/10.1038/s41598-018-36649-w.

Tarca, A.L., Taran, A., Romero, R., Jung, E., Paredes, C., Bhatti, G., Ghita, C., Chaiworapongsa, T., Than, N.G. and Hsu, C. (2022) 'Prediction of preeclampsia throughout gestation with maternal characteristics and biophysical and biochemical markers: A longitudinal study', American Journal of Obstetrics and Gynecology, 226(1), pp. 126.e1–126.e22. Available at: https://doi.org/10.1016/j.ajog.2021.01.020.

TARCA, A.L., TARAN, A., ROMERO, R., JUNG, E.J., PAREDES, C., BHATTI, G., GHITA, C., CHAIWORAPONGSA, T., THAN, N.G. and HSU, C.-D. (2022) 'Prediction of preeclampsia throughout gestation with maternal characteristics and biophysical and biochemical markers: a longitudinal study', American journal of obstetrics and gynecology, 226(1), p. 126.e1-126.e22. Available at: https://doi.org/10.1016/j.ajog.2021.01.020.

Taylor, R.N., Grimwood, J., Taylor, R.S., McMaster, M.T., Fisher, S.J. and North, R.A. (2003) 'Longitudinal serum concentrations of placental growth factor: Evidence for abnormal placental angiogenesis in pathologic pregnancies', American Journal of Obstetrics and Gynecology, 188(1), pp. 177–182. Available at: https://doi.org/10.1067/mob.2003.111.

Taylor-Phillips, S., Freeman, K., Geppert, J., Agbebiyi, A., Uthman, O.A., Madan, J., Clarke, A., Quenby, S. and Clarke, A. (2016) 'Accuracy of non-invasive prenatal testing using cell-free DNA for detection of down, edwards and patau syndromes: A systematic review and meta-analysis', BMJ open, 6(1), pp. e010002.

Tersigni, C., Meli, F., Neri, C., Iacoangeli, A., Franco, R., Lanzone, A., Scambia, G. and Di Simone, N. (2020) 'Role of Human Leukocyte Antigens at the Feto-Maternal Interface in Normal and Pathological Pregnancy: An Update', International Journal of Molecular Sciences, 21(13), p. 4756. Available at: https://doi.org/10.3390/ijms21134756.

Tessema, G.A., Marinovich, M.L., Håberg, S.E., Gissler, M., Mayo, J.A., Nassar, N., Ball, S., Betrán, A.P., Gebremedhin, A.T. and de Klerk, N. (2021) 'Interpregnancy intervals and adverse birth outcomes in high-income countries: An international cohort study', PLoS One, 16(7), pp. e0255000.

Textoris, J., Ivorra, D., Ben Amara, A., Sabatier, F., Ménard, J.-P., Heckenroth, H., Bretelle, F. and Mege, J.-L. (2013) 'Evaluation of Current and New Biomarkers in Severe Preeclampsia: A Microarray Approach Reveals the VSIG4 Gene as a Potential Blood Biomarker', PLoS ONE, 8(12), p. e82638. Available at: https://doi.org/10.1371/journal.pone.0082638.

Thadhani, R., Mutter, W.P., Wolf, M., Levine, R.J., Taylor, R.N., Sukhatme, V.P., Ecker, J. and Karumanchi, S.A. (2004) 'First trimester placental growth factor and soluble fms-like tyrosine kinase 1 and risk for preeclampsia', The Journal of Clinical Endocrinology & Metabolism, 89(2), pp. 770–775.

Than, N.G., Romero, R., Posta, M., Györffy, D., Rossi, S.W., Szilágyi, A., Hupuczi, P., Nagy, S., Török, O., Tarca, A.L., Erez, O., Ács, N. and Papp, Z. (2024) 'Classification of preeclampsia according to molecular clusters with the goal of achieving personalized prevention', Journal of reproductive immunology, 161, p. 104172. Available at: https://doi.org/10.1016/j.jri.2023.104172.

Thomas, P.D. (2017) 'The gene ontology and the meaning of biological function', The gene ontology handbook, , pp. 15–24.

Tian, Y., Liu, M., Sun, J.-Y., Wang, Y., Chen, L., Sun, W. and Zhou, L. (2024) 'Diagnosis of preeclampsia using metabolomic biomarkers', Hypertension in Pregnancy, 43(1), p. 2379386. Available at: https://doi.org/10.1080/10641955.2024.2379386.

Tikhonov, D., Kulikova, L., Rudnev, V., Kopylov, A.T., Taldaev, A., Stepanov, A., Malsagova, K., Izotov, A., Enikeev, D. and Potoldykova, N. (2021) 'Changes in protein structural motifs upon post-translational modification in kidney cancer', Diagnostics, 11(10), pp. 1836.

Tjoa, M.L., Van Vugt, J., Go, A., Blankenstein, M.A., Oudejans, C. and Van Wijk, I.J. (2003) 'Elevated C-reactive protein levels during first trimester of pregnancy

are indicative of preeclampsia and intrauterine growth restriction', Journal of reproductive immunology, 59(1), pp. 29–37.

Tjoa, M.L., van Vugt, J.M., Mulders, M.A., Schutgens, R.B., Oudejans, C.B. and van Wijk, I.J. (2001) 'Plasma placenta growth factor levels in midtrimester pregnancies', Obstetrics & Gynecology, 98(4), pp. 600–607.

Togneri, F.S., Kilby, M.D., Young, E., Williams, D., Griffiths, M.J. and Allen, S.K. (2019) 'Implementation of cell-free DNA-based non-invasive prenatal testing in a national health service regional genetics laboratory', Genetics Research, 101, pp. e11.

Tomczak, A., Mortensen, J.M., Winnenburg, R., Liu, C., Alessi, D.T., Swamy, V., Vallania, F., Lofgren, S., Haynes, W. and Shah, N.H. (2018) 'Interpretation of biological experiments changes with evolution of the gene ontology and its annotations', Scientific reports, 8(1), pp. 5115.

Tomkiewicz, J. and Darmochwał-Kolarz, D.A. (2024) ‘Biomarkers for Early Prediction and Management of Preeclampsia: A Comprehensive Review’, Medical Science Monitor : International Medical Journal of Experimental and Clinical Research, 30, pp. e944104-1-e944104-11. Available at: https://doi.org/10.12659/MSM.944104.

Torres-García, W., Zheng, S., Sivachenko, A., Vegesna, R., Wang, Q., Yao, R., Berger, M.F., Weinstein, J.N., Getz, G. and Verhaak, R.G. (2014) 'PRADA: Pipeline for RNA sequencing data analysis', Bioinformatics, 30(15), pp. 2224–2226.

