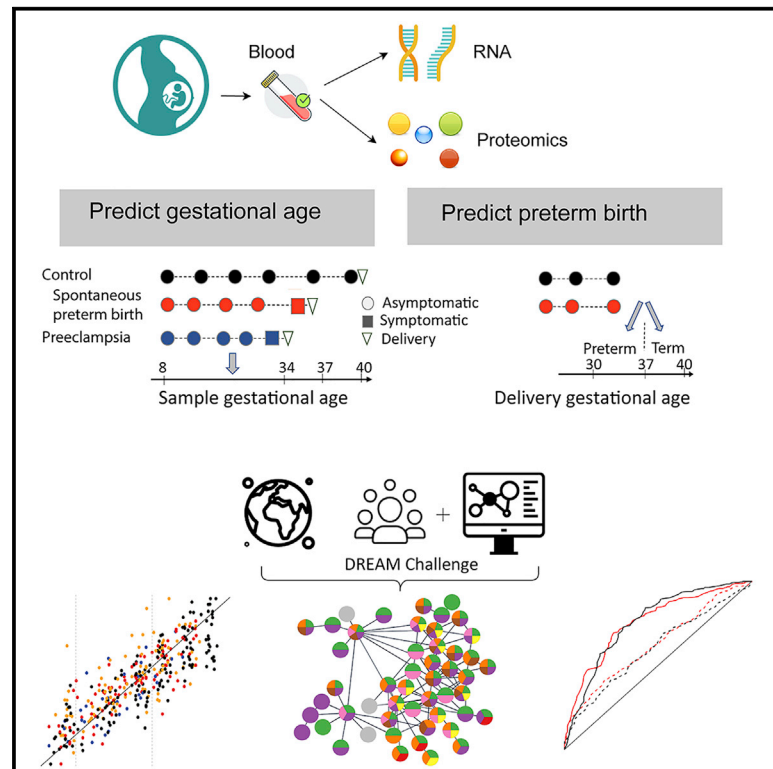**Article**

Crowdsourcing assessment of maternal blood multi-omics for predicting gestational age and preterm birth

Graphical abstract

Authors

Adi L. Tarca, Bálint Ármin Pataki, Roberto Romero, ..., Gustavo Stolovitzky, Istvan Csabai, James C. Costello

Correspondence

atarca@med.wayne.edu (A.L.T.),
prbchiefstaff@med.wayne.edu (R.R.),
gustavo@us.ibm.com (G.S.),
james.costello@cuanschutz.edu (J.C.C.)

In brief

Harnessing the wisdom of crowds in a DREAM Challenge, Tarca et al. developed methods to predict gestational age and preterm birth from longitudinal multi-omics data. The authors show that blood RNAs predict ultrasound-based gestational age, and they identify molecular changes preceding a diagnosis of spontaneous preterm birth in asymptomatic women.

Highlights

- Blood gene expression predicts gestational age in normal and complicated pregnancies

- RNA changes preceding preterm prelabor rupture of the membranes are shared between cohorts

- Plasma proteomic profiles from asymptomatic women predict spontaneous preterm birth

CellPress

Cell Reports Medicine

Article

Crowdsourcing assessment of maternal blood multi-omics for predicting gestational age and preterm birth

Adi L. Tarca,[1,2,3,22,*] Bálint Ármin Pataki,[4] Roberto Romero,[1,5,6,7,8,9,*] Marina Sirota,[10,11] Yuanfang Guan,[12] Rintu Kutum,[13] Nardhy Gomez-Lopez,[1,2,14] Bogdan Done,[1,2] Gaurav Bhatti,[1,2] Thomas Yu,[15] Gaia Andreoletti,[10,11] Tinnakorn Chaiworapongsa,[2] The DREAM Preterm Birth Prediction Challenge Consortium, Sonia S. Hassan,[2,16,17] Chaur-Dong Hsu,[1,2,17] Nima Aghaeepour,[18] Gustavo Stolovitzky,[19,20,*] Istvan Csabai,[4] and James C. Costello[21,*]

[1]Perinatology Research Branch, Division of Obstetrics and Maternal-Fetal Medicine, Division of Intramural Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, US Department of Health and Human Services, Detroit, MI 48201, USA
[2]Department of Obstetrics and Gynecology, Wayne State University School of Medicine, Detroit, MI 48201 USA
[3]Department of Computer Science, Wayne State University College of Engineering, Detroit, MI 48202, USA
[4]Department of Physics of Complex Systems, ELTE Eötvös Loránd University, Budapest, Hungary
[5]Department of Obstetrics and Gynecology, University of Michigan, Ann Arbor, MI 48109, USA
[6]Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI 48824, USA
[7]Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI 48201, USA
[8]Detroit Medical Center, Detroit, MI 48201, USA
[9]Department of Obstetrics and Gynecology, Florida International University, Miami, FL 33199, USA
[10]Department of Pediatrics, University of California, San Francisco, San Francisco, CA 94143, USA
[11]Bakar Computational Health Sciences Institute, University of California, San Francisco, San Francisco, CA 94158, USA
[12]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA
[13]Informatics and Big Data Unit, CSIR-Institute of Genomics and Integrative Biology, New Delhi, India
[14]Department of Biochemistry, Microbiology, and Immunology, Wayne State University School of Medicine, Detroit, MI 48201 USA
[15]Sage Bionetworks, Seattle, WA, USA
[16]Office of Women's Health, Integrative Biosciences Center, Wayne State University, Detroit, MI 48202, USA
[17]Department of Physiology, Wayne State University School of Medicine, Detroit, MI 48201, USA
[18]Department of Anesthesiology, Perioperative, and Pain Medicine, Department of Pediatrics, and Department of Biomedical Data Sciences, Stanford University School of Medicine, Stanford, CA 94305, USA
[19]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[20]IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA
[21]Department of Pharmacology, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA
[22]Lead contact
*Correspondence: atarca@med.wayne.edu (A.L.T.), prbchiefstaff@med.wayne.edu (R.R.), gustavo@us.ibm.com (G.S.), james.costello@cuanschutz.edu (J.C.C.)
https://doi.org/10.1016/j.xcrm.2021.100323

SUMMARY

Identification of pregnancies at risk of preterm birth (PTB), the leading cause of newborn deaths, remains challenging given the syndromic nature of the disease. We report a longitudinal multi-omics study coupled with a DREAM challenge to develop predictive models of PTB. The findings indicate that whole-blood gene expression predicts ultrasound-based gestational ages in normal and complicated pregnancies (r = 0.83) and, using data collected before 37 weeks of gestation, also predicts the delivery date in both normal pregnancies (r = 0.86) and those with spontaneous preterm birth (r = 0.75). Based on samples collected before 33 weeks in asymptomatic women, our analysis suggests that expression changes preceding preterm prelabor rupture of the membranes are consistent across time points and cohorts and involve leukocyte-mediated immunity. Models built from plasma proteomic data predict spontaneous preterm delivery with intact membranes with higher accuracy and earlier in pregnancy than transcriptomic models (AUROC = 0.76 versus AUROC = 0.6 at 27–33 weeks of gestation).

INTRODUCTION

Early identification of patients at risk for obstetrical disease is required to improve health outcomes and develop new therapeu-

tic interventions. One of the "great obstetrical syndromes,"[1] preterm birth, defined as birth before the completion of 37 weeks of gestation, is the leading cause of newborn deaths worldwide. In 2010, 14.9 million babies were born preterm, accounting for

11.1% of all births across 184 countries, the highest preterm birth rates occurring in Africa and North America.[2] In the United States, the rate of prematurity remains fundamentally unchanged in recent years,[3] and it has an annual societal economic burden of at least $26.2 billion.[4] The high incidence of preterm birth is concerning, as 29% of all neonatal deaths worldwide, ~1 million deaths in total, can be attributed to complications of prematurity.[5] Furthermore, children born prematurely are at increased risk for several short- and long-term complications that may include motor, cognitive, and behavioral impairments.[6,7]

Approximately one-third of preterm births are medically indicated for maternal (e.g., preeclampsia) or fetal conditions (e.g., growth restriction); the other two-thirds are categorized as spontaneous preterm births, inclusive of spontaneous preterm labor and delivery with intact membranes (sPTD), and preterm prelabor rupture of the membranes (PPROM).[8] Preterm birth is a syndrome with multiple etiologies,[9] and its complexity makes accurate prediction by a single set of biomarkers difficult. While genetic risk factors for preterm birth have been reported,[10,11] the two most powerful predictors of spontaneous preterm birth are a sonographic short cervix in the midtrimester and a history of spontaneous preterm birth in a prior pregnancy.[12] As for prevention of the syndrome, vaginal progesterone administered to asymptomatic women with a short cervix in the midtrimester reduces the rate of preterm birth before 33 weeks of gestation by 45% and decreases the rate of neonatal complications, including neonatal respiratory distress syndrome.[13–15]

To compensate for the suboptimal prediction of preterm birth by currently used biomarkers, alternative approaches to identify biomarkers have been proposed, such as focusing on fetal and placenta-specific signatures,[16] with the latter eventually refined by single-cell genomics,[16,17] and by expanding the types of data collected via multi-omics platforms.[10,18,19] While molecular profiles have been shown to be strongly modulated by advancing gestation in the maternal blood proteome,[20,21] transcriptome,[16,22] and vaginal microbiome,[23–25] the timing of delivery based on such molecular clocks of pregnancy is still challenging.[16] A recent meta-analysis[26] suggests that specific changes in the maternal whole-blood transcriptome associated with spontaneous preterm birth are largely consistent across studies when both symptomatic and asymptomatic cases are involved and when the samples collected at or near the time of preterm delivery are also included. However, the accuracy of transcriptomic predictive models to make inferences in asymptomatic women early in pregnancy has not been evaluated, and aptamer-based high-throughput plasma proteomics patterns,[27] shown to be comprehensive indicators of health,[28,29] were not assessed in the context of spontaneous preterm birth. This topic is important, since identification of early biomarkers, along with the associated robust assay platform, are necessary to develop treatment strategies that reduce the impact of prematurity.

