

Article

Improved Prediction Accuracy for Late-Onset Preeclampsia Using cfRNA Profiles: A Comparative Study of Marker Selection Strategies

Akiha Nakano ^{1,2}, Kohei Uno ¹ and Yusuke Matsui ^{1,2,*} ¹ Biomedical and Health Informatics Unit, Nagoya University Graduate School of Medicine, Nagoya 461-8671, Japan; uno.kohei.z8@f.mail.nagoya-u.ac.jp (K.U.)² Institute for Glyco-Core Research (iGCORE), Nagoya University, Nagoya 461-8673, Japan

* Correspondence: matsui.yusuke.d4@f.mail.nagoya-u.ac.jp

Abstract: Background: Late-onset pre-eclampsia (LO-PE) remains difficult to predict because placental angiogenic markers perform poorly once maternal cardiometabolic factors dominate. **Methods:** We reanalyzed a publicly available cell-free RNA (cfRNA) cohort (12 EO-PE, 12 LO-PE, and 24 matched controls). After RNA-seq normalization, we derived LO-PE candidate genes using (i) differential expression and (ii) elastic-net feature selection. Predictive accuracy was assessed with nested Monte-Carlo cross-validation (10 × 70/30 outer splits; 5-fold inner grid-search for λ). **Results:** The best LO-PE elastic-net model achieved a mean \pm SD AUROC of 0.88 ± 0.08 and F1 of 0.73 ± 0.17 —substantially higher than an EO-derived baseline applied to the same samples (AUROC ≈ 0.69). Enrichment analysis highlighted immune-tolerance and metabolic pathways; three genes (HLA-G, IL17RB, and KLRC4) recurred across >50% of cross-validation repeats. **Conclusions:** Plasma cfRNA signatures can outperform existing EO-based screens for LO-PE and nominate biologically plausible markers of immune and metabolic dysregulation. Because the present dataset is small ($n = 48$) and underpowered for single-gene claims, external validation in larger, multicenter cohorts is essential before clinical translation.



Academic Editors: Elena Dragioti and Antigoni Sarantaki

Received: 21 March 2025

Revised: 6 May 2025

Accepted: 13 May 2025

Published: 16 May 2025

Citation: Nakano, A.; Uno, K.; Matsui, Y. Improved Prediction Accuracy for Late-Onset Preeclampsia Using cfRNA Profiles: A Comparative Study of Marker Selection Strategies. *Healthcare* **2025**, *13*, 1162. <https://doi.org/10.3390/healthcare13101162>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: late-onset preeclampsia; cell-free RNA; biomarkers; maternal immunity; machine learning; maternal factors

1. Introduction

Preeclampsia is a condition characterized by the new onset of hypertension and proteinuria—or organ dysfunction such as liver or kidney impairment—after 20 weeks of gestation. It is reported to occur in approximately 2–5% of pregnancies worldwide [1]. This disorder significantly increases morbidity and mortality for both mothers and fetuses and can lead to preterm delivery or severe complications (e.g., HELLP syndrome).

Traditionally, preeclampsia has been categorized into early-onset (occurring before 34 weeks of gestation) and late-onset (occurring at or after 34 weeks of gestation) forms. Early-onset preeclampsia is typically associated with marked placental insufficiency and vascular dysfunction and tends to present with more severe clinical outcomes. In contrast, late-onset preeclampsia is thought to be more influenced by maternal factors (obesity, hypertension, metabolic risks, etc.) [2,3]. Although late-onset preeclampsia is often regarded as relatively mild, it still raises the risk of maternal–fetal complications and frequently necessitates cesarean delivery or other medical interventions.

Currently, the only definitive cure for preeclampsia is delivery, and effective pharmacological interventions remain limited—especially for late-onset cases. For instance, low-dose aspirin has been shown to significantly reduce the incidence of early-onset preeclampsia (before 34 weeks), but meta-analyses suggest that this prophylactic effect is less pronounced in late-onset disease [4]. Hence, early risk stratification and management of late-onset preeclampsia remain crucial.

Several screening approaches have been proposed to enable early risk assessment of late-onset preeclampsia, combining maternal background factors (e.g., chronic hypertension, obesity, and history of diabetes), uterine artery Doppler measurements, and serum biomarkers (e.g., PlGF and sFlt-1). However, changes in placental-derived factors are less pronounced in late-onset cases than in early-onset cases, and predictive models relying solely on placental angiogenic factors often reach a sensitivity of around only 40% [2,5] and, in systematic reviews, rarely exceed an AUROC of 0.70 for late-onset disease [3]. Consequently, late-onset preeclampsia has proven more challenging to predict with high accuracy. While it has been noted that maternal factors (e.g., high BMI and advanced maternal age) play a major role in late-onset cases [3,6], the specific molecular mechanisms underlying this subtype have not yet been fully elucidated.

Recent literature points to two converging biological axes in late-onset PE—failure of maternal–fetal immune tolerance and exacerbated third-trimester metabolic stress. Placental and review data highlight diminished HLA-G/KIR signaling and a Th17/IL-17-skewed cytokine milieu in LO-PE [7,8], while large cohort and cfRNA studies report the early enrichment of Allograft Rejection and Estrogen-response pathways [9]. Complementarily, metabolic-syndrome traits—insulin resistance, dyslipidemia, and altered glycolytic flux—have been proposed as principal maternal drivers of LO-PE [10]. Guided by this evidence, we focus our cfRNA feature-selection and downstream interpretation on genes mapping to immune-tolerance and metabolic pathways.

In recent years, machine learning (ML) and artificial intelligence (AI) approaches have gained attention for their potential to integrate these complex risk factors multidimensionally and are now highlighted as promising avenues for PE prediction in dedicated reviews [11]. For example, analyses of large-scale electronic health records (EHRs) incorporating diverse maternal background and laboratory data have demonstrated high-accuracy prediction with an AUC exceeding 0.9 [6]. Nonetheless, most studies to date are retrospective and confined to specific cohorts and thus lack external validation or prospective evaluation. Although attempts to merge multi-omics data (e.g., genetic risk scores, proteomics, and metabolomics) have been reported [12], the cost and clinical feasibility remain significant barriers.

Parallel to the surge in AI-driven risk models, cell-free RNA sequencing (cfRNA-seq) of maternal plasma has become a leading avenue for non-invasive biomarker discovery. Recent high-throughput studies using this technique repeatedly implicate two pathophysiologic axes in late-onset PE—(i) the loss of maternal–fetal immune tolerance, exemplified by downregulated placental HLA-G/KIR checkpoints and a Th17-skewed cytokine milieu [7,8], and (ii) third-trimester metabolic stress, marked by systemic insulin-resistance signatures and enrichment of glycolysis- and estrogen-response gene sets [9,10]. A 2024 systematic review synthesized these observations into a dual-hit model in which rising maternal metabolic load amplifies incipient immune dysfunction to trigger clinical LO-PE [3].

Circulating cfRNA originates from both maternal blood cells and the placenta; trophoblast-derived transcripts are detectable from the first trimester and increase steadily with gestation. Unlike cfDNA, cfRNA captures the moment-to-moment transcriptional state of maternal–fetal tissues, providing time-resolved insight into immune, angiogenic,

and metabolic pathways [9]. Longitudinal, deep-coverage cfRNA-seq has shown that placental, endothelial, and leukocyte signatures begin to diverge weeks before symptom onset. Because LO-PE is driven more by maternal cardiometabolic stress and systemic inflammation than by early placental maldevelopment, cfRNA offers a unique window into these evolving maternal responses—signals that may be missed by placental protein biomarkers alone.