Torres-Torres, J., Villafan-Bernal, J.R., Martinez-Portilla, R.J., Hidalgo-Carrera, J.A., Estrada-Gutierrez, G., Adalid-Martinez-Cisneros, R., Rojas-Zepeda, L., Acevedo-Gallegos, S., Camarena-Cabrera, D.M. and Cruz-Martínez, M.Y. (2024) 'Performance of machine-learning approach for prediction of pre-eclampsia in a middle-income country', Ultrasound in Obstetrics & Gynecology, 63(3), pp. 350–357.

Torres-Vergara, P., Rivera, R. and Escudero, C. (2022) 'How soluble fms-like tyrosine kinase 1 could contribute to blood-brain barrier dysfunction in preeclampsia?', Frontiers in physiology, 12, pp. 2565.

Tranquilli, A.L., Dekker, G., Magee, L., Roberts, J., Sibai, B.M., Steyn, W., Zeeman, G.G. and Brown, M.A. (2014) 'The classification, diagnosis and management of the hypertensive disorders of pregnancy: A revised statement from the ISSHP', Pregnancy Hypertension: An International Journal of Women's Cardiovascular Health, 4(2), pp. 97–104. Available at: https://doi.org/10.1016/j.preghy.2014.02.001.

Trapnell, C. and Salzberg, S.L. (2009) 'How to map billions of short reads onto genomes', Nature Biotechnology, 27(5), pp. 455–457. Available at: https://doi.org/10.1038/nbt0509-455.

Trogstad, L., Magnus, P. and Stoltenberg, C. (2011) 'Pre-eclampsia: Risk factors and causal models', Best practice & research Clinical obstetrics & gynaecology, 25(3), pp. 329–342.

Trogstad, L.I., Eskild, A., Magnus, P., Samuelsen, S.O. and Nesheim, B. (2001) 'Changing paternity and time since last pregnancy; the impact on pre-eclampsia risk. A study of 547 238 women with and without previous pre-eclampsia', International journal of epidemiology, 30(6), pp. 1317–1322.

Tsang, J.C., Vong, J.S., Ji, L., Poon, L.C., Jiang, P., Lui, K.O., Ni, Y., To, K.F., Cheng, Y.K. and Chiu, R.W. (2017) 'Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics', Proceedings of the National Academy of Sciences, 114(37), pp. E7786–E7795.

Tsang, J.C.H., Vong, J.S.L., Ji, L., Poon, L.C.Y., Jiang, P., Lui, K.O., Ni, Y.-B., To, K.F., Cheng, Y.K.Y., Chiu, R.W.K. and Lo, Y.M.D. (2017) 'Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics', Proceedings of the National Academy of Sciences of the United States of America, 114(37), pp. E7786–E7795. Available at: https://doi.org/10.1073/pnas.1710470114.

Tsui, N., Chim, S., Chiu, R., Lau, T.K., Ng, E., Leung, T.N., Tong, Y.K., Chan, K. and Lo, Y. (2004) 'Systematic micro-array based identification of placental mRNA in maternal plasma: Towards non-invasive prenatal gene expression profiling', Journal of medical genetics, 41(6), pp. 461–467.

Tsui, N.B., Jiang, P., Wong, Y.F., Leung, T.Y., Chan, K.A., Chiu, R.W., Sun, H. and Lo, Y.D. (2014) 'Maternal plasma RNA sequencing for genome-wide transcriptomic profiling and identification of pregnancy-associated transcripts', Clinical chemistry, 60(7), pp. 954–962.

Tsui, N.B., Kadir, R.A., Chan, K.A., Chi, C., Mellars, G., Tuddenham, E.G., Leung, T.Y., Lau, T.K., Chiu, R.W. and Lo, Y.D. (2011) 'Noninvasive prenatal diagnosis of hemophilia by microfluidics digital PCR analysis of maternal plasma DNA', Blood, The Journal of the American Society of Hematology, 117(13), pp. 3684–3691.

Turco, M.Y. and Moffett, A. (2019) 'Development of the human placenta', Development (Cambridge, England), 146(22), p. dev163428. Available at: https://doi.org/10.1242/dev.163428.

Turlier, H. and Maître, J. (2015) 'Mechanics of tissue compaction', Seminars in cell & developmental biology, 47-48, pp. 110–117. Available at: https://doi.org/10.1016/j.semcdb.2015.08.001.

Tyrmi, J.S., Kaartokallio, T., Lokki, A.I., Jääskeläinen, T., Kortelainen, E., Ruotsalainen, S., Karjalainen, J., Ripatti, S., Kivioja, A., Laisk, T., Kettunen, J., Pouta, A., Kivinen, K., Kajantie, E., Heinonen, S., Kere, J., Laivuori, H. and FINNPEC Study Group, F.P., and the Estonian Biobank Research Team (2023)

‘Genetic Risk Factors Associated With Preeclampsia and Hypertensive Disorders of Pregnancy’, JAMA Cardiology, 8(7), pp. 674–683. Available at: https://doi.org/10.1001/jamacardio.2023.1312.

Uddin, S., Khan, A., Hossain, M.E. and Moni, M.A. (2019) 'Comparing different supervised machine learning algorithms for disease prediction', BMC medical informatics and decision making, 19(1), pp. 1–16.

Umer, M., Imtiaz, Z., Ullah, D.S., Mehmood, A., Choi, G.S. and On, B. (2020) 'Fake news stance detection using deep learning architecture (CNN-LSTM)', IEEE Access, PP, pp. 1–1. Available at: https://doi.org/10.1109/ACCESS.2020.3019735.

Uptain, S.M., Kane, C.M. and Chamberlin, M.J. (1997) 'Basic mechanisms of transcript elongation and its regulation', Annual Review of Biochemistry, 66(1), pp. 117–172.

Ura, B., Feriotto, G., Monasta, L., Bilel, S., Zweyer, M. and Celeghini, C. (2014) 'Potential role of circulating microRNAs as early markers of preeclampsia', Taiwanese Journal of Obstetrics and Gynecology, 53(2), pp. 232–234.

Urquia, M., Glazier, R., Gagnon, A., Mortensen, L., Nybo Andersen, A.-M., Janevic, T., Guendelman, S., Thornton, D., Bolumar, F., Río Sánchez, I., Small, R., Davey, M.-A. and Hjern, A. (2014) ‘Disparities in pre-eclampsia and eclampsia among immigrant women giving birth in six industrialised countries’, Bjog, 121(12), pp. 1492–1500. Available at: https://doi.org/10.1111/1471-0528.12758.

Urquia, M.L., Glazier, R.H., Gagnon, A.J., Mortensen, L.H., Nybo Andersen, A., Janevic, T., Guendelman, S., Thornton, D., Bolumar, F. and Rio Sanchez, I. (2014) 'Disparities in pre-eclampsia and eclampsia among immigrant women giving birth in six industrialised countries', BJOG: An International Journal of Obstetrics & Gynaecology, 121(12), pp. 1492–1500.

