

RESEARCH

Open Access



Identification of novel cell-free RNAs in maternal plasma as preterm biomarkers in combination with placental RNA profiles

Heyue Jin^{1,2,3,4}, Yimin Zhang^{1,2,3,4}, Zhigang Fan⁵, Xianyan Wang^{6,7}, Chen Rui^{6,7}, Shaozhen Xing⁸, Hongmei Dong⁹, Qunan Wang^{6,7*}, Fangbiao Tao^{1,2,3,4*} and Yumin Zhu^{1,2,3,4*}

Abstract

Background Preterm birth (PTB) is the main driver of newborn deaths. The identification of pregnancies at risk of PTB remains challenging, as the incomplete understanding of molecular mechanisms associated with PTB. Although several transcriptome studies have been done on the placenta and plasma from PTB women, a comprehensive description of the RNA profiles from plasma and placenta associated with PTB remains lacking.

Methods Candidate markers with consistent trends in the placenta and plasma were identified by implementing differential expression analysis using placental tissue and maternal plasma RNA-seq datasets, and then validated by RT-qPCR in an independent cohort. In combination with bioinformatics analysis tools, we set up two protein–protein interaction networks of the significant PTB-related modules. The support vector machine (SVM) model was used to verify the prediction potential of cell free RNAs (cfRNAs) in plasma for PTB and late PTB.

Results We identified 15 genes with consistent regulatory trends in placenta and plasma of PTB while the full term birth (FTB) acts as a control. Subsequently, we verified seven cfRNAs in an independent cohort by RT-qPCR in maternal plasma. The cfRNA *ARHGEF28* showed consistence in the experimental validation and performed excellently in prediction of PTB in the model. The AUC achieved 0.990 for whole PTB and 0.986 for late PTB.

Conclusions In a comparison of PTB versus FTB, the combined investigation of placental and plasma RNA profiles has shown a further understanding of the mechanism of PTB. Then, the cfRNA identified has the capacity of predicting whole PTB and late PTB.

Keywords Preterm birth, Prediction, Cell-free RNA, Placenta, Plasma, Transcriptome

*Correspondence:

Qunan Wang

wqn@ahmu.edu.cn

Fangbiao Tao

fbtao@ahmu.edu.cn

Yumin Zhu

zhuyumin2011@163.com

¹ Department of Maternal, Child and Adolescent Health, School of Public Health, Anhui Medical University, No 81 Meishan Road, Hefei, Anhui, China

² MOE Key Laboratory of Population Health Across Life Cycle, No 81 Meishan Road, Hefei, Anhui, China

³ Anhui Provincial Key Laboratory of Population Health and Aristogenesis, Anhui Medical University, No 81 Meishan Road, Hefei, Anhui, China

⁴ NHC Key Laboratory of Study on Abnormal Gametes and Reproductive Tract, Anhui Medical University, Hefei, Anhui, China

⁵ Department of Neonatology, Ma'anshan Maternal and Child Health Hospital, Ma'anshan, Anhui, China

⁶ Department of Toxicology, School of Public Health, Anhui Medical University, Hefei, Anhui, China

⁷ Key Laboratory of Environmental Toxicology of Anhui Higher Education Institutes, Hefei, Anhui, China

⁸ MOE Key Laboratory of Bioinformatics, Center for Synthetic and Systems Biology, School of Life Sciences, Tsinghua University, Beijing, China

⁹ Department of Obstetrics, Ma'anshan Maternal and Child Health Hospital, Ma'anshan, Anhui, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Preterm birth (PTB) is the leading cause of death in children under 5 years old worldwide. It is estimated that 17.7% of global under-five child deaths and 36.1% of 0–27 day neonatal deaths were due to complications from PTB [1]. In addition, PTB is the leading risk factor that contributes to growth disorders such as cognitive, visual, and learning disabilities [2]. Efforts to reduce both the incidence and mortality of PTB are still crucial [3].

Therefore, the development of predictive tools for identifying the risk of PTB from the antenatal population is of clinical relevance. A newly developed cervical elastography technique has been proposed for screening spontaneous preterm birth [4]. However, this method has not been used widely for early prediction of PTB because standardized baseline values for elastography parameters have not been established [4]. Recent studies have informed the characterization of a wide range of biological changes during pregnancy can be measured by plasma cell-free RNA (cfRNA) transcripts [5], plasma proteome [6], metabolomics [7–9], immunome [8], and microbiome [10, 11]. However, the prediction of PTB based on such molecular profiles is still challenging.

During pregnancy, the placenta is an important organ that connects the mother and the fetus. A placenta with impaired function may lead to reduced blood flow, or the transfer of oxygen and nutrient to the fetus, which could affect the growth and development of the fetus [12]. A recently published study provides a comprehensive assessment of alterations in the placental transcriptome correlated with spontaneous preterm birth, which suggested that we can obtain novel insight into the mechanisms of PTB by discriminating molecular differences in the placenta [13]. The cfRNA is a new class of biomarkers with enormous potential for the non-invasive diagnosis, prognosis and surveillance in many diseases [14], which has been proven that released from maternal, fetal and placental tissue and can be detected in plasma after entering maternal bloodstream [15, 16]. Notably, placental development is closely linked to pregnancy health. The expression level of RNA measurement in maternal plasma may be a useful and convenient approach for reflecting placental gene-expression profiles as previously described. Thus, analysis of plasma RNA may also have critical significance in tracking pregnancy progression and fetal health [17, 18].

To consolidate the abovementioned considerations, our study tried to combine the placental RNA profiles and comprehensively explored the significance of plasma RNA signature in PTB. To evaluate the correlation of RNA expression regulation in plasma and placenta compared with PTB, we integrated a total of 62 RNA-seq datasets from the placenta and 15 cfRNA-seq datasets

from the maternal plasma. Subsequently, we aimed to assess whether these cfRNAs with consistent expression regulation trends can serve as early biomarkers of PTB. Overall, our study offers the novel PTB biomarkers with the clinical significance and further elucidate that the regulation of RNA in the placenta can be revealed from the plasma cfRNA of pregnant women.

Methods

Study design and cohort

Placental samples for 31 PTB infants and 31 paired full term birth (FTB) infants were collected from the Ma'anshan Birth Cohort (MABC) study. We obtained transcriptome data by bulk RNA sequencing from these 62 samples. The plasma cfRNA-seq data were downloaded from Sequence Read Archive (SRA) database, including seven FTB samples and eight PTB samples (SRP130149). In addition, from March to June 2022, we collected 41 maternal plasma samples with recruitment criteria for singleton pregnancy live births as a validation cohort in Ma'anshan Maternal and Child Health Hospital for experimental verification of the identified changes in cfRNA expression. Informed consent was obtained from each participant. The maternal characteristics of participants are presented in Additional file 1: Table S1.

Sample collection

For placenta tissues collected from the MABC study, a piece of placental lobule tissue was separated from the maternal side of the placenta at a distance of 5 cm from the umbilical cord within 30 min of delivery of the placenta. Each piece of tissue, about 1cm³ in volume, was placed in a cryostorage tube with an RNA later and refrigerated overnight at 4 °C, and then stored at –80 °C after the RNA later was absorbed. For plasma samples collected from the validation cohort, based on an estimated due date from the last menstrual period (LMP), the samples of blood were collected before delivery at 37 weeks for the validation cohort. All samples were placed in EDTA tubes. Within 8 h of sample collection, the samples were centrifuged at 3000×g for 6 min to separate plasma and then reposit at –80 °C until assay.

