BMC Pregnancy and Childbirth

RESEARCH

Open Access

Bioinformatics analysis of shared biomarkers and immune pathways of preeclampsia and periodontitis

Fangyi Ruan[1†], Yinan Wang[1†], Xiang Ying[2†], Yadan Liu[1], Jinghui Xu[1], Huanqiang Zhao[3,4], Yawei Zhu[1], Ping Wen[4*], Xiaotian Li[4*], Qiongjie Zhou[1*] and Hefeng Huang[1,5,6,7,8,9*]

Abstract

Background Epidemiological evidence indicates that preeclampsia (PE) is associated with comorbidities such as periodontitis (PD). However, the underlying mechanism remains unclear. To enhance our understanding of their co-pathogenesis, this research investigated the shared biomarkers and pathological mechanisms.

Methods We systematically retrieved transcriptomic datasets from the Gene Expression Omnibus database. These datasets encompass a comparative analysis of the periodontium with and without PD and of the placenta with and without PE. Differentially Expressed Genes Analysis and Weighted Gene Go-expression Network Analysis (WGCNA) were used to identify the key crosstalk genes in patients with PD and PE. The functional characterisation of these genes was performed using enrichment analysis. Protein–protein interaction networks and machine learning methods were leveraged to identify shared hub genes. The XG-Boost algorithm was applied to construct diagnostic models to gain insight into disease aetiology. The identified genes were validated by single-cell RNA sequencing to ensure their robustness and biological relevance.

Results A total of 55 key crosstalk genes were identified, which were primarily enriched in immune-related pathways by using limma and WGCNA. Among them, twenty-four shared hub genes were identified using protein–protein interaction analysis and machine learning methods. The diagnostic model constructed using immune-related genes outperformed the other two models (area under the receiver operating characteristic curve [ROC] = 0.7786 and 0.7454 for PE and PD, respectively). Pathways involving these genes were mapped using the Kyoto Encyclopedia of Genes and Genomes analysis. In addition, single-cell RNA sequencing analysis showed that the expression of *BIN2*, *LYN*, *PIK3AP1*, and *NEDD9* in neutrophils was significantly downregulated, and *LYN* in fibroblasts and endothelial cells was consistently upregulated.

†Fangyi Ruan, Yinan Wang and Xiang Ying contributed equally and share the first authorship.

*Correspondence:
Ping Wen
kjkwenping@smu.edu.cn
Xiaotian Li
xiaotianli555@163.com
Qiongjie Zhou
zhouqiongjie1732@fckyy.org.cn
Hefeng Huang
huanghefg@hotmail.com
Full list of author information is available at the end of the article

Conclusions Shared hub genes and immunologic pathways were identified in PE and PD, characterised by cross-talk between *BIN2*, *LYN*, *NEDD9*, and *PIK3AP1*, suggesting the pathogenesis of PE and PD, which could pave the way for the development of effective diagnostic, treatment, and management strategies.

Keywords Preeclampsia, Periodontitis, Bioinformatics, Machine learning, Single cell, Immune pathway

Background

Preeclampsia (PE) is a pathologic condition, attributing to the second most common cause of maternal deaths worldwide [1, 2]. It is associated with chronic diseases such as periodontitis (PD) [3]. Although an abnormal immune interaction at the maternal–fetal interface may explain the insufficient remodeling of spiral arteries in PE, the imbalanced relationship between PE and other inflammatory comorbidities remains largely uncertain. Thus, it is imperative to understand the association between PE and associated comorbidity of PD, with the aim of determining whether the complication might exacerbate or trigger PE onset.

PD is a common inflammatory disease of the oral cavity caused by periodontal bacterial infections that result in soft and hard tissue loss [4]. It has been estimated that 62% of adults experience PD [5], and pregnant women are more susceptible to gingival problems [6, 7], which may result in adverse pregnancy outcomes, including PE [8]. Increasing epidemiological evidence suggests a possible relationship between PD and PE. Periodontal disease is reportedly associated with PE after adjusting for gestational age and smoking [9]. Women with preeclampsia are at a higher risk of PD compared with their normotensive counterparts [10]. Increased cytokines and endotoxin levels, as well as oxidative stress in PD are related to endothelial cell motivation and trophoblastic syncytialization [11–14]. However, the cellular components involved in this pathological condition remain unclear.

The rapidly growing transcriptomic dataset offers an ideal opportunity to dissect the complexity of diseases and to discover new biomarkers for developing novel diagnostic strategies. Herein, we analysed the publicly available bulk-sequencing and single-cell RNA-sequencing (scRNA-seq) datasets to investigate the co-pathogenesis of PE and PD, and employed multiple bioinformatics analyses to reveal possible hub genes and mechanisms. Additionally, we explored the expression and location of hub genes using scRNA-seq analysis. We aimed to establish shared biomarkers of PD and PE, offering hints for exploring their genetic aetiology for PE and PD.

Methods

Identification of bulk tissue and single-cell transcriptomic datasets

Dataset analysis was systematically performed using the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/). Terms such as 'preeclampsia', 'placenta', and 'homo sapiens' were used to search for PE datasets, and 'periodontitis', 'gingiva', and 'homo sapiens' were used to search for PD datasets. Terms such as 'obstructive sleep apnoea', 'systemic lupus erythematosus', 'obese', and 'metabolic syndrome' were used to search for datasets in homo sapiens respectively.

The inclusion criteria for the bulk tissue datasets in PE were as follows: (1) at least 20 samples, (2) at least 10 controls and 10 cases, and (3) availability of raw data or gene expression data. The inclusion criteria for the bulk tissue datasets in PD were as follows: (1) at least 4 samples, (2) at least 2 controls and 2 cases, and (3) availability of raw data or gene expression data.

The inclusion criteria for the single-cell datasets were as follows: (1) at least 4 samples, (2) at least 2 controls and 2 cases, (3) female patients, and (4) availability of raw data or gene expression data.

Based on these criteria, three PE (GSE75010, GSE35574, and GSE25906) and three PD (GSE16134, GSE106090 and GSE173078) datasets were included in this analysis. single-cell sequencing datasets of PD, including GSE164241 and GSE171213, in addition to single-cell dataset of PE (GSE173193), met our inclusion criteria, and were downloaded from the GEO. To reduce gender bias during bioinformatics analysis, we used female samples as the datasets provide this information. The details of each GEO dataset are shown in Table S1-S5. The analytical workflow is illustrated in Fig. 1.

Differentially expressed genes (DEGs) analysis

The 'limma' package [15] was used to compare the DEGs between the normal and diseased groups with PE, PD, obstructive sleep apnoea (OSA), systemic lupus erythematosus (SLE), obesity, and metabolic syndrome (MetS), and the screening criteria were P value < 0.05 and $|\log_2 FC| > 0.25$. Subsequently, the comparison of the intersecting DEGs with PE was visualised with the 'UpSetR' package [16] to identify the disease that was most relevant to PE. Meanwhile, the expression patterns of DEGs were visualized in the form of volcano plots with the 'ggplot2' package [17]. Furthermore, we visualised the overlapping DEGs that defined as the intersection of PE DEGs and PD DEGs through the 'Venn' R package (https://github.com/dusadrian/venn).