Uusküla, L., Männik, J., Rull, K., Minajeva, A., Kõks, S., Vaas, P., Teesalu, P., Reimand, J. and Laan, M. (2012) ‘Mid-Gestational Gene Expression Profile in Placenta and Link to Pregnancy Complications’, PLOS ONE, 7(11), p. e49248. Available at: https://doi.org/10.1371/journal.pone.0049248.

Uzun, A., Triche, E.W., Schuster, J., Dewan, A.T. and Padbury, J.F. (2016) ‘dbPEC: a comprehensive literature-based database for preeclampsia related genes and phenotypes’, Database, 2016, p. baw006. Available at: https://doi.org/10.1093/database/baw006.

Valdes, G., Luna, J.M., Eaton, E., Simone, C.B., Ungar, L.H. and Solberg, T.D. (2016) 'MediBoost: A patient stratification tool for interpretable decision making in the era of precision medicine', Scientific reports, 6(1), pp. 37854.

Valenzuela, F.J., Pérez-Sepúlveda, A., Torres, M.J., Correa, P., Repetto, G.M. and Illanes, S.E. (2011) 'Pathogenesis of preeclampsia: The genetic component', Journal of Pregnancy, 2012, pp. e632732. Available at: https://doi.org/10.1155/2012/632732.

van der Krogt, L., Ridout, A., Vousden, N., Nathan, H.L., Seed, P.T., Midwife, B.S., Midwife, M.M., Sesay, T., Smart, F., Bellad, M., Vwalika, B. and Shennan, A. (2025) 'Eclampsia among adolescent mothers in low- and middle-Income countries', PLOS Global Public Health, 5(3), p. e0003890. Available at: https://doi.org/10.1371/journal.pgph.0003890.

Van, R., Alvarez, D., Mize, T., Gannavarapu, S., Chintham Reddy, L., Nasoz, F. and Han, M.V. (2024) 'A comparison of RNA-seq data preprocessing pipelines for transcriptomic predictions across independent studies', BMC bioinformatics, 25(1), pp. 181.

Vandepitte, I. (2022) 'No title', The early detection of preeclampsia and intra-uterine growth restriction by noninvasive prenatal testing, .

Varet, H., Brillet-Guéguen, L., Coppée, J. and Dillies, M. (2016) 'SARTools: A DESeq2-and EdgeR-based R pipeline for comprehensive differential analysis of RNA-seq data', PloS one, 11(6), pp. e0157022.

Varughese, B., Bhatla, N., Kumar, R., Dwivedi, S.N. and Dhingra, R. (2010) 'Circulating angiogenic factors in pregnancies complicated by pre-eclampsia', National Medical Journal of India, 23(2), pp. 77.

Venkatesha, S., Toporsian, M., Lam, C., Hanai, J., Mammoto, T., Kim, Y.M., Bdolah, Y., Lim, K., Yuan, H., Libermann, T.A., Stillman, I.E., Roberts, D., D'Amore, P.A., Epstein, F.H., Sellke, F.W., Romero, R., Sukhatme, V.P., Letarte, M. and Karumanchi, S.A. (2006) 'Soluble endoglin contributes to the pathogenesis of preeclampsia', Nature Medicine, 12(6), pp. 642–649. Available at: https://doi.org/10.1038/nm1429.

Vennou, K.E., Kontou, P.I., Braliou, G.G. and Bagos, P.G. (2020) 'Meta-analysis of gene expression profiles in preeclampsia', Pregnancy Hypertension, 19, pp. 52–60.

Verlohren, S., Stepan, H. and Dechend, R. (2012) 'Angiogenic growth factors in the diagnosis and prediction of pre-eclampsia', Clinical science, 122(2), pp. 43–52.

Villalaín, C., Herraiz, I., Olmo, D., Angulo, P., Ayala, J.L. and Galindo, A. (2022) 'Prediction of delivery within 7 days after diagnosis of early onset preeclampsia using machine-learning models', Frontiers in Cardiovascular Medicine, 9, pp. 910701.

Vliet, M.M. van, Schoenmakers, S., Boers, R.G., Meeren, L.E. van der, Gribnau, J. and Steegers-Theunissen, R.P.M. (2025) 'Methylome profiling of cell-free DNA during the early life course in (un)complicated pregnancies using MeD-seq:

Protocol for a cohort study embedded in the prospective Rotterdam periconception cohort', PLOS ONE, 20(1), p. e0310019. Available at: https://doi.org/10.1371/journal.pone.0310019.

Volgin, D.V. (2014) 'Gene expression: Analysis and quantitation'Animal biotechnology Elsevier, pp. 307–325. .

Volik, S., Alcaide, M., Morin, R.D. and Collins, C. (2016) 'Cell-free DNA (cfDNA): Clinical significance and utility in cancer shaped by emerging technologies', Molecular Cancer Research, 14(10), pp. 898–908.

von Dadelszen, P., Payne, B., Li, J., Ansermino, J.M., Pipkin, F.B., Côté, A., Douglas, M.J., Gruslin, A., Hutcheon, J.A. and Joseph, K.S. (2011) 'Prediction of adverse maternal outcomes in pre-eclampsia: Development and validation of the fullPIERS model', The Lancet, 377(9761), pp. 219–227.

Vujović, Ž (2021) 'Classification model evaluation metrics', International Journal of Advanced Computer Science and Applications, 12(6), pp. 599–606.

Wade, J.T. and Struhl, K. (2008) 'The transition from transcriptional initiation to elongation', Current opinion in genetics & development, 18(2), pp. 130–136.

Wagner, L.K. (2004) 'Diagnosis and management of preeclampsia', American Family Physician, 70(12), pp. 2317–2324.

Wainstock, T. and Sheiner, E. (2022) 'Clinical factors associated with preeclampsia recurrence', Pregnancy Hypertension, 30, pp. 31–35. Available at: https://doi.org/10.1016/j.preghy.2022.08.004.

Walker, J.J. (2000) 'Pre-eclampsia', The Lancet, 356(9237), pp. 1260–1265.

Walsh, N., Cooper, A., Dockery, A. and O'Byrne, J.J. (2024) 'Variant reclassification and clinical implications', Journal of Medical Genetics, 61(3), pp. 207–211. Available at: https://doi.org/10.1136/jmg-2023-109488.

Walsh, S.W., Vaughan, J.E., Wang, Y. and Roberts, L.J. (2000) 'Placental isoprostane is significantly increased in preeclampsia', FASEB journal: official publication of the Federation of American Societies for Experimental Biology, 14(10), pp. 1289–1296. Available at: https://doi.org/10.1096/fj.14.10.1289.

Wang, A., Rana, S. and Karumanchi, S.A. (2009) 'Preeclampsia: The role of angiogenic factors in its pathogenesis', Physiology, 24(3), pp. 147–158.