RT-qPCR validation

CfRNAs were extracted from 0.8 ml plasma using the Trizol reagent (Invitrogen, USA). 10.5 µl RNA was reverse transcribed using First Strand cDNA Synthesis Kit (Promega, USA) with adding 1 µl External RNA Controls Consortium (ERCC) according to the manufacturer's protocol. The cDNA was diluted twice and amplified by SYBR green (YESEN, Shanghai, China) in a LightCycler[®] 96 System (Roche).

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is the internal reference for cfRNAs. Primers for each differentially expressed gene were designed by an online database PrimerBank and Integrative Genomics Viewer (IGV) [19, 20]. All sequences are enumerated in Additional file 1: Table S2. The average of the three replicates was performed as the cycle threshold (Ct) value for each cfRNA. Several cfRNAs which considered undetected on the basis of the Ct value limit of 39 were excluded for further analysis [21]. For the evaluation of cfRNA expression levels, we used the $2^{-\Delta Ct}$ method and then normalized.

Differential expression analysis

Differentially expressed genes (DEGs) between the PTB cases and controls were identified using edgeR [22]. The significance cutoff was appraised at $|\log_2(\text{fold-change})| \geq 0.59$ and $p\text{-value} < 0.05$. All information of DEGs was used for principal component analysis (PCA) in each group independently. The comparison of DEGs was performed by the “UpSetR” R package.

Functional enrichment analysis

We applied g: Profiler for performing Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis [23]. The results of enrichment pathways were ranked based on adjusted p-value and plotted in R statistical software.

Protein–protein interaction (PPI) network

We respectively incorporate genes with altered expression levels in plasma and the placenta in PTB to construct the protein–protein interaction (PPI) network. The PPI network was implemented using the STRING database and then the key nodes were screened using the cytohubba in Cytoscape software with four methods including Maximal Clique Centrality (MCC), Maximum Neighborhood Component (MNC), degree, and closeness [24, 25]. The top 25 genes were selected as the key genes in each algorithm based on the score. We identified the final hub genes by intersecting the key genes obtained by these algorithms [26]. Metascape provides the functional annotation of these hub genes [27].

Establishment of a predictive model

To predict PTB (GA at delivery < 37 weeks), we developed a machine learning model of support vector machine (SVM), which relying on a dataset of the measured relative expression levels of cfRNAs by RT-qPCR. The expression levels of cfRNAs whose results of RT-qPCR were consistent with RNA-seq results and showed significantly difference in the preterm and term groups were perceived as input features. All samples we collected were divided into 50% train set and 50% test set

by the createDataPartition method from the caret package in R statistical software. The area under the computing receiver-operating characteristic (ROC) curve (AUC) was calculated to assess the performance of the machine learning model. In addition, maternal age, parity, body-mass index (BMI), platelet distribution width (PDW), neutrophil-to-lymphocyte ratio (NLR), and haemoglobin (HB) were considered as preterm clinical risk factors through literature searching [28–33]. Then, we collected the above clinical data matched to samples used for independent validation by RT-qPCR through electronic medical records.

Statistical analysis

A nonparametric Mann–Whitney U-test was used for all tests designed to compare the expression level of different groups. The significant differences were considered $p\text{-value} < 0.05$.

Results

Discovery of distinct dysregulated genes in RNA-seq data and the pathways involved in preterm birth

We first identified DEGs by analyzing cfRNA-seq in plasma, including a binary grouping of seven PTB (< 37 weeks) versus eight FTB (≥ 37 weeks). Additionally, we applied PCA to our placental and plasma dataset, to explore sample features at gene expression levels (Additional file 1: Fig. S1A, B). The FTB samples were mixed with PTB samples, which may be caused by large heterogeneity. Then we observed broad differences using the FTB group as control with 721 up-regulated genes and 602 down-regulated genes in plasma (Fig. 1A). For the placental transcriptome, we only identified 404 DEGs by comparing 31 PTB samples with 31 FTB samples using tissue RNA-seq (Fig. 1B). We found 11 RNAs were up-regulated both in plasma and placenta of PTB group, which accounted for 1.5% and 4.0% of up-regulated genes in plasma and placenta, respectively (Fig. 1C, and Additional file 1: Fig. S2A); and 4 RNAs were down-regulated both in plasma and placenta of PTB group, which accounted for 0.66% and 3.0% of down-regulated genes in plasma and placenta, respectively (Fig. 1C, and Additional file 1: Fig. S2A).

GO enrichment and pathway analyses were using differently expressed gene-set and the top five pathways in each group measured by adjusted p-value were presented. We found that the DEGs in plasma were involved in distinct biological process (BP) terms from those in placenta. The up-regulated genes in PTB in plasma were mainly involved in developmental and cellular process (Fig. 1D). By contrast, the most significant enrichment pathway of up-regulated genes in PTB in the placenta that was individually involved with immune

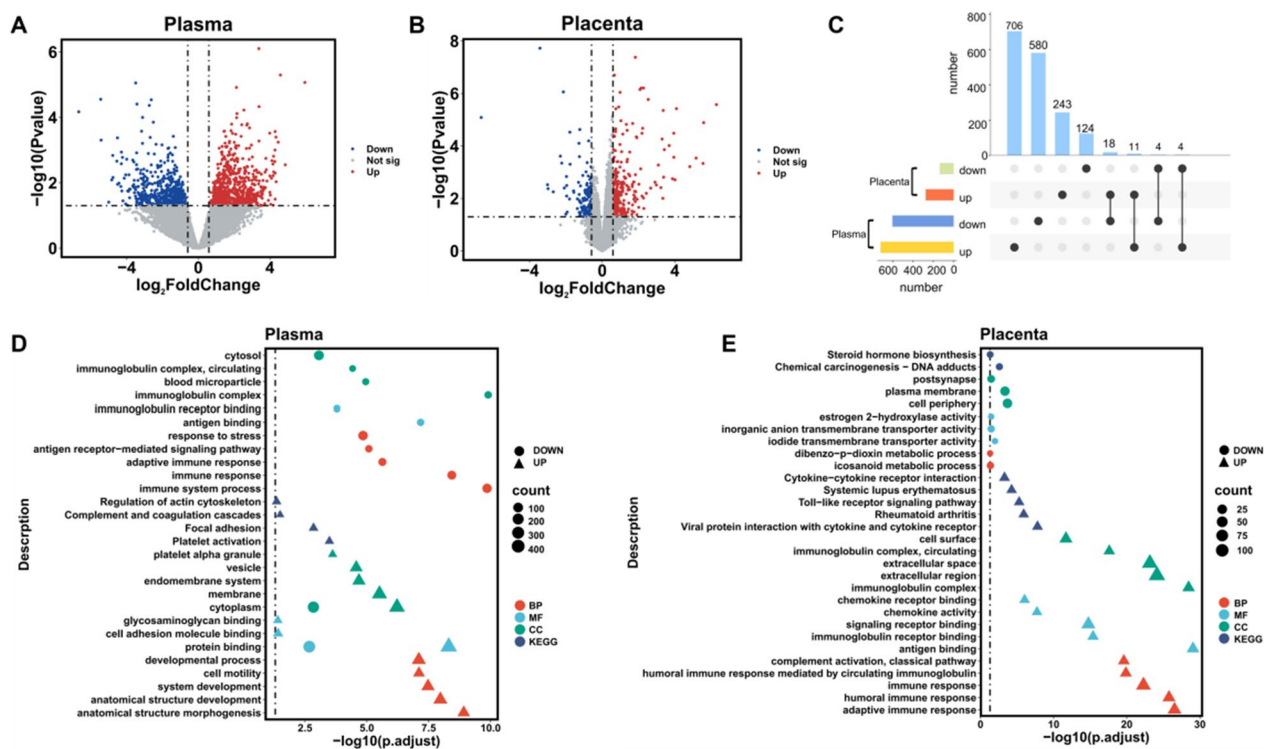


Fig. 1 Differentially expressed genes in preterm birth placenta and plasma. **A** Volcano plots of differentially expressed genes (DEGs) in the comparison (PTB vs. FTB) of RNA-seq data from plasma. **B** Volcano plots of differentially expressed genes (DEGs) in the comparison (PTB vs. FTB) of RNA-seq data from placenta. **C** UpSet plot of the gene symbol mapping overlaps for all sets of DEGs' comparisons. **D** The top five GO-terms and KEGG pathways for the enrichment of DEGs involved in plasma group. **E** The top five GO-terms and KEGG pathways for the enrichment of DEGs involved in placenta group