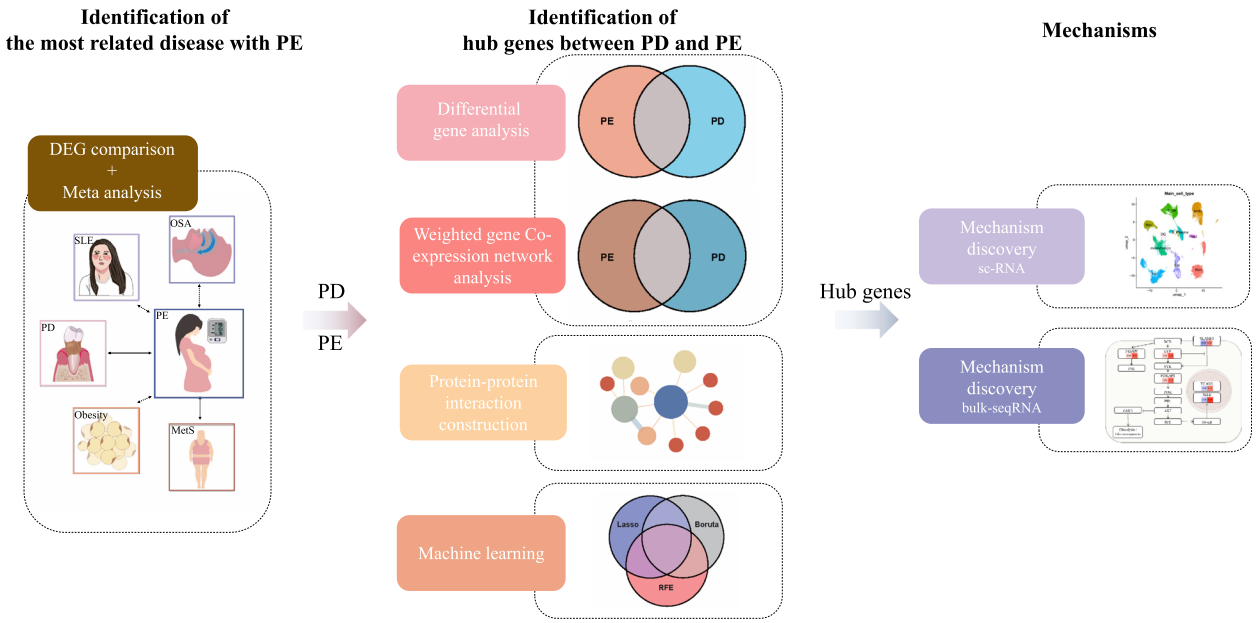


Fig. 1 Study flowchart. DEG Differentially expressed gene, OSA Obstructive sleep apnoea, SLE Systemic lupus erythematosus, PD Periodontitis, PE Preeclampsia, MetS Metabolic syndrome

Meta-analysis of the epidemiological relationship between PD and PE

Studies evaluating the epidemiological relationship between PD and PE were systematically searched from their inception until 23 July 2023. Our search criteria included: 1. analysis of the association between periodontal disease and PE among pregnant women without systemic diseases; 2. observational study; and 3, raw data availability. The meta-analysis was performed using the R package 'meta'. A combined estimate of odds ratios (ORs) was obtained using either fixed- or random-effect models. Model heterogeneity was quantified using the I^2 statistic, which describes the between-study variation as a percentage of the total variation. Funnel plots were used to assess publication bias.

Weighted gene co – expression network analysis (WGCNA) and key module genes identification

To identify gene modules that are important for the phenotype, the R package 'WGCNA' [18] was used to construct a weighted gene co-expression network, using the top 5,000 genes with variance in GSE75010 and GSE16134 as inputs, respectively. The 'cutreeStatic' function was used to delete outliers if needed. The soft-thresholding power was visualized using the 'pickSoftThreshold' function, and the smallest power with a scale-free topology fit index above 0.90 was selected for calculating adjacencies. The adjacencies were then transformed into a topological overlap matrix, and dissimilarity was calculated. A hierarchical clustering method was used to produce a clustering tree structure. Genes with similar expression patterns were categorised into different modules with 'minModuleSize' = 30 and 'mergeCutHeight' = 0.25. The branches of the clustering tree represent different gene modules, and the colours represent the various modules. Notably, the grey module represents genes that cannot be merged. Finally, Pearson's correlation analysis was performed to calculate the correlation between each module and the disease in patients with PE and PD. The modules with the top two most significant positive and negative correlations of module-trait relationships in different diseases were employed for further exploration.

Functional enrichment analysis and protein–protein interaction (PPI) network construction

The intersection of DEGs and the top two most significant module genes of PE and PD were separately identified. These jointly dysregulated genes may be the keys to the links between these two diseases and were called 'crosstalk genes'. To further understand the function of these crosstalk genes, we performed Gene Ontology (GO) biological process enrichment analysis [19], and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis [20] with 'clusterProfiler' R package [21]. Terms with P value < 0.05 were considered to be significantly enriched. Based on this analysis, the top 20 enrichment terms from different databases,

ranked by their *P*-values in descending order, were selected for visualisation. Additionally, we constructed a PPI network of the crosstalk genes using the STRING database [22]. We then imported the network into the Cytoscape software [23] to screen for the highest-connectivity cluster for further analysis, using the Molecular Complex Detection (MCODE) plugin [24] and its default parameter.

Hub genes screening and validation based on the machine learning algorithm

For better analysis, the gene expression of GSE75010 and GSE16134 were merged and corrected for batch effects using the the ComBat function in 'sva' packages [25]. To identify the hub genes among crosstalk genes, we employed several machine learning methods. First, the Boruta algorithm, as a wrapper around the Random Forest classification algorithm, was used to select the candidate biomarkers with the "randomForest" package [26]. Feature selection was repeated using the conventional Recursive Feature Elimination (RFE) algorithm [27]. Additionally, the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm [28], a logistic regression method for filtering variables to enhance the predictive performance, was adopted to screen the candidate biomarkers using the "glmnet" package. Finally, the overlapping genes from these three models were defined as hub genes to develop a diagnostic model. The prediction effectiveness of each hub gene in both training sets was evaluated using receiver operating characteristic (ROC) curves with the pROC package [29].

Construction of the diagnostic model based on the hub genes

Under the Gradient Boosting framework, the Extreme Gradient Boosting (XG-Boost) classifier is an optimised distributed gradient boosting library known for its high efficiency. The R package 'xgboost' [30] was used to construct a model with hub genes, where the expression values of these hub genes served as feature inputs for model training. Furthermore, we employed both internal and external validation techniques and ten-fold cross-validation techniques to evaluate model performance. Using internal and external validation techniques, the PE dataset GSE75010 and PD dataset GSE16134 were divided into training (70%) and test (30%) sets. Subsequently, we used other PE datasets, GSE35574 and GSE25906, and PD datasets, GSE106090 and GSE173078, for external validation. Ten-fold cross-validation techniques were employed in the PE datasets GSE75010 and GSE16134, along with other PE datasets GSE35574 and GSE25906, and PD datasets GSE106090 and GSE173078 for external validation. The diagnostic efficiency was evaluated using the area under the ROC curve and its 95% confidence interval (CI). Meanwhile, the expression patterns of share hub genes were visualised in the form of heatmaps with the 'ComplexHeatmap' package [31].

Single-cell transcriptomic data processing and clustering

scRNA-seq datasets, GSE164241 and GSE171213 for PD, and GSE173173 for PE, met our inclusion criteria. We downloaded raw data and selected female samples for downstream analyses. Seurat R package [32] was used for data processing. Cells with fewer than 200 detected genes, and more than 5,000 detected genes, or for which the total mitochondrial gene expression exceeded 25% were removed. Genes that were expressed in fewer than three cells were also excluded. In addition, we mitigated the effects of cell cycle heterogeneity by calculating cell cycle phase scores based on canonical markers and regressing these out of the data during pre-processing, if needed. The two datasets were integrated using the 'harmony' [33] package. Then, Uniform Manifold Approximation and Projection (UMAP) was constructed using the top 19 principal components of the principal component analyses (PCAs). Unsupervised cell clustering was performed using a graph-based cluster method, and marker genes and DEGs were identified using the FindAllMarkers function with the Wilcox rank-sum test algorithm using the following criteria: 1. logFC > 0.25; 2. *P*_value_adj < 0.05; 3. min.pct > 0.1. Clusters were annotated using canonical cell type markers.

Results

Association between PD and PE

In the comparison of the intersecting DEGs between PE and five other common relevant diseases, including OSA, MetS, SLE, obesity, and PD, 183 DEGs (5.02%) overlapped between PE and PD (Fig. 2A), ranking the first. A meta-analysis was conducted to further verify this relationship between them. A total of 28 studies were included (Fig. 2B), showing that PD was significantly associated with an increased risk for PE (odds ratio [OR] 3.33, 95% CI 2.38–4.65). The concurrences of PD and PE were established.