Wang, B. and Luan, Y. (2023) 'Evaluation of normalization methods for predicting quantitative phenotypes in metagenomic data analysis', bioRxiv, , pp. 2023.10. 25.563937.

Wang, E., Batey, A., Struble, C., Musci, T., Song, K. and Oliphant, A. (2013) 'Gestational age and maternal weight effects on fetal cell-free DNA in maternal plasma', Prenatal diagnosis, 33(7), pp. 662–666.

Wang, G., Yin, H., Li, B., Yu, C., Wang, F., Xu, X., Cao, J., Bao, Y., Wang, L. and Abbasi, A.A. (2019) 'Characterization and identification of long non-coding RNAs based on feature relationship', Bioinformatics, 35(17), pp. 2949–2956.

Wang, H., Zhang, Z., Li, H., Li, J., Li, H., Liu, M., Liang, P., Xi, Q., Xing, Y. and Yang, L. (2023) 'A cost-effective machine learning-based method for preeclampsia risk assessment and driver genes discovery', Cell & bioscience, 13(1), pp. 41.

Wang, J., Wang, Jingxuan, Guan, Y., Liang, J., Tang, D., Yin, T. and Diao, L. (2025) 'Trends and regional disparities in maternal hypertensive disorders among women of childbearing age: a global burden of disease analysis from 1990 to 2021', BMC Pregnancy and Childbirth, 25, p. 812. Available at: https://doi.org/10.1186/s12884-025-07926-0.

Wang, L., Wang, S. and Li, W. (2012) 'RSeQC: Quality control of RNA-seq experiments', Bioinformatics, 28(16), pp. 2184–2185.

Wang, L.Q., Fernandez-Boyano, I. and Robinson, W.P. (2022) 'Genetic variation in placental insufficiency: What have we learned over time?', Frontiers in Cell and Developmental Biology, 10, pp. 1038358. Available at: https://doi.org/10.3389/fcell.2022.1038358.

Wang, Q., Feng, W., Tan, Y., Qiao, J., Liu, H., Liu, Q., Wang, H., Zhang, Q., Tao, J., Li, Z., Yang, B., Xu, Z., Wang, C., Yang, S. and Cui, L. (2025) 'Antiphospholipid antibodies inhibit the migration and invasion of trophoblast cells by suppressing the JNK/C-Jun/MMP1 signaling pathway', Journal of Translational Medicine, 23, p. 581. Available at: https://doi.org/10.1186/s12967-025-06596-y.

Wang, S.S., Lewis, M.J. and Pitzalis, C. (2023) 'DNA methylation signatures of response to conventional synthetic and biologic disease-modifying antirheumatic drugs (DMARDs) in rheumatoid arthritis', Biomedicines, 11(7), pp. 1987.

Wang, W. (2024) 'Ethical issues in personalized medicine: Privacy, consent, and data sharing', GMO Biosafety Research, 15.

Wang, W., Liu, Z., Peng, D., Lin, G.N. and Wang, Z. (2024) 'Genomic insights into genes expressed specifically during infancy highlight their dominant influence on the neuronal system', BMC Genomics, 25(1), pp. 1012. Available at: https://doi.org/10.1186/s12864-024-10911-0.

Wang, W., Xie, X., Yuan, T., Wang, Y., Zhao, F., Zhou, Z. and Zhang, H. (2021) 'Epidemiological trends of maternal hypertensive disorders of pregnancy at the global, regional, and national levels: a population-based study', BMC Pregnancy and Childbirth, 21, p. 364. Available at: https://doi.org/10.1186/s12884-021-03809-2.

Wang, Y., Liu, J., Huang, B.O., Xu, Y., Li, J., Huang, L., Lin, J., Zhang, J., Min, Q. and Yang, W. (2015) 'Mechanism of alternative splicing and its regulation', Biomedical reports, 3(2), pp. 152–158.

Wang, Y., Wu, N. and Shen, H. (2021) 'A Review of Research Progress of Pregnancy with Twins with Preeclampsia', Risk Management and Healthcare Policy, 14, pp. 1999–2010. Available at: https://doi.org/10.2147/RMHP.S304040.

Wang, Z., Gerstein, M. and Snyder, M. (2009) 'RNA-seq: A revolutionary tool for transcriptomics', Nature reviews genetics, 10(1), pp. 57–63.

Wang, Z., Lyu, Z., Pan, L., Zeng, G. and Randhawa, P. (2019) 'Defining housekeeping genes suitable for RNA-seq analysis of the human allograft kidney biopsy tissue', BMC Medical Genomics, 12(1), pp. 86. Available at: https://doi.org/10.1186/s12920-019-0538-z.

Wang, Z., Monteiro, C.D., Jagodnik, K.M., Fernandez, N.F., Gundersen, G.W., Rouillard, A.D., Jenkins, S.L., Feldmann, A.S., Hu, K.S., McDermott, M.G., Duan, Q., Clark, N.R., Jones, M.R., Kou, Y., Goff, T., Woodland, H., Amaral, F.M.R., Szeto, G.L., Fuchs, O., Schüssler-Fiorenza Rose, S.M., Sharma, S., Schwartz, U., Bausela, X.B., Szymkiewicz, M., Maroulis, V., Salykin, A., Barra, C.M., Kruth, C.D., Bongio, N.J., Mathur, V., Todoric, R.D., Rubin, U.E., Malatras, A., Fulp, C.T., Galindo, J.A., Motiejunaite, R., Jüschke, C., Dishuck, P.C., Lahl, K., Jafari, M., Aibar, S., Zaravinos, A., Steenhuizen, L.H., Allison, L.R., Gamallo, P., de Andres Segura, F., Dae Devlin, T., Pérez-García, V. and Ma'ayan, A. (2016) 'Extraction and analysis of signatures from the gene expression omnibus by the crowd', Nature communications, 7(1), pp. 12846. Available at: https://doi.org/10.1038/ncomms12846.

Ward, K., Laivuori, H. and Taylor, R.N. (2022) 'Genetic factors in the etiology of preeclampsia/eclampsia'Chesley's hypertensive disorders in pregnancy Elsevier, pp. 45–69. .

Warden, C.D., Adamson, A.W., Neuhausen, S.L. and Wu, X. (2014) 'Detailed comparison of two popular variant calling packages for exome and targeted exon studies', PeerJ, 2, pp. e600. Available at: https://doi.org/10.7717/peerj.600.

Waszak, S.M., Kilpinen, H., Gschwind, A.R., Orioli, A., Raghav, S.K., Witwicki, R.M., Migliavacca, E., Yurovsky, A., Lappalainen, T., Hernandez, N., Reymond, A., Dermitzakis, E.T. and Deplancke, B. (2014) 'Identification and removal of low-complexity sites in allele-specific analysis of ChIP-seq data', Bioinformatics (Oxford, England), 30(2), pp. 165–171. Available at: https://doi.org/10.1093/bioinformatics/btt667.