response (Fig. 1E). The down-regulated genes in PTB in plasma showed the relationship with immune pathways, whereas most down-regulated genes in PTB in the placenta were associated with metabolic process (Fig. 1D, E). We detected significant enrichment of DEGs of plasma in most molecular function (MF) terms related to binding, such as protein binding, glycosaminoglycan binding, and immunoglobulin receptor binding (Fig. 1D). In the placenta, the enrichment analysis showed that these up-regulated genes were also enriched in signaling receptor binding, and the down-regulated genes were mainly enriched in transmembrane transporter activities (Fig. 1E). AS for the enrichment result of up-regulated genes in plasma, the top five significant cell component (CC) terms were dominated by membrane-bounded organelle and endomembrane system, while the down-regulated genes related to the immunoglobulin complex and intracellular anatomical structure (Fig. 1D). In the placenta, the majority of genes associated with the extracellular region and immunoglobulin complex pathway were up-regulated in PTB (Fig. 1E). Cell periphery was also amongst the significant pathways in down-regulated genes in PTB.

We further performed KEGG pathway analysis for plasma and placental DEGs respectively. Only four KEGG pathways were detected in plasma group, including regulation of actin cytoskeleton, complement and coagulation cascades, focal adhesion, and platelet activation (Fig. 1D). These representative entries of the analysis of placental-derived DEGs mostly contained human disease, steroidogenesis and signaling interaction (Fig. 1E). In total, the immune signaling subgroup was the largest enriched, which consistent with the previous finding on the pathologic mechanisms of PTB [34]. When we compared the pathway enrichment of up-regulated genes in plasma vs. the up-regulated genes in the placenta, 23 common pathways were revealed which accounted for a higher proportion of all plasma and placental pathways (Additional file 1: Fig. S2A). The data reveal that up-regulated genes in PTB have a distinct association in the response to stimulus and glycosaminoglycan binding for the BP and MF term while sharing enrichment in CC related to the extracellular region (Additional file 1: Fig. S2C). The same pathway both in plasma and placenta has not been identified in the down-regulated genes enrichment pathway.

The complexity of RNA regulatory molecules in plasma and placenta

We separately observed the distribution of expression levels of different biotypes of RNA molecules co-detected from plasma and placenta. The expression levels of the different RNA classes in the PTB group were skewed between maternal plasma and the placenta. We found that the majority of annotated mRNA, other non-coding RNAs (ncRNAs), snoRNAs, and snRNAs were expressed weakly in the PTB placenta compared with the PTB plasma (Fig. 2A, B). By contrast, the expression levels of long noncoding RNAs (lncRNAs) and pseudogenes in placenta is higher than that in the plasma for the PTB group (Fig. 2A, B). We also obtained this similar tendency in the FTB group except for the pseudogenes (Additional file 1: Fig. S3A, B). Subsequently, we ranked the total number of RNA biotypes that were involved in the aberrant regulation of placental and plasma respectively (Fig. 2C). Consistent with the plasma, we found that the majority of DEGs in the placenta were distributed in the following types: mRNA, lncRNA, and pseudogene (Fig. 2C).

Previous work in humans has confirmed that single-cell resolution (scRNA-seq) can catch a wide range of cell types that contribute to the placenta and identified

distinct differences in the cell type components of pre-term and full-term pregnancies [35, 36]. To further evaluate the alterations in gene expression in PTB, we attempted to map placental and plasma DEGs to genes with altered expression in different placental cell subtypes and tested for the presence of single cell-derived placental profiles in cfRNA in the maternal circulation [35]. The major populations include cytotrophoblast (CTB), fibroblast, macrophage, monocyte, NK-cell, npiCTB, and T-cell. In the maternal plasma cfRNA profile, the largest number of DEGs between the PTB and FTB groups were observed in the CTB, followed by macrophage (Fig. 2D). Strikingly, the majority of cell marker genes were robustly increased in PTB. Using the placental profile, we found many of macrophage marker genes were significantly altered expression in PTB compared to the FTB group (Fig. 2E). These results validate the macrophage signature changes in preterm pregnancy that were previously reported, and further revealed the single-cell features can be detected non-invasively in the maternal circulation throughout the pregnancy period [35]. Then, we collected 33 PTB-associated genes from several studies, which were identified from different data types (ChIP-seq, RNA-seq, methylation, and others) [5, 37–42]. Meanwhile, 51 genes with differential expression

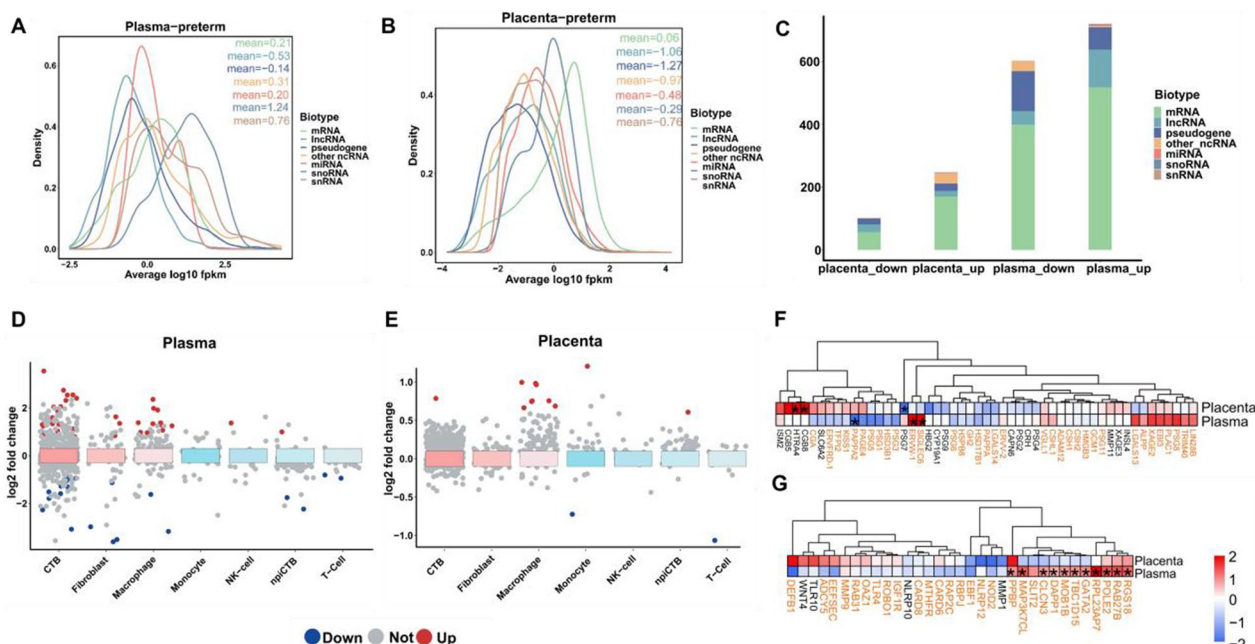


Fig. 2 Assessment of RNA regulatory molecular features detected in plasma and placental transcriptomes. **A, B** The density of expression abundance in preterm birth of different RNA biotypes compared in **A** plasma, and **B** placenta. **C** Numbers of differentially expressed RNAs for different RNA biotypes. **D** Several placental cell type specific genes were differentially expressed in PTB in the maternal plasma cfRNA profile. **E** Several placental cell type specific genes were differentially expressed in PTB in the placental RNA profile. **F** The fold change of placenta-associated genes between PTB and FTB pregnancies. **G** The fold change of PTB-associated genes between PTB and FTB pregnancies. The orange indicates that the value of $|\log_2$ fold change| is higher in plasma, compared to the placenta group. *p-value < 0.05