Crosstalk genes between PE and PD

A search for crosstalk genes based on expression levels and phenotypes can be used to identify key crosstalk genes. There were 183 DEGs that overlapped between the PE and PD groups in terms of expression ($P < 0.05$, $|\log_2 FC| > 0.25$) (Fig. 3A, B, and C). We then employed WGCNA to identify the key phenotypic genes. After constructing a weighted gene co-expression network, we eliminated the outlier samples (Fig. S1A and S1B). As shown in the Fig. S1C and S1D,

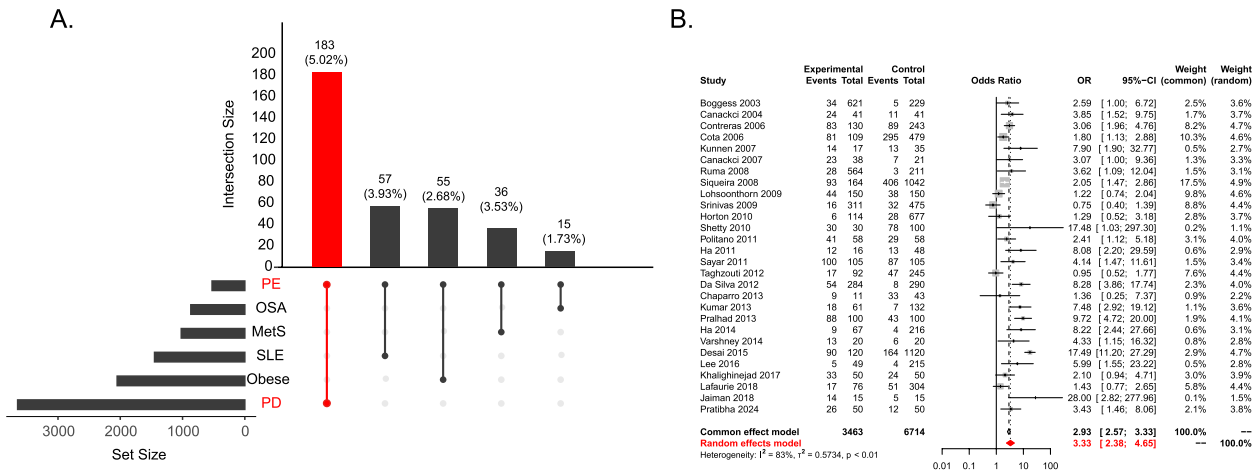


Fig. 2 **A** Upset plot of DEGs in different diseases (*P* value < 0.05 and $|\log_2 FC| > 0.25$). **B** A meta-analysis of pre-pregnancy PD and PE. DEG Differentially expressed gene, OSA Obstructive sleep apnoea, SLE Systemic lupus erythematosus, PD Periodontitis, PE Preeclampsia, MetS Metabolic syndrome

the optimal soft-thresholding power for GSE75010 was 9, whereas that for GSE16134 was 12 (scale-free $R^2 = 0.9$). Figure S1E and S1F shows clustering dendrogram, respectively. Thirteen modules were identified in GSE75010, and 13 in GSE16134. The correlations between the modules and clinical traits were then calculated (Fig. 3D and E). Among these modules, the black module had the strongest positive correlation (142 genes, $r = 0.56$, $p = 6 \times 10^{-14}$), followed by the red module (259 genes, $r = 0.44$, $p = 7 \times 10^{-9}$); the turquoise module had the most negative relation (606 genes, $r = -0.63$, $p = 1 \times 10^{-18}$), followed by purple module (94 genes, $r = -0.35$, $p = 9 \times 10^{-6}$) for PE. As for PD, the pink module displayed the most positive relation (83 genes, $r = 0.64$, $p = 6 \times 10^{-36}$) and the turquoise module had the second positive correlation (949 genes, $r = 0.47$, $p = 2 \times 10^{-18}$); the brown module had the strongest negative correlation (608 genes, $r = -0.64$, $p = 7 \times 10^{-36}$), followed by black module (103 genes, $r = -0.53$, $p = 2 \times 10^{-23}$). Using venn diagrams, we identified 163 overlapping genes between the two datasets in these diseases-relevant modules (Fig. 3F). After DEGs and WGCNA analyses, 55 crosstalk genes were identified.

Enrichment analysis and PPI network construction of crosstalk genes

To further explore the biological functions of dysregulated genes, enrichment analysis was performed. The GO_BP analysis revealed that the significantly enriched terms were the regulation of luekocyte migration, leukocyte migration, ovulation cycle process, regulation of leukocyte chemotaxis, phagocytosis, and leukocyte adhesion to vascular endothelial cells (Fig. 4A). Meanwhile, according to the results of KEGG enrichment analysis database, those genes were significantly involved in lipid and atherosclerosis, ABC transporters, Hippo signaling pathway, Wnt signaling pathway, Prolatin signaling pathway, *PI3K-AKT* signaling pathway, B cell receptor signaling pathway, NF-kappa B signaling pathway, et al. (Fig. 4B). In short, these genes were primarily involved in immunological processes.

Twenty biologically significant hub genes were identified by PPI analysis (Fig. 4C). Interestingly, 20% of these genes were immune related (Fig. 4D). We applied this method to key crosstalk genes (Fig. 4E). To enhance the robustness of the results, we randomly selected genes from 20% of the DEGs in the PE group (Fig. 4F). According to the percentage, immunology was the most significant aspect in this comorbidity compared with the aetiology of PE onset.

Identification and validation of potential hub genes by machine learning methods

To further identify key crosstalk genes, we employed mathematical methods, including Boruta, RFE, and LASSO regression, to select features between GSE75010 and GSE16134 for future diagnostic model construction. Initially, batch effects were removed (Fig. S2). However, there are 3 genes that were not detected in the validation dataset GSE173078. Subsequently, these three machine learning methods were employed based on 52 genes. Boruta was used to identify 22 genes, the RFE method selected 15 features, and the LASSO regression algorithm recommended 27 potential candidate biomarker. By overlapping the genes screened using these three methods, we eventually obtained 6 shared hub genes (Fig. 5A), which were considered the most critical features that significantly