Wataganara, T., LeShane, E.S., Farina, A., Messerlian, G.M., Lee, T., Canick, J.A. and Bianchi, D.W. (2003) 'Maternal serum cell-free fetal DNA levels are increased in cases of trisomy 13 but not trisomy 18', Human genetics, 112, pp. 204–208.

Webster, K., Fishburn, S., Maresh, M., Findlay, S.C. and Chappell, L.C. (2019) 'Diagnosis and management of hypertension in pregnancy: Summary of updated NICE guidance', BMJ (Clinical research ed.), 366, pp. l5119. Available at: https://doi.org/10.1136/bmj.l5119.

Wei, J., Fu, Y., Mao, X., Jing, Y., Guo, J. and Ye, Y. (2019) 'Decreased Filamin b expression regulates trophoblastic cells invasion through ERK/MMP-9 pathway in pre-eclampsia', Ginekologia Polska, 90(1), pp. 39–45. Available at: https://doi.org/10.5603/GP.2019.0006.

Wei, X., Li, Q., Qin, C., Zeng, W., Liu, Y., Lu, J., He, J., Chen, C., Zhang, X. and Lin, Y. (2025) 'Spatial metabolomics and transcriptomics reveal cell type-specific dynamics in the placenta of patients with late-onset preeclampsia', Frontiers in Cell and Developmental Biology, 13. Available at: https://doi.org/10.3389/fcell.2025.1659880.

Wiles, K., Chappell, L., Clark, K., Elman, L., Hall, M., Lightstone, L., Mohamed, G., Mukherjee, D., Nelson-Piercy, C. and Webster, P. (2019) 'Clinical practice guideline on pregnancy and renal disease', BMC nephrology, 20(1), pp. 1–43.

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. and Mons, B. (2016) 'The FAIR Guiding Principles for scientific data management and stewardship', Scientific Data, 3(1), p. 160018. Available at: https://doi.org/10.1038/sdata.2016.18.

Williams, P.J. and Broughton Pipkin, F. (2011) 'The genetics of pre-eclampsia and other hypertensive disorders of pregnancy', Best Practice & Research. Clinical Obstetrics & Gynaecology, 25(4–4), pp. 405–417. Available at: https://doi.org/10.1016/j.bpobgyn.2011.02.007.

Wingrove, J.A., Daniels, S.E., Sehnert, A.J., Tingley, W., Elashoff, M.R., Rosenberg, S., Buellesfeld, L., Grube, E., Newby, L.K. and Ginsburg, G.S. (2008) 'Correlation of peripheral-blood gene expression with the extent of coronary artery stenosis', Circulation: Cardiovascular Genetics, 1(1), pp. 31–38.

Woelkers, D., Barton, J., Dadelszen, P.v. and Sibai, B. (2015) '[71-OR]: The revised 2013 ACOG definitions of hypertensive disorders of pregnancy significantly increase the diagnostic prevalence of preeclampsia', Pregnancy Hypertension: An International Journal of Women's Cardiovascular Health, 5(1), pp. 38. Available at: https://doi.org/10.1016/j.preghy.2014.10.075.

Wójtowicz, A., Zembala-Szczerba, M., Babczyk, D., Kołodziejczyk-Pietruszka, M., Lewaczyńska, O. and Huras, H. (2019) 'Early- and late-onset preeclampsia: A comprehensive cohort study of laboratory and clinical findings according to the new ISHHP criteria', International Journal of Hypertension, 2019, pp. e4108271. Available at: https://doi.org/10.1155/2019/4108271.

Wolf, S., Melo, D., Garske, K.M., Pallares, L.F., Lea, A.J. and Ayroles, J.F. (2023) 'Characterizing the landscape of gene expression variance in humans', PLOS Genetics, 19(7), pp. e1010833. Available at: https://doi.org/10.1371/journal.pgen.1010833.

Wondji, C., Hemingway, J. and Ranson, H. (2007) 'Identification and analysis of single nucleotide polymorphisms (SNPs) in the mosquito anopheles funestus, malaria vector', BMC genomics, 8, pp. 5. Available at: https://doi.org/10.1186/1471-2164-8-5.

Wong, B.C. and Lo, Y.D. (2003) 'Cell-free DNA and RNA in plasma as new tools for molecular diagnostics', Expert Review of Molecular Diagnostics, 3(6), pp. 785–797.

Wong, D., Tageldein, M., Luo, P., Ensminger, E., Bruce, J., Oldfield, L., Gong, H., Fischer, N.W., Laverty, B., Subasri, V., Davidson, S., Khan, R., Villani, A., Shlien, A., Kim, R.H., Malkin, D. and Pugh, T.J. (2024) 'Cell-free DNA from germline TP53 mutation carriers reflect cancer-like fragmentation patterns', Nature Communications, 15(1), pp. 7386. Available at: https://doi.org/10.1038/s41467-024-51529-w.

Wong, Y.K., Chan, W.H., Nies, H.W. and Moorthy, K.A.-L. (2022) 'Multi-stage Feature Selection in Identifying Potential Biomarkers for Cancer Classification', in 2022 2nd International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA). 2022 2nd International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA), pp. 6–11. Available at: https://doi.org/10.1109/ICICyTA57421.2022.10037807.

Woolf, S.H., Grol, R., Hutchinson, A., Eccles, M. and Grimshaw, J. (1999) 'Potential benefits, limitations, and harms of clinical guidelines', BMJ : British Medical Journal, 318(7182), pp. 527–530.

Workman, J.L. and Kingston, R.E. (1998) 'Alteration of nucleosome structure as a mechanism of transcriptional regulation', Annual Review of Biochemistry, 67(1), pp. 545–579.

Wright, A., Wright, D., Syngelaki, A., Georgantis, A. and Nicolaides, K.H. (2019) 'Two-stage screening for preterm preeclampsia at 11–13 weeks' gestation', American Journal of Obstetrics & Gynecology, 220(2), pp. 197.e1–197.e11. Available at: https://doi.org/10.1016/j.ajog.2018.10.092.

Wright, C.F. and Burton, H. (2009) 'The use of cell-free fetal nucleic acids in maternal blood for non-invasive prenatal diagnosis', Human Reproduction

Update, 15(1), pp. 139–151. Available at: https://doi.org/10.1093/humupd/dmn047.

Wright, D., Syngelaki, A., Akolekar, R., Poon, L.C. and Nicolaides, K.H. (2015) 'Competing risks model in screening for preeclampsia by maternal characteristics and medical history', American Journal of Obstetrics and Gynecology, 213(1), pp. 62.e1–62.e10. Available at: https://doi.org/10.1016/j.ajog.2015.02.018.