Therefore, we generated longitudinal whole-blood transcriptomic data at exon-level resolution and plasma proteomic data on 216 women and leveraged the Dialogue for Reverse Engineering Assessments and Methods (DREAM) crowdsourcing framework[30] to engage >500 members of the computational biology community and to robustly assess the value of maternal blood multi-omics data in two sub-challenges. In sub-challenge 1, we assessed maternal whole-blood transcriptomic data for prediction of gestational age in normal and complicated pregnancies using the last menstrual period (LMP) and ultrasound estimate as the gold standard, and showed that predictions are robust to disease-related perturbations. To avoid potential biases in the gold standard, in a post-challenge analysis, we also predicted delivery dates in women with spontaneous birth (Figure 1) and found similar prediction performance. In sub-challenge 2, we evaluated within- and cross-cohort prediction of preterm birth leveraging longitudinal transcriptomic data in asymptomatic women generated herein and by Heng et al.[31] in a cohort in Calgary, Canada. The separate consideration of both spontaneous preterm birth phenotypes (i.e., sPTD and PPROM), allowed us to pinpoint that previously reported leukocyte activation-related RNA changes preceding preterm birth are shared across the racially diverse cohorts for the PPROM phenotype but not for sPTD. Moreover, the evaluation of highly reproducible plasma proteomic assays[32] and blood multi-omics data to determine the earliest stage in gestation when biomarkers have predictive value (Figure 1) also make this study unique and led to the conclusion that changes in plasma proteomics can be detected earlier and are more accurate than whole-blood transcriptomics for prediction of preterm birth. In addition to the transcriptomic signatures of gestational age and the multi-omics signatures of preterm birth that were identified here, this work sets a benchmark for the evaluation of longitudinal omics data in pregnancy research. The computational lessons and algorithms for risk prediction from longitudinal omics data can be used to develop future studies.

RESULTS

Prediction of gestational age by maternal whole-blood transcriptomics

We have generated and shared with the community exon-level gene expression data profiled in 703 maternal whole-blood samples collected from 133 women enrolled in a longitudinal study at the Center for Advanced Obstetrical Care and Research of the Perinatology Research Branch, National Institute of Child Health and Human Development/National Institutes of Health/Department of Health and Human Services (NICHD/NIH/DHHS); the Detroit Medical Center; and the Wayne State University School of Medicine. The Human Transcriptome Arrays platform was chosen based on its favorable performance compared to RNA sequencing (RNA-seq), especially for quantifying short and low abundant genes,[33] and it was previously used for detecting gestational age- and parturition-related changes in maternal whole blood.[22] The patient population included women with a normal pregnancy who delivered at term ($\geq$37 weeks) (controls, N = 49), women who delivered before 37 completed weeks of gestation by sPTD (N = 34) or PPROM (N = 37) , and women who experienced an indicated delivery before 34 weeks due to early preeclampsia (N = 13) (Figure 2A). After including data from 16 additional normal pregnancies obtained from the same population[22] and using the same microarray platform (Gene Expression Omnibus: GSE113966, 32 transcriptomes), the resulting set of 149 pregnancies (see demographic characteristics
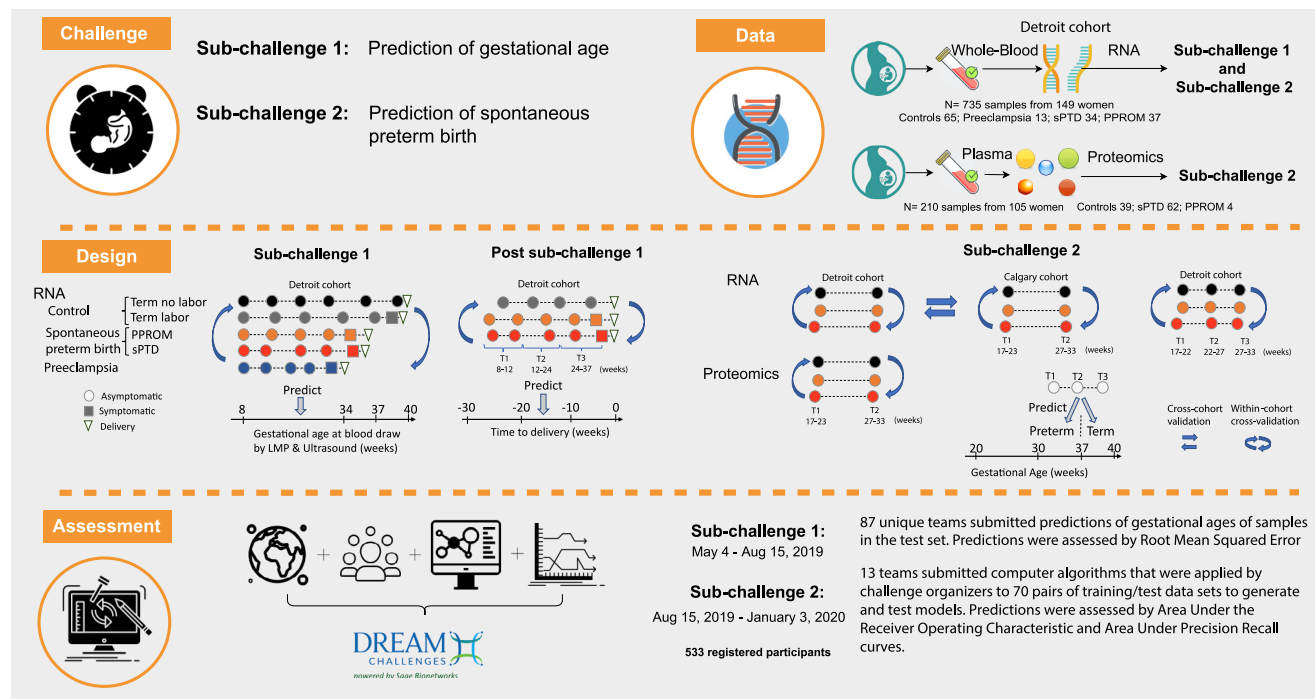
Figure 1. Study overview

Whole-blood transcriptomic and/or plasma proteomic profiles were generated from 216 women with either normal pregnancy, spontaneous preterm birth with intact (sPTD) or ruptured membranes (PPROM), or preeclampsia. Sub-challenge 1: transcriptomic data were generated from samples collected in normal pregnancies without labor at term (black dots) and spontaneous labor at term (gray dots), and those complicated by sPTD (red dots), PPROM (orange dots), or preeclampsia (blue dots). Participating teams were provided gene expression data to develop prediction models for gestational age at blood draw defined by last menstrual period (LMP) and ultrasound (gold standard). Participants submitted predictions on a blinded test set (see Figure S1 for training/test partition). In a post-challenge analysis, the approach of the top team (smallest test set root mean square error [RMSE]) was applied to predict time to delivery. Sub-challenge 2: participants submitted risk prediction algorithms designed to use as input omics data at ≥2 time points and patient outcomes (control, sPTD, or PPROM) for a subset of them (training set), and return disease risk scores for women with blinded outcomes (test set). The algorithms were applied to 70 training/test pairs of datasets (see Figure 4A and Table S5) to assess within- and across-cohort predictions of preterm birth by transcriptomics and within-cohort prediction by multi-omics data. Predictions were assessed by area under the receiver operating characteristic and area under the precision recall curves and aggregated across datasets and prediction scenarios (see STAR Methods).

in Table S1), totaling 735 transcriptomes, was divided randomly into training (N = 367) and test (N = 368) sets; the latter set excludes publicly available data to avoid the possibility that the models are trained with data to be used for testing (Figure S1). All of the longitudinal samples for the same patient were assigned to the training set or the test set; thus, samples were not split between training/test sets (see Figure S1 and Video S1). The research community was challenged to use data from the training set to develop gene expression prediction models for gestational age, as defined by the LMP and ultrasound fetal biometry, and to make predictions based only on gene expression in the test set. The clinical diagnosis and sample-to-patient assignments were not disclosed to the challenge participants, while gestational age at the time of sampling was also blinded for the test set. Teams were allowed to submit up to 5 predictions using the test samples, and the best submission (smallest root mean square error [RMSE]) was retained for each unique team. We received 331 submissions for this sub-challenge from 87 participating teams, 37 teams of which provided the required details on the computational methods used to be qualified for the final team ranking in this sub-challenge (Table S2).

Robustness analysis of team rankings (see STAR Methods) suggested that the predictions of the top-ranked team (authors B.A.P. and I.C., abbreviated as team 1) were significantly better (Bayes factor >3) than those of the second- (author Y.G., abbreviated as team 2) and third-ranked teams (Figure S1). Among the top 20 teams, the most frequent methods used to select predictor genes included univariate gene ranking and meta-gene building via principal-component analysis, as well as literature-based gene selection. Common prediction models included neural networks, random forest, and regularized regression (LASSO and ridge regression), with the latter being used by the top-ranked team in this sub-challenge.

