International Journal of Women's Health

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Integrative Multi-Omics Analysis Unveils Biomarkers Linking the Gut Microbiota, Blood Metabolites, and Recurrent Pregnancy Loss

Weiming Hao[1], Ruigao Song[1], Huimin Lv[1], Chunying Song[2]

[1]Reproductive Medicine Center, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Third Hospital of Shanxi Medical University, Tongji Shanxi Hospital, Taiyuan, People's Republic of China; [2]Human Sperm Bank, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Third Hospital of Shanxi Medical University, Tongji Shanxi Hospital, Taiyuan, People's Republic of China

Correspondence: Weiming Hao, Reproductive Medicine Center, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Third Hospital of Shanxi Medical University, Tongji Shanxi Hospital, Taiyuan, People's Republic of China, Email haoweiming@sxbqeh.com.cn

Purpose: Emerging evidence suggests that dysbiosis of gut microbiota and metabolic disturbance can adversely affect reproductive health. This study seeks to shed light on the connection between gut microbiota, blood metabolites, and recurrent pregnancy loss (RPL), and identify potential biomarkers linking them.

Patients and Methods: The associations of gut microbiota and blood metabolites with RPL were explored through Mendelian randomization (MR) and mediation analyses. Differential expression analysis combined with three machine learning algorithms was then used to identify biomarkers that link the gut microbiota-blood metabolite network in RPL. Additionally, a nomogram was constructed to evaluate their predictive performance for RPL. On this basis, immune infiltration analysis and single-cell RNA sequencing (scRNA-seq) were further conducted to gauge the immune characteristics of RPL.

Results: A total of 28 gut microbiota and 82 blood metabolites/metabolite ratios showed significant potential associations with RPL. Among them, mediation analysis revealed that 3-amino-2-piperidone amplified the hazardous effect of *Photobacterium* abundance on RPL (mediation proportion = 14.4%, β = 0.017, $P = 0.0478$), whereas *CAG-495* attenuated the protective effect of cysteine-glutathione disulfide levels on RPL (mediation proportion = 15.5%, β = 0.003, $P = 0.0497$). ASH1L, G6PD, SETDB1, and LAP3 were identified as biomarkers linking the gut microbiota-blood metabolites network in RPL. The nomogram constructed based on these biomarkers exhibited excellent ability to discriminate RPL, with an area under the curve (AUC) value of 0.972. Finally, scRNA-seq demonstrated an increasing proportion of decidual macrophages and enhanced cell-cell communication in the RPL group.

Conclusion: Significant potential links were observed between the gut microbiota, blood metabolites, and RPL. Integrative multi-omics analysis further identified key biomarkers linking gut microbiota, blood metabolites, and RPL, and highlighted the role of the gut microbiota-metabolites-immune axis in the pathogenesis of RPL.

Keywords: gut microbiota, blood metabolites, recurrent pregnancy loss, multi-omics analysis, biomarkers, immune

Introduction

Recurrent pregnancy loss (RPL) is a problem that still troubles women of childbearing age to date. According to the newest European Society of Human Reproduction and Embryology (ESHRE) guidelines, it is defined as the loss of two or more consecutive pregnancies (excluding ectopic pregnancy and molar pregnancy) and affects approximately 1–2% of couples worldwide.[1] It has been confirmed that many factors can result in RPL, such as embryonic chromosomal abnormalities, autoimmune disorders, endometrial dysfunction and infections.[2,3] Nonetheless, the cause of more than half of RPL remains to be clarified.[2,3]

The microbial community that colonizes within the human digestive system is known as the gut microbiota, and it is vital to human health.[4,5] Studies have demonstrated that patients with pregnancy loss exhibit significantly reduced gut microbial diversity, along with decreased relative abundance of *Prevotellaceae* and *Selenomonas*.[6] In patients with RPL,

https://doi.org/10.2147/IJWH.S598767
Received: 26 January 2026
Accepted: 18 April 2026
Published: 30 April 2026

those positive for antiphospholipid antibodies and antinuclear antibodies show higher gut microbial richness and diversity, as well as increased proportions of *Megasphaera* and *Enterococcus*.[7] Thus, gut microbial dysbiosis—regardless of whether it presents as decreased or abnormally elevated diversity—adversely affects pregnancy outcomes and increases the risk of RPL. During pregnancy, the embryo acts as a semi-allograft, and the establishment and maintenance of maternal-fetal immune tolerance serve as the prerequisite for successful pregnancy.[8,9] In this process, the gut microbiota contributes to the maintenance of normal pregnancy by systemically regulating immune homeostasis and maternal-fetal immune tolerance[10,11] through a gut-placenta immune axis, mainly by inducing myeloid-derived suppressor cells (MDSCs) and gut-derived RORγt+ regulatory T cells (Tregs) to suppress excessive IFN-γ+ and IL-17A+ T cell responses at the maternal-fetal interface.[9] Gut microbial dysbiosis disrupts immune balance, triggers a systemic pro-inflammatory state, and indirectly impairs embryo implantation, decidualization, angiogenesis, spiral artery remodeling, and placental development, thereby representing a crucial potential mechanism underlying RPL.[12]

Additionally, the embryo may alter maternal metabolic pathways through a series of complex regulatory mechanisms during pregnancy.[6,8,13] Notably, a metabolomic profiling has shown significant differences between RPL patients and the control group.[14] It also remains unclear whether the gut microbiota is involved in and interacts with these metabolic alterations to affect RPL. Hence, clarifying the role of gut microbiota and blood metabolites in RPL is worth exploring. However, conducting clinical trials to address these questions is rather difficult due to the constraints of sample size and the interference of many confounders. By contrast, Mendelian Randomization (MR) has emerged as a powerful research methodology for assessing the potential associations between exposures and specific diseases as it can minimize the interference of confounders and the bias caused by reverse causality while providing reliable and robust estimations.[15,16]

Herein, we explored the potential causal associations of gut microbiota and blood metabolites with RPL, and the potential interaction/mediating relationship between them through bidirectional two-sample MR and mediation analyses. Then, biomarkers linking the gut microbiota-blood metabolites network in RPL, as well as their roles in the RPL immune microenvironment were further identified via integrative bioinformatics analysis. This study aimed to improve the understanding of RPL pathogenesis and offer fresh perspectives and strategies for its clinical treatment.

Materials and Methods

MR Analysis Exploring the Associations of Gut Microbiota and Blood Metabolites with RPL

Study Design

This MR study was carried out in accordance with the STROBE-MR statement (Table S1).[17] First and foremost, to eliminate bias and guarantee the reliability of causal inference, the genetic variants that were utilized as instrumental variables (IVs) in the MR study must comply with three fundamental assumptions: (1) there should be a significant association between IVs and the exposures under investigation; (2) there is no association between IVs and confounders of the exposures and outcomes; (3) the effect of IVs on the outcomes should be fully mediated by the exposures (Figure 1). Based on the above, the two-sample MR study was initially implemented to explore the potential causal associations of gut microbiota and blood metabolites with RPL. Subsequently, a two-step mediation MR analysis was implemented to further explore the effect of gut microbiota on RPL via blood metabolites, as well as that of blood metabolites on RPL via gut microbiota. Finally, reverse MR analysis was implemented to gauge the effect of RPL on both gut microbiota and blood metabolites. The steps involved and the results were accompanied by a series of rigorous statistical analyses and verifications. The flowchart of the overall study design was shown in Figure 2.

Data Sources

Data regarding gut microbiota, blood metabolites, and RPL were obtained from the NHGRI-EBI GWAS Catalog database (https://www.ebi.ac.uk/gwas/). The human gut microbiota data (GCST90032172-GCST90032644) used in this study were derived from the largest GWAS published to date, which were based on fecal samples from 5959 Finnish individuals, involving 7,967,866 single nucleotide polymorphisms (SNPs) and covering 473 taxonomic units, including 11 phyla, 19 classes, 24 orders, 62 families, 146 genera, and 209 species.[18] The metabolite data utilized in this analysis were from one of the most comprehensive metabolite studies, which encompassed 1091 blood metabolites and 309

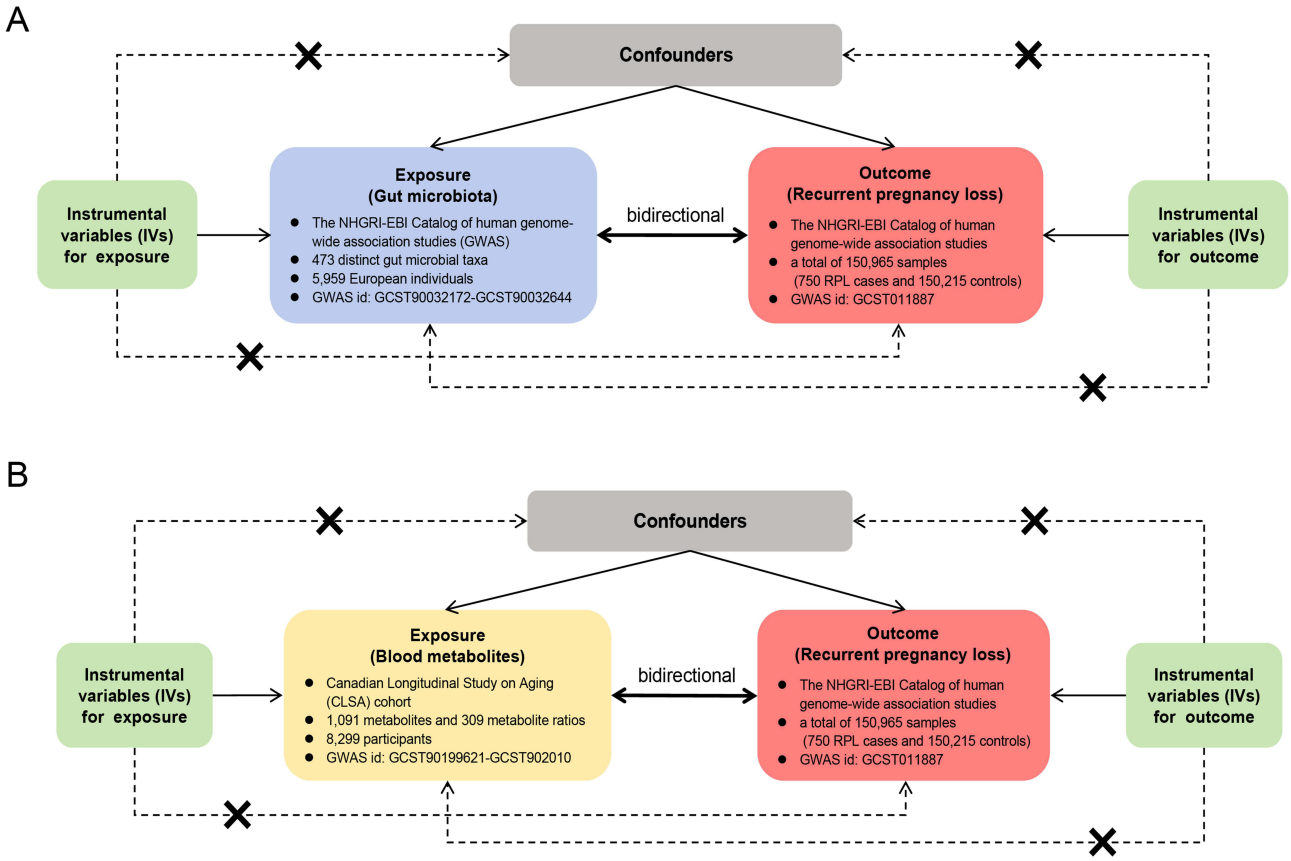

Figure 1 Fundamental assumptions and data of genetic variants for the bidirectional two-sample MR study. (**A**) MR study for gut microbiota and RPL; (**B**) MR study for blood metabolites and RPL. The black X symbol indicates the exclusion of potential confounding pathways, fulfilling the core exclusion assumptions of the MR study. The bold font represents important factors in the analysis: exposure and outcome. The label "bidirectional" indicates a bidirectional two-sample MR design, which simultaneously tests the potential causal effects in both directions (exposure factor → outcome and outcome → exposure factor) to separate the direction of causality.
Abbreviations: MR, Mendelian randomization; RPL, recurrent pregnancy loss; GWAS, genome-wide association study.