Wright, M.L., Goin, D.E., Smed, M.K., Jewell, N.P., Nelson, J.L., Olsen, J., Hetland, M.L., Zoffmann, V. and Jawaheer, D. (2023) 'Pregnancy-associated systemic gene expression compared to a pre-pregnancy baseline, among healthy women with term pregnancies', Frontiers in immunology, 14, pp. 1161084.

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W. and Zhan, L.I. (2021) 'clusterProfiler 4.0: A universal enrichment tool for interpreting omics data', The innovation, 2(3).

Wu, Y., Werlang, A., Cheng, W., Lanes, A., Wen, S.W. and Walker, M. (2021) 'Association between levels of total cell-free DNA and development of Preeclampsia—A literature review', AJP Reports, 11(1), pp. e38–e48. Available at: https://doi.org/10.1055/s-0040-1721674.

Wynn, J., Rego, S., Chandler-Brown, D., Carter, R., Talati, A., Zaretsky, M. and Trimble, A. (2025) ‘Routine cell-free DNA prenatal screening identifies pregnancies at high risk for cystic fibrosis that may benefit from fetal therapy’, Journal of Cystic Fibrosis [Preprint]. Available at: https://doi.org/10.1016/j.jcf.2025.08.004.

Xiong, Y., Chen, J., Wu, Y., Zhao, P., Liao, M., Guo, J., Liu, C., Zheng, M., Ren, Y., Zou, K., Sun, X. and Tan, J. (2025) ‘The effect of maternal pre-pregnancy body mass index on hypertensive disorders of pregnancy (HDP): a systematic review and dose-response meta-analysis of cohort studies involving 50 million pregnancies’, eClinicalMedicine, 86, p. 103395. Available at: https://doi.org/10.1016/j.eclinm.2025.103395.

Xu, Chenming, Li, J., Chen, S., Cai, X., Jing, R., Qin, X., Pan, D., Zhao, X., Ma, D., Xu, X., Liu, X., Wang, C., Yang, B., Zhang, L., Li, S., Chen, Y., Pan, N., Tang, P., Song, J., Liu, N., Zhang, C., Zhang, Z., Qiu, X., Lu, W., Ying, C., Li, X., Xu, Congjian, Wang, Y., Wu, Y., Huang, H.-F. and Zhang, J. (2022) ‘Genetic deconvolution of fetal and maternal cell-free DNA in maternal plasma enables next-generation non-invasive prenatal screening’, Cell Discovery, 8(1), p. 109. Available at: https://doi.org/10.1038/s41421-022-00457-4.

Xu, H. and Shuttleworth, K.M.J. (2024) ‘Medical artificial intelligence and the black box problem: a view based on the ethical principle of “do no harm”’, Intelligent Medicine, 4(1), pp. 52–57. Available at: https://doi.org/10.1016/j.imed.2023.08.001.

Xu, Q., Xing, L., Zhang, T. and Liu, G. (2025) ‘Development and validation of a preeclampsia prediction model for the first and second trimester pregnancy based on medical history’, BMC pregnancy and childbirth, 25(1), p. 616. Available at: https://doi.org/10.1186/s12884-025-07733-7.

Xu, Y., Zu, Y., Zhang, Y., Liang, Z., Xu, X. and Yan, J. (2025) ‘A Multi-Algorithm Machine Learning Model for Predicting the Risk of Preterm Birth in Patients with Early-Onset Preeclampsia’, International Journal of General Medicine, 18, pp. 4195–4207. Available at: https://doi.org/10.2147/IJGM.S521763.

Xu, Z., Wu, C., Liu, Y., Wang, N., Gao, S., Qiu, S., Wang, Z., Ding, J., Zhang, L., Wang, H., Wu, W., Wan, B., Yu, J., Fang, J., Yang, P. and Shao, Q. (2020) ‘Identifying key genes and drug screening for preeclampsia based on gene expression profiles’, Oncology Letters, 20(2), pp. 1585–1596. Available at: https://doi.org/10.3892/ol.2020.11721.

Yadama, A.P., Maiorino, E., Carey, V.J., McElrath, T.F., Litonjua, A.A., Loscalzo, J., Weiss, S.T. and Mirzakhani, H. (2020) ‘Early-pregnancy transcriptome signatures of preeclampsia: from peripheral blood to placenta’, Scientific Reports, 10(1), p. 17029. Available at: https://doi.org/10.1038/s41598-020-74100-1.

Yagel, S. and Verlohren, S. (2020) ‘Role of placenta in development of pre-eclampsia: revisited’, Ultrasound in Obstetrics & Gynecology, 56(6), pp. 803–808. Available at: https://doi.org/10.1002/uog.22040.

Yang, T.Y. (2015) 'A simple rank product approach for analyzing two classes', Bioinformatics and Biology Insights, 9, pp. BBI.S26414. Available at: https://doi.org/10.4137/BBI.S26414.

Yang, Y. and Wu, N. (2022) ‘Gestational Diabetes Mellitus and Preeclampsia: Correlation and Influencing Factors’, Frontiers in Cardiovascular Medicine, 9, p. 831297. Available at: https://doi.org/10.3389/fcvm.2022.831297.

Yang, Y., Le Ray, I., Zhu, J., Zhang, J., Hua, J. and Reilly, M. (2021) 'Preeclampsia prevalence, risk factors, and pregnancy outcomes in sweden and china', JAMA network open, 4(5), pp. e218401.

Yi, H., Raman, A.T., Zhang, H., Allen, G.I. and Liu, Z. (2018) 'Detecting hidden batch factors through data-adaptive adjustment for biological effects', Bioinformatics, 34(7), pp. 1141–1147.

Ying, S., Chang, D.C. and Lin, S. (2008) 'The microRNA (miRNA): Overview of the RNA genes that modulate gene function', Molecular biotechnology, 38, pp. 257–268.

Yogev, Y., Melamed, N., Bardin, R., Tenenbaum-Gavish, K., Ben-Shitrit, G. and Ben-Haroush, A. (2010) 'Pregnancy outcome at extremely advanced maternal age', American Journal of Obstetrics and Gynecology, 203(6), pp. 558. e1–558. e7.

Young, B.C., Levine, R.J. and Karumanchi, S.A. (2010) 'Pathogenesis of preeclampsia', Annual review of pathology, 5, pp. 173.

Youssef, L., Testa, L., Crovetto, F. and Crispi, F. (2024) '10. role of high dimensional technology in preeclampsia (omics in preeclampsia)', Best Practice & Research. Clinical Obstetrics & Gynaecology, 92, pp. 102427. Available at: https://doi.org/10.1016/j.bpobgyn.2023.102427.

Yu, G., Wang, L., Han, Y. and He, Q. (2012) 'clusterProfiler: An R package for comparing biological themes among gene clusters', Omics: a journal of integrative biology, 16(5), pp. 284–287.