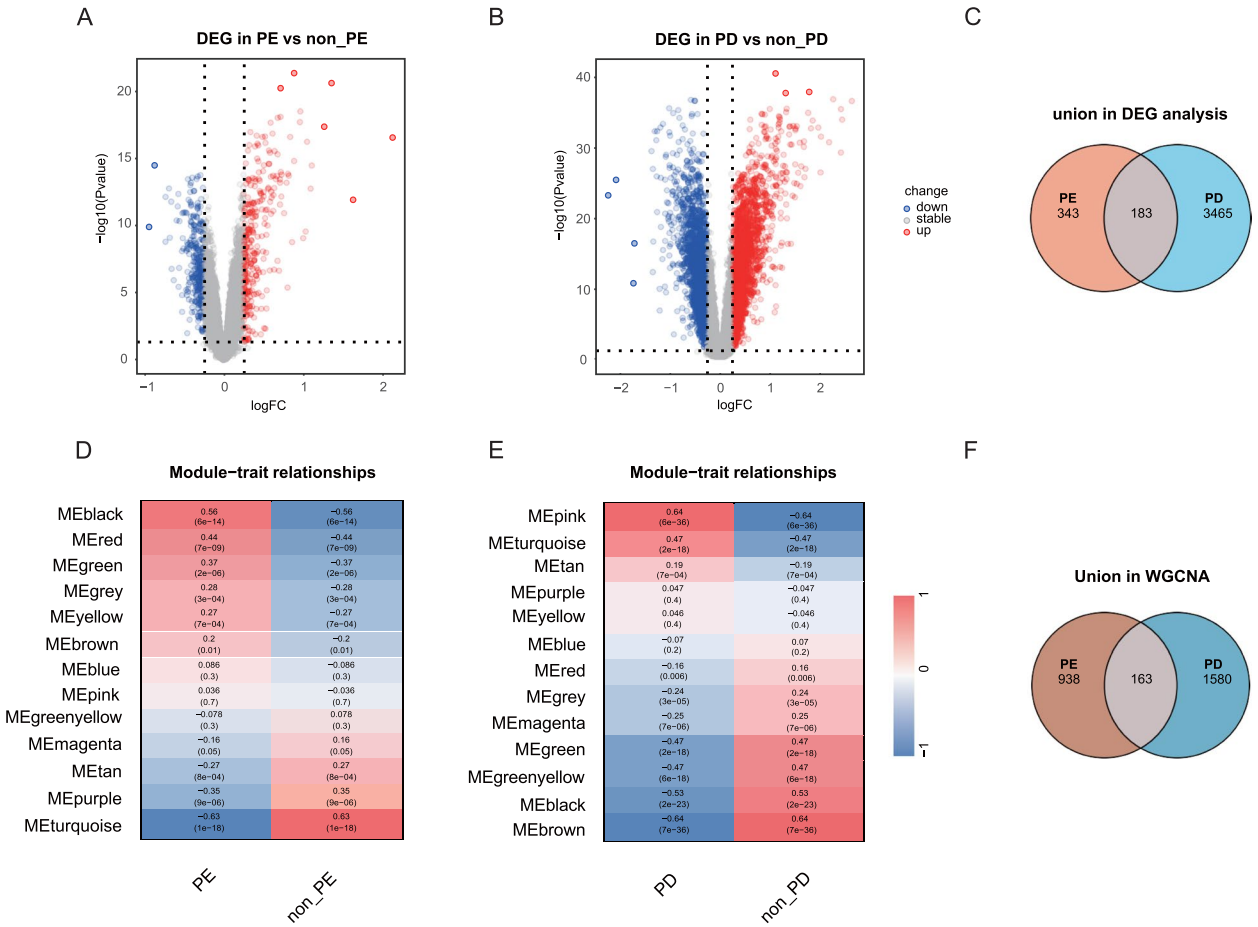


Fig. 3 **A** Volcano plot of DEGs in GSE75010 ($|\log_2 FC| > 0.25$) and **B** GSE16134 ($|\log_2 FC| > 0.25$). Red: upregulated; blue: downregulated. **C** Overlapping DEGs of GSE75010 and GSE16134. **D** Heat map of the correlation between module eigengenes and the occurrence of PE. **E** Heat map of the correlation between module eigengenes and the occurrence of PD. **F** Overlapping module genes of GSE75010 and GSE16134. DEG Differentially expressed gene, PD Periodontitis, PE Preeclampsia, WGCNA Weighted Gene Co-expression Network Analysis

influenced the predictive value. To verify the prediction of hub genes using ROC curve, GSE75010 and GSE16134 were analysed separately, and the results were shown in the Fig. 5B.

Using the above selection procedure, we deleted three genes to identify potential hub genes. From the perspective of the biological functions of these genes, it was necessary to verify whether there was a significant difference in the enrichment analysis between the 52 genes and the original 55 genes. We performed GO_BP and KEGG enrichment analyses on these 52 genes and found that the results were almost the same as those for 55 genes (Fig. S3A and S3B). Nineteen biologically significant genes were identified using PPI analysis. Only *NECTIN4* was excluded because it was not detected in the GSE173078 dataset (Fig. S3C). Subsequent comparison of the proportions revealed the same trend. These results indicated that these 52 genes were functionally similar to 55 genes. Therefore, 24 shared hub genes were identified by intersecting these 19 genes using PPI analysis and 6 genes through machine learning methods.

Construction of prognostic model based on XG-Boost

A comprehensive diagnostic model to gain insights into disease aetiology was constructed based on these hub genes using another machine learning method: the XG-Boost algorithm, which was proven to have excellent performance in machine learning classifiers. Here, 24 shared hub genes were acquired, 9 of which were immunological. Figure 6A describes the workflow of model construction and evaluation. The training and testing sets were used to confirm the effectiveness and reliability of the prognostic model using the ROC curve. Using different types of genes, the performance in the PE external validation set showed that the AUCs were 0.7786, 0.6543, and 0.6439, whereas the performance in the PD external validation set showed that the AUCs were 0.7454, 0.5694, and 0.588, respectively in the immune, non-immune,

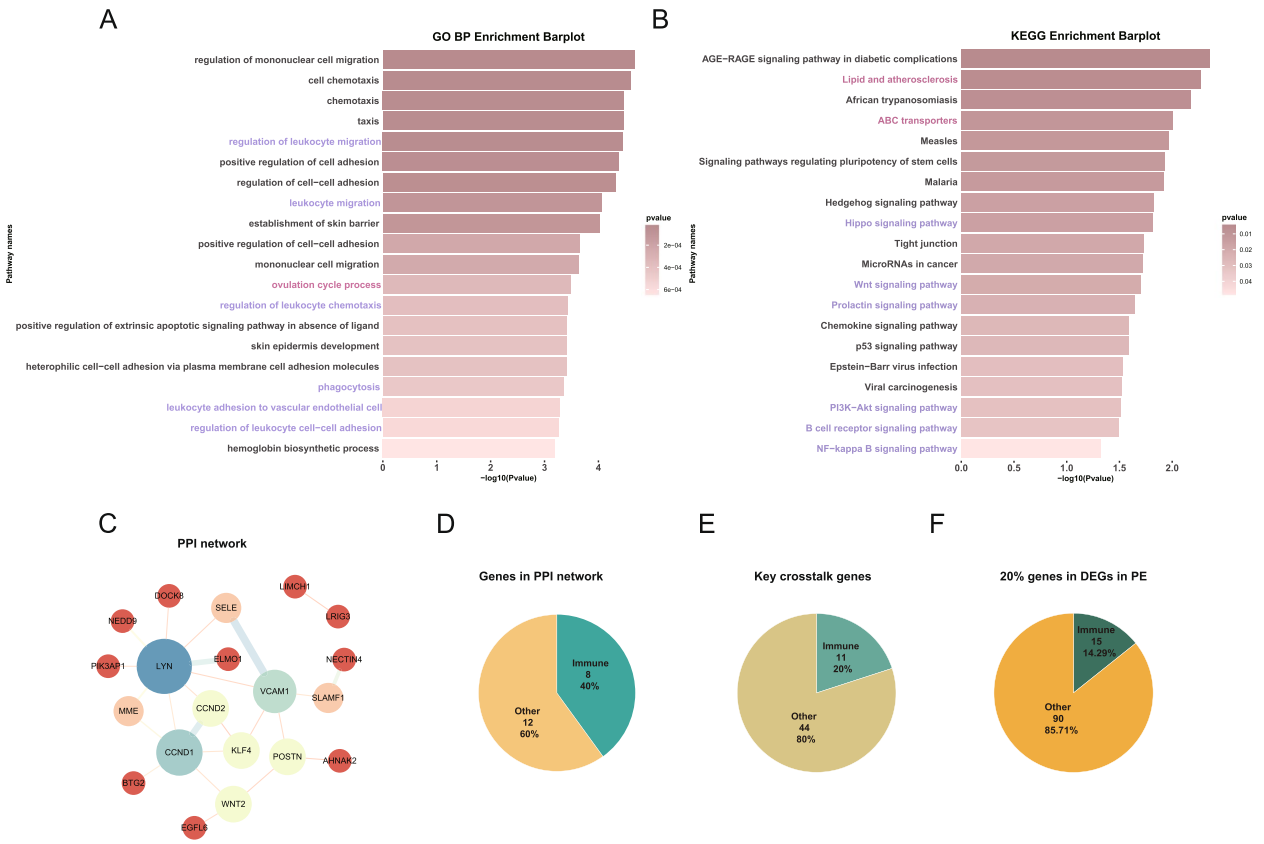


Fig. 4 **A** The top 20 enriched Gene Ontology biological processes across feature genes, colored by *P*-values. **B** The top 20 enriched Kyoto Encyclopedia of Genes and Genomes across feature genes, coloured by *P*-values. The darker the colour, the stronger is the enrichment of the gene in that pathway. **C** PPI network reveals that 20 genes are mostly interact with each other. **D**, **E** and **F** Venn plot reveals that the proportion of immunological genes in 55 overlapped genes, in the PPI network and in the 20% DEGs in PE. GO Gene Ontology, KEGG Kyoto Encyclopedia of Genes and Genomes, DEG Differentially expressed gene, PE Preeclampsia, PPI Protein–protein interaction

and all genes (Fig. 6B). The results of the ten-fold cross-validation model were similar (Fig. S4). Consistent with previous results, these nine immune genes built a better diagnostic model to distinguish patients with PE and PD from healthy controls. To further verify the ability of these genes to distinguish between healthy individuals and those with diseases, heat maps were generated (Fig. 6C).