- **Early-Onset Preeclampsia (EO-PE):** This form is predominantly driven by placental abnormalities and immune dysregulation that begin early in gestation; distinct differential expression of cfRNA has been reported. For example, Moufarrej et al. demonstrated a high-accuracy model ($AUC \approx 0.9$) using cfRNA derived from maternal plasma, suggesting an impairment of immune response and angiogenic pathways [9].
- **Late-Onset Preeclampsia (LO-PE):** Maternal comorbidities such as obesity or chronic hypertension play a substantial role, often diminishing the utility of purely placental biomarkers for high-sensitivity prediction. Indeed, many studies investigating cfRNA- or metabolite-based tests focus on overall PE risk and do not provide separate metrics (e.g., AUC) for LO-PE alone. For example, while Maric et al. [13] report robust performance in predicting PE, their models do not isolate late-onset cases. As a result, the true accuracy for LO-PE remains unclear, and some data even suggest that maternal factors may overshadow direct placental signals, leading to potentially lower AUCs for late-onset compared to early-onset PE. Moving forward, it will be crucial to refine LO-PE-specific molecular signatures—possibly through multi-omics approaches integrated with maternal clinical data—and validate such signatures in large external cohorts. This line of research is expected to clarify whether dedicated LO-PE models can outperform current one-size-fits-all approaches and ultimately improve risk stratification in this patient population.

This study focuses on LO-PE and aims to (1) identify cfRNA-based biomarker candidates specific to LO-PE and (2) develop and evaluate machine learning models using these markers. More specifically, our objectives are:

1. To characterize cfRNA profiles in LO-PE and compare them with known markers predominantly associated with EO-PE.
2. To apply two feature selection strategies—(A) an approach based on differential expression analysis and (B) an approach leveraging prediction errors (via the elastic-net solution path)—and then assess LO-PE prediction performance in terms of AUC, sensitivity, and specificity.
3. To examine the performance trade-offs involved in simultaneously predicting both EO- and LO-PE, and to investigate how immune tolerance and metabolic pathways might be affected.

Ultimately, this study seeks to elucidate the mechanisms underlying late-onset preeclampsia—particularly those related to immune modulation and placental invasion—by leveraging cfRNA signatures, with the goal of informing future clinical management of preeclampsia.

2. Materials and Methods

This study analyzed cfRNA sequencing data from a total of 48 samples, comprising EO-PE, LO-PE, and corresponding control groups for each subtype. Our goal was to identify potential biomarkers specifically associated with LO-PE and then construct and evaluate a diagnostic prediction model. The overall analytical workflow is illustrated in Figure 1.

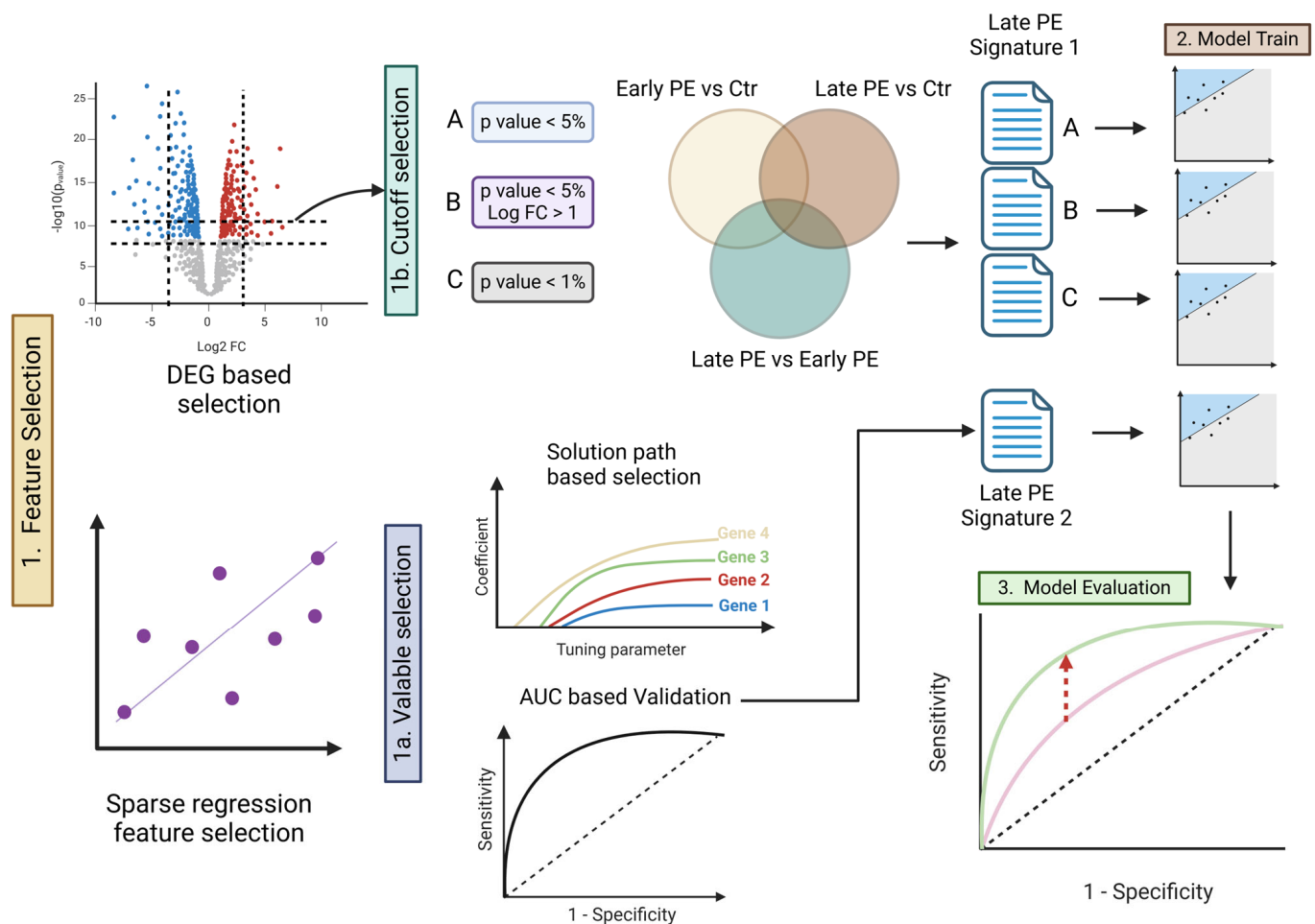


Figure 1. Schematic of the analytical workflow. Panel 1 (feature selection): cfrRNA features were prioritized by two complementary routes. 1a Sparse-regression (elastic net) solution paths identified genes whose coefficients remained non-zero across a 50-value λ grid, and the resulting models were validated by AUC. 1b Differential expression (DEG) analysis compared Early-PE vs. Control, Late-PE vs. Control, and Early- vs. Late-PE; three cut-off schemes were explored—A $p < 0.05$, B $p < 0.05$ and $|\log \text{FC}| > 1$, and C $p < 0.01$ —yielding signature candidates shown in the Venn diagram. In the volcano plot, red dots indicate significantly up-regulated genes, blue dots down-regulated genes, and grey dots non-significant genes. Panel 2 (model training): for each DEG-derived signature (A–C) and for the elastic-net signature, separate elastic-net classifiers were trained on the Late-PE training data. Panel 3 (model evaluation): nested Monte-Carlo cross-validation supplied independent test sets; ROC curves illustrate that Late-PE-specific signatures (green) outperform naïvely transferred Early-PE signatures (pink), while the diagonal denotes chance.