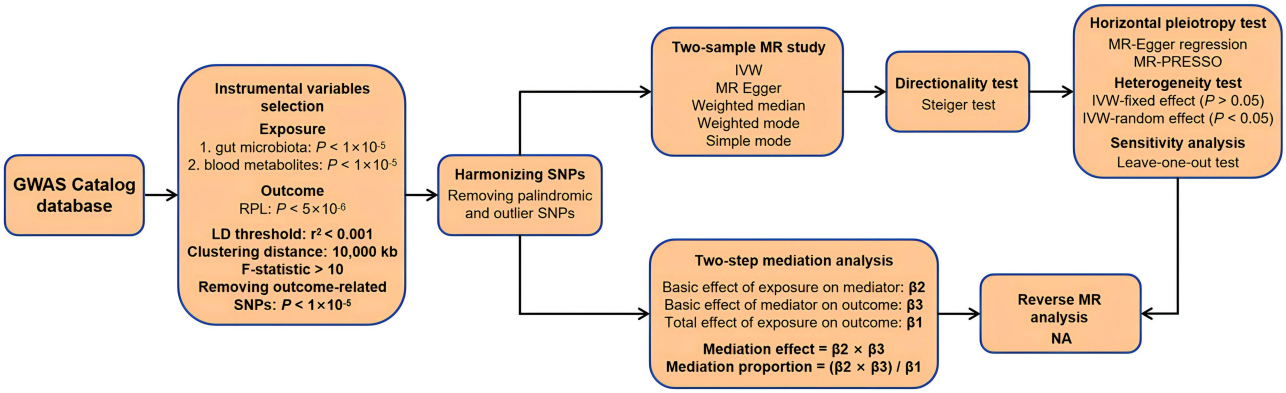

Figure 2 Flowchart of the overall design for the bidirectional two-sample MR study and two-step mediation analysis. The dark black arrow indicates the sequential analysis process of the research, showing the direction of data processing and analysis from left to right. The bold font indicates important factors, analysis methods, and critical thresholds.
Abbreviations: SNPs, single nucleotide polymorphisms; LD, linkage disequilibrium; IVW, inverse variance weighted; MR-PRESSO, MR pleiotropy residual sum and outlier; NA, no association.

https://doi.org/10.2147/IJWH.S598767

metabolite ratios, deriving from the analysis of 8299 European participants in the Canadian Longitudinal Study on Aging cohort and involving approximately 15.4 million SNPs.[19] The full GWAS summary statistics for these 1400 blood biomarkers were accessible to the public in the GWAS Catalog (GCST90199621-GCST90201020). The RPL dataset was identified as GCST011887 and consisted of 150,965 samples, including 750 RPL cases and 150,215 control samples, all of which were of European descent and encompassed 23,125,249 SNPs.[20]

Selection of IVs

To fulfill the three assumptions of MR analysis (Figure 1), SNPs used as IVs need to adhere to the following criteria: Firstly, a relatively lenient statistical threshold of $P < 1\times10^{-5}$[21,22] was employed to recognize SNPs associated with 473 gut microbiota and 1400 blood metabolites, to avoid excessive filtering and retain sufficient statistical power. For RPL, a threshold of $P < 5\times10^{-6}$ was utilized. Secondly, to avoid biased results caused by LD in this study, the "ieugwasr" package (v 1.0.0) was utilized to remove inapplicable IVs with LD, with a threshold of $r^2 < 0.001$ and a clustering distance of 10,000 kb. Thirdly, the F-statistic was applied to measure the strength of the IVs, those IVs exhibiting the F-statistic below 10 were deemed weak instruments and thus removed using the formula $F = (N\text{-}k\text{-}1)/k \times [R^2/(1\text{-}R^2)]$, where N represents the sample size of the exposure GWAS, k is the number of IVs, and R^2 is the proportion of variance explained.[23] Fourthly, the IVs strongly associated with the outcome GWAS traits ($P < 1\times10^{-5}$) were excluded to minimize potential confounding factors (Figure 2). Subsequently, to ensure the unidirectionality of causalities, the Steiger directionality test was used to further screen IVs.[24] Finally, the SNPs containing palindromic sequences were eliminated after matching the results to ensure the effects of SNPs on the exposure consistent with the same allele as those on the outcome.

Bidirectional Two-Sample MR Study and Mediation Analysis

To explore the impact of gut microbiota and blood metabolites on RPL, the TwoSampleMR package (v 0.6.0) and MRPRESSO package (v 1.0) were utilized.[25,26] The inverse variance weighted (IVW) method was selected as the primary method for MR analysis as it could integrate information from all IVs to estimate the causal effect through a weighted average. On this basis, to further confirm the robustness of the MR analysis, weighted median, MR Egger, weighted mode, and simple mode were employed to assess the potential causal effects.[27–29] Moreover, reverse MR analysis was also conducted, in which RPL served as the exposure while gut microbiota or blood metabolites as the outcome, to assess whether RPL exerted a causal effect on gut microbiota or blood metabolites.

To illustrate whether the mediator mediated the link between exposure and outcome, two-step mediation analyses were further executed. The effect estimates of gut microbiota (blood metabolites) on RPL (β1), the effect estimates of gut microbiota (blood metabolites) on blood metabolites (gut microbiota) (β2), and the effect estimates of blood metabolites (gut microbiota) on RPL (β3) were calculated. The mediation effect was quantified as (β2×β3), while the direct effect of exposure on outcome was represented as (β1-β2×β3). The mediation proportion was calculated as (β2×β3/β1) (Figure 2). Standard errors for the mediation effect, direct effect, and mediation proportion were estimated using the delta method.

Integrative Transcriptome Analysis of the Mechanisms Underlying Mediation Effects

Data Source

The training (GSE165004), validation (GSE26787), and scRNA-seq (GSE214607) datasets were acquired from the Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/). Dataset selection was based on the following criteria: the species was restricted to Homo sapiens, all samples were human tissue specimens, subjects were clearly divided into RPL patients and healthy controls, and all datasets were authentic and reliable. GSE165004 dataset (GPL16699) comprised endometrial samples from 24 RPL patients and 4 controls. GSE26787 (GPL570) dataset comprised 5 RPL and 5 control endometrial samples. GSE214607 dataset (GPL24676) comprised 6 RPL samples and 10 control decidual and villous samples. The transcriptomic data were obtained as preprocessed and standardized expression matrices. No further upstream processing or data merging was performed, and all analyses were conducted independently in each dataset. Additionally, 94 metabolic pathway-related genes (KEGG_GLUTATHIONE_METABOLISM and KEGG_LYSINE_DEGRADATION) were gathered via the MSigdb database (http://www.gsea-MSigdb.org/gsea/msigdb).

Differential Expression and Enrichment Analyses

In the GSE165004 dataset, genes corresponding to blood metabolites with mediating effects were subjected to the Wilcoxon test between the RPL and control groups, with genes having *P*-values less than 0.05 being recognized as key genes. Subsequently, the clusterProfiler package (v 4.6.2)[30] was implemented to execute Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses on these key genes ($P < 0.05$). Finally, the STRING database was employed to build protein-protein interaction (PPI) network, setting the low-confidence threshold at 0.150.

Identification of Biomarkers via Machine Learning

Candidate biomarkers were identified using three machine learning algorithms: least absolute shrinkage and selection operator (LASSO), Boruta, and support vector machine-recursive feature elimination (SVM-RFE), which were implemented using the glmnet (v 4.1–8),[31] Boruta (v 8.0.0),[32] and e1071 (v 1.7–14)[33] packages, respectively. LASSO excels in handling high-dimensional data and aids in variable selection and model interpretation. Boruta can select any feature associated with the response variable without lowering the cost function of the model. To optimize the margin between classes and facilitate efficient classification and regression, SVM-RFE finds the best hyperplane. Both Lasso and SVM-RFE analyses were performed with 10-fold cross-validation, while maxRuns = 500 was set in the Boruta analysis. Candidate biomarkers were determined by identifying overlapping genes from the results of these three algorithms. In the GSE165004 and GSE26787 datasets, genes that showed significant differences and consistent expression trends between the RPL and controls were designated as the final biomarkers.

Construction of the Nomogram

Based on the expression levels of biomarkers, the nomogram was established in the GSE165004 dataset via the rms package (v 6.7–0)[34] to predict RPL risk. The predictive accuracy of the nomogram was gauged by generating a receiver operating characteristic (ROC) curve, and its reliability was validated using a calibration curve.

Gene Set Enrichment Analysis (GSEA)

Within the GSE165004 dataset, Spearman analysis was executed between each biomarker and all genes using correlation coefficients as the ranking metric. Then GSEA was conducted using the clusterProfiler package ($|NES| > 1$ and $P < 0.05$), basing on the "KEGG.v2024.1.Hs.symbols.gmt" file downloaded from the MSigdb database.

Immune Infiltration Analysis

The CIBERSORT algorithm, in conjunction with the LM22 gene signature,[35] was implemented to analyze the proportions of 22 immune cells in RPL and controls from the GSE165004 dataset. Wilcoxon test was implemented to contrast variations in immune infiltration between the groups. Additionally, Pearson analysis was conducted to evaluate the relationships between biomarkers and immune cells.

Creation of the Regulatory Network

The TFs targeting the biomarkers were predicted by the ChEA3 database (https://maayanlab.cloud/chea3/), and those supported by ChIP-seq data from the ENCODE (https://www.encodeproject.org/) were selected. The microRNAs (miRNAs) regulating biomarkers were predicted in the miRDB (https://mirdb.org/) (score > 70). Subsequently, lncRNAs interacting with predicted miRNAs were predicted by the starbase database (https://rnasysu.com/encori/) (clipExpNum > 10). Finally, the networks of biomarker-TF and biomarker-miRNA-lncRNA were constructed using the Cytoscape software.

scRNA-Seq Analysis

The Seurat package (v 5.1.0)[36] was implemented to process the GSE214607 dataset. As part of the initial quality control, cells with fewer than 200 genes and genes present in fewer than three cells were excluded. Cells with mitochondrial gene ratios exceeding 25%, as well as those with gene numbers ≤ 200 and ≥ 8000, and total counts ≤ 200 and ≥ 100,000, were also removed. To identify the top 2000 highly variable genes, the vst method within the FindVariableFeatures function was applied. Data were standardized using ScaleData and subjected to principal component analysis with RunPCA, selecting the top 30 principal components for subsequent analysis. Cells were clustered using the FindNeighbors and

FindClusters functions (resolution = 0.2), and the clustering outcomes were showed using UMAP. The expression of each cell type's marker genes as documented in the literatures was used to annotate the various cell types.[37–41] DoubletFinder (v 2.0.4)[42] was used to identify and remove potential doublets. The proportions of different cell types in RPL and controls were analyzed, and the distribution of biomarkers within these cells, as well as expression differences between RPL and controls, were explored. Finally, the CellChat package (v 1.6.1)[43] was employed to visualize cell communication networks, revealing ligand-receptor interactions and signaling patterns to elucidate communication mechanisms among various cells in the RPL microenvironment.