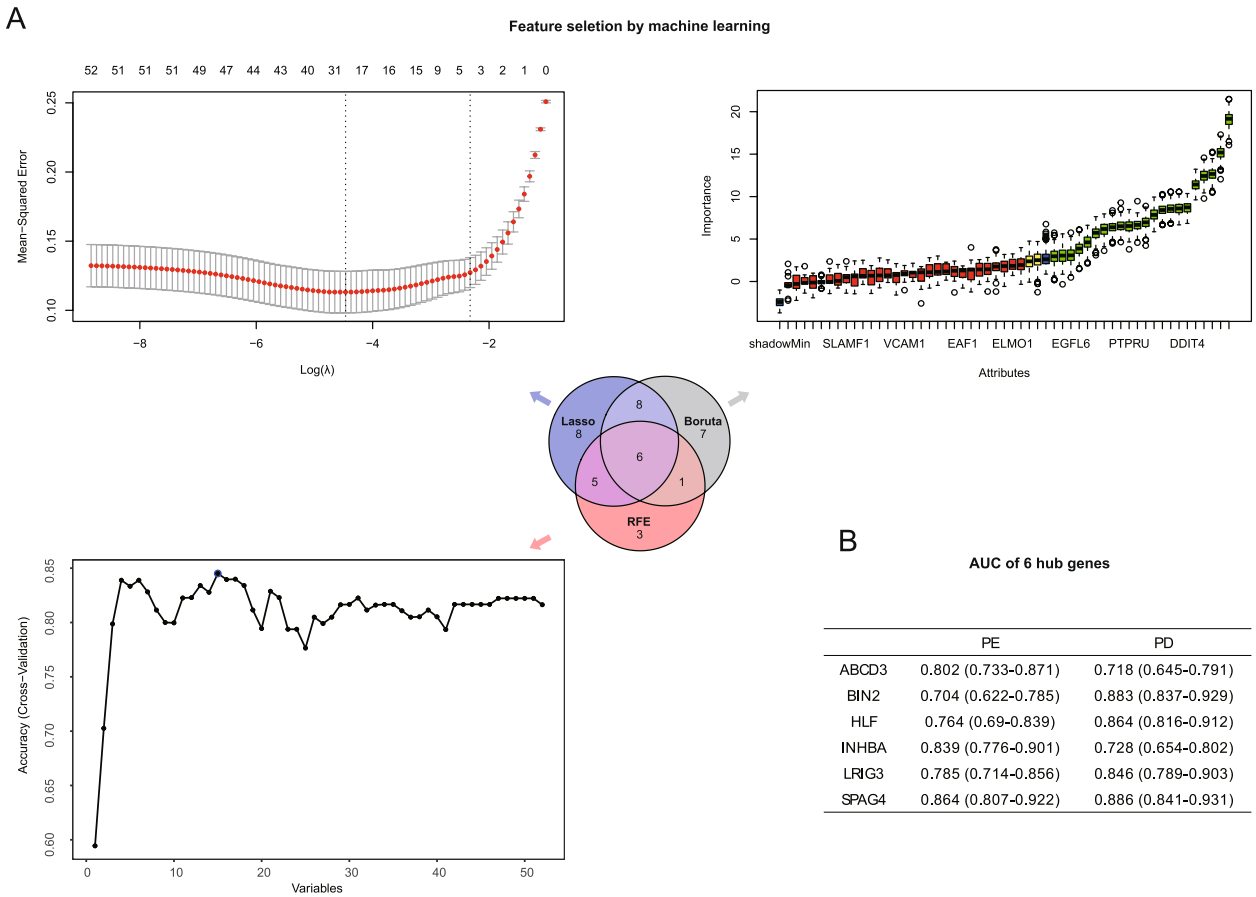
Single-cell analysis for the location of hub genes

To further understand the shared hub gene expression levels under different conditions, we performed single-cell transcriptome profiling of the placenta and periodontium, following a systematic search of public datasets. After quality control, we combined the single cell datasets from PE and PD. We clustered the 108,362 cells into 10 clusters (Fig. 7A) through using the marker genes *COL1A1*, *COL1A2*, *COL3A1*, *PECAM1*, *VWF*, *S100A8*, *SOD2*, *CSF3R*, *CD3D*, *TRAC*, *C1QA*, *RNASE1*, *CD163*, *CD79A*, *SDC1*, *HPGDS*, *KIT*, *GZMB*, *IRF7* for annotation (Fig. 7B). There were 37,356 cells and 71,006 cells in the PE and PD single cell datasets, respectively. In the comparison of expression of 9 hub immunological genes between normal and diseased samples in heatmap (Fig. 7C), the pathways related to these genes are demonstrated in Fig. 7D.

Discussion

Although previous studies have illustrated the epidemiological connections between PD and PE, the shared mechanisms underlying these conditions remain unclear. In this study, we demonstrated through comprehensive analysis that immune dysregulation is a common molecular feature involved in the pathogenesis of PE and PD. In addition, several immunological pathways mediate this comorbidity, including *PI3K-AKT*, B cell receptor, and NF-kappa B signaling pathways. Thus, our findings provide new perspectives on the pathogeneses of PE and PD.

Several studies have reported a possible association between PD and PE [3]. Here, we used bioinformatics analysis to reveal that immune dysfunction is an essential molecular feature of PE and PD. Immune dysfunction



	PE	PD
ABCD3	0.802 (0.733-0.871)	0.718 (0.645-0.791)
BIN2	0.704 (0.622-0.785)	0.883 (0.837-0.929)
HLF	0.764 (0.69-0.839)	0.864 (0.816-0.912)
INHBA	0.839 (0.776-0.901)	0.728 (0.654-0.802)
LRIG3	0.785 (0.714-0.856)	0.846 (0.789-0.903)
SPAG4	0.864 (0.807-0.922)	0.886 (0.841-0.931)

Fig. 5 **A** Gene selection via Boruta algorithm, RFE algorithm, and random Forest. **B** AUC of 6 hub genes. The left line shows the genes in PE, while the right shows those in PD. RFE Recursive Feature Elimination, AUC Area under the curve, PD Periodontitis, PE Preeclampsia

plays a pivotal role in the occurrence and progression of both PE and PD, with potential modulatory mechanisms related to inflammatory reactions, aberrant phenotypes and functions of immune cells, and cytokine and immune imbalances [11, 34, 35]. We performed bioinformatics analysis to systematically confirm, for the first time, that immune dysregulation serves as a shared pathophysiological mechanism in both PE and PD. Therefore, the underlying regulatory role of immune dysregulation in these two conditions merits further investigation, highlighting the necessity of exploring the role of immunology in these comorbidities, which could facilitate the development of more effective clinical prevention and treatment strategies. Additionally, studies have found that the BCR pathway can modulate PE [36] and PD [37] through processes such as B cell activation, antibody production, immune tolerance, and inflammatory factor secretion. Besides, NF-kappa B pathway could mediate common mechanisms of PE and PD in pathophysiological processes such as inflammatory responses [38–40]. We therefore speculate that *PI3K-AKT*, B cell receptor, and NF-kappa B signaling pathways might regulate inflammatory reactions and innate and adaptive immunity, contributing to the common pathogenesis mechanisms in PE and PD. Further research is necessary to clarify the specific mechanisms through which these pathways regulate the pathophysiological processes of PE and PD, thereby paving the way for the development of effective diagnostic, treatment, and management strategies.