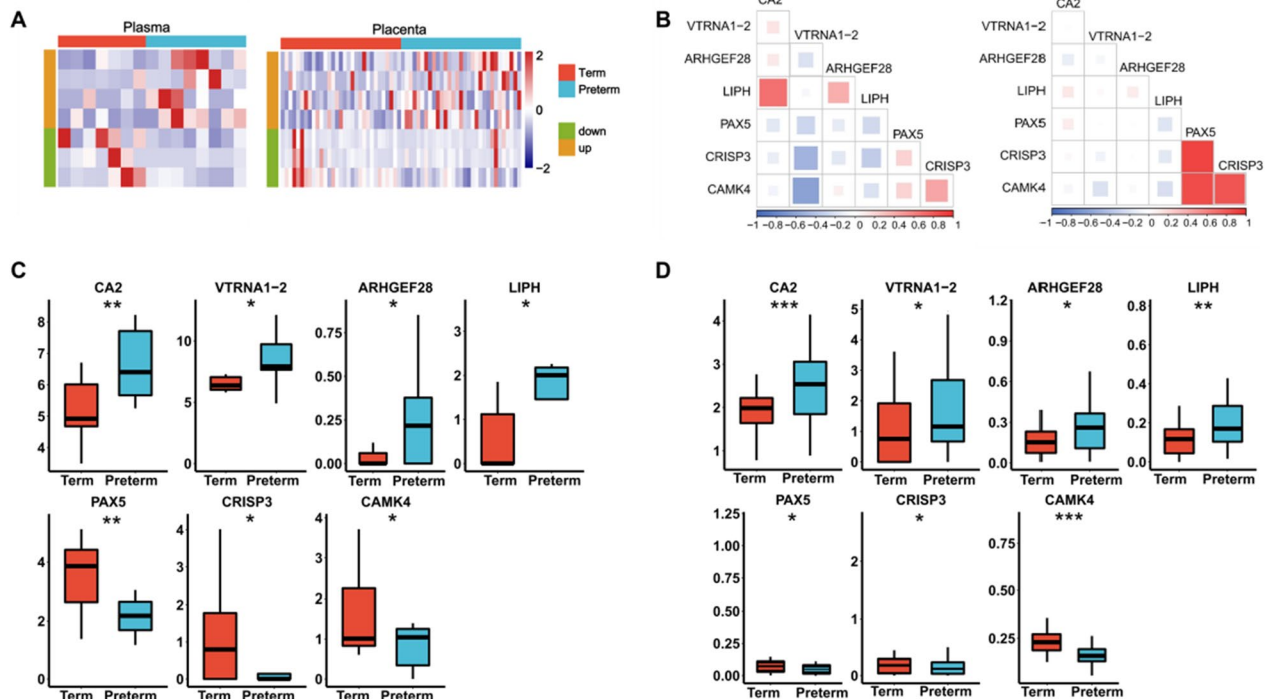


Fig. 4 Candidate cfRNAs for predicting preterm birth. **A** Heatmap showing expression level in each samples of candidate genes in plasma and placental RNA-seq datasets (plasma samples: seven FTB vs. eight PTB, placenta samples: 31 FTB vs. 31 PTB). **B** Correlation heatmap displaying the inter connectivity among candidate genes. The size of the squares and the color scale correlate to the correlation of gene expression in RNA-seq data including plasma and placenta. Left: plasma, right: placenta. **C** Expression level for differentially expressed genes in the discovery based on plasma cRNA-seq datasets. **D** Expression level for differentially expressed genes in the discovery based on placental RNA-seq datasets. *p-value < 0.05, **p-value < 0.01, ***p-value < 0.001

required to be consistent with the regulatory trend results obtained by edgeR. Taken together, we generated a candidate marker set of seven cfRNAs for subsequent validation.

Then we explored the co-expression signatures of these genes in the placenta and plasma using Pearson correlation coefficient, which suggested that the occurrence of the high synergy of these genes is skewed in the plasma transcriptome compared to the placental transcriptome (Fig. 4B). Two sets of moderately related genes were present in the data from plasma dataset (*LIPH* and *CA2*, *CAMK4* and *VTRNA1-2*, $0.5 \leq |R| < 0.8$, p-value < 0.05; Fig. 4B). In placenta dataset, we found a strong positive correlation between *PAX5* and the two genes (*CRISP3*, *CAMK4*), which revealed that they may play a role in similar biological processes ($R > 0.8$, p-value < 0.05; Fig. 4B). All of these seven cfRNAs had the same expression patterns in both sets of samples from the placenta and plasma RNA-seq data. Among them, four cfRNAs (*CA2*, *VTRNA1-2*, *ARHGEF28*, and *LIPH*) were observed to be up-regulated and three cfRNAs (*PAX5*, *CRISP3*, and *CAMK4*) were down-regulated in PTB group compared with FTB group (Fig. 4C, D).

Independent validation of the selected cfRNA biomarkers for preterm birth and predictive modeling of PTB

To further validate the potential PTB cfRNA biomarkers selected by the analyses described above, we used RT-qPCR for experimental verification in an independent cohort recruiting 41 plasma samples in mid to late pregnancy (Fig. 5A, Additional file 1: Table S1). In the PTB group, the 24 women delivered at 34.7 ± 1.9 weeks (average \pm SD), while in the FTB group, the 17 women delivered at 39.7 ± 0.8 weeks. Based on the results of RT-qPCR, we found the *ARHGEF28* gene was significantly up-regulated in the plasma of PTB group, suggesting that *ARHGEF28* is a reliable biomarker for PTB prediction (Fig. 5B). In our study, most of FTB samples were collected later than PTB samples. Therefore, to investigate whether the expression level of *ARHGEF28* is altered with the progression of pregnancy, we built a subset of plasma samples collected at 35 ± 1 weeks and compared the expression level of *ARHGEF28* between preterm and full-term groups (13 PTB samples, 3 FTB samples). We found the *ARHGEF28* gene was still significantly up-regulated in PTB (Additional file 1: Fig. S6A), which suggested that the gene still has the potential to be