Statistical Analysis

To confirm the robustness of the causal effects, sensitivity test was executed, including heterogeneity test, horizontal pleiotropy test, and leave-one-out (LOO) test. The mr_heterogeneity function was used to inspect the heterogeneity level. When heterogeneity was present ($P < 0.05$), the IVW-random effects model was selected; otherwise, the IVW-fixed effects model was implemented. To gauge potential horizontal pleiotropy of the results, MR-Egger regression intercept tests and MR-PRESSO global tests were conducted. Any result demonstrating substantial pleiotropy ($P < 0.05$) was omitted from further analysis. In addition, the effect of each SNP on the results was measured by the LOO test. This involved sequentially excluding each SNP to evaluate its effect on the overall causal estimation.[44] The Wilcoxon test was employed to gauge group differences in bioinformatics analysis. The R software was employed for all statistical analyses. A statistically significant *P*-value was less than 0.05.

Results

Potential Associations of Gut Microbiota and Blood Metabolites with RPL

For the IVs corresponding to each exposure and outcome, the F-statistic ranged from 19.56 to 41.22 for gut microbial SNPs, from 19.52 to 999.88 for blood metabolites/metabolite ratios, and from 21.07 to 97.06 for RPL. All F-statistics exceeded the conventional threshold of 10, indicating the absence of weak instrument bias and ensuring the reliability of subsequent causal inference. Of the 473 gut microbiota, 28 gut microbiota were found to be significantly associated with RPL. Among them, the abundances of *Brevibacillaceae, Faecalicatena sp001517425, Photobacterium, Staphylococcus aureus, Syntrophomonadia, Alloprevotella*, etc were connected with an increasing risk of RPL [odds ratio (OR) > 1.000, $P < 0.05$]. The others such as *Fibrobacterales, Gluconobacter, Pararhizobium, Ruminococcus, CAG-495*, and *UBA2658 sp002841545* were identified as potential protective factors for RPL (OR < 1.000, $P < 0.05$) (Figure S1).

Among 1091 blood metabolites and 309 metabolite ratios examined, suggestive causal associations with RPL were identified for 82 traits at a nominal threshold ($P < 0.05$). Among them, 37 blood metabolites and 12 metabolite ratios were identified as potential protective factors for RPL, such as levels of trimethylamine n-oxide, 3-hydroxyisobutyrate, phosphate, caprate (10:0), 12,13-DiHOME, and cysteine-glutathione disulfide (OR < 1.000, $P < 0.05$); while 27 blood metabolites and 6 metabolite ratios were recognized as risk factors for RPL, such as levels of 3-hydroxylaurate, 4-ethylphenylsulfate, 3-amino-2-piperidone, N-acetylvaline, X-23587, and cytidine to N-acetylneuraminate ratio (OR > 1.000, $P < 0.05$) (Figure S2). No heterogeneity or horizontal pleiotropy was observed in the findings (Table S2).

Potential Mediation Effect of Blood Metabolites in the Association Between Gut Microbiota and RPL

Based on the above results, an investigation was conducted to further explore the potential association between the 28 gut microbiota and 82 blood metabolites/metabolite ratios. A total of 114 significant associations were screened, comprising 60 potential risk factors and 54 potential protective factors (Figure S3). Subsequently, mediation analysis revealed that 3-amino-2-piperidone levels (GCST90200188) had a suggestive mediation effect ($\beta = 0.017$, $P = 0.0478$) on *Photobacterium* abundance (GCST90032511) and RPL with a mediation proportion of 14.4% (Figure 3A and Table 1). This suggestive mediation supports a potential mechanistic pathway linking *Photobacterium* abundance to RPL risk.

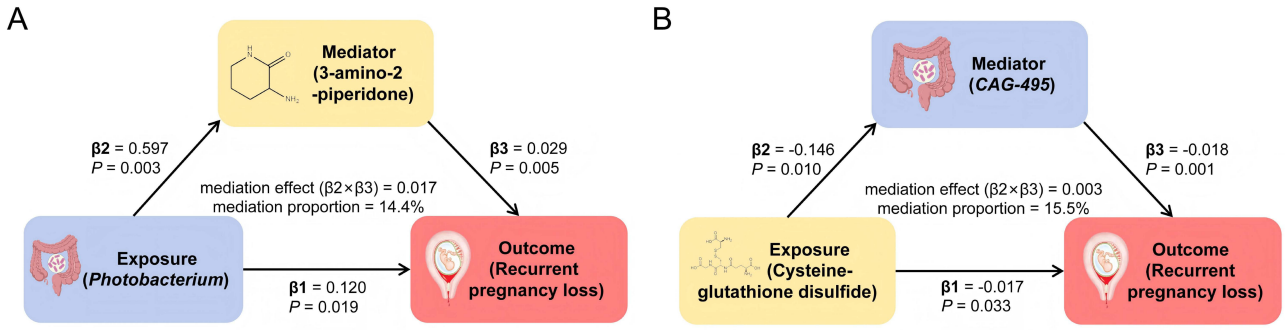

Figure 3 The illustrative diagram of mediation analysis. (**A**) Mediation effect of 3-amino-2-piperidone on the causality between *Photobacterium* abundance and RPL. (**B**) Mediation effect of *CAG-495* on the causality between cysteine-glutathione disulfide levels and RPL. Arrows indicate potential causal effect pathways: β2 (exposure → mediator), β3 (mediator → outcome), and β1 (exposure → outcome, total effect). The bold font indicates exposure factors, mediating variables, and outcomes.

Potential Mediation Effect of Gut Microbiota in the Association Between Blood Metabolites and RPL

The 82 blood metabolites/metabolite ratios were utilized as exposures, and 28 gut microbiota were designated as outcomes. A total of 120 nominally significant associations were identified, of which 63 showed potential protective associations with gut microbiota and 57 showed potential risk associations (Figure S4). Mediation analysis suggested that *CAG-495* abundance (GCST90032290) exhibited a suggestive mediating role in the pathway between cysteine-glutathione disulfide levels (GCST90199784) and RPL (β = 0.003, $P = 0.0497$), with a mediation proportion of 15.5% (Figure 3B and Table 2).

Reverse MR Analysis

Reverse MR analysis was implemented to study the potential reverse associations between RPL and gut microbiota as well as blood metabolites. The analysis showed no reverse associations between them (Figures S5 and S6).

Transcriptome Analysis Identified Key Genes Linked to Metabolites in RPL

To further elucidate the roles of 3-amino-2-piperidone and cysteine-glutathione disulfide in RPL, integrative transcriptome analysis was conducted. Multiple families of secondary metabolites originate from basic amino acids, such as lysine and ornithine, as well as from intermediates or derivatives of their biosynthetic pathways.[45] Considering that 3-amino-2-piperidone is a product of ornithine metabolism and that ornithine shares high structural similarity with lysine, differing only by one fewer carbon atom in its side chain,[46,47] we investigated 94 genes associated with the glutathione metabolism and lysine degradation pathways. Among these, 20 genes showed significant differences between RPL and control samples and were identified as the key genes (Table S3). Enrichment analysis unveiled that these key genes were associated with 241 GO terms, including glutathione metabolic process, cellular modified amino acid metabolic process, and antioxidant activity (Figure S7A). KEGG analysis identified 21 pathways, such as glutathione metabolism, lysine degradation, and arginine and proline metabolism, further supporting the metabolic roles suggested by the mediation analysis at the transcriptomic level (Figure S7A). Additionally, the PPI network of these key genes uncovered 20 nodes and 78 interaction pairs, with ALDH9A1, G6PD, and GCLM showing more interactions with other proteins, suggesting their potential key roles (Figure S7B).

Determination and Functional Analysis of Biomarkers (ASH1L, G6PD, SETDB1, and LAP3)

Three machine learning algorithms were integrated to simplify the most critical feature variables. LASSO identified 10 feature genes (Lambda.min = 0.006) including G6PD, GCLM, GSTO2, GSTP1, etc (Figure 4A). Boruta analysis determined 11 of the most important feature genes, such as G6PD, GCLM, GSTO1, etc (Figure 4B). The SVM-RFE algorithm identified 16 genes when the error rate was lowest at 0.0672, including SETDB1, G6PD, GSTP1, etc

https://doi.org/10.2147/IJWH.S598767

Table 1 Mediation Effect of 3-Amino-2-Piperidone on the Causality Between *Photobacterium* Abundance and RPL

Outcome (Y)	Exposure (X)	Mediator (M)	id. exposure	id. outcome	beta	nsnp	*P* value	OR (95% CI)	pleio_*P*	Mediation Effect	Direct Effect	Mediation Proportion
RPL	*Photobacterium* abundance in stool	3-amino-2-piperidone levels	X	Y	β1	13	0.019	1.128 (1.020 to 1.247)	0.402	β2×β3	β1-β2×β3	β2×β3/β1
GCST011887	GCST90032511	GCST90200188	X	M	β2	14	0.003	1.816 (1.227 to 2.690)	0.080	1.017 (1.000 to 1.035)	1.109 (1.001 to 1.228)	0.142 (0.001 to 0.282)
			M	Y	β3	23	0.005	1.029 (1.009 to 1.050)	0.809	*P* value=0.0478	*P* value=0.0475	*P* value=0.0478

Abbreviations: RPL, recurrent pregnancy loss; nsnp, number of single nucleotide polymorphisms; OR, odds ratio; 95% CI, 95% confidence interval.

https://doi.org/10.2147/IJWH.S599767

Table 2 Mediation Effect of *CAG-495* on the Causality Between Cysteine-Glutathione Disulfide Levels and RPL

Outcome (Y)	Exposure (X)	Mediator (M)	id.exposure	id.outcome	beta	nsnp	*P* value	OR (95% CI)	pleio_*P*	Mediation Effect	Direct effect	Mediation Proportion
RPL	Cysteine-glutathione disulfide levels	*CAG-495* abundance in stool	X	Y	β1	26	0.033	0.984 (0.969 to 0.999)	0.898	β2×β3	β1-β2×β3	β2×β3/β1
GCST011887	GCST90199784	GCST90032290	X	M	β2	25	0.010	0.865 (0.774 to 0.966)	0.594	1.003 (1.000 to 1.005)	0.981 (0.966 to 0.996)	0.160 (0.000 to 0.320)
			M	Y	β3	23	0.001	0.982 (0.971 to 0.993)	0.559	*P* value=0.0497	*P* value=0.0148	*P* value=0.0497

Abbreviations: RPL, recurrent pregnancy loss; nsnp, number of single nucleotide polymorphisms; OR, odds ratio; 95% CI, 95% confidence interval.