Finally, we performed machine learning analyses and single-cell sequencing to determine whether *LYN, BIN2, NEDD9*, and *PIK3AP1* are hub regulatory genes associated with similar mechanisms in PE and PD. Previous studies have revealed that *LYN* is expressed in both immune and stromal cells [41], which is consistent with our findings. Additionally, we found that *LYN* levels were downregulated in neutrophils within the pathological tissues of PE and PD, whereas they were upregulated in fibroblasts and endothelial cells. Previous studies suggested that *LYN* plays an essential role in regulating the BCR signaling pathway, inflammatory responses, T cell immune function, and cell fate. Hence, we speculate that

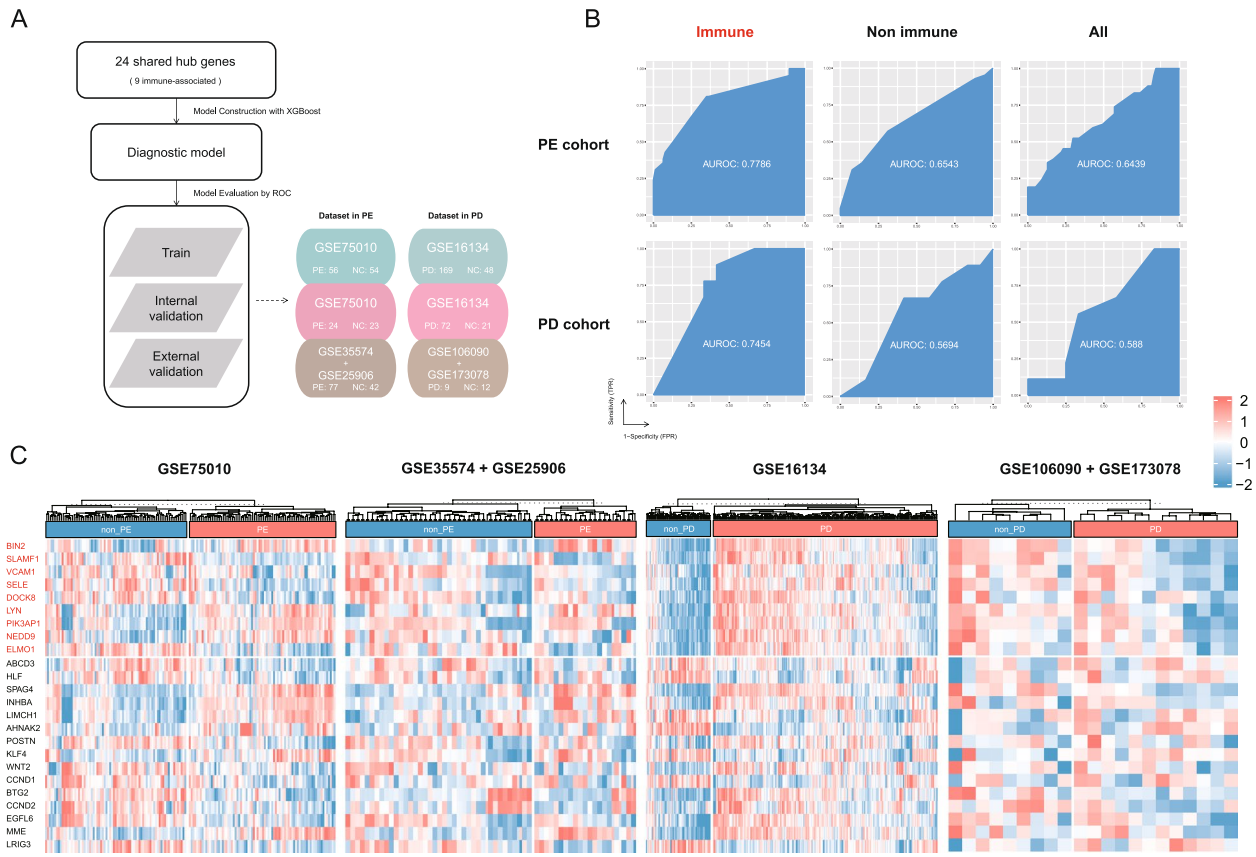


Fig. 6 **A** Overview of the workflow in model construction. **B** Diagnostic performance in the PE and the PD cohorts using different kinds of genes. **C** Heat map of these genes in the PE cohort and the PD cohorts PD Periodontitis, PE Preeclampsia

LYN may be pivotal in the common pathogenesis of PE and PD by modulating various mechanisms, including B cell and T cell immune functions and inflammatory responses. Similarly, we found that the levels of *BIN2*, *NEDD9*, and *PIK3AP1* were reduced in neutrophils, indicating their potential involvement in the shared pathophysiological mechanisms of PE and PD by affecting signal transduction [42], immune cell functions, and phenotypic mechanisms [43–45]. It is essential to elucidate the underlying mechanisms of these key regulatory genes, aiding the understanding of immunological mechanisms in PE and PD while providing novel insights for clinical management and intervention.

Strengths and limitation

We demonstrated possible mechanisms underlying the co-pathogenesis of PE and PD. Although we applied several novel tools, such as machine learning, to make this study more comprehensive, some potential confounders given the limited clinical information in public datasets could not be excluded [46, 47]. Sex, ethnicity, and other information such as in vitro fertilisation were not mentioned. Thus, we only analysed female samples in the PD datasets if they contained gender information. In addition, the datasets included multiple ethnicities (Table S1—S5). Therefore, we confirmed whether ethnic differences or pathological conditions accounted for

(See figure on next page.)

Fig. 7 **A** UMAP visualization of the annotated cells in the combined PE and PD sc RNA-seq dataset; different colours indicate distinct annotated cell types. **B** Dotplot of cell population annotation based on different signatures. **C** Heat map of 9 hub immunological genes expressions between groups with different diseases. Red: upregulated; gray: downregulated. **D** Diagrams summarizing these immunological genes involved in *PI3K-AKT* signal pathways, B cell receptor signal pathway, and NF-κB signal pathways. Alterations are defined by significant upregulation or downregulation of mRNA expression in the PE or PD cohort. Alteration scores of each gene are depicted as log ratios (fold-change, expressed as log2 [ratio of average mRNA expression in diseased subtype versus the paired normal group]). Red: upregulated genes; blue: downregulated genes. PD, periodontitis; PE, preeclampsia