2.1. Dataset

The dataset is based on a cfrRNA cohort described in Reference [14], which includes 12 subjects with LO-PE, 12 subjects with EO-PE, and 12 controls for each group. All participants carried singleton pregnancies without structural fetal anomalies and were recruited prospectively at Stanford University Medical Center between 2014 and 2017 under IRB protocol #28979. Maternal blood was drawn at the time of clinical diagnosis (mean \pm SD gestational age: EO-PE 29.2 ± 2.3 weeks; matched EO-controls 29.3 ± 2.3 weeks; LO-PE 35.6 ± 1.3 weeks; matched LO-controls 35.9 ± 0.8 weeks). The 24 normotensive controls were selected from a single pool and randomly assigned 1:1 to the EO and LO subgroups so that gestational age matching with their respective case groups was preserved. Written informed consent for cfrRNA sequencing and data deposition in dbGaP (accession phs002017.v1) was obtained from every participant in accordance with the Declaration of

Helsinki. Blood samples were collected at the time of PE diagnosis, and cfRNA (cell-free RNA) was extracted from maternal plasma for Next-Generation Sequencing (NGS). Because the resulting RNA reads may include transcripts of both placental and maternal origin, it offers the intriguing possibility of capturing both maternal and placental factors. Given that our analysis involves a relatively small sample of 48 total specimens, special attention must be paid to sample-size limitations, the risk of overfitting, and the need for further external validation when constructing prediction models. Because the published study supplies an already aligned gene-count matrix, our workflow starts from this matrix; additional read-level QC or alignment was not required. Participant demographics are summarized in Table S1.

2.2. Strategy for Selecting Signature Genes

To identify biomarkers specific to LO-PE, we employed two approaches: differential expression gene (DEG) and an elastic-net-based machine learning method leveraging prediction error. Differential expression (DE) analysis tests whether the mean read count for each gene differs between conditions. First, we used RNA-seq count data to conduct three intergroup comparisons: “early-onset vs. control”, “late-onset vs. control”, and “early-onset vs. late-onset”. We then used edgeR [15,16] and limma [17,18] to extract genes showing statistically significant differential expression. This process involved adjusting the p -values via the Benjamini–Hochberg method [19] to control the false discovery rate (FDR) and using log2 fold change (logFC) values as an additional criterion for candidate gene selection. By excluding genes that were differentially expressed in both early- and late-onset groups, we obtained a set of candidate genes more specific to LO-PE.

Next, we applied an elastic-net regression model using the glmnet [20] package, which implements coordinate-descent optimization [20], to tackle two classification tasks—“control vs. early-onset” and “control vs. late-onset”. We optimized the model’s hyperparameter, λ , through cross-validation to maximize prediction performance (AUC) and simultaneously minimize the number of genes used. By examining the solution path, we extracted genes that contributed most significantly to predicting LO-PE and designated them as late-onset-specific signatures for subsequent functional analysis. Since the elastic net combines both L_1 (Lasso) and L_2 (Ridge) regularization, it effectively prevents overfitting and performs variable selection automatically. This makes it particularly useful for scenarios, such as ours, where one must narrow down important features from a large pool of genes.

2.3. Building and Evaluating the Predictive Model

Using the selected signatures, we constructed models to predict LO-PE and evaluated their classification accuracy. Model performance was assessed with a nested Monte-Carlo cross-validation (MC-CV) procedure. A stratified 70/30 split was repeated 10 times (outer loop); within each outer training set, a 5-fold inner CV tuned the elastic-net λ across a 50-value grid. We trained an elastic-net model separately on the “early-onset signature” and the “late-onset signature” and then computed the AUC to assess performance for the classifications “control vs. late-onset” and “control vs. early-onset”. In addition, we evaluated its performance when combining the early-onset and late-onset signatures to investigate whether handling both simultaneously would induce any performance trade-off. The AUC (Area Under the ROC Curve) serves as a comprehensive measure of a model’s ability to discriminate between true positives and false positives, with 1.0 indicating perfect accuracy and 0.5 indicating performance equivalent to random guessing. Where necessary, we also considered sensitivity and specificity to gain insight into the balance between false positives and false negatives.

2.4. Searching for Biomarker Candidates

We further investigated the late-onset signature genes extracted via prediction-error analysis by conducting gene set and pathway analyses to clarify their functional characteristics. Specifically, we cross-referenced the gene lists with databases such as Gene Ontology and KEGG [21] to statistically evaluate the enrichment of pathways related to metabolism, immunity, and other processes, using Fisher's exact test. Of particular interest were genes involved in immune tolerance or placental invasion, such as HLA-G and IL17RB; their inclusion in the signature could suggest associations with maternal immune dysregulation or trophoblast (EVT) dysfunction in LO-PE. We compared such findings against previous studies to explore their biological significance. Ultimately, this functional validation of late-onset-specific gene groups helps lay the groundwork for determining their potential clinical utility as diagnostic biomarkers in future research.

2.5. Performance Metrics

Classifier performance was evaluated with the following standard measures:

$$\text{AUROC} = \int_0^1 \text{TPR}(\alpha) d\alpha \quad (1)$$

$$\text{Sensitivity (Recall)} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad F_1 = \frac{2 \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (3)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. AUROC provides an overall measure of discrimination; the F_1 -score complements AUROC by balancing Precision and Sensitivity, which is useful when class sizes are imbalanced.

2.6. Model Evaluation

Finally, to benchmark the proposed elastic net against alternative learners, we repeated the entire nested Monte-Carlo pipeline with Random Forest, linear SVM, and XGBoost (hyper-parameters tuned in the inner loop). Mean \pm SD AUROC and F1 are reported in Supplementary Tables S3 and S4, and pooled ROC curves are shown in Supplementary Figures S2 and S3.

3. Results

3.1. Identification of Signature Genes and Feature Selection

Table S1 summarizes maternal age, body mass index (BMI), and gestational age at sampling for each study group. Groups did not differ in age or gestational week (all $p > 0.5$), whereas BMI was significantly higher in the LO-PE group compared with its matched controls ($p = 0.036$) (Supplementary Table S2). First, we performed differential expression analyses (DEG) for three comparisons—(1) early-onset PE vs. control, (2) late-onset PE vs. control, and (3) early-onset PE vs. late-onset PE—and generated lists of signature candidates by systematically varying the thresholds for p -values and log fold change (logFC) (Figure 1, panel 1a-1). We tested three cutoff conditions: (A) $p < 0.05$, (B) $p < 0.05$ and $|\log\text{FC}| > 1$, and (C) $p < 0.01$ (1a-1 in Figure 1). As shown in Figure 2A, the late-onset-specific signatures comprised 64 genes under condition A, 1 gene under condition B, and 7 genes under condition C. In contrast, the early-onset-specific signatures included 1334 genes (A), 11 genes (B), and 295 genes (C).