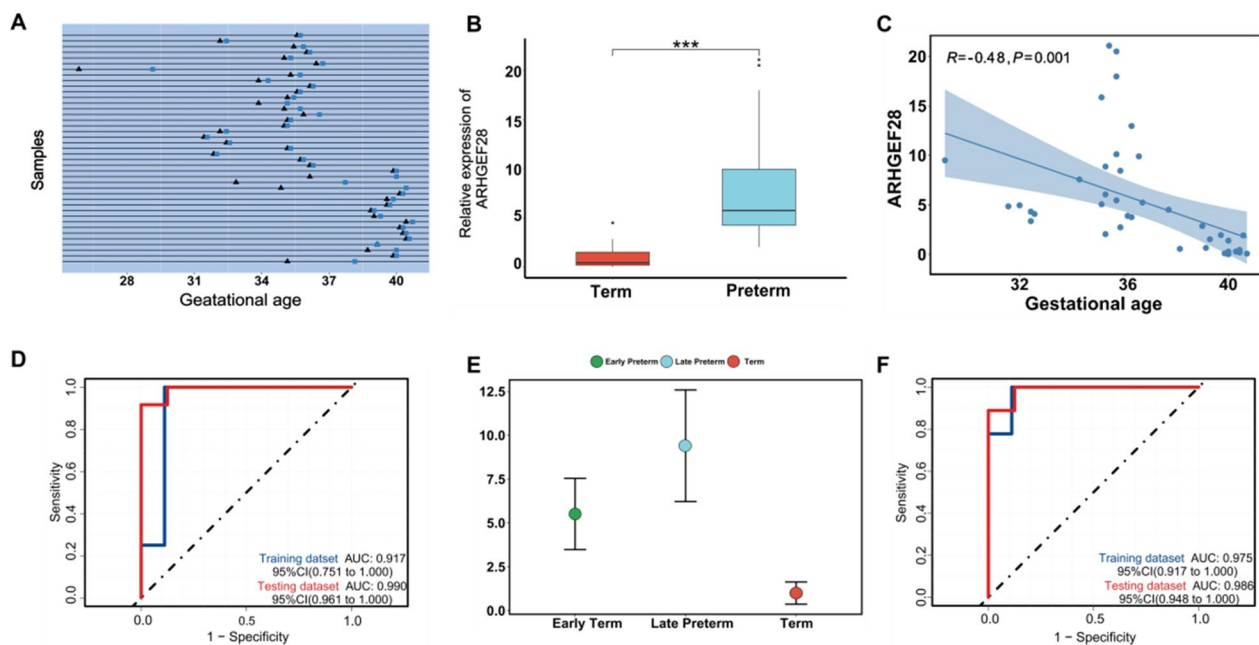


Fig. 5 The validation of candidate markers and the performance of predictive models. **A** Blood sample collection timelines in the validation cohort. triangles: GA at blood collection, squares: GA at delivery. **B** Means \pm SD for the relative expression level of *ARHGEF28* validated using RT-qPCR. ***p-value < 0.001 (Mann–Whitney U-test). **C** The correlation of the relative expression level of *ARHGEF28* with gestational age. **D** Receiver operating characteristic (ROC) curve representing prediction of preterm birth by the gene *ARHGEF28* across the training dataset and testing dataset using SVM algorithm. **E** Altered in the relative expression level of *ARHGEF28* in the three subgroups including early preterm, late preterm, and term. **F** ROC curves representing prediction of late preterm by the gene *ARHGEF28* across the training dataset and testing dataset using SVM algorithm

a predictive biomarker for preterm birth if all sampling times are controlled to 35 ± 1 weeks. In addition, we tried to divide the 17 FTB samples into two groups according to the time point of collection (Before 37 weeks: $n=4$, After 37 weeks: $n=13$). Then we found no significant differences in the expression level changes of *ARHGEF28* between these two groups of samples (Additional file 1: Fig. S6B), which suggested that the gene may not be used as a biomarker for gestational development. Gestational age was considered to be a partial mediator between biological factors of PTB and neonatal outcomes, associated with neonatal morbidity in late preterm and early term birth [45]. Additionally, we also observed the significant correlation between gene expression level of *ARHGEF28* and gestational age: the decreased expression of *ARHGEF28* along with the gestational age increased gradually (p-value = 0.001, $R = -0.48$) (Fig. 5C).

For the prediction of PTB, we divided the RT-qPCR data into training datasets with 21 samples and testing datasets with 20 samples (training dataset: 9 FTB samples, 12 PTB samples, testing dataset: 8 FTB samples, 12 PTB samples) at first. Next, we selected the cRNA *ARHGEF28* as a feature applied to the PTB classifier. For the SVM model, the AUC was 0.917 in the training dataset and 0.990 in the testing dataset (Fig. 5D). Multiple clinical

risk factors contribute to PTB, so we also collected other traditional risk factors that have been reported [28–33]. To assess the prediction capability for PTB of traditional risk factors, we integrated a base model depending on the following factors: maternal age, parity, BMI, PDW, NLR, HB. We presented the AUC of traditional risk factors with and without the addition of *ARHGEF28* applying the SVM machine model. After the exclusion of samples with missing clinical data, the training dataset included 19 samples, while the testing dataset included 17 samples. The lower AUC was yielded in the base model and then the AUC increased with the addition of *ARHGEF28* (Additional file 1: Fig. S7A–C). We calculated Pearson correlation coefficients to further assess the affiliation between the expression level of *ARHGEF28*, gestational age, and PTB-associated clinical characteristics, but no significant correlation was detected, which may cause by the small amount of clinical data (Additional file 1: Figs. S8A–F and S9A–F). Compared to the predictive model with *ARHGEF28* as an independent factor, the combination of *ARHGEF28* expression levels and clinical risk factors may have a more valid predictive performance by integrating more samples in the future work.

At last, we performed a separate analysis of PTB subgroups defined using the guideline from the American

College of Obstetricians and Gynecologists (ACOG) [46]. The result reveals that there has been a rise in the relative expression level of *ARHGEF28* between the early-preterm group (28–34 weeks) and the late-preterm group (34–37 weeks), especially in the late-preterm group (Fig. 5E). We constructed a predictive model of late preterm using the similar approach, which still achieved an excellent performance (training dataset, AUC=0.975; testing dataset, AUC=0.986) (Fig. 5F). Late preterm is the largest group accounting for nearly three-fourths of all preterm infants, with the risk of increased neonatal morbidities as well as the risk of long-term adverse outcomes [47, 48]. Therefore, the detection of the risk of late preterm clinically remains necessary for infant development. The findings from our study confirmed the potential of *ARHGEF28* in the prediction of preterm birth and especially late preterm.

Discussion

In this study, for the first time, we systematically delineated the expression profiles of aberrant regulatory genes associated with PTB in plasma and placenta, then identified a novel cfRNA as a biomarker for predicting PTB. By integrating and analyzing the RNA-seq data of placental tissue and the cfRNA-seq data of maternal plasma, we found that the differentially expressed RNA in placentas of preterm infants could be detected in the plasma of mothers. Although the main result here was the prediction of PTB, we further supported the altered expression level of genes in the placenta can be detected in maternal plasma. This also accords with the observations from earlier study which focused that the placental miRNA profiles combined with matched profiles from maternal plasma reflecting physiological changes occurring at early to middle gestation [49].

To explore the possible molecular mechanisms of PTB, we searched enriched pathways for the differentially expressed genes of PTB. More enrichment of immune pathways is in line with those of previous findings that pregnancy and parturition involve widespread changes in the maternal immune system [50]. Different from PTB-associated pathways in plasma, we detected distinct pathways in placenta group about metabolism. These are similarities between the observation described by a previous placental transcriptomic study of PTB [13]. Accumulating evidence indicates that abnormalities in maternal or fetal membrane metabolism stimulate inflammatory cytokines, which may drive PTB [51]. However, there were a few pathways associated with metabolic processes in our observations. The estrogen metabolism pathway, fetal stress, and fetal anomalies are the regularly reported pathways associated with PTB [51], although these pathways were not present in our result.