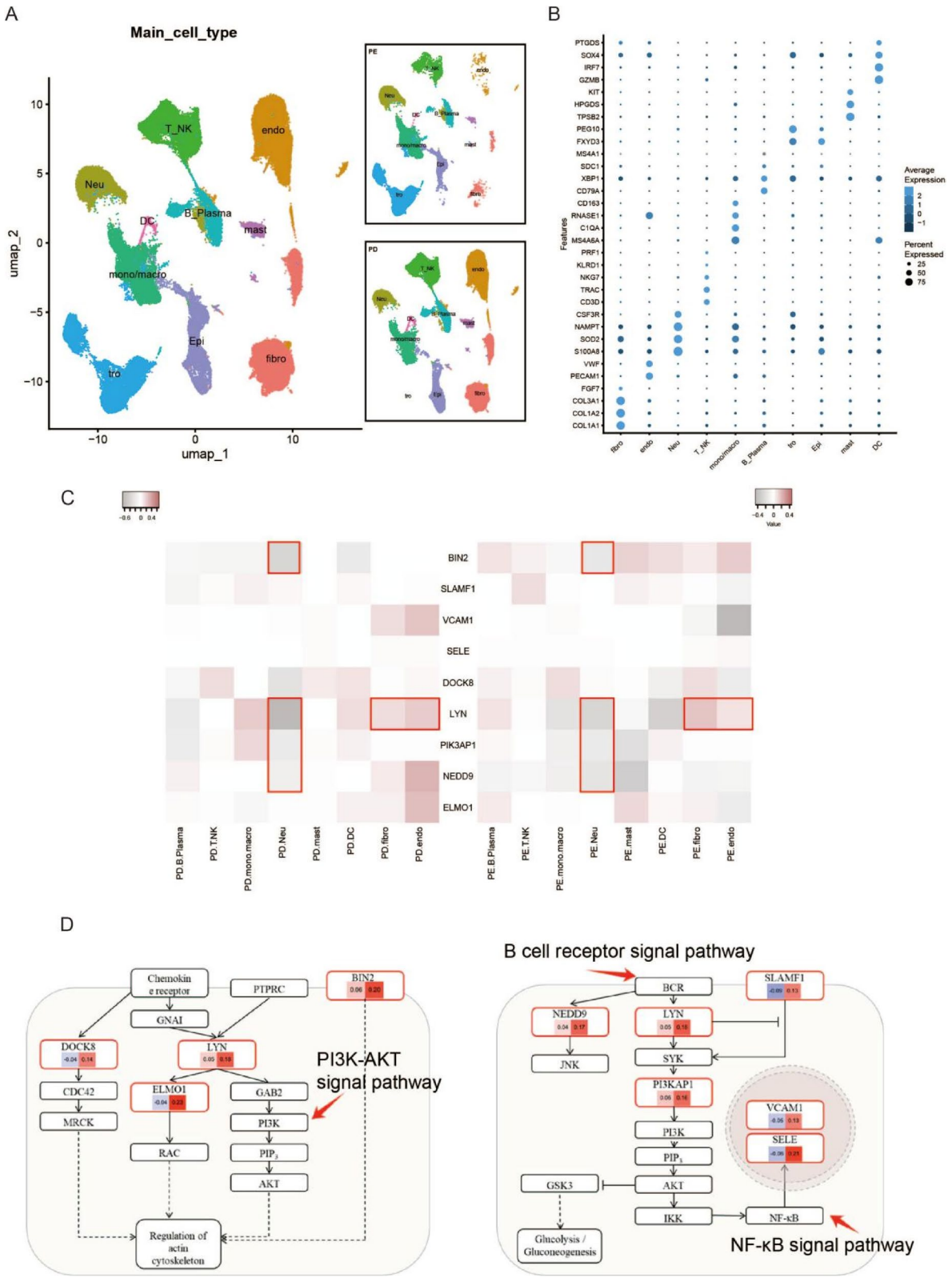


Fig. 7 (See legend on previous page.)

most variability in the data using PCA (Fig. S5), which is a common technique used for projecting variables and objects on the principal component axes [48]. Moreover, similar to a study that confirmed genetic and epigenetic changes in the periodontium of patients with PD with and without gestational diabetes [49], clinical studies are required to discover their relationship and mechanisms and human tissues for experimental validation..

Conclusions

In summary, this study has investigated shared hub genes and immunological pathways for the coexistence of PE and PD. In addition, the key crosstalk genes *BIN2*, *LYN*, *NEDD9* and *PIK3AP1* for PE and PD were indicated in the crosstalk through immune pathways. A new perspective on the pathogenesis of PE and PD comorbidities is suggesting, facilitating the development of more effective clinical prevention and treatment strategies.

Abbreviations

PE	Preeclampsia
PD	Periodontitis
SLE	Systemic lupus erythematosus
OSA	Obstructive sleep apnoea
MetS	Metabolic syndrome
GEO	Gene Expression Omnibus
DEGs	Differentially expressed genes
ORs	Odds ratios
WGCNA	Weighted Gene Co – Expression Network Analysis
PPI	Protein–protein interaction
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
RFE	Recursive feature elimination
LASSO	Least Absolute Shrinkage and Selection Operator
Umap	Uniform Manifold Approximation and Projection
PCA	Principal component analysis
ROC	Receiver operating characteristic
CI	Confidence interval
XG-Boost	Extreme Gradient Boosting
LYN	*LYN* Proto-oncogene
BIN2	Bridging Integrator 2
NEDD9	Neural Precursor Cell Expressed, Developmentally Down-Regulated 9
PIK3AP1	Phosphoinositide-3-Kinase Adaptor Protein 1

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12884-025-07277-w.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6
Supplementary Material 7
Supplementary Material 8
Supplementary Material 9
Supplementary Material 10

Acknowledgements

We thank the authors of the GSE75010, GSE35574, GSE25906, GSE16134, GSE106090, GSE173078, GSE164241, GSE171213, GSE164241, GSE171213, GSE173193, GSE179455, GSE200744, GSE72326, and GSE75097 datasets for their contribution. We deeply appreciate the support of Dr. Feng Zhang (Shanghai Jiao Tong University School of Medicine) and Mr. Jiezhong Guan (Sun Yat-sen University). We thank the reviewers and editors of this journal for their assistance with the manuscript preparation. We also thank the use of the Biorstudio High Performance Computing Cluster (https://biorstudio.cloud) at Biotrainee and the Shanghai HS Biotech Co., Ltd., for conducting the research reported in this study.

Authors' contribution

Hefeng Huang, Qiongjie Zhou, Xiaotian Li, and Ping Wen supervised the study. Fangyi Ruan, Yinan Wang, and Xiang Ying conceived the study and performed the analysis. Yadan Liu, Jinghui Xu, Huanqiang Zhao, and Yawei Zhu reviewed the datasets. Fangyi Ruan, Yinan Wang, Xiang Ying, Yadan Liu, Jinghui Xu, Huanqiang Zhao, Yawei Zhu, Ping Wen, Xiaotian Li, Qiongjie Zhou, and Hefeng Huang drafted and revised the manuscript. All the authors have read and approved the final version of the manuscript.

Funding

This study was supported by the National Key Research and Development Program (2021YFC2701600, 2021YFC2701601, 2021YFC2701603, and 2024YFC2707500), National Natural Science Foundation of China (82371698), Shenzhen Key Laboratory of Maternal and Child Health and Diseases (ZDSYS20230626091559006), Science Foundation of Shanghai (21ZR1410600), Clinical Research Plan of SHDC (SHDC2024CRI077), Shenzhen Medical Research Special Fund Project (C2401035), and Sanming Project of Medicine in Shenzhen (No.SZSM202211032, No.SZSM202311005).

Data availability

The datasets analyzed during the current study are available in the GEO (https://www.ncbi.nlm.nih/).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Obstetrics and Gynecology Hospital of Fudan University, Shanghai, China. [2]Department of Prenatal Diagnostic Center, The International Peace Maternity & Child Health Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [3]Institute of Maternal and Child Medicine, Shenzhen Maternity and Child Healthcare Hospital, Shenzhen, Guangdong, China. [4]Shenzhen Maternity and Child Healthcare Hospital, Southern Medical University, Shenzhen, Guangdong, China. [5]Obstetrics and Gynecology Hospital, Institute of Reproduction and Development, Fudan University, Shanghai, China. [6]Key Laboratory of Reproductive Genetics (Ministry of Education), Zhejiang University School of Medicine, Hangzhou, Zhejiang, China. [7]Research Units of Embryo Original Diseases, Chinese Academy of Medical Sciences, Shanghai, China. [8]Shanghai Key Laboratory of Reproduction and Development, Shanghai, China. [9]Department of Reproductive Endocrinology, Women's Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China.

Received: 8 September 2024 Accepted: 3 February 2025
Published online: 27 February 2025

References

1. Say L, Chou D, Gemmill A, et al. Global causes of maternal death: a who systematic analysis. Lancet Glob Health. 2014;2(6):e323–33. https://doi.org/10.1016/S2214-109X(14)70227-X.