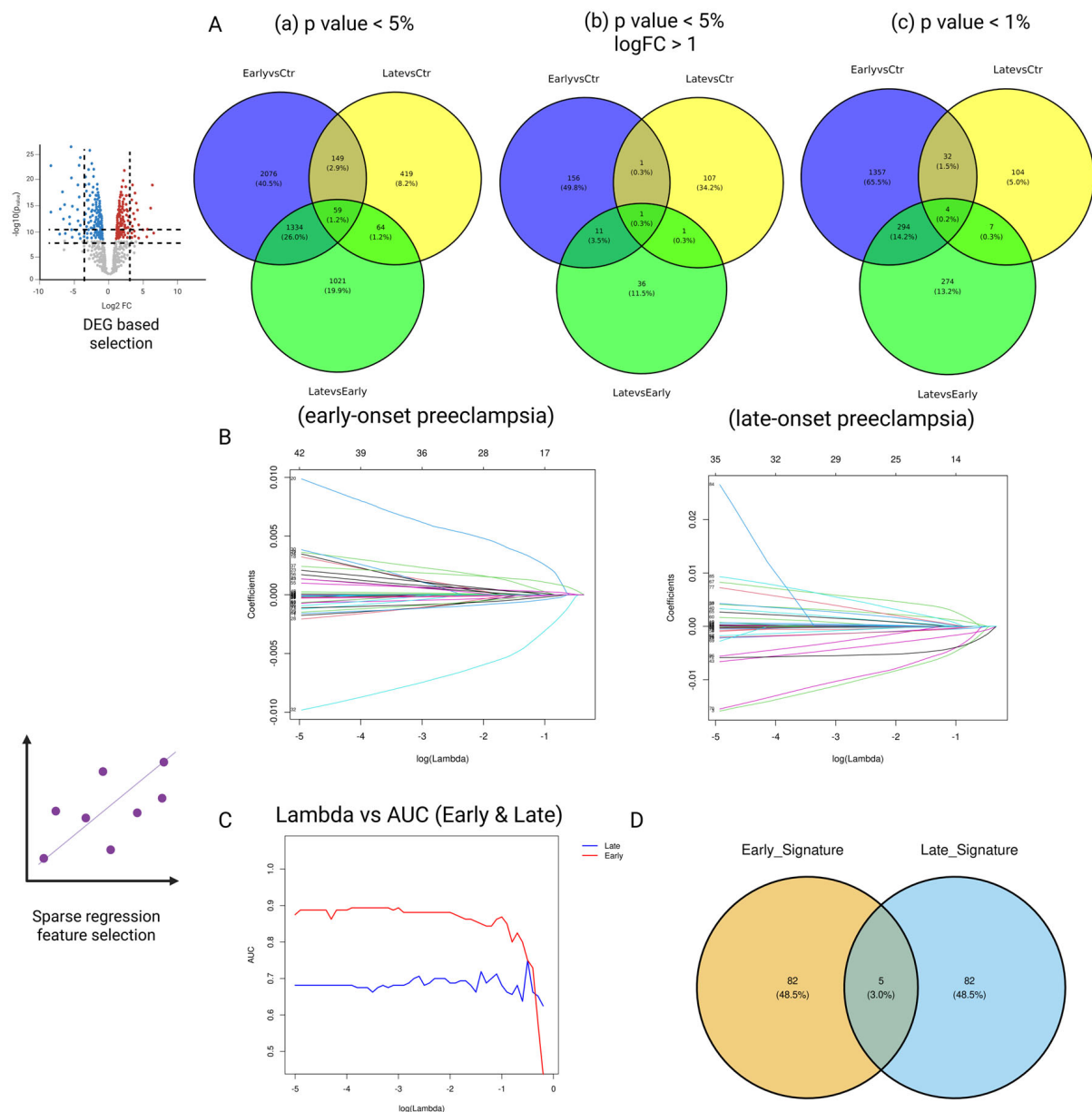


Figure 2. Multi-stage feature selection and signature derivation for early- (EO-PE) and late-onset pre-eclampsia (LO-PE). **(A)** Differential expression (DEG) stage. Venn diagrams show EO- and LO-specific gene sets obtained under three progressively stringent cut-offs: **(a)** $p < 0.05$; **(b)** $p < 0.05$ and $|\log_2 FC| > 1$; **(c)** $p < 0.01$. Numbers indicate genes unique to, or shared between, the three pairwise contrasts (Early vs. Ctrl, Late vs. Ctrl, Late vs. Early). **(B)** Elastic-net solution paths. For each subtype, the coefficients of all candidate genes are traced across the $\log_{10}(\lambda)$ grid; colored curves enter the model as λ decreases, illustrating automatic sparsification. **(C)** Model-selection curve. Mean outer-loop AUROC vs. $\log_{10}(\lambda)$ for EO- (blue) and LO-PE (red). Plateaus mark the λ values giving maximal discrimination with minimal complexity. **(D)** Final signatures. Intersecting the optimal EO and LO coefficient sets yields five shared genes (center), with 82 genes retained exclusively in each subtype-specific signature.

When we used these results to train an elastic-net model, the model that relied solely on the single gene KLRC4 (from condition B) attained an extremely high predictive performance for LO-PE ($AUC = 1.0$). A logistic model based on the single gene KLRC4 yielded an apparent $AUROC = 0.875$. A 1000-iteration label-permutation test produced a null $AUROC$ distribution centered at 0.874 ± 0.004 ; the observed $AUROC$ of 0.875 lay well within this

range (one-sided $p = 0.997$). The full distribution is visualized in Supplementary Figure S1, underscoring that the single-gene signal is indistinguishable from chance.

Benchmarking against three conventional classifiers confirmed the superiority of the elastic net: for EO-PE the elastic net reached AUROC 0.91 vs. 0.87 (RF), 0.83 (SVM), and 0.74 (XGBoost); for LO-PE, the corresponding figures were 0.83, 0.71, 0.73, and 0.61, respectively (Supplementary Tables S3 and S4; Supplementary Figures S2 and S3).

3.2. Candidate Selection via Prediction Error (Elastic-Net Solution Path)

Next, we performed cross-validation on the elastic-net model while varying the hyperparameter λ in 50 increments, adopting the parameter setting that maximized predictive performance (AUC) while minimizing the number of selected genes (1b in Figure 1). This approach extracted 52 genes as late-onset-specific signatures and 5 genes as early-onset-specific signatures (Figure 2B–D). These sets exhibited very little overlap with the signatures identified via DEG analysis. As shown in the Venn diagram in Figure 3, most genes from the solution path (SP)-based signatures and those from the DEG-based approach did not overlap for both early- and late-onset PE, indicating that the two methods complement each other.

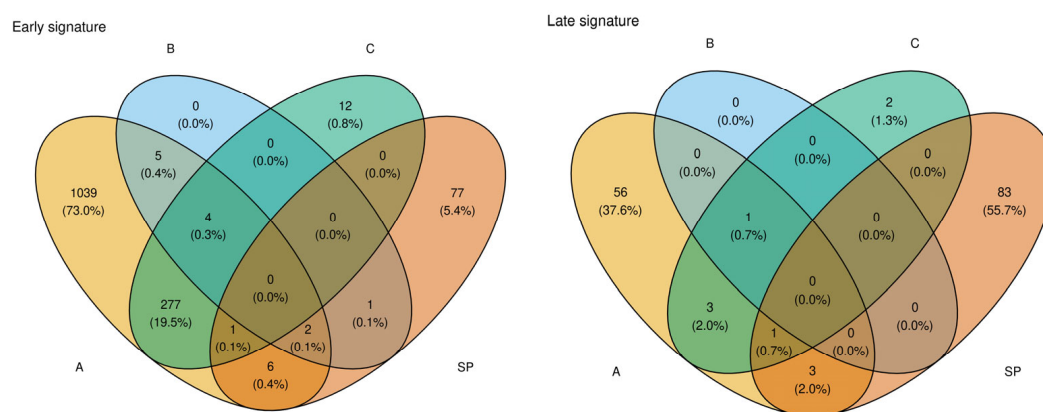


Figure 3. Overlap of gene signatures obtained with four selection strategies for (left) early-onset and (right) late-onset pre-eclampsia. Each Venn diagram contrasts the differential expression cut-offs A ($p < 0.05$), B ($p < 0.05$ and $|\log_2FC| > 1$), and C ($p < 0.01$) with the sparsity-optimized elastic-net solution path (SP). Numbers denote the absolute gene count and, in parentheses, the percentage of the total signature size for that subtype. The minimal intersection—five shared genes in the early set and three in the late set—underscores that the early- and late-onset signatures are largely distinct, supporting divergent molecular mechanisms between the two PE subtypes.