https://doi.org/10.2147/IJWH.S598767

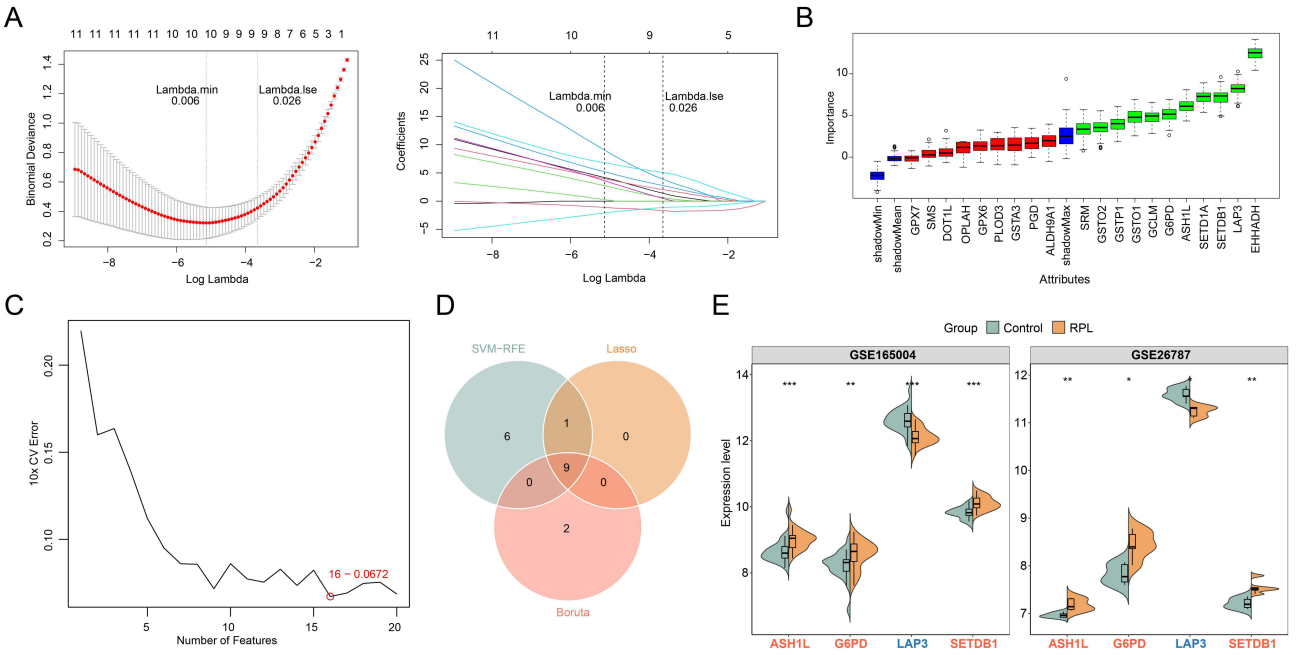

Figure 4 Machine learning algorithms identified biomarkers for RPL. (**A**) Feature gene selection using the LASSO algorithm. (**B**) Feature gene selection using the Boruta algorithm. (**C**) Feature gene selection using the SVM-RFE algorithm. (**D**) Cross-referencing the results from the three algorithms yielded 9 candidate biomarkers. (**E**) Validation of biomarker expression in independent datasets. Genes labeled in red were upregulated in the RPL group, and genes labeled in blue were downregulated in the RPL group. *P < 0.05; **P < 0.01; ***P < 0.001.

Abbreviations: LASSO, least absolute shrinkage and selection operator; SVM-RFE, support vector machine-recursive feature elimination.

(Figure 4C). By cross-referencing the results from the three algorithms, nine overlapping genes (G6PD, GCLM, GSTO2, GSTP1, LAP3, ASH1L, EHHADH, SETD1A, and SETDB1) were identified (Figure 4D). In GSE165004 and GSE26787, ASH1L, G6PD, and SETDB1 were significantly upregulated in RPL, while LAP3 was downregulated, showing consistent expression trends and serving as biomarkers (Figure 4E). Results from GSEA further indicated that ribosome and complement and coagulation cascades were commonly enriched among these biomarkers (Figure 5A–D and Table S4). Additionally, FC gamma R mediated phagocytosis and B-cell receptor signaling pathways were enriched in ASH1L, G6PD, and SETDB1, suggesting that these biomarkers may participate in the RPL process by regulating immune responses and cellular homeostasis.

Development of the Nomogram Based on Biomarkers

The nomogram was developed based on the four biomarkers to predict the risk of RPL. Each feature variable in this nomogram was assigned a unique score, and the total of all feature scores in each sample indicated the likelihood of RPL occurrence (Figure 6A). ROC analysis demonstrated that the nomogram had an area under the curve (AUC) of 0.972, indicating good diagnostic value (Figure 6B). The calibration curve unveiled that the nomogram ensured a high consistency between the predictions and actual observations (Figure 6C). These findings suggested that the nomogram based on these biomarkers could serve as an effective tool for predicting RPL risk.

Immune Cell Infiltration Patterns in RPL

To understand the immune microenvironment characteristics of RPL, CIBERSORT was implemented to analyze the infiltration patterns of 22 immune cell types (Figure 7A). Notably, regulatory T cells (Tregs) and M1 macrophages exhibited markedly variations between RPL and controls (Figure 7B). Correlation analysis further showed that SETDB1 had the strongest positive correlation with Tregs (cor = 0.473, P < 0.001), while ASH1L had the strongest negative correlation with M2 macrophages (cor = −0.457, P < 0.01), providing clues regarding the connection between biomarkers and immune regulation (Figure 7C).

Regulatory Landscape of Biomarkers

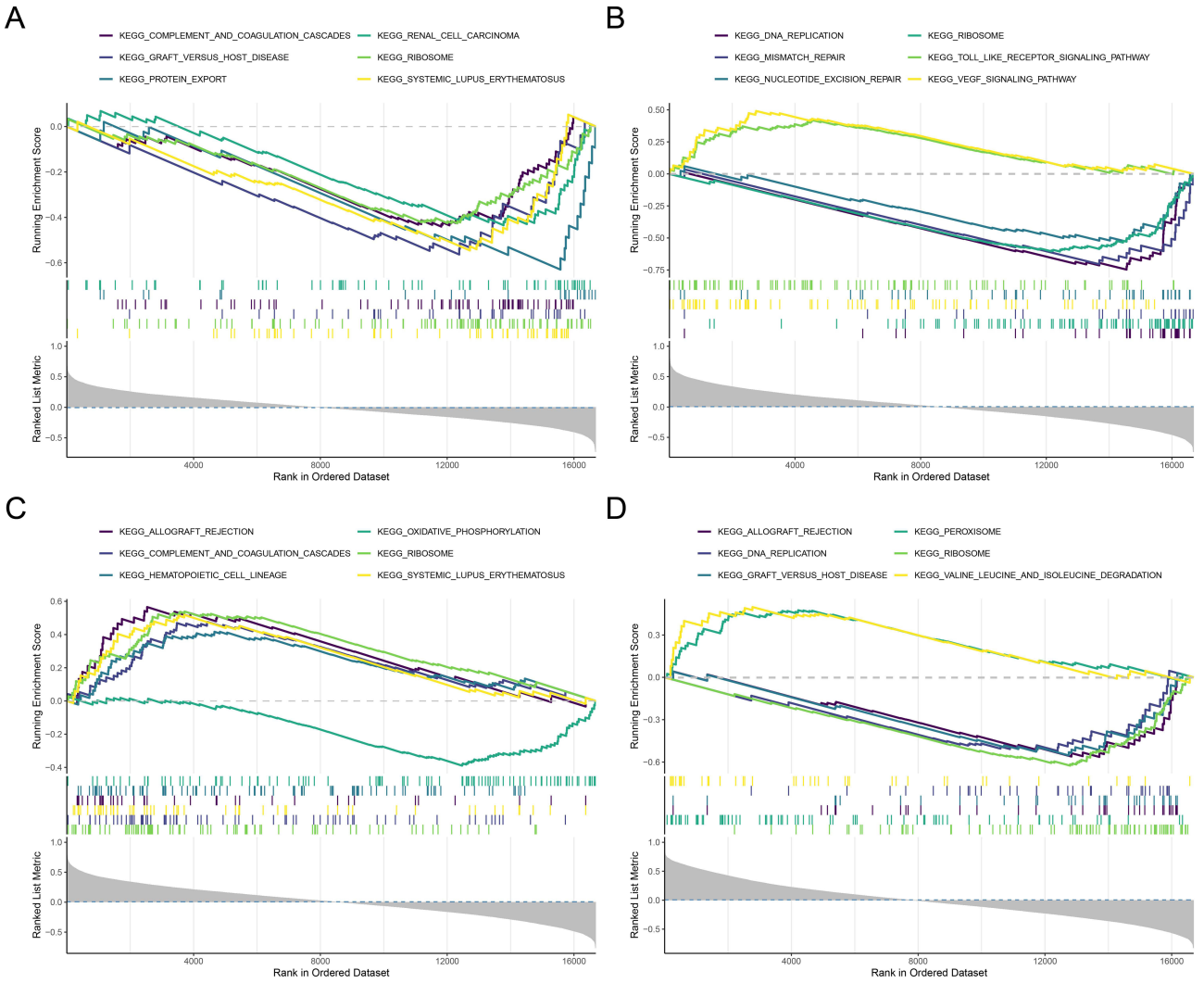

Figure 5 GSEA of the biomarkers. (**A-D**) GSEA of ASH1L (**A**), G6PD (**B**), LAP3 (**C**), and SETDB1 (**D**), respectively.
Abbreviation: GSEA, gene set enrichment analysis.

To scout the regulatory mechanisms of the biomarkers, TFs were predicted via the ChEA3 database, and a biomarker-TF regulatory network comprising 89 nodes and 116 relationships was constructed (Figure S8A). GABPA was predicted to regulate G6PD, SETDB1, and LAP3; while CTCF, EBF1, and EGR1 were predicted to jointly regulate SETDB1 and LAP3. Additionally, 54 miRNAs corresponding to 4 biomarkers and 125 miRNA-lncRNA interactions were predicted, and a biomarker-miRNA-lncRNA network involving 4 biomarkers, 19 lncRNAs, and 54 miRNAs was constructed (Figure S8B). This network revealed relationships such as ASH1L-hsa-miR-139-5p-AC084082.1, LAP3-hsa-miR-1297-MALAT1, and SETDB1-hsa-miR-1296-5p-LZTS1-AS1, depicting the complex regulatory landscape of the biomarkers.

Single-Cell Level Analysis of RPL and Revelation of Complex Communication

scRNA-seq was performed to further scout the features of RPL at the single-cell level. After quality control, 29,299 genes and 125,973 cells were included (Figure S9A). Subsequently, all cells were classified into 20 clusters using UMAP, identifying 14 distinct cell types [B cells, T cells, dendritic cells, decidual macrophages (dM), mast cells, endothelial cells, epithelial cells, syncytiotrophoblast (SCT), villous cytotrophoblast (VCT), extravillous trophoblast (EVT), perivascular (PV), decidual stromal cells (DSC), decidual natural killer cells (dNK), and red blood cells (RBC)] (Figures S9B, C and 8A). Moreover, after removing 9448 (7.5%) high-confidence doublets, the cell boundaries became clearer, reducing the interference of technical noise with biological signals (Figure S9D). Among the 14 annotated cell types, dM