2. Dimitriadis E, Rolnik DL, Zhou W, et al. Pre-eclampsia: 1. Nat Rev Dis Primers. 2023;9(1):1–22. https://doi.org/10.1038/s41572-023-00417-6.
3. Jung E, Romero R, Yeo L, et al. The etiology of preeclampsia. Am J Obstet Gynecol. 2022;226(2S):S844–66. https://doi.org/10.1016/j.ajog.2021.11.1356.
4. Helmi MF, Huang H, Goodson JM, et al. Prevalence of periodontitis and alveolar bone loss in a patient population at harvard school of dental medicine. BMC Oral Health. 2019;19(1):254. https://doi.org/10.1186/s12903-019-0925-z.
5. Trindade D, Carvalho R, Machado V, et al. Prevalence of periodontitis in dentate people between 2011 and 2020: a systematic review and meta-analysis of epidemiological studies. J Clin Periodontol. 2023;50(5):604–26. https://doi.org/10.1111/jcpe.13769.
6. Chen P, Hong F, Yu X. Prevalence of periodontal disease in pregnancy: a systematic review and meta-analysis. J Dent. 2022;125. https://doi.org/10.1016/j.jdent.2022.104253.
7. Raju K, Berens L. Periodontology and pregnancy: an overview of biomedical and epidemiological evidence. Periodontology 2000. 2021;87(1):132–42. https://doi.org/10.1111/prd.12394.
8. Daalderop LA, Wieland BV, Tomsin K, et al. Periodontal disease and pregnancy outcomes: overview of systematic reviews. JDR Clinical & Translational Research. 2018;3(1):10–27. https://doi.org/10.1177/2380084417731097.
9. Ruma M, Boggess K, Moss K, et al. Maternal periodontal disease, systemic inflammation, and risk for preeclampsia. Am J Obstet Gynecol. 2008;198(4):389.e1-5. https://doi.org/10.1016/j.ajog.2007.12.002.
10. Jaiman G, Nayak PA, Sharma S, et al. Maternal periodontal disease and preeclampsia in jaipur population. Journal of Indian Society of Periodontology. 2018;22(1):50–4. https://doi.org/10.4103/jisp.jisp_363_15.
11. Hirano E, Sugita N, Kikuchi A, et al. The association of aggregatibacter actinomycetemcomitans with preeclampsia in a subset of japanese pregnant women. J Clin Periodontol. 2012;39(3):229–38. https://doi.org/10.1111/j.1600-051X.2011.01845.x.
12. Horton AL, Boggess KA, Moss KL, et al. Periodontal disease, oxidative stress, and risk for preeclampsia. J Periodontol. 2010;81(2):199–204. https://doi.org/10.1902/jop.2009.090437.
13. Contreras A, Herrera JA, Soto JE, et al. Periodontitis is associated with preeclampsia in pregnant women. J Periodontol. 2006;77(2):182–8. https://doi.org/10.1902/jop.2006.050020.
14. Güncü G, Tözüm T, Çaglayan F. Effects of endogenous sex hormones on the periodontium — review of literature. Aust Dent J. 2005;50(3):138–45. https://doi.org/10.1111/j.1834-7819.2005.tb00352.x.
15. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for rna-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. https://doi.org/10.1093/nar/gkv007.
16. Conway JR, Lex A, Gehlenborg N. UpSetR: an r package for the visualization of intersecting sets and their properties. Bioinformatics. 2017;33(18):2938–40. https://doi.org/10.1093/bioinformatics/btx364.
17. Hadley W. Ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2016.
18. Langfelder P, Horvath S. WGCNA: an r package for weighted correlation network analysis. BMC Bioinformatics. 2008;9(1):559. https://doi.org/10.1186/1471-2105-9-559.
19. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology: 1. Nat Genet. 2000;25(1):25–9. https://doi.org/10.1038/75556.
20. Ogata H, Goto S, Sato K, et al. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 1999;27(1):29–34.
21. Wu T, Hu E, Xu S, et al. ClusterProfiler 4.0: a universal enrichment tool for interpreting omics data. The Innovation. 2021;2(3). https://www.cell.com/the-innovation/abstract/S2666-6758(21)00066-7. https://doi.org/10.1016/j.xinn.2021.100141.
22. Szklarczyk D, Gable AL, Nastou KC, et al. The string database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 2020;49(D1):D605–12. https://doi.org/10.1093/nar/gkaa1074.
23. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504. https://doi.org/10.1101/gr.1239303.
24. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics. 2003;4:2. https://doi.org/10.1186/1471-2105-4-2.
25. Zhang Y, Parmigiani G, Johnson W E. ComBat-seq: batch effect adjustment for rna-seq count data. NAR Genomics and Bioinformatics, 2020;2(3):lqaa078. https://doi.org/10.1093/nargab/lqaa078.
26. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324.
27. Ambroise C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. Proc Natl Acad Sci USA. 2002;99(10):6562–6. https://doi.org/10.1073/pnas.102102699.
28. Tibshirani R. Regression shrinkage and selection via the lasso. J Roy Stat Soc: Ser B (Methodol). 1996;58(1):267–88. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
29. Robin X, Turck N, Hainard A, et al. PROC: an open-source package for r and s+ to analyze and compare roc curves. BMC Bioinformatics. 2011;12(1):77. https://doi.org/10.1186/1471-2105-12-77.
30. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [2024–01–09]. http://arxiv.org/abs/1603.02754. https://doi.org/10.1145/2939672.2939785.
31. Gu Z. Complex heatmap visualization. IMeta. 2022;1(3):e43. https://doi.org/10.1002/imt2.43.
32. Stuart T, Butler A, Hoffman P, et al. Comprehensive integration of single-cell data. Cell. 2019;177(7):1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.
33. Korsunsky I, Fan J, Slowikowski K, et al. Fast, sensitive, and accurate integration of single cell data with harmony. bioRxiv. 2018;461954(2018–11–05)[2024–01–09]. https://www.biorxiv.org/content/10.1101/461954v2. https://doi.org/10.1101/461954.
34. Li Y, Chen Y, Cai G, et al. Roles of trained immunity in the pathogenesis of periodontitis. J Periodontal Res. 2023;58(5):864–73. https://doi.org/10.1111/jre.13158.
35. Walker KA, Basisty N, Wilson DM, et al. Connecting aging biology and inflammation in the omics era. J Clin Investig. 2022;132(14):e158448. https://doi.org/10.1172/JCI158448.
36. Liu J-C, Zeng Q, Duan Y-G, et al. B cells: roles in physiology and pathology of pregnancy. Front Immunol. 2024;15:1456171. https://doi.org/10.3389/fimmu.2024.1456171.
37. Liu L, Chen Y, Wang L, et al. Dissecting b/plasma cells in periodontitis at single-cell/bulk resolution. J Dent Res. 2022;101(11):1388–97. https://doi.org/10.1177/00220345221099442.
38. Wang Y, Li B, Zhao Y. Inflammation in preeclampsia: genetic biomarkers, mechanisms, and therapeutic strategies. Front Immunol. 2022;13:883404. https://doi.org/10.3389/fimmu.2022.883404.
39. Yu H, Lin L, Zhang Z, et al. Targeting nf-κb pathway for the therapy of diseases: mechanism and clinical study. Signal Transduct Target Ther. 2020;5(1):209. https://doi.org/10.1038/s41392-020-00312-6.
40. Li Y, Xu C, Mao J, et al. ZIF-8-based nanoparticles for inflammation treatment and oxidative stress reduction in periodontitis. ACS Appl Mater Interfaces. 2024;16(28):36077–94. https://doi.org/10.1021/acsami.4c05722.
41. Brian B F, Freedman T S. The src-family kinase lyn in immunoreceptor signaling. Endocrinology, 2021;162(10):bqab152. https://doi.org/10.1210/endocr/bqab152.
42. Couture C, Caron M, St-Onge P, et al. Identification of divergent placental profiles in clinically distinct pregnancy complications revealed by the transcriptome. Placenta. 2024;154:184–92. https://doi.org/10.1016/j.placenta.2024.07.008.
43. Wang X, Jin Y, Xu L, et al. Integrating single-cell rna-seq and bulk rna-seq to construct a novel γδt cell-related prognostic signature for human papillomavirus-infected cervical cancer. Cancer Control: Journal of the Moffitt Cancer Center. 2024;31:10732748241274228. https://doi.org/10.1177/10732748241274228.
44. Zhang Y, Murphy S, Lu X. Cancer-cell-intrinsic mechanisms regulate mdscs through cytokine networks. Int Rev Cell Mol Biol. 2023;375:1–31. https://doi.org/10.1016/bs.ircmb.2022.09.001.
45. Lin J, Liu C, Hu E. Elucidating sleep disorders: a comprehensive bioinformatics analysis of functional gene sets and hub genes. Front Immunol. 2024;15:1381765. https://doi.org/10.3389/fimmu.2024.1381765.
46. Meijsen J, Hu K, Krebs MD, et al. Quantifying the relative importance of genetics and environment on the comorbidity between mental and

cardiometabolic disorders using 17 million scandinavians. Nat Commun. 2024;15(1):5064. https://doi.org/10.1038/s41467-024-49507-3.
47. Zamani Esteki M, Viltrop T, Tšuiko O, et al. In vitro fertilization does not increase the incidence of de novo copy number alterations in fetal and placental lineages. Nat Med. 2019;25(11):1699–705. https://doi.org/10.1038/s41591-019-0620-2.
48. Souza AS, Bezerra MA, Cerqueira UMFM, et al. An introductory review on the application of principal component analysis in the data exploration of the chemical analysis of food samples. Food Science and Biotechnology. 2024;33(6):1323–36. https://doi.org/10.1007/s10068-023-01509-5.
49. Kang J, Lee H, Joo J-Y, et al. Comparison of genetic and epigenetic profiles of periodontitis according to the presence of type 2 diabetes. MedComm. 2024;5(7):e620. https://doi.org/10.1002/mco2.620.

Publisher's Note