The model generated by team 1 in sub-challenge 1 predicted gestational ages at blood draw with an RMSE of 4.5 weeks in the test set (Pearson correlation between actual and predicted values, r = 0.83, p < 0.001) (Figure 2B). The correlation between predicted and actual gestational ages was also significant after accounting for repeated observations from the same patients in the test set via linear mixed-effects modeling (slope 0.76, likelihood ratio test p < 0.001; see STAR Methods). The prediction model of team 1 (M_GA_Team1) was based on ridge regression,

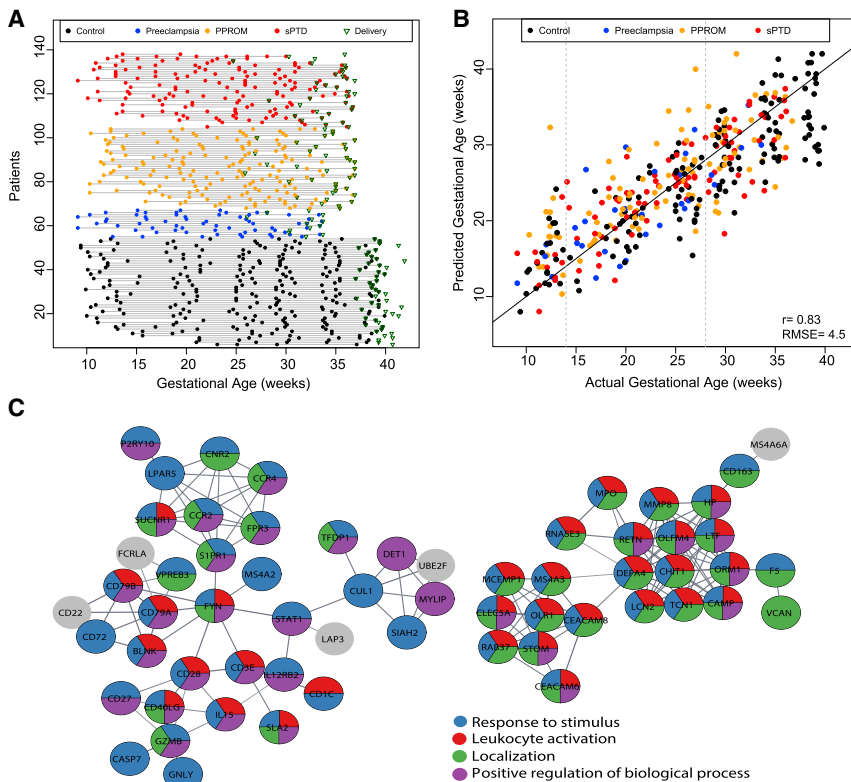**Figure 2. Prediction of gestational age by whole-blood transcriptomics**

(A) Detroit cohort transcriptomics study design. Each line corresponds to 1 patient and each dot to 1 sample. Gestational ages at delivery are marked by a triangle. The set includes 703 samples from 133 women: controls (N = 49), women who delivered before 37 completed weeks of gestation by sPTD (N = 34) or PPROM (N = 37) and women who experienced an indicated delivery before 34 weeks due to early preeclampsia (N = 13).

(B) Test set prediction of gestational age by the model of the top-ranked team (M_GA_Team1). The 368 samples are colored according to the phenotypic group of patients. r, Pearson correlation coefficient. RMSE: root mean squared error.

(C) Protein-protein interaction network modules for genes part of the 249-gene core transcriptome predicting gestational age (M_GA_Core). A select group of biological processes enriched among these genes are shown in the pie charts.

MMP8, *CECAM8*, and *DEFA4*) were most highly modulated in the normal pregnancy group used here,[35] and others have shown the same to be true at a cell-free RNA level in a Danish cohort.[16] In addition, these data are consistent with the concept that pregnancy is characterized by a systemic cellular inflammatory response.[36–40] In this study, we also show that these mediators correlate with gestational age in both normal and complicated pregnancies, and the latter group contributed more than half of the transcriptomes used to fit and evaluate the models (Figure S1; Table S1).

and the predictors were meta-genes derived by principal-component analysis from the expression data of 6,106 genes. As shown in Figure 2B, the gestational age predictions showed little bias in the second trimester (14–28 weeks) samples (mean error 0.6 weeks); however, gestational ages of first-trimester samples were overestimated (mean error 3.7 weeks), while the third-trimester samples were underestimated (mean error −1.96 weeks). This finding can be understood, in part, by the larger number of second-trimester samples relative to first- and third-trimester samples available for training of the model. Of interest, the prediction errors for complicated pregnancies were similar to those of normal pregnancies (ANOVA, p > 0.1), suggesting that this model of gestational age, in general, was robust for obstetrical disease- and parturition-related perturbations in gene expression data (Figure S2).

To identify a core transcriptome predicting gestational age in normal and complicated pregnancies that captures most of the predictive power of the full model (M_GA_Team1) that involved >6,000 predictor genes, in a post-challenge analysis, we combined linear mixed effects modeling for longitudinal data[34] to prioritize gene expression and then used these features as input in a LASSO regression model. The resulting 249 gene regression models (M_GA_Core) (Figure S2; Table S3) had an RMSE of 5.1 weeks (r = 0.80) and involved 2 tightly connected modules related to immune response, leukocyte activation, inflammation- and development-related Gene Ontology biological processes (Figure 2C; Table S4). We previously reported that several member genes of these networks (e.g.,

Comparison of gene expression models and the clinical standard in predicting time to delivery (TTD) in women with spontaneous term or preterm birth

To enable a direct comparison with a previous landmark study of pregnancy dating by targeted cell-free RNA profiling,[16] in a post-challenge analysis, we used the same methods as described above for model M_GA_Team1 (see STAR Methods), except for the use of a time variable defined backward from delivery, hence, independent of LMP and ultrasound estimations as response [TTD = date at sample − date at delivery, (weeks)]. As in the study by Ngo et al.,[16] only those patients with spontaneous term delivery were included in this analysis, thus omitting the subset of normal pregnancies that had been truncated by elective cesarean delivery. The training set in this analysis included 74 transcriptomes from 18 women and the test set included 64 transcriptomes from 11 women with a spontaneous term delivery based on the original data split (Figure S1). As shown in Figure 3A, the gene expression model significantly predicted TTD with the same accuracy (RMSE, 4.5 weeks, r = 0.86, p < 0.001) as when predicting LMP and ultrasound-based gestational age in the full cohort of normal and complicated pregnancies. The predicted TTD values were then averaged over multiple samples per patient in a given gestational age
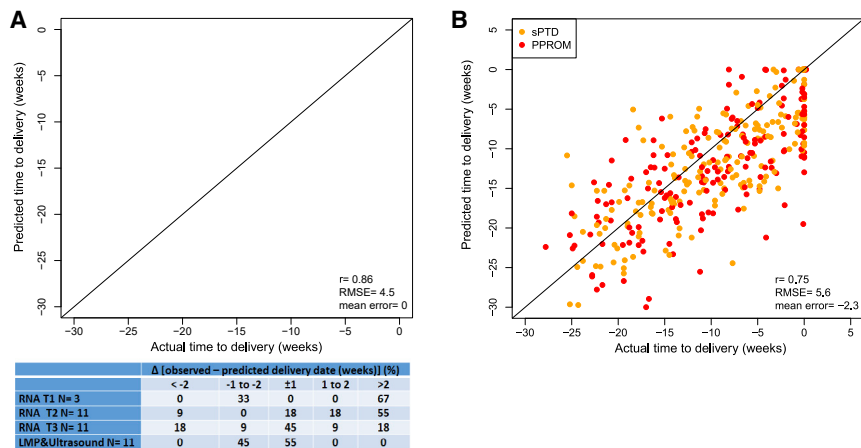
Figure 3. Prediction of time to delivery (TTD) by whole-blood transcriptomics

(A) The top panel shows the test set TTD estimates from the M_sTD_TTD model plotted against actual values for 64 transcriptomes from 11 women. The bottom panel shows the distribution of prediction errors (TTD observed–TTD predicted). A negative error means that delivery occurred sooner than expected/predicted, while positive values indicate the opposite. TTD was estimated using RNA measurements from the first- (T1), second- (T2), and third- (T3) trimester samples separately. For comparison, trimesters are defined as in Ngo et al.[16] T1: <12 weeks; T2 = 12–24 weeks, and T3 = 24–37 weeks of gestation.

(B) Prediction of TTD in women with spontaneous preterm birth by a gene expression model established in women with spontaneous term delivery (M_sTD_TTD). Predictions are shown for 355 longitudinal transcriptomes from 71 women with preterm prelabor rupture of membranes (PPROM, N = 37) and spontaneous preterm delivery with intact membranes (sPTD, N = 34). r, Pearson correlation coefficient. RMSE: root mean squared error.

interval to calculate accuracy, defined as predicting delivery within 1 week of the actual date and previously reported to be 55.1%.[41] Results were also compared to the LMP and ultrasound fetal biometry, with the latter predicting delivery at 40 weeks of gestation. The test set accuracy of the gene expression model was 45% (5/11) based on third-trimester samples, which is comparable to the LMP and ultrasound estimate based on first- or second-trimester fetal biometry (55%) (Figure 3A, bottom panel). Of note, 45% accuracy was also reported by Ngo et al.[16] using cell-free RNA based on second- and third-trimester samples.