https://doi.org/10.2147/IJWH.S598767

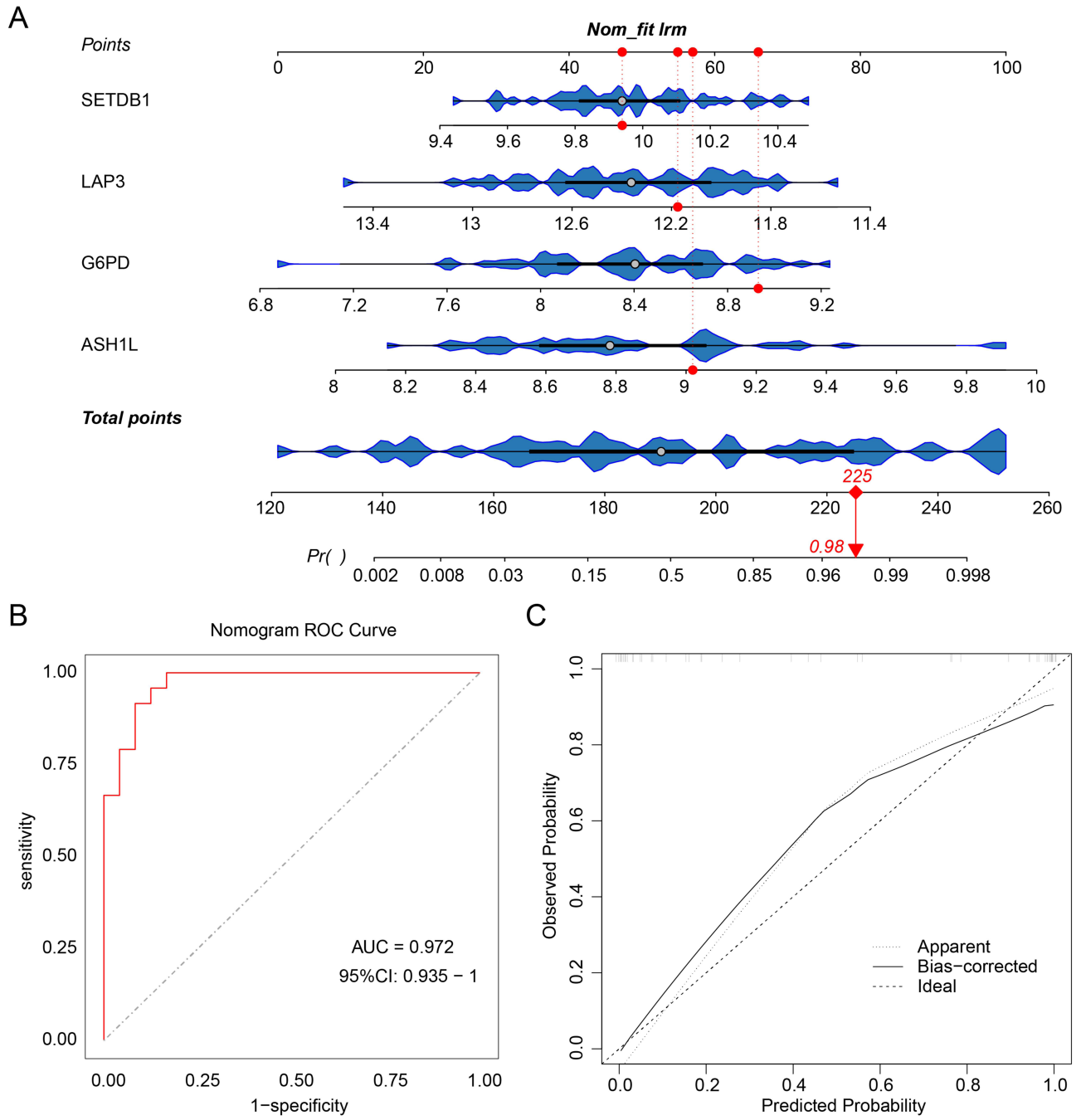

Figure 6 Development and validation of the nomogram for RPL risk prediction. (**A**) The nomogram was developed based on 4 biomarkers to predict the risk of RPL occurrence. The blue area shows the overall distribution of each variable. The gray dots are marked on the black median line, precisely indicating the specific value of the median. The red solid points represent the scores corresponding to the expression levels. The red quadrilateral and arrow at the bottom indicate the total score and its corresponding prediction probability. (**B**) ROC curve of the nomogram. (**C**) Calibration curve of the nomogram. **Abbreviations**: ROC, receiver operating characteristic; AUC, area under the curve.

and dNK had higher proportions in RPL (Figure 8B). Additionally, ASH1L was distributed across almost all cell types, while LAP3 was mainly found in dM (Figure 8C). Importantly, in dM, SCT, VCT, EVT, DSC, dNK, and RBC, significant expression differences of these biomarkers were observed between RPL and control groups (Figure 8D).

Cell communication patterns unveiled that in comparison with the control group, interactions and interaction frequencies between dM and VCT increased in RPL (Figure 9A). Overall, RPL exhibited stronger signal input, output,

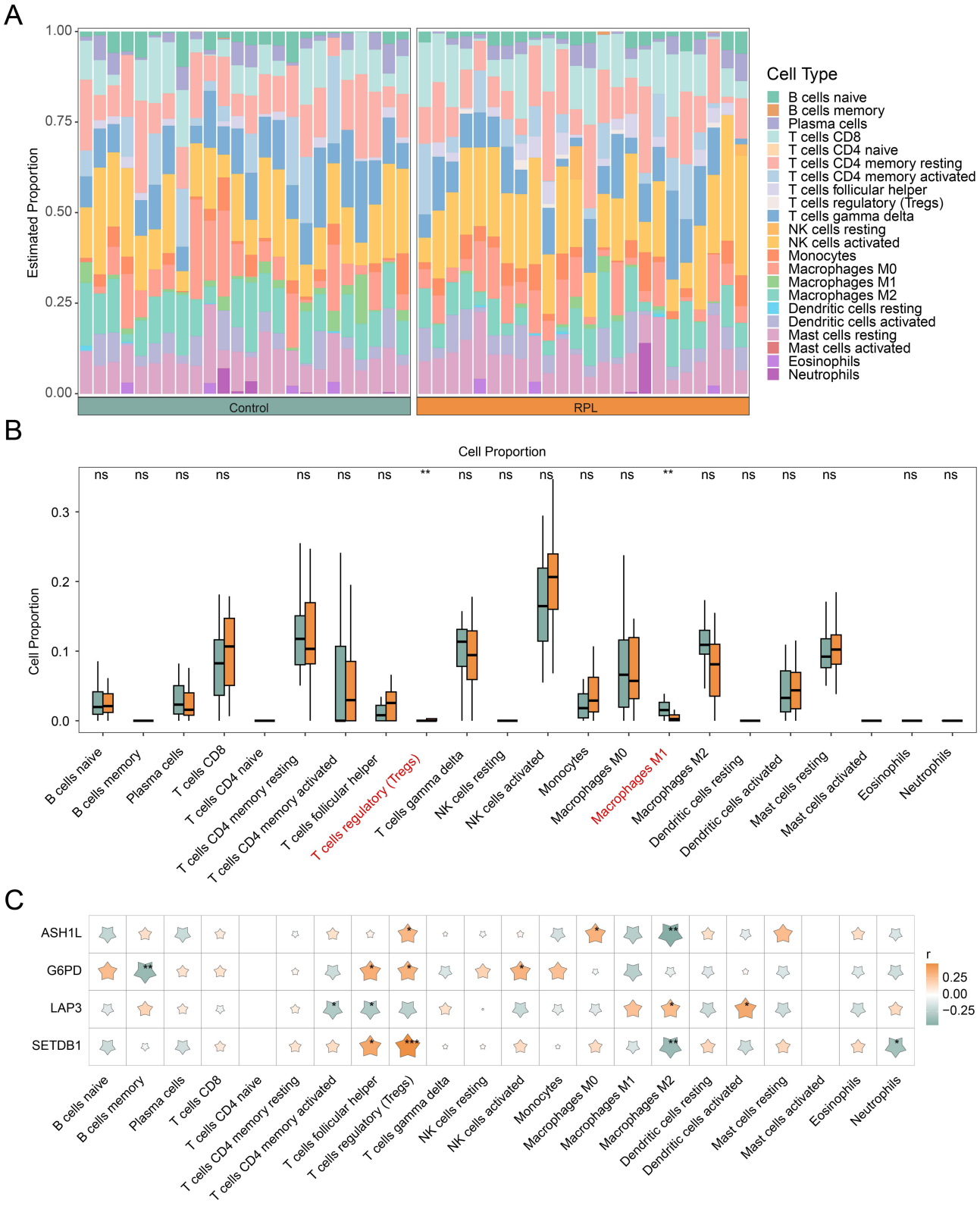

Figure 7 Immune cell infiltration landscape and its correlation with biomarkers in RPL. (**A**) Composition of 22 immune cell types in RPL and control samples. (**B**) Differences in immune cell infiltration between RPL and control samples. Green represents the control group and Orange represents the RPL group. Red text indicates cell subsets with significant differences in proportion between the two groups. ns, no significance; $^{**}P < 0.01$. (**C**) Correlation between biomarkers and immune cells. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

https://doi.org/10.2147/IJWH.S598767

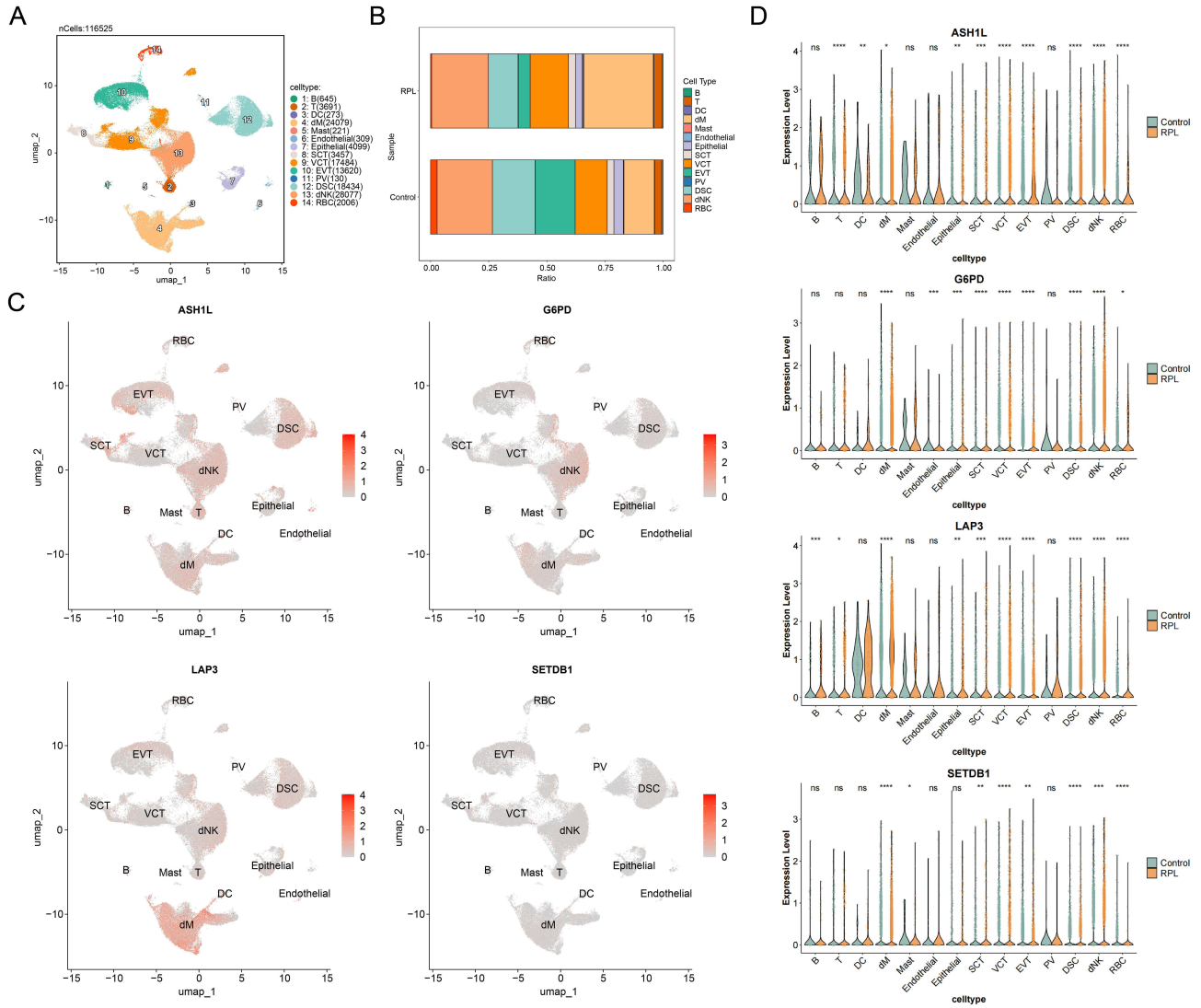

Figure 8 Single-cell analysis in RPL. (**A**) Annotation of cell types. (**B**) Proportion of cells in the RPL and control sample. (**C**) Distribution of biomarkers across cell types. (**D**) Expression of biomarkers across cell types in RPL vs. controls. ns, no significance; *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

and total communication patterns, which might lead to immune tolerance imbalance at the maternal-fetal interface (Figure 9B).