We next explored the regulation of PTB by different types of RNA molecules detected in placenta and plasma. Existing research revealed the critical association between placental gene expression levels and the abundance of the genes in maternal plasma [18]. We observed significant correlation between altered expression of mRNA and snoRNA in placenta and plasma in our data, this correlation pattern of mRNA is weak and the mismatch between placenta and plasma samples may be an important reason. In our study, we found the significant correlation in snoRNAs. The snoRNAs were considered to belong to the abundant small non-coding RNA (sncRNA) species. The sncRNA molecules can cross the placenta barrier and then be discharged into the maternal circulation, on account of their stable structure and small size [52]. Thus, these sncRNAs could possibly consider to be prime candidates for placental and pregnancy diagnostic, and this potential was reflected again in our study. Up to date, researches have not treated the specific roles within the placenta and pregnancy of sncRNAs in much detail. In total, the contribution of sncRNAs is still an area worth exploring.

The novel cfRNA marker *ARHGEF28* validated by RT-qPCR on independent samples showed the increased expression level in maternal plasma during pregnancy and was significantly associated with increased risk of PTB. As far as we know, characterization of *ARHGEF28* was measured in the prediction of PTB firstly in our study. Previous reports demonstrated the role of *ARHGEF28* in modulating neuronal function or maintenance, and the formation of *ARHGEF28* aggregates is involved in the pathogenesis of motor neuron disease [53, 54]. Accumulating evidence has been pointed out that PTB may leave the nervous vulnerable to dysfunction [55–58]. These findings perhaps imply the *ARHGEF28* involves a role in neuron function for PTB children. We subsequently tackled the issue of predicting PTB from plasma samples collected while women were asymptomatic before 37 weeks of gestation. Collectively, the machine learning models based on *ARHGEF28* have shown excellent performance in predicting PTB, especially late PTB.

To our knowledge, this is the first study to generate cfRNA profiles as PTB biomarkers using maternal plasma in combination with RNA expression from the placenta. Analysis based on the framework proposed by us, we obtained a more comprehensive understanding of the mechanism about PTB and the regulatory role of RNA in PTB. Consistent with the previous pattern of analysis [36, 59], our profiling is based on a sample of different population cohorts which may facilitate further validation of selected biomarkers. In particular, we focused on the predictive performance of the validated markers for late PTB. We noted that late preterm infants

with an increased birth rate composed now 75% of all preterm infants [60]. It is still an important public health issue. Although the validation study specifically included late PTB pregnancies, we were not assuming that late PTB was an initial independent event. In other words, discovery of potential markers that can be used to predict late PTB unaffected by previously proposed hypothetical mechanisms. As the etiology of PTB is not fully definitive, the clinical support decision model is crucial in helping doctors provide early intervention for women at high risk of PTB [61]. In this setting, we tested to garner a predictive model combining cfRNA markers and clinical factors associated with PTB, which provide a novel strategy for the development of PTB prediction model with clinical benefit. Our work further supports the application of non-invasive blood testing techniques for monitoring the risk of PTB, nevertheless, it is limited by several factors. At first, we collected placenta and plasma samples from different datasets and were not paired between two sample types, which may have led to the heterogeneity in our results and presented bias in correlation analysis. The small sample size in our study may weaken the reliability of the predictive model for the heterogeneity are more pronounced. Although we attempted to demonstrate that PTB was the only factor contributing to the increasing trend in *ARHGEF28* by selecting plasma samples collected at similar gestation weeks for additional analysis and comparison, it may be an inadequate clarification also due to the small sample size. Subsequent cohort studies on a large scale are required to resolve these concerns. In addition, PTB can be classified into three clinical subtypes including spontaneous preterm birth with intact fetal membranes, preterm premature rupture of membranes (PROM) before the onset of labor, and medically indicated preterm birth [62]. Each subtype of PTB may cause by differ pathways [63]. We were lack of adequately performing specific analysis of each subtype of PTB respectively in this study, which leads to a poor cognition of the etiological mechanisms of PTB and limits the development of predictive markers for specific subtypes. Among the RT-qPCR experiment, we found that the expression level of *ARHGEF28* was significantly increased in PTB plasma, suggesting aberrant expression of *ARHGEF28* may be related to the occurrence of PTB. However, there is no clear validation in placental tissues for the gene *ARHGEF28*. The examination of placental pathology was considered to be the major phenotype in an assessment of PTB and provided important insight into subtypes of PTB [64, 65]. We failed to extract the histopathological data of these placental specimens in this cohort study, which limited understanding of pathological features of the preterm placenta. Integration of other omics data and consideration of other clinical factors of

PTB may also generate a more robust RNA profile that reveals the signature of PTB in future studies.

In summary, this study provided molecular-level evidence that RNA expression regulation is relevant in the placenta and maternal plasma in a comparison of preterm and term. The identification of novel biomarkers from plasma for preterm birth revealed the ability of cfRNA to identify the risk of PTB in advance. Combined analysis of different transcriptomic profiles of PTB can contribute to a deep mechanistic understanding of early parturition, which provides a wider perspective into the efficiency of clinical non-invasive diagnostic methods.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-023-04083-w>.

Additional file 1: Table S1. Maternal characteristics of participants in this study involved in RT-qPCR. **Table S2.** Gene primer sequences used in RT-qPCR. **Figure S1.** Principal component analysis (PCA) for sample clustering using RNA-seq data. **Figure S2.** Comparison of plasma and placental transcriptome analysis. **Figure S3.** Complexity of RNA regulatory molecules biotype. **Figure S4.** Construction of the PPI network. **Figure S5.** The key genes in plasma and placental PPI network. **Figure S6.** Comparison of the relative expression levels of *ARHGEF28*. **Figure S7.** Predictive models for preterm birth combined with clinical factors. **Figure S8.** The exploration of relatedness between clinical factors for preterm birth and the expression level of *ARHGEF28*. **Figure S9.** The exploration of relatedness between clinical factors for preterm birth and gestational age.

Additional file 2: Functional annotation of hub genes in plasma and placental group.

Acknowledgements

This study was supported by Department of Neonatology, Ma'anshan Maternal and Child Health Hospital. We also thank local nurses and clinicians for collecting blood and aggregating clinical data.

Author contributions

HJ and YZ prepared the manuscript. HJ and YZ analysed the data. HJ, XW, YZ, SX, and CR performed the experiments. ZF, HD, QW, and FT contributed the data and collected the samples. All authors read and approved the final manuscript.

Funding

This work was supported by grants from the National Natural Science Foundation of China (82103870) and Natural Science Foundation of Anhui Province (2108085QH360).

Availability of data and materials

RNA sequencing data have been deposited in the Sequence Read Archive (SRA) under study accession number SRP410951.