When data from all of the pregnancies with spontaneous term delivery were used to train a transcriptomic model of time to delivery and the model was applied to data from women with spontaneous preterm birth, the prediction was found to be statistically significant. However, the error increased (RMSE = 5.6) (Figure 3B) relative to the estimate (RMSE = 4.5) for prediction of TTD in women with spontaneous term delivery (Figure 3A). The additional preterm parturition-specific perturbations in gene expression explain, in part, the added uncertainty in prediction estimates of TTD in spontaneous preterm birth cases compared to spontaneous term pregnancies. Moreover, as expected, the term pregnancy TTD model overestimated the duration of pregnancy of women who were destined to experience preterm birth (Figure 3B). The overestimation (mean prediction error) was 2.3 weeks compared to the 5-week gap between the LMP and ultrasound-based gestational ages at delivery in the term (mean = 39 weeks) and preterm (mean = 34 weeks) birth groups. The significant correlation of predicted and actual delivery dates in spontaneous preterm birth (sPTB) cases suggests that the M_sTD_TTD model captured both gene expression changes related to immune- and development-related processes establishing the age of pregnancy and the effects of the common pathway of parturition.[42,43] Hence, the prediction model generalized to the set of women with spontaneous preterm birth when

samples at or near delivery were included and genome-wide gene expression data were available.

Prediction of preterm birth by maternal blood omics data collected in asymptomatic women (sub-challenge 2)

Post-challenge analyses of sub-challenge 1 demonstrated that a whole-blood transcriptomic model derived from the data of women with spontaneous term delivery (M_sTD_TTD) predicted delivery date in spontaneous preterm birth cases based on data collected throughout pregnancy, including near or at the time of preterm parturition up to 37 weeks. With sub-challenge 2 of the DREAM Preterm Birth Prediction Challenge, we addressed the more difficult task of predicting preterm birth from data collected up to 33 weeks of gestation, while the women were asymptomatic. Of importance, the development of interventions to prevent preterm birth requires pregnant women at risk to be identified as early as possible before the onset of preterm parturition. Moreover, to enable future targeted studies of candidate biomarkers, we limited the maximum number of molecular predictors in this sub-challenge to be 50 per outcome considered (see Figure 4A).

We drew from the Detroit cohort longitudinal study (Figure 2A) only samples collected at specific gestational age intervals while women were asymptomatic (i.e., before an eventual diagnosis of sPTD or PPROM). Two scenarios of prediction of preterm birth were devised: (1) to include cases and controls with available samples collected at 17–23 and 27–33 weeks (Figure S3A), and (2) to include patients with available samples collected at 3 gestational age intervals (17–22, 22–27, and 27–33 weeks) (Figure S3B). The selection of the 17- 23- and 27- 33-week intervals enabled cross-study model development and testing, with the microarray gene expression study of Heng et al.[31] derived from a cohort in Calgary, Canada. Furthermore, we also included the profiles of 1,125 maternal plasma proteins measured by using an aptamer-based technology[27,44] in samples collected at
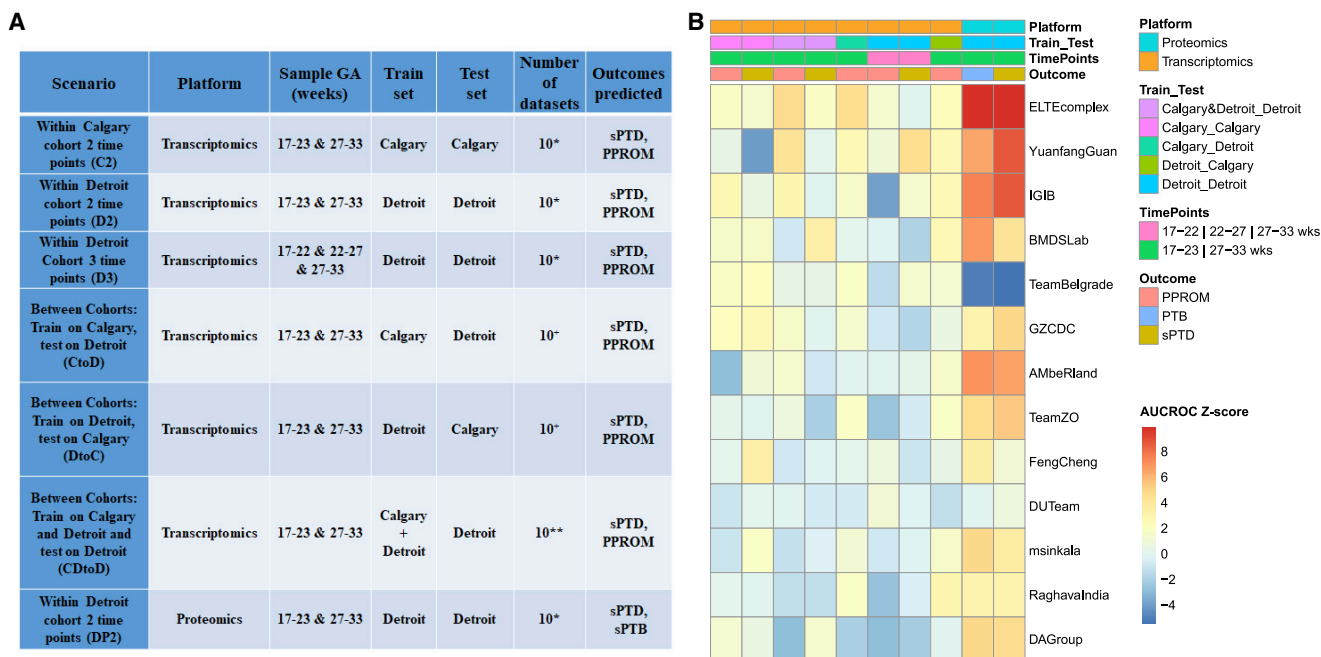
A

B

Figure 4. Sub-challenge 2 design and results

(A) Scenarios of spontaneous preterm birth model training and testing using multi-omics data. *Subjects in the original cohort were randomly split into equally sized groups that were balanced with respect to the phenotypes. **One-fifth of patients from the Detroit cohort (balanced with respect to the phenotypes) were randomly selected in the training set, while the remaining four-fifths were used as the test set. +Training set subjects were sampled with replacements from the original cohort to create different versions of the training set, and the trained model was then applied to the original test cohort. Sample sizes of training and test sets are shown in Table S5.

(B) Prediction performance for preterm birth-related outcomes based on algorithms submitted by 13 teams. AUROC values were converted into Z scores and shown as a heatmap for scenarios/outcome combinations shown in (A) that led to a significant prediction. sPTB, spontaneous preterm birth.

17–23 and 27–33 weeks of gestation from 66 women before the diagnosis of preterm birth (62 sPTD and 4 PPROM). These samples were profiled in the same experimental batch with samples from 39 normal pregnancies that we previously described,[21,45] which served here as controls (Figure S3C). The characteristics of pregnancies with available proteomics profiles are shown in Table S1.

The prediction algorithms generated by 13 teams that participated in sub-challenge 2 were applied by the Challenge organizers to train and test models on 70 pairs of training/test datasets generated under 7 scenarios (Figure 4A; Table S5; Video S1). The scenarios differed in terms of omics data type, number of longitudinal measurements per patient, the outcome being predicted, and the patient cohorts used for training/testing (Figure 4A). In all of the cases, there were no differences in terms of number of samples and gestational age at sampling between the cases and controls (Figure S3).

To assess the prediction performance in each of the 70 test sets in sub-challenge 2, we used both the area under the receiver operating characteristic curve (AUROC) as well as the area under the precision-recall curve (AUPRC), the latter being especially suited to imbalanced datasets (e.g., the proteomics set that features more cases than controls) (Figure S3C).

AUROC and AUPRC metrics were averaged over the 10 test sets of each prediction scenario and the result for each team was converted into a Z score. Final team rankings were obtained

by aggregating the ranks over all of the scenarios and outcome combinations that were significant, according to at least one team, after multiple testing correction (Table S6; see STAR Methods). Robustness analysis of team ranks (Figure S3D) determined that the top-ranked team (authors B.A.P. and I.C., abbreviated as team 1) outperformed the second-ranked team (author Y.G. abbreviated as team 2) and that the second- and third-ranked teams outperformed the fifth-ranked team (Bayes factor > 3). For all of the scenarios (Figure 4A), the models of team 1 involved data from 50 molecules (RNA or proteins) collected at the last available measurement (closest to delivery), while team 2 used data collected at the last 2 available time points for 50 molecules selected based on overall expression as opposed to correlation with the outcome. Among other differences in their approaches, team 1 treated the outcome as a binary variable, while team 2 used a continuous variable derived from gestational age at delivery (see STAR Methods). Of note, the two top-ranked teams were the same in both sub-challenges 1 and 2.

A summary of prediction scenarios and outcome combinations with significant predictions based on the approach of at least one team in sub-challenge 2 is depicted in Figure 4B. These results suggest overall higher prediction accuracy based on proteomics compared to transcriptomic data. We next highlight some of the prediction results for the top team.