Discussion

Pregnancy is a long and arduous task for women, which involves a series of complex immune and metabolic regulatory mechanisms and is characterized by major shifts in maternal biology.[13] Currently, studies regarding RPL, especially unexplained RPL, are more focused on the failure of maternal-fetal crosstalk caused by immune factors.[6,8,16] Here, from the perspectives of microbiota and metabolome, our MR analysis suggested potential links within the gut microbiota–blood metabolites axis in RPL and further identified four biomarkers (ASH1L, G6PD, SETDB1, and LAP3) at the transcriptomic level, providing an entero-metabolic axis perspective to understand the pathogenesis of RPL.

Short-chain fatty acids and other metabolites synthesized by gut microbiota can enter the bloodstream and regulate the host's immune system together with endogenous metabolites,[48] while changes in metabolite abundance can also affect the composition and function of gut microbiota.[13] This interaction may play a unique and more significant role in the occurrence of RPL. Surprisingly, our mediation analysis suggested that *Photobacterium* might putatively increase the risk of RPL through the mediation of 3-amino-2-piperidone (mediation proportion = 14.4%, P = 0.0478).

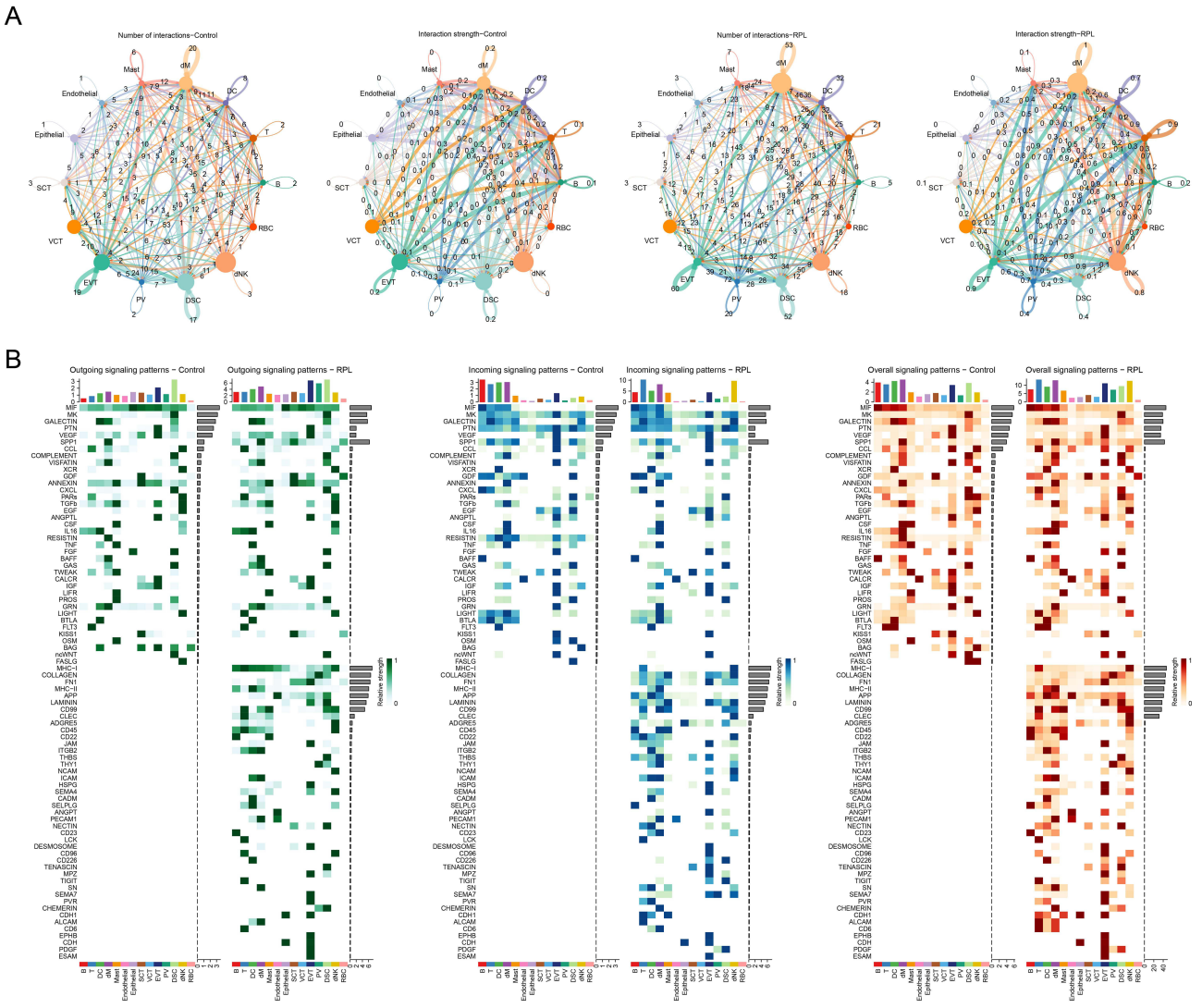

Figure 9 Cell communication analysis. (**A**) Cell communication across cell types. (**B**) RPL exhibited stronger signal input, output, and total communication patterns compared to the controls.

Photobacterium is a Gram-negative facultative anaerobic coccobacillus that is widely distributed in marine habitats and can infect marine humans through the consumption of fish, causing various primary diseases.[49,50] 3-amino-2-piperidone, an ornithine cycle-related metabolite which is probably involved in lysine metabolic pathways, has been found to promote the release of inflammatory factors.[47,51,52] Based on data from European populations, particularly fish-consuming Finns, our findings suggest that *Photobacterium* may enter the human body through diet and putatively influence RPL via immune regulation mediated by this metabolite. On the other hand, cysteine-glutathione disulfide, a non-endogenous metabolite formed under oxidative stress, has demonstrated protective properties in non-alcoholic fatty liver disease by influencing lipid metabolism.[53] Similarly, as an oxidized form of glutathione, it may putatively reduce RPL risk by lowering oxidative stress and inflammation,[54,55] although this protection was suggested to be partially weakened by the mediation of *CAG-495* (mediation proportion = 15.5%, $P = 0.0497$). Studies have found that the synthesis of the disulfide is regulated by gut microbiota in mice with adenomyosis,[56] and *CAG-495* was found to be increased during the remission phase of ulcerative colitis,[57] suggesting its putative regulatory role under certain pathological conditions.

To further explore the biological mechanisms of blood metabolites with mediating effects, we focused on the lysine degradation and glutathione metabolism pathways associated with these metabolites. Women with recurrent miscarriage

https://doi.org/10.2147/IJWH.S598767

have been reported to exhibit markedly reduced erythrocyte glutathione levels, particularly in those with autoimmune, unexplained, or luteal phase defect etiologies, indicating impaired antioxidant defense and elevated oxidative stress.[58] Glutathione depletion may promote lysine acetylation,[59] whereas lysine uptake enhances glutathione metabolism and reduces oxidative stress.[60] Accordingly, upregulation of lysine and significantly reduced levels of glutathione in placental tissues have been observed in the RPL group,[61,62] suggesting that lysine and glutathione metabolism probably play an indispensable role in RPL.

Subsequently, four biomarkers—ASH1L, G6PD, SETDB1, and LAP3 were identified. ASH1L is essential for controlling transcription and chromatin remodeling, as well as for promoting the methylation of certain histone lysine residues.[63] Recent studies have emphasized the pathogenic role of ASH1L in congenital malformations of the female genital tract.[64] In mouse models, ASH1L mutation leads to partial postnatal death, while the surviving mutant mice exhibit growth dysfunction and infertility due to defects in epididymis and uterus development.[65] In addition, ASH1L regulates the expression of p63 and p-CHK2 during early meiosis in mice, thereby protecting oocyte genome integrity and eliminating oocytes with severe DNA damage.[66] G6PD is the rate-limiting enzyme of the pentose phosphate pathway and is a key molecule for cells to resist oxidative damage.[67,68] Maternal G6PD deficiency has been proven to be fatal to embryos and causes severe placental abnormalities.[69] Additionally, children with G6PD deficiency showed increased oxidative damage to embryonic DNA, fetal mortality, and birth abnormalities when treated with the anticonvulsant medication phenytoin.[70] SETDB1 was originally thought to H3K9 in the nucleus, where it regulates chromatin function.[71] It is crucial for maintaining embryonic stem cell pluripotency and inhibiting trophoblast over-differentiation.[72] Maternal SETDB1 deficiency leads to preimplantation developmental arrest of embryos, accompanied by cell cycle and chromosome segregation abnormalities.[73] LAP3 is participated in the processing of bioactive peptides and the presentation of MHC-I antigens in mammals.[74] Its expression level significantly affects the development of sheep embryonic myoblasts.[75] In conclusion, these findings link specific biomarkers to metabolic disorders and RPL pathology, providing a gene-level understanding of the gut microbiota-metabolic axis mechanism of this disease.

GSEA revealed that the four biomarkers were commonly enriched in the complement and coagulation cascade pathways. Over-activation of this pathway has been reported to be related to RPL.[76] By expressing various regulatory proteins, human trophoblasts facilitate controlled complement activation, which is beneficial for spiral artery remodeling and the clearance of cell debris.[77] Crosstalk between the complement and coagulation cascades can prevent maternal rejection of the embryo and is favorable for maintaining normal pregnancy. However, excessive activation triggers an intrinsic immune feedback loop that simultaneously induces a compensatory anti-inflammatory response, which rapidly amplifies other targeted responses, thereby causing inflammation at the maternal-fetal interface or systemically and ultimately leading to miscarriage.[78] These results suggested that abnormal activation of the complement and coagulation cascades might significantly contribute to the pathogenesis of RPL.

Moreover, the interconnection between the coagulation system and immune system is particularly prominent under pathological conditions, further suggesting that immunity may be involved. During pregnancy, the maternal immune system undergoes significant changes to maintain maternal-fetal tolerance.[79] Our immune infiltration analysis unveiled marked variations in Tregs and M1 macrophages between RPL and controls. Notably, at the single-cell level, the dM showed more complex cell communication patterns in RPL. Abnormal polarization of dM is known to be associated with RPL,[80] and M1/M2 macrophage imbalance is considered one of the important causes of spontaneous abortion.[81] M1 macrophage-dominated pro-inflammatory responses (such as high expression of TNF-α) are markedly connected with the occurrence of RPL.[82,83] On the other hand, Tregs multiply following exposure to fetal antigens and are crucial for sustaining maternal-fetal immunological tolerance.[84] They often proliferate at the maternal-fetal contact and in peripheral circulation during parturition.[85] Certain studies indicate that both elevated and diminished levels of Tregs may increase the incidence of miscarriage, exhibiting a U-shaped impact curve.[86] These findings highlight the immune imbalance in RPL. It is noteworthy that the gut microbiota and blood metabolites screened in this study may participate in the pathogenesis of RPL by modulating the immune microenvironment.

The advantages and innovations of this study are reflected in the following aspects. Firstly, the effect of microbiota and metabolism on human reproductive health is currently a widely-concerned topic, and to the best of our knowledge, this is the first MR study examining microbiota and metabolome to investigate potential causes of RPL. In this study, we

identified dozens of gut microbiota and blood metabolites as risk or protective factors for RPL. Secondly, mediation analysis provided support for potential mediating effects involving gut microbiota and blood metabolites. Lastly, by combining mediation MR analysis with transcriptomics, we not only identified potential biomarkers but also clarified that these molecules might mediate their effects through immunomodulatory mechanisms, thereby establishing a comprehensive gut microbiota-metabolites-immune axis in RPL.