3.3. Comparative Performance of Prediction Models

Table 1 and Figure 4 (ROC curves) present the results of elastic-net regression models constructed and validated using each signature set (DEG-based: A, B, C; elastic-net-based: SP). As a baseline, we also examined models without any signature genes:

- Training on early-onset samples yielded an AUC of 0.9375 for predicting early-onset PE but only 0.6875 for predicting late-onset PE.
- Training on late-onset samples resulted in an AUC of 0.6875 for predicting late-onset PE.

Table 1. Discriminatory performance of gene-signature models for early- (EO-PE) and late-onset pre-eclampsia (LO-PE). Models were trained on either the EO or LO subset (columns 2–3) and evaluated on both subtypes (columns 4–5). “Prediction without signature” uses all expressed genes; the three DEG signatures correspond to cut-offs (A) $p < 0.05$, (B) $p < 0.05$ and $|\log_2\text{FC}| > 1$, and (C) $p < 0.01$; “Elastic-net signatures” are the sparse sets selected by nested Monte-Carlo cross-validation.

Prediction Without Signature		Training Data	Early	Late	Early	Late
		Early prediction AUC	0.9375	0.6875		
		Late prediction AUC	0.6875	0.6875		
		Signature	Early	Late	Early and Late	
DEG Signature						
(a) p value < 0.05						
Number of Early Signature	1334	Early prediction AUC	0.9944	0.6826	0.9938	0.6493
Number of Late Signature	64	Late prediction AUC	0.6625	0.9563	0.65	0.6979
(b) p value < 0.05 and logFC > 1						
Number of Early Signature	11	Early prediction AUC	0.915	0.653	0.905	0.738
Number of Late Signature	1	Late prediction AUC	0.687	0.736	0.68	0.786
(c) p value < 0.01						
Number of Early Signature	295	Early prediction AUC	0.997	0.66	0.995	0.606
Number of Late Signature	7	Late prediction AUC	0.67	0.755	0.656	0.651
Elastic Net–Based Signatures						
Number of Early Signature	87	Early prediction AUC	0.988	0.828	0.869	0.856
Number of Late Signature	87	Late prediction AUC	0.71	0.988	0.687	0.869

These findings indicate that the baseline approach is insufficient for predicting late-onset PE with high accuracy. In contrast, introducing late-onset-specific signatures (e.g., the 64 genes from condition A or the single gene from condition B) increased the AUC for late-onset PE prediction to around 0.88–1.0, highlighting the importance of markers specific to the late-onset subtype. Notably, the model using only KLRC4 (condition B) achieved an AUC of 1.0, but its generalizability remains uncertain. Additionally, when early-onset and late-onset signatures were used together, late-onset AUC sometimes decreased, and a drop in early-onset performance was also observed—indicative of a trade-off.

As shown in Table 1, the major differences in pathophysiology and molecular mechanisms between early- and late-onset PE make it challenging for a single signature to achieve high accuracy for both subtypes. Indeed, while adding late-onset-specific signatures improved the AUC for late-onset PE, it sometimes slightly reduced performance for early-onset predictions. These findings underscore previously reported observations that without tailored markers for late-onset PE, it is difficult to achieve high prediction accuracy.

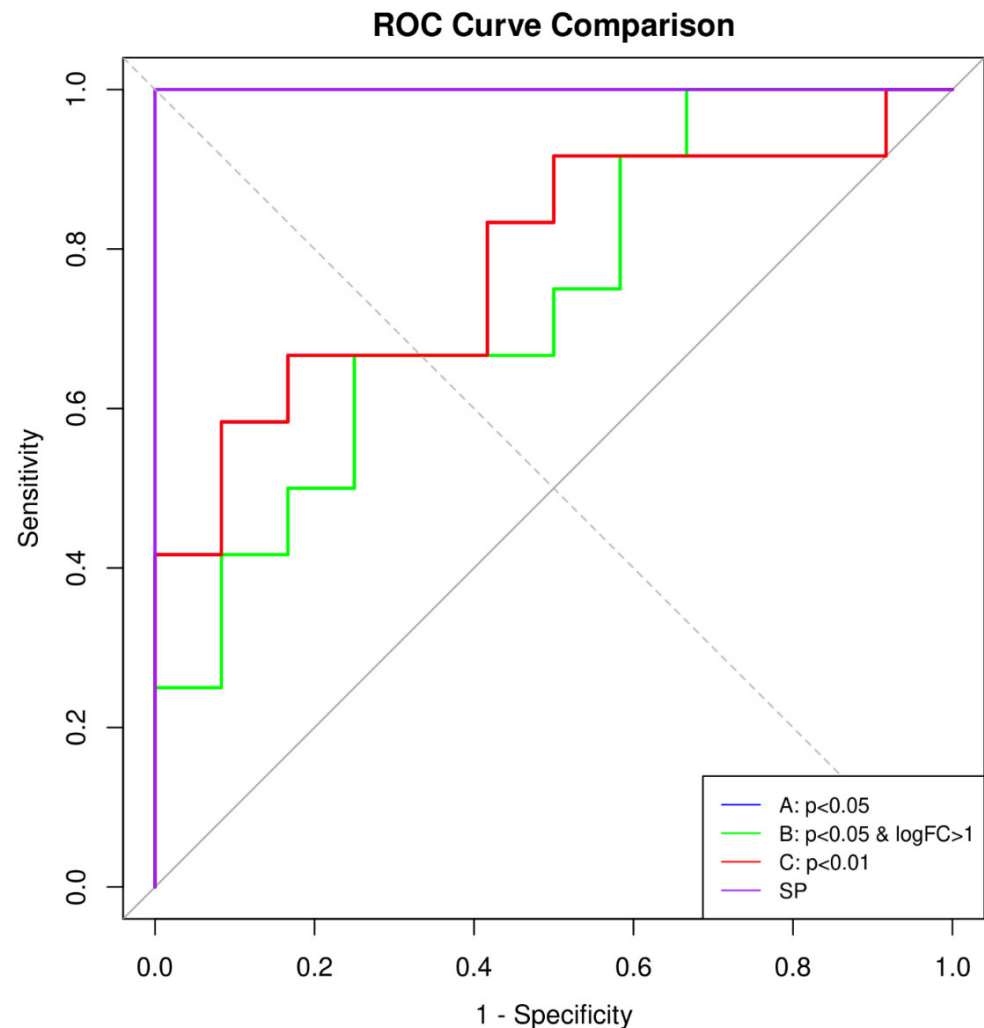


Figure 4. ROC curves from 10 repeats of Monte-Carlo cross-validation comparing three late-onset PE gene-signature models. The green, red, and violet lines correspond to signatures derived from cut-off B ($p < 0.05$ and $|\log_2FC| > 1$), C ($p < 0.01$; identical to A and therefore shown only once), and the sparsity-optimized elastic-net solution path (SP), respectively. Curves are obtained by aggregating predictions from the outer 70/30 test folds ($n = 10$). The elastic-net model (violet) maintains the highest sensitivity across the entire false-positive-rate range, confirming its superior AUROC in Table 1. Abbreviations: FPR, false-positive rate; TPR, true-positive rate.