With the approach of team 1, one transcriptomic profile at 27–33 weeks of gestation from asymptomatic women predicted
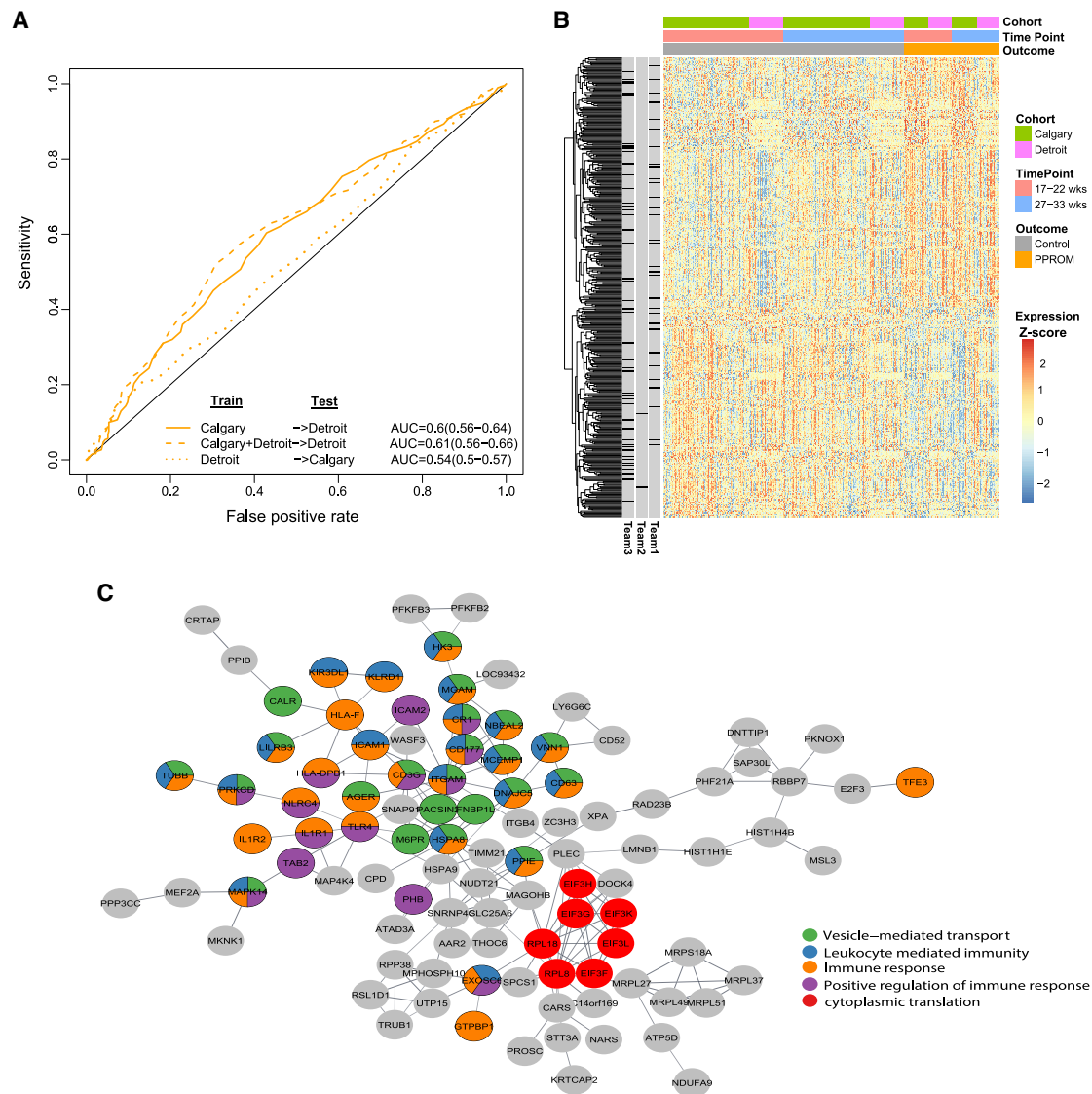
Figure 5. Prediction of preterm prelabor rupture of the membranes from samples collected in asymptomatic women

(A) Receiver operating characteristic (ROC) curve representing prediction of PPROM by 50 genes across the cohorts and microarray platforms using the team 1 approach. Sample sizes of test sets used to derive the ROC curves are shown in Table S5. AUC: area under the curve is given with 95% DeLong confidence intervals.

(B) Heatmap of 402 genes differentially expressed in PPROM across the cohorts and time points. Bars on the left indicate gene inclusion as a predictor by the methods of the top 3 teams in sub-challenge 2.

(C) STRING network constructed from among the 402 genes with differential expression in PPROM. Select significantly enriched biological processes are highlighted.

PPROM across the cohorts and microarray platforms with an AUCROC of ~0.6, depending on the prediction scenario (Figure 5A). Although separate differential expression analyses of the data from each cohort and time point failed to reach statistical significance after multiple testing correction, the consistency across cohorts and time points of gene expression changes preceding the diagnosis of PPROM was demonstrated by a post-challenge individual patient meta-analysis, which identified 402 differentially expressed genes after adjusting for cohort and

time point (moderated t test; q < 0.1) (Figure 5B; Table S7). A highly connected protein-protein interaction sub-network corresponding to genes significant in this meta-analysis is shown in Figure 5C, and it illustrates some of the Gene Ontology biological processes significantly enriched in PPROM. These included vesicle-mediated transport and leukocyte- (myeloid and lymphocyte) mediated immunity, among others (Table S4). Enrichment analysis based on canonical pathways and custom gene sets curated in the Molecular Signatures database
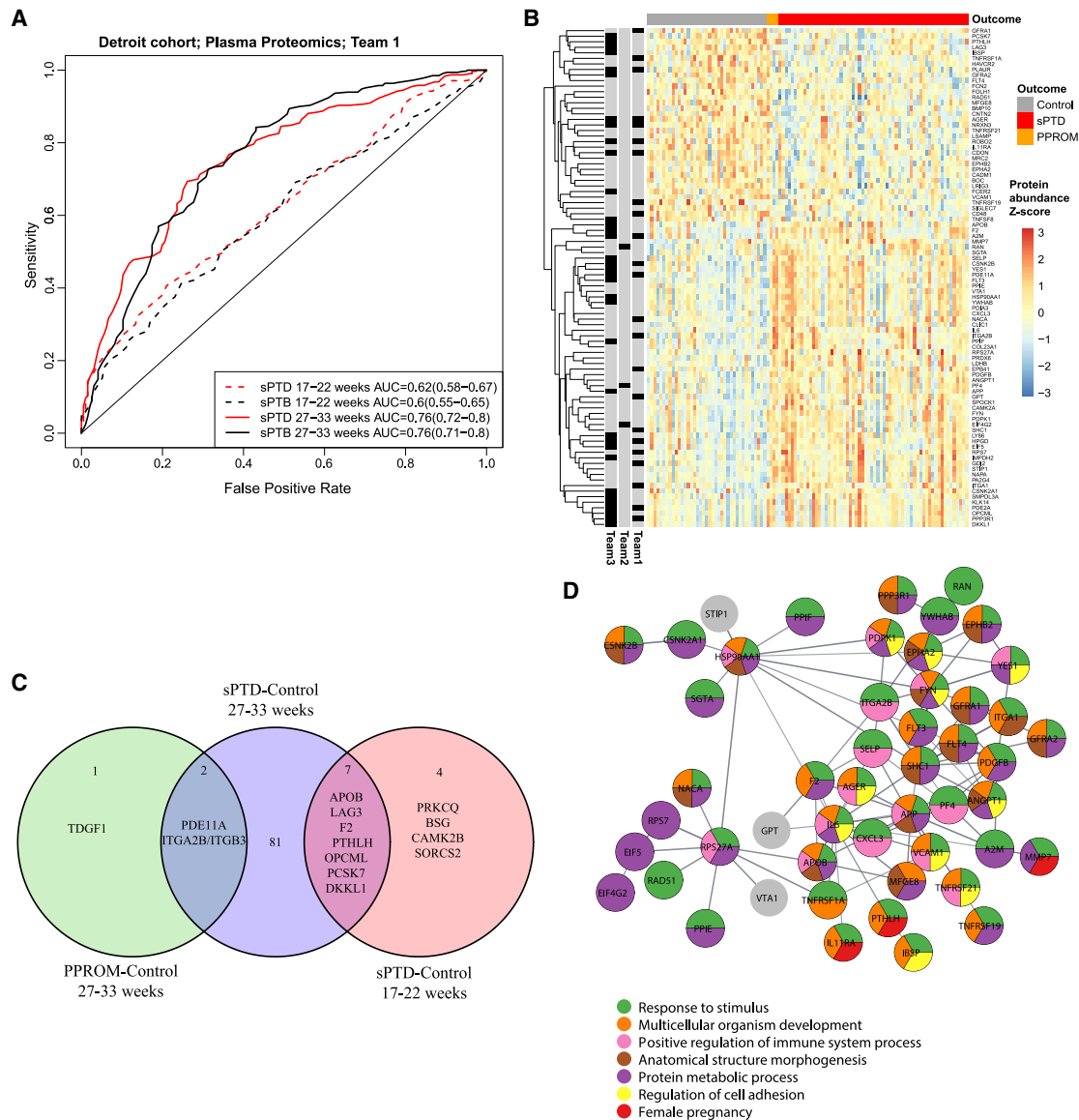
Figure 6. Prediction of spontaneous preterm delivery by plasma proteomic data

(A) ROC curve for sPTD and sPTB (which includes sPTD and PPROM) for team 1. The ROC curves were obtained from pooled predictions over 10 test sets each test set including 20 controls versus 31 sPTD cases and 20 controls versus 33 sPTB cases (see Table S5). AUC: area under the curve is given with 95% DeLong confidence intervals.

(B) Plasma protein abundance for all proteins deemed significant according to a moderated t test (q < 0.1); those selected as predictors by the top teams in their models are marked on the left side of the heatmap.

(C) Overlap of protein changes with sPTD at 17–22 and 27–33 weeks, and with PPROM at 27–33 weeks. See also Table S8.

(D) Network of proteins among those shown in (B): each protein node is annotated to biological processes based on corresponding Gene Ontology.