Declarations

Ethics approval and consent to participate

The samples collected during the study come from the relevant studies where written informed consent was received. To protect personal privacy, all the personal details were not obtained.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 7 January 2023 Accepted: 25 March 2023

Published online: 12 April 2023

References

- Perin J, Mulick A, Yeung D, Villavicencio F, Lopez G, Strong KL, Prieto-Merino D, Cousens S, Black RE, Liu L. Global, regional, and national causes of under-5 mortality in 2000–19: an updated systematic analysis with implications for the Sustainable Development Goals. *Lancet Child Adolesc Health*. 2022;6:106–15.
- Chang HH, Larson J, Blencowe H, Spong CY, Howson CP, Cairns-Smith S, Lackritz EM, Lee SK, Mason E, Serazin AC, et al. Preventing preterm births: analysis of trends and potential reductions with interventions in 39 countries with very high human development index. *Lancet*. 2013;381:223–34.
- Cao G, Liu J, Liu M. Global, regional, and national incidence and mortality of neonatal preterm birth, 1990–2019. *JAMA Pediatr*. 2022;176:787–96.
- Patberg ET, Wells M, Vahanian SA, Zavala J, Bhattacharya S, Richmond D, Akerman M, Demishev M, Kinzler WL, Chavez MR, Vintzileos AM. Use of cervical elastography at 18 to 22 weeks' gestation in the prediction of spontaneous preterm birth. *Am J Obstet Gynecol*. 2021;225:525.e1–525.e9.
- Ngo TTM, Moufarrej MN, Rasmussen MH, Camunas-Soler J, Pan W, Okamoto J, Neff NF, Liu K, Wong RJ, Downes K, et al. Noninvasive blood tests for fetal development predict gestational age and preterm delivery. *Science*. 2018;360:1133–6.
- Tarca AL, Pataki BA, Romero R, Sirota M, Guan Y, Kutum R, Gomez-Lopez N, Done B, Bhatti G, Yu T, et al. Crowdsourcing assessment of maternal blood multi-omics for predicting gestational age and preterm birth. *Cell Rep Med*. 2021;2: 100323.
- Jehan F, Sazawal S, Baqui AH, Nisar MI, Dhingra U, Khanam R, Ilyas M, Dutta A, Mitra DK, Mehmood U, et al. Multiomics characterization of preterm birth in low- and middle-income countries. *JAMA Netw Open*. 2020;3: e2029655.
- Stelzer IA, Ghaemi MS, Han X, Ando K, Hedou JJ, Feyaerts D, Peterson LS, Rumer KK, Tsai ES, Ganio EA, et al. Integrated trajectories of the maternal metabolome, proteome, and immunome predict labor onset. *Sci Transl Med*. 2021;13: eabd9898.
- Liang L, Rasmussen MH, Piening B, Shen X, Chen S, Rost H, Snyder JK, Tibshirani R, Skotte L, Lee NC, et al. Metabolic dynamics and prediction of gestational age and time to delivery in pregnant women. *Cell*. 2020;181:1680–1692.e15.
- Fettweis JM, Serrano MG, Brooks JP, Edwards DJ, Girerd PH, Parikh HI, Huang B, Arodz TJ, Edupuganti L, Glascock AL, et al. The vaginal microbiome and preterm birth. *Nat Med*. 2019;25:1012–21.
- Tabatabaei N, Eren AM, Barreiro LB, Yotova V, Dumaine A, Allard C, Fraser WD. Vaginal microbiome in early pregnancy and subsequent risk of spontaneous preterm birth: a case-control study. *BJOG*. 2019;126:349–58.
- Maltepe E, Fisher SJ. Placenta: the forgotten organ. *Annu Rev Cell Dev Biol*. 2015;31:523–52.
- Paquette AG, MacDonald J, Bammler T, Day DB, Loftus CT, Buth E, Mason WA, Bush NR, Lewinn KZ, Marsit C, et al. Placental transcriptomic signatures of spontaneous preterm birth. *Am J Obstet Gynecol*. 2022;228:73.e1–73.e18.
- Drag MH, Kilpelainen TO. Cell-free DNA and RNA-measurement and applications in clinical diagnostics with focus on metabolic disorders. *Physiol Genom*. 2021;53:33–46.
- Edlow AG, Bianchi DW. Tracking fetal development through molecular analysis of maternal biofluids. *Biochim Biophys Acta*. 2012;1822:1970–80.
- Rasmussen M, Reddy M, Nolan R, Camunas-Soler J, Khodursky A, Scheller NM, Cantonwine DE, Engelbrechtsen L, Mi JD, Dutta A, et al. RNA profiles reveal signatures of future health and disease in pregnancy. *Nature*. 2022;601:422–7.
- Tsui NB, Jiang P, Wong YF, Leung TY, Chan KC, Chiu RW, Sun H, Lo YM. Maternal plasma RNA sequencing for genome-wide transcriptomic profiling and identification of pregnancy-associated transcripts. *Clin Chem*. 2014;60:954–62.
- Tsui NB, Lo YM. Placental RNA in maternal plasma: toward noninvasive fetal gene expression profiling. *Ann N Y Acad Sci*. 2006;1075:96–102.
- Wang X, Spandidos A, Wang H, Seed B. PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res*. 2012;40:D1144–1149.
- Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14:178–92.
- Joglekar MV, Wong WKM, Ema FK, Georgiou HM, Shub A, Hardikar AA, Lappas M. Postpartum circulating microRNA enhances prediction of future type 2 diabetes in women with previous gestational diabetes. *Diabetologia*. 2021;64:1516–26.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–40.
- Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*. 2019;47:W191–8.
- Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. 2021;49:D605–12.
- Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol*. 2014;8(Suppl 4):S11.
- Han Y, Wu J, Gong Z, Zhou Y, Li H, Wang B, Qian Q. Identification and development of a novel 5-gene diagnostic model based on immune infiltration analysis of osteoarthritis. *J Transl Med*. 2021;19:522.
- Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10:1523.
- Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. *Lancet*. 2008;371:75–84.
- Hendler I, Goldenberg RL, Mercer BM, Iams JD, Meis PJ, Moawad AH, MacPherson CA, Caritis SN, Miodovnik M, Menard KM, et al. The preterm prediction study: association between maternal body mass index and spontaneous and indicated preterm birth. *Am J Obstet Gynecol*. 2005;192:882–6.
- Jiang M, Mishu MM, Lu D, Yin X. A case control study of risk factors and neonatal outcomes of preterm birth. *Taiwan J Obstet Gynecol*. 2018;57:814–8.
- Koullali B, van Zijl MD, Kazemier BM, Oudijk MA, Mol BWJ, Pakjrt E, Ravelli ACJ. The association between parity and spontaneous preterm birth: a population based study. *BMC Pregnancy Childbirth*. 2020;20:233.
- Ma M, Zhu M, Zhuo B, Li L, Chen H, Xu L, Wu Z, Cheng F, Xu L, Yan J. Use of complete blood count for predicting preterm birth in asymptomatic pregnant women: a propensity score-matched analysis. *J Clin Lab Anal*. 2020;34: e23313.
- Zhang X, Xu Q, Yang Y, Wang L, Liu F, Li Q, Ji M, He Y, Wang Y, Zhang Y, et al. Preconception Hb concentration and risk of preterm birth in over 2.7 million Chinese women aged 20–49 years: a population-based cohort study. *Br J Nutr*. 2018;120:508–16.
- Green ES, Arck PC. Pathogenesis of preterm birth: bidirectional inflammation in mother and fetus. *Semin Immunopathol*. 2020;42:413–29.
- Pique-Regi R, Romero R, Tarca AL, Sandler ED, Xu Y, Garcia-Flores V, Leng Y, Luca F, Hassan SS, Gomez-Lopez N. Single cell transcriptional signatures of the human placenta in term and preterm parturition. *Elife*. 2019;8: e52004.
- Tsang JCH, Vong JSL, Ji L, Poon LCY, Jiang P, Lui KO, Ni YB, To KF, Cheng YKY, Chiu RWK, Lo YMD. Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. *Proc Natl Acad Sci USA*. 2017;114:E7786–95.
- Bhattacharjee E, Maitra A. Spontaneous preterm birth: the underpinnings in the maternal and fetal genomes. *NPJ Genom Med*. 2021;6:43.
- Knijnenburg TA, Vockley JG, Chambwe N, Gibbs DL, Humphries C, Huddleston KC, Klein E, Kothiyal P, Tasseff R, Dhankani V, et al. Genomic and molecular characterization of preterm birth. *Proc Natl Acad Sci USA*. 2019;116:5819–27.
- Sakabe NJ, Aneas I, Knoblauch N, Sobreira DR, Clark N, Paz C, Horth C, Ziffra R, Kaur H, Liu X, et al. Transcriptome and regulatory maps of