3.4. Candidate Biomarkers and Functional Analysis

Based on an over-representation analysis of the 87 late-onset-specific genes (Figure 5), the most significantly enriched themes were (i) pro-inflammatory and innate-immune signaling—notably interferon- γ/α and TNF- $\alpha \rightarrow$ NF- κ B cascades, together with adaptive-immune terms such as allograft-rejection/graft-vs.-host pathways—and (ii) extracellular-matrix remodeling and cell-metabolic processes, including elastic-fiber assembly and heme metabolism.

Notably, HLA-G, IL17RB, and KLRC4—genes previously implicated in immune tolerance and trophoblast invasion—showed marked expression differences in the LO-PE group compared with controls. This observation suggests a potential role for impaired maternal-placental interactions.

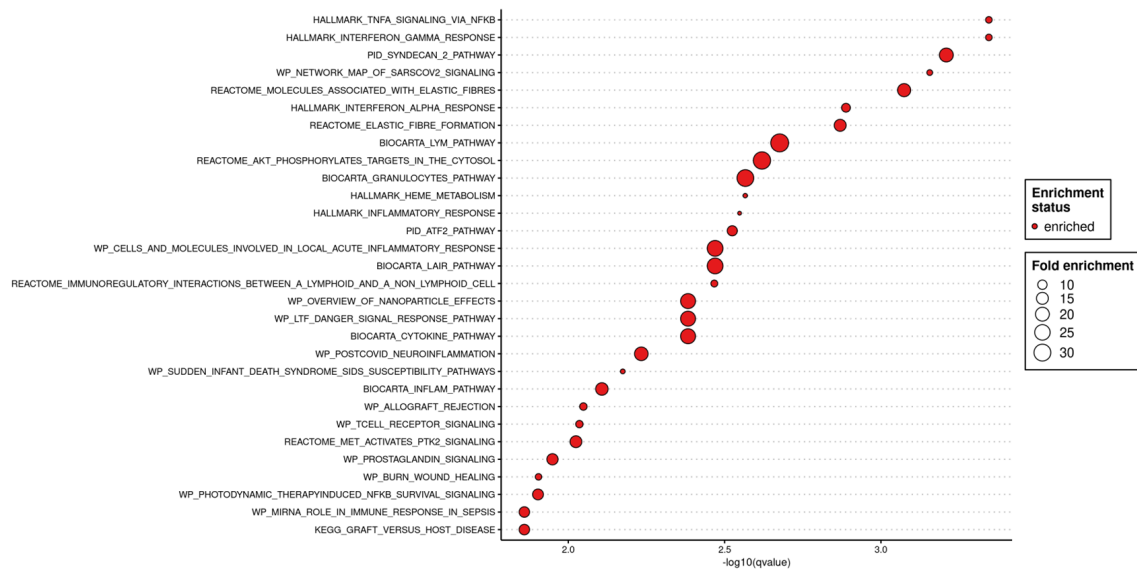


Figure 5. Results of pathway enrichment analysis for the 87 late-onset signature genes identified by the elastic-net model. The horizontal axis represents $-\log_{10}(p\text{-value})$, while bubble size corresponds to fold enrichment. Estrogen response, rejection-response pathways, and glycolysis/ketone-body metabolism pathways appear among the top enriched categories.

4. Discussion

This study is constrained by its modest cohort size (12 LO-PE, 12 EO-PE, and 24 controls). An a priori power calculation shows that for 12 vs. 12 samples and a Cohen's $d = 0.8$, a two-sided $\alpha = 0.05$ yields power ≈ 0.47 ($\beta \approx 0.53$), confirming that the dataset is under-powered for stable single-gene inference. Consistent with this, a 1000-iteration permutation test indicated that the KLRC4 single-gene model does not outperform chance ($p = 0.997$; Supplementary Figure S1), underscoring the risk of over-fitting. No public cfRNA dataset with late-onset PE labels is currently available, so external replication remains an essential future step.

When the early-onset cfRNA signature was naively applied to the 12 LO-PE samples, discrimination was modest (AUROC ≈ 0.69); this serves as our internal baseline. Incorporating a late-onset-specific signature raised performance to AUROC = 0.88–1.00 under nested MC-CV, demonstrating a clear gain over the EO-baseline despite the small cohort.

For clinical context, protein/Doppler screens reach lower or comparable accuracy: in a first-trimester cohort, Tan et al. [5] reported AUROC = 0.744 (0.776 with MAP), while a third-trimester screen by Andrietti et al. [22] achieved AUROC = 0.881 (0.902 with MAP). These figures derive from different cohorts, time-points, and analytes (PIGF \pm UtA-PI \pm MAP) and therefore do not constitute a head-to-head comparison, but they indicate the present clinical performance ceiling. Genomics-assisted ML models that blend polygenic risk scores with routine clinical factors reach a similar range (AUROC ≈ 0.83) [12]. Taken together, these benchmarks suggest that cfRNA signatures—once validated in larger cohorts—could provide a non-invasive alternative with competitive, and potentially earlier, discrimination for late-onset PE.

Among the immune- and hormone/metabolic pathways, HLA-G and IL17RB appear particularly relevant in LO-PE. Wedenoja et al. [8] showed that HLA-G is significantly downregulated in preeclamptic placentas, indicating impaired fetal immune tolerance and reduced EVT infiltration—both hallmarks of shallow placental invasion. Likewise, IL17RB (the IL-25 receptor) fosters trophoblast proliferation; Liu et al. [23] reported that diminished IL-17RB expression in PE placentas correlates with suboptimal placental development. Our findings reinforce that a late-onset immune “collapse” may be tied to early disruptions

in maternal–fetal tolerance. Moreover, Ma et al. [24] identified maternal KIR2DL4–fetal HLA-G genotype combinations that modulate preeclampsia risk, underscoring the genetic dimension of immune tolerance. Altogether, these data underscore how dysregulated HLA-G, IL17RB, and related genes (e.g., KLRC4) may drive LO-PE pathophysiology.

Previous placental-tissue studies have linked reduced HLA-G expression to inadequate EVT invasion in early-onset PE, but evidence in late-onset disease is scant and restricted to small immunohistochemistry series [8]. Our plasma-cfRNA data extend these observations by showing a systemic downregulation of HLA-G transcripts in LO-PE, detectable at the time of diagnosis. Similarly, IL17RB has been reported as a trophoblast-proliferation receptor in experimental work [23], yet has not, to our knowledge, been quantified in circulating RNA from LO-PE pregnancies. Finally, KLRC4 (NKG2F) appears only once in the PE literature (as a placental mRNA outlier in EO-PE); its repetition across five of the ten Monte-Carlo signature extractions suggests it may serve as a novel peripheral marker of late-stage maternal immune activation. Together, these results support the recently proposed ‘dual-hit’ model—immune tolerance collapse compounded by metabolic stress—in LO-PE [10], and they illustrate how cfRNA can surface candidate genes that traditional biomarker panels overlook.