Yu, T., Wang, G., Xu, X. and Yan, J. (2025) 'Identification and validation of key biomarkers associated with immune and oxidative stress for preeclampsia by WGCNA and machine learning', Frontiers in genetics, 16, pp. 1500061. Available at: https://doi.org/10.3389/fgene.2025.1500061.

Yu, Y., Jenkins, A.J., Nankervis, A.J., Hanssen, K.F., Scholz, H., Henriksen, T., Lorentzen, B., Clausen, T., Garg, S.K. and Menard, M.K. (2009) 'Anti-angiogenic factors and pre-eclampsia in type 1 diabetic women', Diabetologia, 52, pp. 160–168.

Yu, Y., Xu, W., Zhang, S., Feng, S., Feng, F., Dai, J., Zhang, X., Tian, P., Wang, S. and Zhao, Z. (2024) 'Non-invasive prediction of preeclampsia using the maternal plasma cell-free DNA profile and clinical risk factors', Frontiers in Medicine, 11, pp. 1254467.

Yuan, Y., Gong, Y., Zhong, L., Ding, X., Yang, Z., Su, X., Chen, M., Zhang, F. and Yang, L. (2022) 'Circular RNA expression profile and competing endogenous RNA regulatory network in preeclampsia', Placenta, 119, pp. 32–38. Available at: https://doi.org/10.1016/j.placenta.2022.01.013.

Yun-Qing, Z., Xing-Yu, Y., Hua, L. and Cong, Z. (2021) 'Insights into the pathogenesis of preeclampsia based on the features of placentation and tumorigenesis', Reproductive and Developmental Medicine, 05(02), pp. 97–106. Available at: https://doi.org/10.4103/2096-2924.320886.

Yusupova, G.Z., Yusupov, M.M., Cate, J. and Noller, H.F. (2001) 'The path of messenger RNA through the ribosome', Cell, 106(2), pp. 233–241.

Zakiyah, N., Postma, M.J., Baker, P.N. and van Asselt, A.D. (2015) 'Pre-eclampsia diagnosis and treatment options: A review of published economic assessments', PharmacoEconomics, 33(10), pp. 1069–1082.

Zakour Khadari, M., Zainal, H., Daud, N.A.A. and Sha'aban, A. (2025) 'Expectant Management of Severe Preeclampsia in Advanced Maternal Age With Multiple Risk Factors: A Case Report', Cureus [Preprint]. Available at: https://doi.org/10.7759/cureus.79600.

Zeisler, H., Llurba, E., Chantraine, F., Vatish, M., Staff, A.C., Sennström, M., Olovsson, M., Brennecke, S.P., Stepan, H. and Allegranza, D. (2016) 'Predictive

value of the sFlt-1: PlGF ratio in women with suspected preeclampsia', New England Journal of Medicine, 374(1), pp. 13–22.

Zeng, X., Gan, Y., Zhao, J., Zhang, L., Zhang, Q., Shen, L., Wang, X., Sun, K. and Zhang, J. (2025) ‘Molecular profiles of the great obstetrical syndromes reveal common features and dynamic changes in early pregnancy’, Communications Medicine, 5, p. 369. Available at: https://doi.org/10.1038/s43856-025-01103-2.

Zhai, J., Xiao, Z., Wang, Y. and Wang, H. (2022) 'Human embryonic development: From peri-implantation to gastrulation', Trends in Cell Biology, 32(1), pp. 18–29. Available at: https://doi.org/10.1016/j.tcb.2021.07.008.

Zhang, J., Ma, X., Hao, L., Liu, L., Lv, J. and Zhang, H. (2015) 'A systematic review and meta-analysis of outcomes of pregnancy in CKD and CKD outcomes in pregnancy', Clinical journal of the American Society of Nephrology: CJASN, 10(11), pp. 1964.

Zhang, S., Bamakan, S.M.H., Qu, Q. and Li, S. (2018) 'Learning for personalized medicine: A comprehensive review from a deep learning perspective', IEEE reviews in biomedical engineering, 12, pp. 194–208.

Zhang, S., Lin, H., Kong, S., Wang, S., Wang, H., Wang, H. and Armant, D.R. (2013) 'Physiological and molecular determinants of embryo implantation', Molecular Aspects of Medicine, 34(5), pp. 939–980. Available at: https://doi.org/10.1016/j.mam.2012.12.011.

Zhang, S., Mao, M., Lv, Y., Yang, Y., He, W., Song, Y., Wang, Y., Yang, Y., Al Abo, M. and Freedman, J.A. (2022) 'A widespread length-dependent splicing dysregulation in cancer', Science advances, 8(33), pp. eabn9232.

Zhang, X., Chen, Y., Salerno, S., Li, Y., Zhou, L., Zeng, X. and Li, H. (2022) 'Prediction of severe preeclampsia in machine learning', Medicine in Novel Technology and Devices, 15, pp. 100158.

Zhang, Y., Deng, Q., Liang, W. and Zou, X. (2018) 'An efficient feature selection strategy based on multiple support vector machine technology with gene expression data', BioMed Research International, 2018, pp. 7538204. Available at: https://doi.org/10.1155/2018/7538204.

Zhang, Y., Parmigiani, G. and Johnson, W.E. (2020) 'ComBat-seq: Batch effect adjustment for RNA-seq count data', NAR genomics and bioinformatics, 2(3), pp. lqaa078.

Zhao, L., Triche, E.W., Walsh, K.M., Bracken, M.B., Saftlas, A.F., Hoh, J. and Dewan, A.T. (2012) 'Genome-wide association study identifies a maternal copy-number deletion in PSG11 enriched among preeclampsia patients', BMC Pregnancy and Childbirth, 12(1), pp. 61. Available at: https://doi.org/10.1186/1471-2393-12-61.

Zhao, Q., Li, J., Diao, Z., Zhang, X., Feng, S., Hou, G., Xu, W., Zhao, Z., Qiu, Z., Yang, W., Zhou, S., Tian, P., Zhang, Q., Chen, W., Li, H., Xiao, G., Qin, J., Hu,

L., Li, Z., Lin, L., Wang, S., Gao, R., Huang, W., Ruan, X., Zhang, S., Zhang, J., Zhao, L. and Zhang, R. (2025) ‘Early prediction of preeclampsia from clinical, multi-omics and laboratory data using random forest model’, BMC Pregnancy and Childbirth, 25(1), p. 531. Available at: https://doi.org/10.1186/s12884-025-07582-4.

Zhao, T. and Peng, Y. (2025) ‘Identification of immune-related genes and molecular subtypes associated with preeclampsia via bioinformatics analysis and experimental validation’, Hereditas, 162, p. 89. Available at: https://doi.org/10.1186/s41065-025-00458-9.