(MSigDB)[46] revealed perturbations associated with PPROM in 59 pathways such as interleukin-12 (IL-12), membrane trafficking, cytokine signaling in immune system, cellular senescence, and integrin cell surface interactions, among others (q < 0.1; see Table S4). These data are consistent with the hypothesis that circulating myeloid (monocytes and neutrophils) and lymphoid (T cells) cells are especially activated in women who experience pregnancy complications such as preterm labor[47–50] and PPROM.[51]

Although participating teams in sub-challenge 2 did not have access to the longitudinal preterm birth plasma proteomics when they developed prediction algorithms, their algorithms resulted in prediction performances that surpassed those obtained by using transcriptomic data (Figures 4 and 6A; Table S6) when applied to training and test sets derived from the plasma proteomics set (Figure S3C). Prediction of sPTD by the approach of team 1 involved 50 plasma proteins selected by random forest model importance from the panel of 1,125 available proteins.
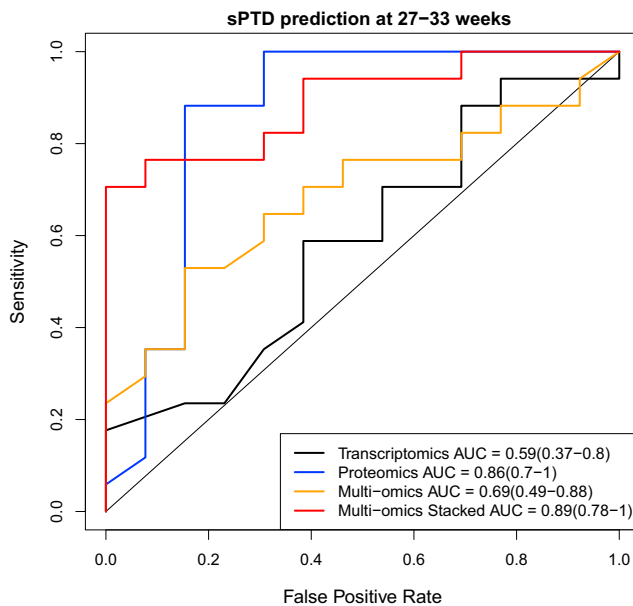
Figure 7. Comparison of prediction performance of spontaneous preterm delivery between platforms

ROC curve for prediction of sPTD by models obtained with the approach of team 1 based on a subset of samples for which data from both platforms were available. Leave-one-out cross-validation was used to generate the ROC curves from a set of 13 controls and 17 sPTD cases. The multi-omics model was obtained by applying the same approach on a concatenated set of proteomic and transcriptomic features. The multi-omics stacked generalization approach involved combining predictions from models based on each platform via LASSO logistic regression. AUC: area under the curve is given with 95% DeLong confidence intervals.

The test set accuracy was the highest when using data collected at 27–33 weeks of gestation (AUROC = 0.76 [0.72–0.8]). However, importantly, even one proteome profile at 17–22 weeks of gestation predicted spontaneous preterm delivery significantly (AUROC = 0.62 [0.58–0.67]) (Figure 6A), suggesting that this approach has value in the early identification of women at risk. The addition of four cases with PPROM to those with sPTD did not affect the prediction performance of the proteomics models of team 1, suggesting that this approach could generalize to both preterm birth phenotypes. The increase in plasma protein abundance of PDE11A and ITGA2B preceded the diagnosis of both sPTD and PPROM in the Detroit cohort at 27–33 weeks of gestation (Figures 6B and 6C). The tightly interconnected network of proteins built from differential profiles with sPTD in asymptomatic women at 27–33 weeks of gestation (q < 0.1; Figure 6B; Table S8) included not only several previously known markers of preterm delivery (IL-6, ANGPT1) but also MMP7 and ITGA2B, which we previously described as dysregulated in women with preeclampsia.[52] Member proteins of this network perturbed before a diagnosis of spontaneous preterm delivery are annotated to biological processes such as regulation of cell adhesion, response to stimulus, and development (Figure 6D). Differentially expressed proteins preceding diagnosis with sPTD also included mediators annotated to biological processes found by transcriptomic analysis in PPROM, such as leukocyte-mediated immunity

(AGER, PDPK1, LAG3, HAVCR2, IL-6, FCER2, CADM1), neutrophil-mediated immunity (PLAUR, IMPDH2, PRDX6, PA2G4, F2, IL-6, PPIE, GDI2), and regulation of vesicle-mediated transport (NAPA, PDPK1, MFGE8, ANGPT1, CAMK2A); however, enrichment of these biological processes did not reach statistical significance. In contrast, the pathway enrichment analysis based on MSigDB identified AMB2 neutrophils and cell surface interactions at the vascular wall pathways as significantly enriched based on plasma proteomic dysregulation preceding diagnosis with sPTD (q < 0.1; Table S4). Other top-ranked pathways included nervous system development, developmental biology, focal adhesion, VEGFA/VEGFR2 signaling, and membrane trafficking pathways (p < 0.05; Table S4), with the latter two being in common with those involved in PPROM (Table S4).

Given that differences in the patient characteristics could have contributed to the higher prediction performance of spontaneous preterm delivery by plasma proteomics as compared to maternal whole-blood transcriptomics, the approach of team 1 was also evaluated via leave-one-out cross-validation on a subset of 13 controls and 17 sPTD cases for which both types of data originated from the same blood draw. The prediction performance for spontaneous preterm delivery by plasma proteomics remained high (AUROC = 0.86 [0.7–1.0]), while prediction by transcriptomic data remained non-significant (Figure 7), confirming the superior value of proteomics relative to transcriptomics for this endpoint. Of note, for a fixed number of 50 predictors allowed, a stacked generalization[53] approach combining predictions from individual platform models via a LASSO logistic regression led to a higher leave-one-out cross-validation performance estimate (AUROC = 0.89 [0.78–1.0]) compared to building a single model from the combined transcriptomic and proteomic features (Figure 7).

To extract further insights from the computational approaches best suited to predict preterm birth from longitudinal omics data in sub-challenge 2, we investigated which computational aspects explained the higher performances of the top two teams. Given that team 1 relied only on omics data at the last available time point (T2), we kept all of the aspects of this method except for the temporal information considered among the following: (1) first point (T1), (2) change in expression between T2 and T1 (slope), or (3) a combined approach in which slopes for all genes and measurements at T2 compete for inclusion in the 50 allowed predictors for a given outcome (PPROM or sPTD) (see STAR Methods). As shown in Figure S4, none of these approaches would have improved prediction performance relative to the baseline approach of team 1, which considered only the data from the last time point (T2). We then considered several key aspects of the approach of team 2 and have subsequently incorporated them in the approach of team 1 to determine whether such hybrid approaches could translate into higher performances relative to the baseline approach. In particular, we have modified the approach of team 1: (1) to start with only the top half of the most highly abundant features on each platform, (2) to convert the binary classification (preterm versus term) into a regression of gestational age at delivery, and (3) given the selected 50 predictor genes based on the correlation of T2 expression values with the outcome, to add the expression of those genes at the previous time point as independent predictors in the random

forest model. Of these three scenarios, the last, which expands the number of predictors from 50 to 100 without increasing the number of molecules, slightly outperformed the approach of the overall prediction performance of team 1 across scenarios (Figure S4) and led to the consistent prediction of PPROM in all cross-study analyses (see improvement in prediction from Figures 5A to S4). Interestingly, simply doubling the number of molecules profiled at T2 that were allowed as predictors in the model (from 50 to 100) led to a worse overall prediction performance relative to the approach of team 1 that used only 50 molecules at T2 (Figure S4). This finding suggests that for preterm birth prediction, it is more important to measure the right biomarkers at one additional time point than to double the number of markers at the most recent time point.

DISCUSSION

In this study, we evaluated maternal blood omics data to predict gestational age in normal and complicated pregnancies, as well as the risk of preterm birth. Although the main interest here was the prediction of spontaneous preterm birth, the correlation of omics data with advancing gestation was relevant not only to serve as a positive control for the evaluation of omics data but also to possibly provide relevant information for the development of more affordable tools to date pregnancy. We chose the DREAM collaborative competition framework[30] to identify the best computational methods for making inferences and to assess them in an unbiased and robust way based on longitudinal omics data that we and others have generated. DREAM Challenges have been used to establish unbiased performance benchmarks across a wide array of prediction tasks.[54–58] Moreover, the results gained from these challenges define community standards and advancements in many scientific fields.[59,60]

Collectively, sub-challenge 1 and the additional post-challenge analyses demonstrated that models based on the maternal whole-blood transcriptome (1) significantly predict LMP and ultrasound-defined gestational age at venipuncture in both normal and complicated pregnancies (RMSE = 4.5) and (2) predict a delivery date within ±1 week in women with spontaneous term delivery with an accuracy (45%) comparable to the clinical standard (55%). The accuracy of gestational age prediction was likely understated based on sub-challenge 1 results due to the inclusion of cases with early preeclampsia and spontaneous preterm birth at much higher rates than expected in the general population. Disease-specific perturbations, especially close to the time of delivery in early preeclampsia and spontaneous preterm birth cases, are expected to have contributed additional variation to gene expression patterns establishing gestational age.