Although various immune and metabolic pathways have been proposed in late-onset preeclampsia (LO-PE), direct evidence for specific processes—such as Allograft Rejection, Estrogen Response, or Glycolysis—and for genes like HLA-G, IL17RB, and KLRC4 remains limited. Recent cfRNA-based studies nonetheless suggest that maternal–placental signaling abnormalities can be detected earlier than clinical onset, potentially offering a broader diagnostic window for LO-PE risk [9,13]. However, the current data primarily indicate general immune dysregulation rather than a uniquely “late-onset-specific” mechanism. Further validation—for example, in multi-ethnic cohorts and via single-cell or multi-omics approaches—will be crucial to pinpoint precisely which genes or pathways diverge in LO-PE.

This study observed a trade-off wherein including both early- and late-onset subtypes in a single model caused decreased accuracy for at least one subtype. That outcome likely stems from the distinct etiological underpinnings of EO-PE vs. LO-PE [25]. As noted by Moufarrej et al. [9], while EO-PE exhibits prominent signals from the placental formation stage, LO-PE often entails maternal metabolic and immune dysfunction that becomes clinically apparent later in gestation. Accordingly, future research should focus on (1) dynamic risk models incorporating longitudinal data by gestational week or (2) algorithms that screen early- and late-onset cases separately and then generate an integrated risk score.

The present study is subject to several limitations and suggests directions for future research. Consequently, prospective cohorts will be required to verify whether circulating HLA-G, IL17RB, and KLRC4 decline before symptom onset, thereby establishing them as early-warning markers rather than late epiphenomenal markers. First, each group comprised only 12 samples, emphasizing the need for validation in larger cohorts and multi-center collaborations to ensure the generalizability of these findings. Second, the predictive signatures identified in this study are sensitive to the choice of thresholds in differential expression analyses (*p*-values) and the setting of hyperparameters (λ in the elastic-net model). Comparative evaluations under multiple conditions are therefore necessary to confirm the robustness of the proposed signatures.

Finally, there are challenges to clinical implementation, as measuring cfRNA and conducting NGS analyses remain expensive and require specialized equipment. Standardizing sample processing and streamlining analytic protocols will be critical for broader clinical adoption. Moreover, integrating other omics data (e.g., metabolomics and epigenomics)

to provide a more comprehensive assessment of both maternal and placental status is an important next step.

5. Conclusions

This study has three principal limitations: (i) the cohort is small ($n = 48$), limiting statistical power; (ii) model selection on a modest dataset raises a non-trivial risk of over-fitting; and (iii) no external cfRNA dataset with late-onset PE labels is yet available for validation. These caveats temper generalizability but also define clear priorities for future work. This study identified late-onset-specific cfRNA signatures and demonstrated that incorporating them into an elastic-net model substantially boosts predictive performance. Abnormalities in immune tolerance and metabolic systems—beyond what conventional early-onset markers can detect—may underlie the pathology of LO-PE. At the same time, challenges remain regarding sample size and model generalizability, pointing to the need for large-scale longitudinal studies and multi-omics integration. Ultimately, leveraging cfRNA-seq-based composite maternal–placental biomarkers in tandem with AI could significantly advance the early diagnosis and management of LO-PE.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/healthcare13101162/s1>, Figure S1. Label-permutation test for the KLRC4 single-gene model. The histogram shows the distribution of AUROC values obtained after 1 000 random shuffles of the case-control labels (grey bars). The vertical red dashed line marks the observed AUROC of the original logistic-regression model (0.875); the blue solid line marks the mean AUROC across permutations (0.874 ± 0.044). The one-sided empirical p -value (0.997) indicates that the model's discrimination is statistically indistinguishable from chance, supporting the conclusion that perfect or near-perfect performance is likely an over-fit of the small dataset. Figure S2. Pooled ROC curves for early-onset PE. The red line (Elastic-Net) dominates the plot with the highest AUROC (~ 0.91), followed by green (Random Forest, ~ 0.87), cyan (linear SVM, ~ 0.83) and violet (XGBoost, ~ 0.74); the grey dashed diagonal denotes chance performance. Figure S3. pooled ROC curves for late-onset PE. Elastic-Net (red) again leads with an AUROC of ~ 0.83 , while Random Forest (green, ~ 0.71), linear SVM (cyan, ~ 0.73) and XGBoost (violet, ~ 0.61) trail closer to the 45° chance line (grey dashes). Table S1. Maternal demographic characteristics of the cfRNA cohort. Mean \pm SD values and unpaired two-tailed t -test p -values compare each pre-eclampsia subtype with its gestational-age-matched controls. Groups did not differ in age or sampling week (all $p > 0.5$); maternal BMI was significantly higher in the LO-PE group than in its matched controls ($p = 0.036$). Table S2. Baseline characteristics of the two normotensive control groups. Values are mean \pm SD; p -values derive from two-sided Welch's t -tests except gestational age, which was predetermined by matching to the corresponding PE cases (therefore no statistical test was performed). No significant differences were detected in maternal age or BMI between EO- and LO-controls (both $p > 0.5$). Table S3. Ten-repeat stratified Monte-Carlo cross-validation performance for early-onset pre-eclampsia: mean \pm SD AUROC and F1 for the proposed Elastic-Net model versus three conventional classifiers (Random Forest, linear SVM and XGBoost). Table S4. Ten-repeat stratified Monte-Carlo cross-validation results for late-onset pre-eclampsia: mean \pm SD AUROC and F1 comparing the proposed Elastic-Net model with Random Forest, linear SVM and XGBoost baselines.

Author Contributions: Conceptualization, Y.M.; methodology, Y.M. and K.U.; formal analysis, A.N.; data curation, Y.M. and A.N.; writing—original draft preparation, A.N.; writing—review and editing, Y.M. and K.U. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Japan Society for the Promotion of Science (JSPS), grant numbers JP20H04282 and JP23K18505.

Institutional Review Board Statement: Ethical review and approval were waived for this study, as it used only publicly available, de-identified data and did not involve direct interaction with human participants.

Informed Consent Statement: Not applicable because this study did not involve the collection of new data from human participants.