However, our study has certain limitations. Firstly, although we utilized the largest GWAS datasets that currently available, all the data were from individuals of European ancestry. This reduced heterogeneity and restricted the generalizability of our findings to other populations at the same time. Further validation in diverse groups is needed to confirm their broader generalizability. Secondly, due to the limited number of SNPs reaching genome-wide significance, we adopted a relatively lenient *P*-value threshold. Although this approach is well-established in similar studies[53,87] and was supplemented multiple validation methods to ensure reliability, further large-scale randomized controlled trials are still warranted to enhance the validity of the findings. Thirdly, as with all MR investigations, our study cannot completely exclude horizontal pleiotropy or weak instrument bias, which may affect the stability of causal estimates. Additionally, the nomogram had potential overfitting risk due to the small control sample size in the training set. The scarcity of public RPL datasets limits further sample expansion and external validation, requiring cautious interpretation and future large-scale independent validation. Finally, the specific mechanisms of the obtained biomarkers also need to be functionally verified through in vivo and in vitro experiments.

Conclusion

MR and mediation analyses unveiled the potential associations of gut microbiota and blood metabolites with RPL and two mediation pathways: *Photobacterium*-3-amino-2-piperidone-RPL and cysteine-glutathione disulfide-*CAG-495*-RPL. Integrative transcriptome analysis further identified four candidate key biomarkers involved in these pathways: ASH1L, G6PD, SETDB1, and LAP3. Additionally, our study revealed that these biomarkers and metabolic pathways converged to regulate immune homeostasis. Notably, this study is hypothesis-generating rather than definitive, and all findings should be considered preliminary and suggestive. In summary, our study identifies candidate microbial, metabolic, and candidate biomarkers with putative diagnostic and therapeutic potential for RPL, supporting the role of the gut microbiota–metabolites–immune axis. These hypothesis-generating findings may enhance the translational relevance of RPL research but require validation in future studies.

Abbreviations

RPL, recurrent pregnancy loss; MR, Mendelian randomization; GSEA, gene set enrichment analysis; scRNA-seq, single-cell RNA sequencing; ESHRE, European Society of Human Reproduction and Embryology; PCOS, polycystic ovary syndrome; IVW, inverse variance weighted; GEO, Gene Expression Omnibus; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; LASSO, least absolute shrinkage and selection operator; SVM-RFE, support vector machine-recursive feature elimination; ROC, receiver operating characteristic; miRNAs, microRNAs; LOO, leave-one-out; OR, odds ratio; AUC, area under the curve; dM, decidual macrophages; SCT, syncytiotrophoblast; VCT, villous cytotrophoblast; EVT, extravillous trophoblast; PV, perivascular; DSC, decidual stromal cells; dNK, decidual natural killer cells; RBC, red blood cells.

Data Sharing Statement

Data supporting the findings of this study are publicly available from the NHGRI-EBI GWAS Catalog database (https://www.ebi.ac.uk/gwas/) and GEO database (http://www.ncbi.nlm.nih.gov/geo/). The involved accession numbers are shown in the main text and supplementary tables.

Ethical Approval and Informed Consent

Based on national legislation guidelines in China, especially the item 1 and 2 of Article 32 of the Measures for Ethical Review of Life Science and Medical Research Involving Human Subjects dated February 18, 2023, the studies that (1) use legally obtained, publicly available data or data generated through observation without interference with public

behavior; (2) involve the use of anonymized information or data, are exempt from ethical approval. Hence, ethical approval and informed consent were waived for this study as it was conducted via utilizing summary GWAS and transcriptome data from published studies and consortia that have been made publicly available, which was in accordance with the aforementioned provisions.

Acknowledgments

The authors acknowledge the researchers and staff of the NHGRI-EBI GWAS Catalog database for the publicly available GWAS summary data. At the same time, we would like to express our gratitude to Dr. Weijing Meng for her contributions during the initial stage of this research, which included participation in the preliminary data collection and collation.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This study received no external funding.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Bender Atik R, Christiansen OB, Elson J, et al; The ESHRE Guideline Group on RPL. ESHRE guideline: recurrent pregnancy loss: an update in 2022. *Hum Reprod Open*. 2022;2023(1):hoad002. doi:10.1093/hropen/hoad002
2. Branch DW, Gibson M, Silver RM. Clinical practice. Recurrent miscarriage. *N Engl J Med*. 2010;363(18):1740–1747. doi:10.1056/NEJMcp1005330
3. Rai R, Regan L. Recurrent miscarriage. *Lancet*. 2006;368(9535):601–611. doi:10.1016/S0140-6736(06)69204-0
4. Li B, Xiong Y, Guo D, Deng G, Wu H. The gut-reproductive axis: bridging microbiota balances to reproductive health and fetal development. *Int Immunopharmacol*. 2025;144:113627. doi:10.1016/j.intimp.2024.113627
5. Chadchan SB, Singh V, Kommagani R. Female reproductive dysfunctions and the gut microbiota. *J Mol Endocrinol*. 2022;69(3):R81–R94. doi:10.1530/JME-21-0238
6. Liu Y, Chen H, Feng L, Zhang J. Interactions between gut microbiota and metabolites modulate cytokine network imbalances in women with unexplained miscarriage. *npj Biofilms Microbiomes*. 2021;7(1):24. doi:10.1038/s41522-021-00199-3
7. Jin M, Li D, Ji R, Liu W, Xu X, Feng X. Changes in gut microorganism in patients with positive immune antibody-associated recurrent abortion. *Biomed Res Int*. 2020;2020. doi:10.1155/2020/4673250
8. Yao Y, Ye Y, Chen J, Zhang M, Cai X, Zheng C. Maternal-fetal immunity and recurrent spontaneous abortion. *Am J Reprod Immunol*. 2024;91(5):e13859. doi:10.1111/aji.13859
9. Brown JA, Amir M, Yu S, et al. Gut microbiota promotes immune tolerance at the maternal-fetal interface. *Cell*. 2026;189(1):196–214.e24. doi:10.1016/j.cell.2025.11.022
10. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. *Cell*. 2014;157(1):121–141. doi:10.1016/j.cell.2014.03.011
11. Ivanov II, Atarashi K, Manel N, et al. Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell*. 2009;139(3):485–498. doi:10.1016/j.cell.2009.09.033
12. Vomstein K, Krog MC, Wrønding T, Nielsen HS. The microbiome in recurrent pregnancy loss - A scoping review. *J Reprod Immunol*. 2024;163:104251. doi:10.1016/j.jri.2024.104251
13. Amato KR, Pradhan P, Mallott EK, Shirola W, Lu A. Host-gut microbiota interactions during pregnancy. *Evol Med Public Health*. 2024;12(1):7–23. doi:10.1093/emph/eoae001
14. Li X, Yin M, Gu J, Hou Y, Tian F, Sun F. Metabolomic profiling of plasma samples from women with recurrent spontaneous abortion. *Med Sci Monit*. 2018;24:4038–4045. doi:10.12659/MSM.907653
15. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol*. 2015;181(4):251–260. doi:10.1093/aje/kwu283
16. Wu J, Cao Q, Liao J, et al. Immunological indicators of recurrent pregnancy loss: a mendelian randomization study. *Reprod Sci*. 2024;31(9):2783–2793. doi:10.1007/s43032-024-01555-2
17. Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomization: the STROBE-MR statement. *JAMA*. 2021;326(16):1614. doi:10.1001/jama.2021.18236

18. Qin Y, Havulinna AS, Liu Y, et al. Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. *Nat Genet*. 2022;54(2):134–142. doi:10.1038/s41588-021-00991-z
19. Chen Y, Lu T, Pettersson-Kymmer U, et al. Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. *Nat Genet*. 2023;55(1):44–53. doi:10.1038/s41588-022-01270-1
20. Laisk T, Soares ALG, Ferreira T, et al. The genetic architecture of sporadic and multiple consecutive miscarriage. *Nat Commun*. 2020;11(1):5980. doi:10.1038/s41467-020-19742-5
21. Z Y, L W, W M, W X, W S. Causal roles of skin and gut microbiota in skin appendage disorders suggested by genetic study. *Front Immunol*. 2024;15. doi:10.3389/fimmu.2024.1427276
22. Zhu A, Li P, Chu Y, et al. Causal effects of gut microbiota on the prognosis of ischemic stroke: evidence from a bidirectional two-sample Mendelian randomization study. *Front Microbiol*. 2024;15:1346371. doi:10.3389/fmicb.2024.1346371
23. Burgess S, Thompson SG; CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol*. 2011;40(3):755–764. doi:10.1093/ije/dyr036
24. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLOS Genet*. 2017;13(11):e1007081. doi:10.1371/journal.pgen.1007081
25. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife*. 2018;7:e34408. doi:10.7554/eLife.34408
26. Rasooly D, Patel CJ. Conducting a reproducible mendelian randomization analysis using the R analytic statistical environment. *Curr Protoc Hum Genet*. 2019;101(1):e82. doi:10.1002/cphg.82
27. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;40(4):304–314. doi:10.1002/gepi.21965
28. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *Int J Epidemiol*. 2016;45(6):dyw220. doi:10.1093/ije/dyw220
29. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*. 2017;46(6):1985–1998. doi:10.1093/ije/dyx102
30. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–287. doi:10.1089/omi.2011.0118
31. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1–22.
32. Kong C, Zhu Y, Xie X, Wu J, Qian M. Six potential biomarkers in septic shock: a deep bioinformatics and prospective observational study. *Front Immunol*. 2023;14:1184700. doi:10.3389/fimmu.2023.1184700
33. Sun L, Ren C, Leng H, et al. Peripheral blood mononuclear cell biomarkers for major depressive disorder: a transcriptomic approach. *Depress Anxiety*. 2024;2024:1089236. doi:10.1155/2024/1089236
34. Xing Y, Qin Y, Li X, et al. Enhancing prognostic guidance in renal light-chain amyloidosis: a new staging system incorporating pathological characters. *Int Urol Nephrol*. 2025;57(1):275–283. doi:10.1007/s11255-024-04182-7
35. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–457. doi:10.1038/nmeth.3337
36. Hao Y, Stuart T, Kowalski MH, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*. 2024;42(2):293–304. doi:10.1038/s41587-023-01767-y
37. Zhu W, Gu Y, Li M, et al. Integrated single-cell RNA-seq and DNA methylation reveal the effects of air pollution in patients with recurrent spontaneous abortion. *Clin Clin Epigenet*. 2022;14(1):105. doi:10.1186/s13148-022-01327-2
38. He YB, Li JY, Chen SL, et al. TRAF3 as a potential diagnostic biomarker for recurrent pregnancy loss: insights from single-cell transcriptomics and machine learning. *BMC Pregnancy Childbirth*. 2025;25(1):637. doi:10.1186/s12884-025-07742-6
39. Lin ZJ, Zhu L, Dong Y, et al. Integrated analysis of WES and scRNA-seq data reveals the genetic basis of immune dysregulation in unexplained recurrent pregnancy loss. *J Clin Lab Anal*. 2025;39(6):e70011. doi:10.1002/jcla.70011
40. Du L, Deng W, Zeng S, et al. Single-cell transcriptome analysis reveals defective decidua stromal niche attributes to recurrent spontaneous abortion. *Cell Prolif*. 2021;54(11):e13125. doi:10.1111/cpr.13125
41. Wei P, Dong M, Bi Y, et al. Identification and validation of a signature based on macrophage cell marker genes to predict recurrent miscarriage by integrated analysis of single-cell and bulk RNA-sequencing. *Front Immunol*. 2022;13:1053819. doi:10.3389/fimmu.2022.1053819
42. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst*. 2019;8(4):329–337.e4. doi:10.1016/j.cels.2019.03.003
43. Huang D, Jiao X, Huang S, et al. Analysis of the heterogeneity and complexity of murine extraorbital lacrimal gland via single-cell RNA sequencing. *Ocul Surf*. 2024;34:60–95. doi:10.1016/j.jtos.2024.06.005
44. Liu Z, Zhou X, Kuang L, et al. Novel insights into immune-gut microbiota interactions in colorectal cancer: a Mendelian randomization study. *Infect Agent Cancer*. 2025;20(1):27. doi:10.1186/s13027-025-00653-3
45. Liras P, Martín JF. Interconnected set of enzymes provide lysine biosynthetic intermediates and ornithine derivatives as key precursors for the biosynthesis of bioactive secondary metabolites. *Antibiotics*. 2023;12(1):159. doi:10.3390/antibiotics12010159
46. Simsek B, Çakatay U. Could ornithine supplementation be beneficial to prevent the formation of pro-atherogenic carbamylated low-density lipoprotein (c-LDL) particles? *Med Hypotheses*. 2019;126:20–22. doi:10.1016/j.mehy.2019.03.004
47. Li T, Ning N, Li B, et al. Longitudinal metabolomics reveals ornithine cycle dysregulation correlates with inflammation and coagulation in COVID-19 severe patients. *Front Microbiol*. 2021;12:723818. doi:10.3389/fmicb.2021.723818
48. Wastyk HC, Fragiadakis GK, Perelman D, et al. Gut-microbiota-targeted diets modulate human immune status. *Cell*. 2021;184(16):4137–4153.e14. doi:10.1016/j.cell.2021.06.019
49. Moi IM, Roslan NN, Leow ATC, et al. The biology and the importance of Photobacterium species. *Appl Microbiol Biotechnol*. 2017;101(11):4371–4385. doi:10.1007/s00253-017-8300-y
50. Hayano S, Masaki T, Tadakuma R, Kashima M. Photobacterium damselae subsp. damselae bacteraemia in a patient with liver cirrhosis. *BMJ Case Rep*. 2021;14(6):e242580. doi:10.1136/bcr-2021-242580