Zhao, Y., Li, M., Konaté, M.M., Chen, L., Das, B., Karlovich, C., Williams, P.M., Evrard, Y.A., Doroshow, J.H. and McShane, L.M. (2021) 'TPM, FPKM, or normalized counts? A comparative study of quantification measures for the analysis of RNA-seq data from the NCI patient-derived models repository', Journal of translational medicine, 19(1), pp. 1–15.

Zhao, Y., Wong, L. and Goh, W.W.B. (2020) 'How to do quantile normalization correctly for gene expression data analyses', Scientific Reports, 10(1), pp. 1–11. Available at: https://doi.org/10.1038/s41598-020-72664-6.

Zhaoer, Y., Mingming, G., Wei, Z., Dan, Y., Yating, Q. and Ruizhe, J. (2022) 'Extracellular vesicles for the treatment of preeclampsia', Tissue and Cell, , pp. 101860.

Zheng, D., Hao, X., Khan, M., Wang, L., Li, F., Xiang, N., Kang, F., Hamalainen, T., Cong, F., Song, K. and Qiao, C. (2022) ‘Comparison of machine learning and logistic regression as predictive models for adverse maternal and neonatal outcomes of preeclampsia: A retrospective study’, Frontiers in Cardiovascular Medicine, 9. Available at: https://doi.org/10.3389/fcvm.2022.959649.

Zheng, Y., Fang, Z., Wu, X., Zhang, H. and Sun, P. (2024) ‘Identification of hub genes, diagnostic model, and immune infiltration in preeclampsia by integrated bioinformatics analysis and machine learning’, BMC Pregnancy and Childbirth, 24(1), p. 847. Available at: https://doi.org/10.1186/s12884-024-07028-3.

Zheng, Y., Ma, C., Liu, X., Wu, S., Zhang, W. and Zhao, S. (2020) ‘Association between HLA-A gene polymorphism and early-onset preeclampsia in Chinese pregnant women early-onset’, BMC Pregnancy and Childbirth, 20(1), p. 656. Available at: https://doi.org/10.1186/s12884-020-03340-w.

Zhong, Q., Xiao, X., Qiu, Y., Xu, Z., Chen, C., Chong, B., Zhao, X., Hai, S., Li, S. and An, Z. (2023) 'Protein posttranslational modifications in health and diseases: Functions, regulatory mechanisms, and therapeutic implications', MedComm, 4(3), pp. e261.

Zhong, Y., Zhu, F. and Ding, Y. (2015) 'Serum screening in first trimester to predict pre-eclampsia, small for gestational age and preterm delivery: Systematic

review and meta-analysis', BMC Pregnancy and Childbirth, 15. Available at: https://doi.org/10.1186/s12884-015-0608-y.

Zhou, M., Feng, Y., Zhang, C., Tian, X., Li, M. and Zheng, Y. (2025) ‘Relationship Between Glucose/Lipid Metabolism and Placental Biomarkers in Gestational Diabetes and Preeclampsia’, Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy, 18, pp. 691–702. Available at: https://doi.org/10.2147/DMSO.S504653.

Zhou, Q., Madala, N.S. and Huang, C. (2025) ‘Pathway-guided architectures for interpretable AI in biological research’, Computational and Structural Biotechnology Journal, 27, pp. 4779–4791. Available at: https://doi.org/10.1016/j.csbj.2025.10.063.

Zhou, Q., Su, X., Jing, G., Chen, S. and Ning, K. (2018) 'RNA-QC-chain: Comprehensive and fast quality control for RNA-seq data', BMC genomics, 19(1), pp. 1–10.

Zhou, Q., Su, X., Wang, A., Xu, J. and Ning, K. (2013) 'QC-chain: Fast and holistic quality control method for next-generation sequencing data', PloS one, 8(4), pp. e60234. Available at: https://doi.org/10.1371/journal.pone.0060234.

Zhou, S., Li, J., Yang, W., Xue, P., Yin, Y., Wang, Y., Tian, P., Peng, H., Jiang, H. and Xu, W. (2023) 'Noninvasive preeclampsia prediction using plasma cell–free RNA signatures', American Journal of Obstetrics and Gynecology, 229(5), pp. 553. e1–553. e16.

Zhou, S., Li, J., Yang, W., Xue, P., Yin, Y., Wang, Y., Tian, P., Peng, H., Jiang, H., Xu, W., Huang, S., Zhang, R., Wei, F., Sun, H.-X., Zhang, J. and Zhao, L. (2023) ‘Noninvasive preeclampsia prediction using plasma cell–free RNA signatures’, American Journal of Obstetrics and Gynecology, 229(5), p. 553.e1-553.e16. Available at: https://doi.org/10.1016/j.ajog.2023.05.015.

Zhou, W., Chen, Y., Zheng, Y., Bai, Y., Yin, J., Wu, X.-X., Hong, M., Liang, L., Zhang, J., Gao, Y., Sun, N., Li, J., Zhang, Y., Wu, L., Jin, X. and Niu, J. (2024) ‘Characterizing immune variation and diagnostic indicators of preeclampsia by single-cell RNA sequencing and machine learning’, Communications Biology, 7(1), p. 32. Available at: https://doi.org/10.1038/s42003-023-05669-2.

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., Bork, P., Jensen, L.J. and von Mering, C. (2023) ‘The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest’, Nucleic Acids Research, 51(D1), pp. D638–D646. Available at: https://doi.org/10.1093/nar/gkac1000.

Zhu, G., Guo, Y.A., Ho, D., Poon, P., Poh, Z.W., Wong, P.M., Gan, A., Chang, M.M., Kleftogiannis, D., Lau, Y.T., Tay, B., Lim, W.J., Chua, C., Tan, T.J., Koo,

S., Chong, D.Q., Yap, Y.S., Tan, I., Ng, S. and Skanderup, A.J. (2021) 'Tissue-specific cell-free DNA degradation quantifies circulating tumor DNA burden', Nature Communications, 12(1), pp. 2229. Available at: https://doi.org/10.1038/s41467-021-22463-y.

Zhu, J., He, F., Song, S., Wang, J. and Yu, J. (2008) 'How many human genes can be defined as housekeeping with current expression data?', BMC genomics, 9, pp. 1–11.

Zolotareva, O., Khakabimamaghani, S., Isaeva, O.I., Chervontseva, Z., Savchik, A. and Ester, M. (2021) 'Identification of differentially expressed gene modules in heterogeneous diseases', Bioinformatics, 37(12), pp. 1691–1698. Available at: https://doi.org/10.1093/bioinformatics/btaa1038.

Zyprych-Walczak, J., Szabelska, A., Handschuh, L., Górczak, K., Klamecka, K., Figlerowicz, M. and Siatkowski, I. (2015) 'The impact of normalization methods on RNA-seq data analysis', BioMed Research International, 2015, pp. e621690. Available at: https://doi.org/10.1155/2015/621690.