Of interest, the accuracy of dating gestation in women with spontaneous term delivery was similar to the report by Ngo et al.,[16] who used cell-free RNA profiling in a Danish cohort, although that study involved more frequent (weekly) sampling of fewer genes (about 50 immune, placental, and fetal liver specific) instead of the genome-wide data used here. In the study by Ngo et al.,[16] the TTD transcriptomic model derived from samples of women with normal pregnancy failed to predict delivery dates on independent cohorts of women with preterm birth, while in this study, such a model resulted in the significant prediction

of delivery dates in cases with spontaneous preterm birth (r = 0.75; Figure 3B). A possible explanation, in addition to the cohort differences between the training and testing sets in the previous study, is that our model of normal pregnancy captured not only gene changes establishing the gestational age but also those changes involved in the common pathway of labor. While the prediction of the delivery date of women with spontaneous preterm birth by omics data collected up to <37 weeks, including samples taken when women were symptomatic, was demonstrated above without using any data from preterm birth cases to establish the model, it was also previously shown by others who used data from both cases and controls.[10,19,61,62]

In the context of sub-challenge 2, we tackled the issue of predicting spontaneous preterm birth from samples collected while women were asymptomatic before 33 weeks of gestation. Overall, transcriptomic-based prediction performance for PPROM was low (AUROC = 0.6 at 27–33 weeks of gestation); however, the sub-challenge and post-challenge analyses provided evidence of changes in maternal whole-blood gene expression that precede a diagnosis of PPROM and are shared across gestational age time points and racially diverse cohorts and different microarray platforms. These transcriptomic changes involved immune-, inflammation-, and metabolism-related biological processes and pathways (Table S4). Plasma protein changes preceding a diagnosis with sPTD were larger at 27–33 weeks and led to higher prediction performance (AUROC = 0.76). The 90 proteins differentially abundant with sPTD (Table S8) were encoded by genes annotated to some of the same biological processes found by transcriptomic analysis in PPROM, with vascular endothelial growth factor A/ VEGF receptor 2 (VEGFA/VEGFR2) signaling and membrane trafficking pathways being top-ranked pathways based on both proteomics changes with sPTD and transcriptomic changes with PPROM. The involvement of membrane trafficking within the secretory membrane system, which includes the endoplasmic reticulum (ER), is in line with previous observations that ER stress is increased after spontaneous labor in gestational tissues, where it regulates the expression of prolabor mediators.[63,64] The involvement of the VEGF family of proteins in early placentation and of the abnormalities in maternal plasma and placental expression of angiogenic factors was also reported in adverse pregnancy outcomes.[65] Moreover, proteins annotated to endocrine system development (PDPK1, IL-6), a pathway associated with parturition,[66] were increased in the maternal plasma before the onset of sPTD. The inflammatory cytokine IL-6, known to play a central role during pregnancy and its complications,[67,68] was increased in the amniotic fluid of women having a preterm labor with intra-amniotic infection.[69] IL-6 and ANGPT4, a member of the angiopoietins family, were highlighted as predictors of preterm birth based on proteomics analysis of maternal plasma at 8–20 weeks of gestation in a population of women from low- and middle-income countries.[70]

Regarding the comparison between omics platforms for the prediction of preterm birth, this study demonstrated evidence of superior performance by plasma proteomics compared to whole-blood transcriptomics in the prediction of spontaneous preterm delivery (AUROC = 0.76 versus 0.6 at 27–33 weeks) (see Figure 7 for analyses in the same samples). This is in line with a recent multi-omics analysis in preterm birth using cell-

free RNA and plasma proteomics in preterm birth.[70] After we and other investigators reported the value of aptamer-based SomaLogic assays to predict early[52] and late preeclampsia,[45] in this study, we evaluated this platform to predict preterm birth and found the SomaLogic assay to be of superior value when compared to whole-blood transcriptomics in predicting spontaneous preterm delivery. The plasma proteomics signatures of spontaneous preterm birth identified here have direct implications for the development of future SomaSignal Tests that were demonstrated in other health applications by combining reproducible proteomic signals with machine learning.[28,29]

The use of crowdsourcing to evaluate computational approaches and longitudinal multi-omics data to predict preterm birth is a major strength of this study. Many independent approaches to solve this challenge were implemented by the data science community. Coincidentally, the first and second best-performing teams were the same for both sub-challenges, which is indicative of the team's skill, as opposed to chance, a fact that has been observed in several other crowdsourcing initiatives (e.g., sbv IMPROVER,[22,71,72] CAGI,[73–76] and DREAM[55,77,78]). Another advantage of the DREAM Challenge framework is that the model development and the prediction assessment are separate; thus, the risk of overstating the prediction performance is reduced. As with other similar crowdsourcing initiatives, we investigated the key factors that could explain the higher prediction performance of the top teams relative to the other teams. Given the multitude of differences in prediction pipelines among teams, it is challenging to single out individual key components that explain prediction performance variability. Therefore, in post-challenge analyses of sub-challenge 2, we modified the approach of team 1 to include a single new element borrowed from the approach of team 2. Based on this strategy, we have identified that the reliance of team 1 on the last-available snapshot of molecular activity was a key methodological aspect that was superior to the use of all of the available time points or the rate of change across points as implemented by other lower-ranked teams. This finding is in agreement with previous observations that the closer the sampling to the clinical diagnosis, the higher the predictive value of the biomarkers.[45,52,79] Once the molecular signature was reliably selected based on the last available time point, also including the measurements at the second-to-last available time point as independent predictors into the model would have been beneficial to improve prediction of preterm birth (Figure S4).

This robust evaluation of prediction performance, combined with a separate consideration of preterm birth phenotypes (sPTD and PPROM), of time points at sampling, and multi-omic platforms, makes this work one of the most comprehensive longitudinal omics studies in preterm birth. Importantly, this study provides omics data in a majority African-American cohort in which the rate of prematurity is higher than that observed in other populations[80] and omics data are scarce. Data collected in diverse populations are needed since some disease-related molecular changes can be cohort specific, as it was reported for other pregnancy complications such as preeclampsia.[81] Finally, the work herein has resulted in computational algorithms with associated code made available to the community with an open-source license, allowing for reproducible research and applications to other similar research questions based on longitudinal omics data.

Limitations of the study

Two possible limitations of the comparison between omics platforms are the lower sample size used to analyze the same blood draws and the much larger number of transcriptomic than proteomic features, which made the "needle in the haystack" problem more difficult for the transcriptomic platform. This curse of dimensionality was noted when transcriptomic and proteomic features were combined, resulting in a lower performance estimate for the multi-omics model obtained with the approach of team 1, than for proteomics data alone. Although here the remedy to this issue was to combine the predictions of each platform into a meta-model (stacked generalization) (Figure 7), alternative approaches focus on biologically plausible sets of features derived by single-cell genomics. This latter category of methods was demonstrated to predict preeclampsia[79,82] and to distinguish between women with spontaneous preterm labor and the gestational age-matched controls.[49,50] Another limitation of the study is that the RNA data collection was limited to genes present on the Human Transcriptome Array 2.0 microarray platform as opposed to sequencing-based methods that could provide a more comprehensive snapshot of the transcriptome.[83]

CONSORTIA

The DREAM Preterm Birth Prediction Challenge Consortium is listed below in alphabetical order, and author affiliations are available in Table S9.

Benan Bardak, Madhuchhanda Bhattacharjee, Michael Blair, Huiyuan Chen, Feng Cheng, Changje Cho, Junseok Choe, Mohit Choudhary, Yang Dai, Ophilia Daniel, Bikram K. Das, Francisco de Abreu e Lima, Anjali Dhall, Işıksu Ekşioğlu, Bogdan N. Gavrilovic, Akshay Gupta, Romeharsh Gupta, Rohan Gurve, Dániel Győrffy, Eric D. Hill, Jinseub Hwang, Yuguang F. Ipsen, Rıza Işık, Priyansh Jain, Pratheepa Jeganathan, Sujae Jeong, Chan-Seok Jeong, Anshul Jha, JinZhu Jia, Jaewoo Kang, Hyojin Kang, Gaurang A. Karwande, Harpreet Kaur, Hannah Kim, Keonwoo Kim, Sunkyu Kim, Dohyang Kim, Junseok Kim Jongtae Kim, Min-Jeong Kim, Amrit Koirala, Adriana N. König, Prachi Kothiyal, Vladimir B. Kovacevic, Aleksandra V. Kovacevic, Shiu Kumar, Chandrani Kumari, Christoph F. Kurz, Taeyong Kwon, Thuc D. Le, Kyeongjun Lee, Hyungyu Lee, Dawoon Leem, Shuya Li, Weng Khong Lim, Xinyue Liu, Yunan Luo, Bahattin C. Maral, Suyash Mishra, Yeongeun Nam, Leelavati Narlikar, Thin Nguyen, Zoran Obradovic, Hyeju Oh, Kousuke Onoue, Hyojung Paik, Wenchu Pan, Bogyu Park, Sumeet Patiyal, Jian Peng, Dimitri Perrin, Kaike Ping, Alidivinas Prusokas, Augustinas Prusokas, Peng Qiu, Gajendra P.S. Raghava, Derek Reiman, Renata Retkute, Nay Min Min Thaw Saw, Neelam Sharma, Alok Sharma, Ronesh Sharma, Rahul Siddharthan, Musalula Sinkala, Alex Soupir, Marija Stanojevic, Yufeng Su, Alexander M. Sutherland, András Szilágyi, Mehmet Tan, Nandor G. Than, Buu Truong, Edwin Vans, Fangping Wan, Rohan B.H. Williams, Wendy S.W. Wong, Jeong Woong, Li Xiaomei, Dongchan Yang, Sanghoo Yoon, Dakota York, James Young, and Wei Zhu.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

AUTHOR CONTRIBUTIONS

A.L.T., R.R., M.S., N.A., G.S., and J.C.C. designed the challenge. A.L.T., T.Y., and J.C.C. developed tools to receive and evaluate participant submissions. The top-performing approach was designed by I.C. and B.A.P. Data analysis for the top-performing approach was conducted by B.A.P. The DREAM Preterm Birth Prediction Challenge Consortium provided predictions for sub-challenge 1 and method implementations and descriptions for sub-challenges 1 and 2. A.L.T. and B.D. applied methods to datasets for sub-challenge 2 and designed and applied hybrid methods between approaches of the top two teams. A.L.T., B.D., B.A.P., G.B., G.A., and J.C.C. performed the post-challenge analyses. A.L.T., R.R., M.S., N.A., J.C.C., and G.S. interpreted the results of the challenge. A.L.T., R.R., N.G.-L., T.C., S.S.H., and C.-D.H. designed the protocol in which patients were enrolled and coordinated the collection and generation of the clinical and omics data. A.L.T., N.G.-L., M.S., and J.C.C. wrote the manuscript. A.L.T., R.R., J.C.C., and G.S. supervised the project.