51. Luo S, Yin J, Zhang J, et al. Genetically predicted leucine level mediates association between CD4/CD8br T lymphocytes and insomnia. *Cell Mol Neurobiol*. 2025;45(1):15. doi:10.1007/s10571-025-01533-5
52. Sun S, Chen M, Zhang T, et al. Identification of key factors in cartilage tissue during the progression of osteoarthritis using a non-targeted metabolomics strategy. *Phenomics*. 2024;4(3):227–233. doi:10.1007/s43657-023-00123-z
53. Ouyang C, Liu P, Liu Y, Lan J, Liu Q. Metabolites mediate the causal associations between gut microbiota and NAFLD: a Mendelian randomization study. *BMC Gastroenterol*. 2024;24(1):244. doi:10.1186/s12876-024-03277-w
54. Wu G, Lupton JR, Turner ND, Fang YZ, Yang S. Glutathione metabolism and its implications for health. *J Nutr*. 2004;134(3):489–492. doi:10.1093/jn/134.3.489
55. Ballatori N, Krance SM, Notenboom S, Shi S, Tieu K, Hammond CL. Glutathione dysregulation and the etiology and progression of human diseases. *Bchm*. 2009;390(3):191–214. doi:10.1515/BC.2009.033
56. Chen P, Wang K, Zhuang M, et al. An insight into gut microbiota and metabolites in the mice with adenomyosis. *Front Cell Infect Microbiol*. 2023;13:1075387. doi:10.3389/fcimb.2023.1075387
57. Zhu R, Tang J, Xing C, et al. The distinguishing bacterial features from active and remission stages of ulcerative colitis revealed by paired fecal metagenomes. *Front Microbiol*. 2022;13:883495. doi:10.3389/fmicb.2022.883495
58. Vural P, Akgül C, Yildirim A, Canbaz M. Antioxidant defence in recurrent abortion. *Clin Chim Acta*. 2000;295(1–2):169–177. doi:10.1016/s0009-8981(99)00255-7
59. Pehar M, Ball LE, Sharma DR, et al. Changes in protein expression and lysine acetylation induced by decreased glutathione levels in astrocytes. *Mol Cell Proteomics*. 2016;15(2):493–505. doi:10.1074/mcp.M115.049288
60. Olin-Sandoval V, Jsl Y, Miller-Fleming L, et al. Lysine harvesting is an antioxidant strategy and triggers underground polyamine metabolism. *Nature*. 2019;572(7768):249–253. doi:10.1038/s41586-019-1442-6
61. Ye X, Ma C, Guo W, et al. Metabolomic analysis reveals potential role of immunometabolism dysregulation in recurrent pregnancy loss. *Front Endocrinol*. 2024;15:1476774. doi:10.3389/fendo.2024.1476774
62. Ghneim HK, Alshebly MM. Biochemical markers of oxidative stress in saudi women with recurrent miscarriage. *J Korean Med Sci*. 2016;31(1):98. doi:10.3346/jkms.2016.31.1.98
63. Zhang C, Xu L, Zheng X, Liu S, Che F. Role of Ash1l in Tourette syndrome and other neurodevelopmental disorders. *Dev Neurobiol*. 2021;81(2):79–91. doi:10.1002/dneu.22795
64. Qiu JJ, Chang XY, Zhang N, et al. Genetic variation and molecular profiling of congenital malformations of the female genital tract based on whole-genome sequencing. *World J Pediatr*. 2024;20(11):1179–1195. doi:10.1007/s12519-024-00839-6
65. Brinkmeier ML, Geister KA, Jones M, Waqas M, Maillard I, Camper SA. The histone methyltransferase gene absent, small, or homeotic discs-1 like is required for normal hox gene expression and fertility in mice. *Biol Reprod*. 2015;93(5):121. doi:10.1095/biolreprod.115.131516
66. Zhang T, Ren T, Lin H, et al. ASH1L contributes to oocyte apoptosis by regulating DNA damage. *Am J Physiol Cell Physiol*. 2022;323(4):C1264–C1273. doi:10.1152/ajpcell.00196.2022
67. Meng Q, Zhang Y, Hao S, et al. Recent findings in the regulation of G6PD and its role in diseases. *Front Pharmacol*. 2022;13:932154. doi:10.3389/fphar.2022.932154
68. Luzzatto L, Arese P. Favism and glucose-6-phosphate dehydrogenase deficiency. *N Engl J Med*. 2018;378(1):60–71. doi:10.1056/nejmra1708111
69. Longo L, Vanegas OC, Patel M, et al. Maternally transmitted severe glucose 6-phosphate dehydrogenase deficiency is an embryonic lethal. *EMBO J*. 2002;21(16):4229–4239. doi:10.1093/emboj/cdf426
70. Nicol CJ, Zielenski J, Tsui LC, Wells PG. An embryoprotective role for glucose-6-phosphate dehydrogenase in developmental oxidative stress and chemical teratogenesis. *FASEB J*. 2000;14(1):111–127. doi:10.1096/fasebj.14.1.111
71. Rapone R, Del Maestro L, Bouyioukos C, et al. The cytoplasmic fraction of the histone lysine methyltransferase Setdb1 is essential for embryonic stem cells. *iScience*. 2023;26(8):107386. doi:10.1016/j.isci.2023.107386
72. Lohmann F, Loureiro J, Su H, et al. KMT1E mediated H3K9 methylation is required for the maintenance of embryonic stem cells by repressing trophectoderm differentiation. *Stem Cells*. 2010;28(2):201–212. doi:10.1002/stem.278
73. Eymery A, Liu Z, Ozonov EA, Stadler MB, Peters AHFM. The methyltransferase Setdb1 is essential for meiosis and mitosis in mouse oocytes and early embryos. *Development*. 2016;143(15):2767–2779. doi:10.1242/dev.132746
74. Matsui M, Fowler JH, Walling LL. Leucine aminopeptidases: diversity in structure and function. *Biol Chem*. 2006;387(12):1535–1544. doi:10.1515/BC.2006.191
75. Ge L, Su P, Wang S, et al. New insight into the role of the leucine aminopeptidase 3 (LAP3) in cell proliferation and myogenic differentiation in sheep embryonic myoblasts. *Genes*. 2022;13(8):1438. doi:10.3390/genes13081438
76. Girardi G, Lingo JJ, Fleming SD, Regal JF. Essential role of complement in pregnancy: from implantation to parturition and beyond. *Front Immunol*. 2020;11:1681. doi:10.3389/fimmu.2020.01681
77. Abeln M, Albers I, Peters-Bernard U, et al. Sialic acid is a critical fetal defense against maternal complement attack. *J Clin Invest*. 2019;129(1):422–436. doi:10.1172/JCI99945
78. Liu K, Xu X, Sun L, et al. Proteomics profiling reveals lipid metabolism abnormalities during oogenesis in unexplained recurrent pregnancy loss. *Front Immunol*. 2024;15:1397633. doi:10.3389/fimmu.2024.1397633
79. Sun F, Wang S, Du M. Functional regulation of decidual macrophages during pregnancy. *J Reprod Immunol*. 2021;143:103264. doi:10.1016/j.jri.2020.103264
80. Wang L, Wang H, Luo J, Xie T, Mor G, Liao A. Decorin promotes decidual M1-like macrophage polarization via mitochondrial dysfunction resulting in recurrent pregnancy loss. *Theranostics*. 2022;12(17):7216–7236. doi:10.7150/thno.78467
81. Tsao FY, Wu MY, Chang YL, Wu CT, Ho HN. M1 macrophages decrease in the deciduae from normal pregnancies but not from spontaneous abortions or unexplained recurrent spontaneous abortions. *J Formos Med Assoc*. 2018;117(3):204–211. doi:10.1016/j.jfma.2017.03.011
82. Feng J, Gao P, Wu T, Hou W, Zhang Y, Li L. Imbalance polarization of M1/M2 macrophages in miscarried uterus. *PLoS One*. 2024;19(7):e0304590. doi:10.1371/journal.pone.0304590
83. Xu Y, Romero R, Miller D, et al. An M1-like macrophage polarization in decidual tissue during spontaneous preterm labor that is attenuated by rosiglitazone treatment. *J Immunol*. 2016;196(6):2476–2491. doi:10.4049/jimmunol.1502055

84. Wang W, Zhou X, Zhang Y, et al. The characteristics of antigenic specificity of memory regulatory t cells in women with unexplained recurrent pregnancy loss. *J Reprod Immunol*. 2022;154:103694. doi:10.1016/j.jri.2022.103694
85. Salvany-Celades M, van der Zwan A, Benner M, et al. Three types of functional regulatory T cells control T cell responses at the human maternal-fetal interface. *Cell Rep*. 2019;27(9):2537–2547.e5. doi:10.1016/j.celrep.2019.04.109
86. Liu PC, Li JB, Huang YP, Zhang M, Yu SJ, Wu R. Overexpression of regulatory T cells in patients with unexplained recurrent pregnancy loss: friend or foe? *Front Med*. 2023;10:1244424. doi:10.3389/fmed.2023.1244424
87. Zhu F, Zhang P, Liu Y, et al. Mendelian randomization suggests a causal relationship between gut dysbiosis and thyroid cancer. *Front Cell Infect Microbiol*. 2023;13:1298443. doi:10.3389/fcimb.2023.1298443