- decidua-derived stromal cells inform gene discovery in preterm birth. *Sci Adv.* 2020;6: eabc8696.
40. Strauss JF 3rd, Romero R, Gomez-Lopez N, Haymond-Thornburg H, Modi BP, Teves ME, Pearson LN, York TP, Schenkein HA. Spontaneous preterm birth: advances toward the discovery of genetic predisposition. *Am J Obstet Gynecol.* 2018;218:294-314.e2.
 41. Tiensuu H, Haapalainen AM, Karjalainen MK, Pasanen A, Huusko JM, Marttila R, Ojaniemi M, Muglia LJ, Hallman M, Ramet M. Risk of spontaneous preterm birth and fetal growth associates with fetal SLIT2. *PLoS Genet.* 2019;15: e1008107.
 42. Zhang G, Feenstra B, Bacelis J, Liu X, Muglia LM, Juodakis J, Miller DE, Litterman N, Jiang PP, Russell L, et al. Genetic associations with gestational duration and spontaneous preterm birth. *N Engl J Med.* 2017;377:1156-67.
 43. Gong S, Gaccioli F, Dopierala J, Sovio U, Cook E, Volders PJ, Martens L, Kirk PDW, Richardson S, Smith GCS, Charnock-Jones DS. The RNA landscape of the human placenta in health and disease. *Nat Commun.* 2021;12:2639.
 44. Hong S, Park KH, Kim YM, Lee YE, Park Y, Lee JE. A protein microarray analysis of plasma proteins for the prediction of spontaneous preterm delivery in women with preterm labor. *Reprod Sci.* 2020;27:1187-96.
 45. Brown HK, Speechley KN, Macnab J, Natale R, Campbell MK. Neonatal morbidity associated with late preterm and early term birth: the roles of gestational age and biological determinants of preterm birth. *Int J Epidemiol.* 2014;43:802-14.
 46. American College of Obstetricians and Gynecologists. ACOG Committee Opinion No 579: definition of term pregnancy. *Obstet Gynecol.* 2013;122:1139-40.
 47. Baron IS, Litman FR, Ahronovich MD, Baker R. Late preterm birth: a review of medical and neuropsychological childhood outcomes. *Neuropsychol Rev.* 2012;22:438-50.
 48. Mitha A, Chen R, Altman M, Johansson S, Stephansson O, Bolk J. Neonatal morbidities in infants born late preterm at 35-36 weeks of gestation: a Swedish nationwide population-based study. *J Pediatr.* 2021;233:43-50.e45.
 49. Smith MD, Pillman K, Jankovic-Karasoulos T, McAninch D, Wan Q, Bogias KJ, McCullough D, Bianco-Miotto T, Breen J, Roberts CT. Large-scale transcriptome-wide profiling of microRNAs in human placenta and maternal plasma at early to mid gestation. *RNA Biol.* 2021;18:507-20.
 50. Valentine CJ, Khan AQ, Brown AR, Sands SA, Defranco EA, Gajewski BJ, Carlson SE, Reber KM, Rogers LK. Higher-dose DHA supplementation modulates immune responses in pregnancy and is associated with decreased preterm birth. *Nutrients.* 2021;13:4248.
 51. Ansari A, Bose S, You Y, Park S, Kim Y. Molecular mechanism of microbiota metabolites in preterm birth: pathological and therapeutic insights. *Int J Mol Sci.* 2021;22:8145.
 52. Telkar N, Stewart GL, Pewarchuk ME, Cohn DE, Robinson WP, Lam WL. Small non-coding RNAs in the human placenta: regulatory roles and clinical utility. *Front Genet.* 2022;13: 868598.
 53. Droppelmann CA, Wang J, Campos-Melo D, Keller B, Volkening K, Hegele RA, Strong MJ. Detection of a novel frameshift mutation and regions with homozygosity within ARHGEF28 gene in familial amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Frontotemporal Degener.* 2013;14:444-51.
 54. Wells HRR, Freidin MB, Zainul Abidin FN, Payton A, Dawes P, Munro KJ, Morton CC, Moore DR, Dawson SJ, Williams FMK. GWAS identifies 44 independent associated genomic loci for self-reported adult hearing difficulty in UK Biobank. *Am J Hum Genet.* 2019;105:788-802.
 55. Allotey J, Zamora J, Cheong-See F, Kalidindi M, Arroyo-Manzano D, Asztalos E, van der Post J, Mol BW, Moore D, Birtles D, et al. Cognitive, motor, behavioural and academic performances of children born preterm: a meta-analysis and systematic review involving 64 061 children. *BJOG.* 2018;125:16-25.
 56. Lindstrom K, Lindblad F, Hjerna A. Preterm birth and attention-deficit/hyperactivity disorder in schoolchildren. *Pediatrics.* 2011;127:858-65.
 57. Miller JV, Chau V, Synnes A, Miller SP, Grunau RE. Brain development and maternal behavior in relation to cognitive and language outcomes in preterm-born children. *Biol Psychiatry.* 2022;92(8):663-73.
 58. Paulés C, Pueyo V, Martí E, de Vilchez S, Burd I, Calvo P, Oros D. Threatened preterm labor is a risk factor for impaired cognitive development in early childhood. *Am J Obstet Gynecol.* 2017;216:157.e1-157.e7.
 59. Bao S, Zhou T, Yan C, Bao J, Yang F, Chao S, Zhou M, Xu Z. A blood-based miRNA signature for early non-invasive diagnosis of preeclampsia. *BMC Med.* 2022;20:303.
 60. Torres-Munoz J, Jimenez-Fernandez CA, Ortega RR, Cuero DJM, Mendoza DM. Factors associated with late prematurity in the university hospital of Valle Cali, Colombia during 2013-2014. *Front Public Health.* 2020;8:200.
 61. Sterckx L, Vandewiele G, Dehaene I, Janssens O, Ongenaes F, De Backere F, De Turck F, Roelens K, Decruyenaere J, Van Hoecke S, Demeester T. Clinical information extraction for preterm birth risk prediction. *J Biomed Inform.* 2020;110: 103544.
 62. Chen A, Feresu SA, Barsoom MJ. Heterogeneity of preterm birth subtypes in relation to neonatal death. *Obstet Gynecol.* 2009;114:516-22.
 63. Li Z, Ye R, Zhang L, Li H, Liu J, Ren A. Periconceptional folic acid supplementation and the risk of preterm births in China: a large prospective cohort study. *Int J Epidemiol.* 2014;43:1132-9.
 64. Catov JM, Scifres CM, Caritis SN, Bertolet M, Larkin J, Parks WT. Neonatal outcomes following preterm birth classified according to placental features. *Am J Obstet Gynecol.* 2017;216:411.e1-411.e14.
 65. Stanek J. Comparison of placental pathology in preterm, late-preterm, near-term, and term births. *Am J Obstet Gynecol.* 2014;210:234.e1-234.e6.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