Data Availability Statement: The sequencing data reanalyzed in this study are publicly available at NCBI dbGaP (accession number phs002017.v1.p1) and in the Supplementary Materials of [<https://doi.org/10.1126/scitranslmed.aaz0131>]. All newly derived data (including gene signature lists and statistical outputs) are presented in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors gratefully acknowledge the support provided by the Japan Society for the Promotion of Science (JSPS), Grant Numbers JP20H04282 and JP23K18505.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AUC	Area Under the Receiver Operating Characteristic Curve
BMI	Body Mass Index
cfRNA	Cell-Free RNA
DEG	Differentially Expressed Gene
EHR	Electronic Health Record
EVT	Extravillous Trophoblast
HLA-G	Human Leukocyte Antigen-G
IL17RB	Interleukin-17 Receptor B
EO-PE	Early-Onset Preeclampsia
LO-PE	Late-Onset Preeclampsia
ML	Machine Learning
NGS	Next-Generation Sequencing
PE	Preeclampsia
PIGF	Placental Growth Factor
ROC	Receiver Operating Characteristic
sFlt-1	Soluble fms-like Tyrosine Kinase-1

References

1. Liu, M.; Yang, X.; Chen, G.; Ding, Y.; Shi, M.; Sun, L.; Huang, Z.; Liu, J.; Liu, T.; Yan, R.; et al. Development of a Prediction Model on Preeclampsia Using Machine Learning: A Retrospective Cohort Study in China. *Front. Physiol.* **2022**, *13*, 896969. [[CrossRef](#)] [[PubMed](#)]
2. Erez, O.; Romero, R.; Maymon, E.; Chaemsaitong, P.; Done, B.; Pacora, P.; Panaitescu, B.; Chaiworapongsa, T.; Hassan, S.S.; Tarca, A.L. The Prediction of Late-Onset Preeclampsia: Results from a Longitudinal Proteomics Study. *PLoS ONE* **2017**, *12*, e0181468. [[CrossRef](#)]
3. Baylis, A.; Zhang, W.; Menkhorst, E.; Dimitriadis, E. Prediction and Prevention of Late-Onset Pre-Eclampsia: A Systematic Review. *Front. Med.* **2024**, *10*, 1459289. [[CrossRef](#)]
4. Roberge, S.; Bujold, E.; Nicolaides, K.H. Aspirin for the Prevention of Preterm and Term Preeclampsia: Systematic Review and Meta-Analysis. *Am. J. Obstet. Gynecol.* **2018**, *218*, 287–293.e1. [[CrossRef](#)]
5. Tan, M.Y.; Syngelaki, A.; Poon, L.C.; Rolnik, D.L.; Syngelaki, A.; de Alvarado, M.; Carbone, I.F.; Dutemeyer, V.; Fiolna, M.; Frick, A.; et al. Screening for Pre-Eclampsia by Maternal Factors and Biomarkers at 11–13 Weeks' Gestation. *Ultrasound Obstet. Gynecol.* **2018**, *52*, 186–195. [[CrossRef](#)]
6. Jhee, J.H.; Lee, S.; Park, Y.; Lee, S.E.; Kim, Y.A.; Kang, S.-W.; Kwon, J.-Y.; Park, J.T. Prediction Model Development of Late-Onset Preeclampsia Using Machine Learning-Based Methods. *PLoS ONE* **2019**, *14*, e0221202. [[CrossRef](#)] [[PubMed](#)]
7. Rana, S.; Lemoine, E.; Granger, J.P.; Karumanchi, S.A. Preeclampsia: Pathophysiology, Challenges, and Perspectives. *Circ. Res.* **2019**, *124*, 1094–1112. [[CrossRef](#)]
8. Wedenoja, S.; Yoshihara, M.; Teder, H.; Sariola, H.; Gissler, M.; Katayama, S.; Wedenoja, J.; Häkkinen, I.M.; Ezer, S.; Linder, N.; et al. Fetal HLA-G Mediated Immune Tolerance and Interferon Response in Preeclampsia. *EBioMedicine* **2020**, *59*, 102872. [[CrossRef](#)] [[PubMed](#)]

9. Moufarrej, M.N.; Vorperian, S.K.; Wong, R.J.; Campos, A.A.; Quaintance, C.C.; Sit, R.V.; Tan, M.; Detweiler, A.M.; Mekonen, H.; Neff, N.F.; et al. Early Prediction of Preeclampsia in Pregnancy with Cell-Free RNA. *Nature* **2022**, *602*, 689–694. [\[CrossRef\]](#)
10. Robillard, P.-Y.; Dekker, G.; Scioscia, M.; Saito, S. Progress in the Understanding of the Pathophysiology of Immunologic Maladaptation Related to Early-Onset Preeclampsia and Metabolic Syndrome Related to Late-Onset Preeclampsia. *Am. J. Obstet. Gynecol.* **2022**, *226*, S867–S875. [\[CrossRef\]](#)
11. Layton, A.T. Artificial Intelligence and Machine Learning in Preeclampsia. *Arterioscler. Thromb. Vasc. Biol.* **2025**, *45*, 165–171. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Kovacheva, V.P.; Eberhard, B.W.; Cohen, R.Y.; Maher, M.; Saxena, R.; Gray, K.J. Preeclampsia Prediction Using Machine Learning and Polygenic Risk Scores from Clinical and Genetic Risk Factors in Early and Late Pregnancies. *Hypertension* **2024**, *81*, 264–272. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Marić, I.; Contrepois, K.; Moufarrej, M.N.; Stelzer, I.A.; Feyaerts, D.; Han, X.; Tang, A.; Stanley, N.; Wong, R.J.; Traber, G.M.; et al. Early Prediction and Longitudinal Modeling of Preeclampsia from Multiomics. *Patterns* **2022**, *3*, 100655. [\[CrossRef\]](#)
14. Munchel, S.; Rohrback, S.; Randise-Hinchliff, C.; Kinnings, S.; Deshmukh, S.; Alla, N.; Tan, C.; Kia, A.; Greene, G.; Leety, L.; et al. Circulating Transcripts in Maternal Blood Reflect a Molecular Signature of Early-Onset Preeclampsia. *Sci. Transl. Med.* **2020**, *12*, eaaz0131. [\[CrossRef\]](#)
15. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* **2010**, *26*, 139–140. [\[CrossRef\]](#)
16. Robinson, M.D.; Oshlack, A. A Scaling Normalization Method for Differential Expression Analysis of RNA-seq Data. *Genome Biol.* **2010**, *11*, R25. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. Limma Powers Differential Expression Analyses for RNA-sequencing and Microarray Studies. *Nucleic Acids Res.* **2015**, *43*, e47. [\[CrossRef\]](#)
18. Smyth, G.K. Limma: Linear Models for Microarray Data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*; Gentleman, R., Carey, V., Huber, W., Irizarry, R.A., Dudoit, S., Eds.; Springer: New York, NY, USA, 2005; pp. 397–420.
19. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. B* **1995**, *57*, 289–300. [\[CrossRef\]](#)
20. Friedman, J.H.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [\[CrossRef\]](#)
21. Kanehisa, M.; Furumichi, M.; Sato, Y.; Ishiguro-Watanabe, M.; Tanabe, M. KEGG: Integrating Viruses and Cellular Organisms. *Nucleic Acids Res.* **2021**, *49*, D545–D551. [\[CrossRef\]](#)
22. Andrietti, S.; Silva, M.; Wright, A.; Wright, D.; Nicolaides, K.H. Competing-Risks Model in Screening for Pre-Eclampsia by Maternal Factors and Biomarkers at 35–37 Weeks' Gestation. *Ultrasound Obstet. Gynecol.* **2016**, *48*, 72–79. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Liu, S.; Sun, Y.; Tang, Y.; Hu, R.; Zhou, Q.; Li, X. IL-25 Promotes Trophoblast Proliferation and Invasion via Binding with IL-17RB and Associated with PE. *Hypertens. Pregnancy* **2021**, *40*, 209–217. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Ma, Y.; Qian, Y.; Jiang, H.; Meng, H.; Wang, Y.; Yang, Y. Combined Maternal KIR2DL4 and Fetal HLA-G Polymorphisms Were Associated with Preeclampsia in a Han Chinese Population. *Front. Genet.* **2024**, *15*, 1442938. [\[CrossRef\]](#)
25. Aisagbonhi, O.; Morris, G.P. Human Leukocyte Antigens in Pregnancy and Preeclampsia. *Front. Genet.* **2022**, *13*, 884275. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.