DECLARATION OF INTERESTS

A.L.T., R.R., S.S.H., and T.C. are listed as co-inventors on the US 10,802,030 B2 patent, which involves the prediction of preterm birth using proteomics data. All of the other authors declare no competing interests.

REFERENCES

1. Brosens, I., Pijnenborg, R., Vercruysse, L., and Romero, R. (2011). The "Great Obstetrical Syndromes" are associated with disorders of deep placentation. Am. J. Obstet. Gynecol. 204, 193–201.

2. Blencowe, H., Cousens, S., Oestergaard, M.Z., Chou, D., Moller, A.B., Narwal, R., Adler, A., Vera Garcia, C., Rohde, S., Say, L., and Lawn, J.E. (2012). National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. Lancet 379, 2162–2172.

3. Martin, J.A., Hamilton, B.E., and Osterman, M.J.K. (2019). Births in the United States, 2018. NCHS Data Brief (346), 1–8.

4. R.E. Behrman and A.S. Butler, eds. (2007). Preterm Birth: Causes, Consequences, and Prevention (The National Academies Press).

5. Lawn, J.E., Kerber, K., Enweronu-Laryea, C., and Cousens, S. (2010). 3.6 million neonatal deaths–what is progressing and what is not? Semin. Perinatol. 34, 371–386.

6. Mwaniki, M.K., Atieno, M., Lawn, J.E., and Newton, C.R. (2012). Long-term neurodevelopmental outcomes after intrauterine and neonatal insults: a systematic review. Lancet 379, 445–452.

7. Spittle, A., Orton, J., Anderson, P.J., Boyd, R., and Doyle, L.W. (2015). Early developmental intervention programmes provided post hospital discharge to prevent motor and cognitive impairment in preterm infants. Cochrane Database Syst. Rev. (11), CD005495.

8. Goldenberg, R.L., Culhane, J.F., Iams, J.D., and Romero, R. (2008). Epidemiology and causes of preterm birth. Lancet 371, 75–84.

9. Romero, R., Dey, S.K., and Fisher, S.J. (2014). Preterm labor: one syndrome, many causes. Science 345, 760–765.

10. Knijnenburg, T.A., Vockley, J.G., Chambwe, N., Gibbs, D.L., Humphries, C., Huddleston, K.C., Klein, E., Kothiyal, P., Tasseff, R., Dhankani, V., et al. (2019). Genomic and molecular characterization of preterm birth. Proc. Natl. Acad. Sci. USA 116, 5819–5827.

11. Zhang, G., Feenstra, B., Bacelis, J., Liu, X., Muglia, L.M., Juodakis, J., Miller, D.E., Litterman, N., Jiang, P.-P., Russell, L., et al. (2017). Genetic Associations with Gestational Duration and Spontaneous Preterm Birth. N. Engl. J. Med. 377, 1156–1167.

12. Romero, R., Yeo, L., Miranda, J., Hassan, S.S., Conde-Agudelo, A., and Chaiworapongsa, T. (2013). A blueprint for the prevention of preterm birth: vaginal progesterone in women with a short cervix. J. Perinat. Med. 41, 27–44.

13. Romero, R., Nicolaides, K., Conde-Agudelo, A., Tabor, A., O'Brien, J.M., Cetingoz, E., Da Fonseca, E., Creasy, G.W., Klein, K., Rode, L., et al.

(2012). Vaginal progesterone in women with an asymptomatic sonographic short cervix in the midtrimester decreases preterm delivery and neonatal morbidity: a systematic review and metaanalysis of individual patient data. Am. J. Obstet. Gynecol. 206, 124.e1–124.e19.

14. Conde-Agudelo, A., Romero, R., Da Fonseca, E., O'Brien, J.M., Cetingoz, E., Creasy, G.W., Hassan, S.S., Erez, O., Pacora, P., and Nicolaides, K.H. (2018). Vaginal progesterone is as effective as cervical cerclage to prevent preterm birth in women with a singleton gestation, previous spontaneous preterm birth, and a short cervix: updated indirect comparison meta-analysis. Am. J. Obstet. Gynecol. 219, 10–25.

15. Hassan, S.S., Romero, R., Vidyadhari, D., Fusey, S., Baxter, J.K., Khandelwal, M., Vijayaraghavan, J., Trivedi, Y., Soma-Pillay, P., Sambarey, P., et al.; PREGNANT Trial (2011). Vaginal progesterone reduces the rate of preterm birth in women with a sonographic short cervix: a multicenter, randomized, double-blind, placebo-controlled trial. Ultrasound Obstet. Gynecol. 38, 18–31.

16. Ngo, T.T.M., Moufarrej, M.N., Rasmussen, M.H., Camunas-Soler, J., Pan, W., Okamoto, J., Neff, N.F., Liu, K., Wong, R.J., Downes, K., et al. (2018). Noninvasive blood tests for fetal development predict gestational age and preterm delivery. Science 360, 1133–1136.

17. Pavličev, M., Wagner, G.P., Chavan, A.R., Owens, K., Maziarz, J., Dunn-Fletcher, C., Kallapur, S.G., Muglia, L., and Jones, H. (2017). Single-cell transcriptomics of the human placenta: inferring the cell communication network of the maternal-fetal interface. Genome Res. 27, 349–361.

18. Ghaemi, M.S., DiGiulio, D.B., Contrepois, K., Callahan, B., Ngo, T.T.M., Lee-McMullen, B., Lehallier, B., Robaczewska, A., Mcilwain, D., Rosenberg-Hasson, Y., et al. (2019). Multiomics modeling of the immunome, transcriptome, microbiome, proteome and metabolome adaptations during human pregnancy. Bioinformatics 35, 95–103.

19. Romero, R., Espinoza, J., Gotsch, F., Kusanovic, J.P., Friel, L.A., Erez, O., Mazaki-Tovi, S., Than, N.G., Hassan, S., and Tromp, G. (2006). The use of high-dimensional biology (genomics, transcriptomics, proteomics, and metabolomics) to understand the preterm parturition syndrome. BJOG 113 (Suppl 3), 118–135.

20. Aghaeepour, N., Lehallier, B., Baca, Q., Ganio, E.A., Wong, R.J., Ghaemi, M.S., Culos, A., El-Sayed, Y.Y., Blumenfeld, Y.J., Druzin, M.L., et al. (2018). A proteomic clock of human pregnancy. Am. J. Obstet. Gynecol. 218, 347.e1–347.e14.

21. Romero, R., Erez, O., Maymon, E., Chaemsaithong, P., Xu, Z., Pacora, P., Chaiworapongsa, T., Done, B., Hassan, S.S., and Tarca, A.L. (2017). The maternal plasma proteome changes as a function of gestational age in normal pregnancy: a longitudinal study. Am. J. Obstet. Gynecol. 217, 67.e1–67.e21.

22. Tarca, A.L., Romero, R., Xu, Z., Gomez-Lopez, N., Erez, O., Hsu, C.D., Hassan, S.S., and Carey, V.J. (2019). Targeted expression profiling by RNA-Seq improves detection of cellular dynamics during pregnancy and identifies a role for T cells in term parturition. Sci. Rep. 9, 848.

23. Romero, R., Hassan, S.S., Gajer, P., Tarca, A.L., Fadrosh, D.W., Bieda, J., Chaemsaithong, P., Miranda, J., Chaiworapongsa, T., and Ravel, J. (2014). The vaginal microbiota of pregnant women who subsequently have spontaneous preterm labor and delivery and those with a normal delivery at term. Microbiome 2, 18.

24. DiGiulio, D.B., Callahan, B.J., McMurdie, P.J., Costello, E.K., Lyell, D.J., Robaczewska, A., Sun, C.L., Goltsman, D.S., Wong, R.J., Shaw, G., et al. (2015). Temporal and spatial variation of the human microbiota during pregnancy. Proc. Natl. Acad. Sci. USA 112, 11060–11065.

25. Aagaard, K., Riehle, K., Ma, J., Segata, N., Mistretta, T.A., Coarfa, C., Raza, S., Rosenbaum, S., Van den Veyver, I., Milosavljevic, A., et al. (2012). A metagenomic approach to characterization of the vaginal microbiome signature in pregnancy. PLoS ONE 7, e36466.

26. Vora, B., Wang, A., Kosti, I., Huang, H., Paranjpe, I., Woodruff, T.J., MacKenzie, T., and Sirota, M. (2018). Meta-Analysis of Maternal and Fetal Transcriptomic Data Elucidates the Role of Adaptive and Innate Immunity in Preterm Birth. Front. Immunol. 9, 993.

27. Gold, L., Ayers, D., Bertino, J., Bock, C., Bock, A., Brody, E.N., Carter, J., Dalby, A.B., Eaton, B.E., Fitzwater, T., et al. (2010). Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS ONE 5, e15004.

28. Ganz, P., Heidecker, B., Hveem, K., Jonasson, C., Kato, S., Segal, M.R., Sterling, D.G., and Williams, S.A. (2016). Development and Validation of a Protein-Based Risk Score for Cardiovascular Outcomes Among Patients With Stable Coronary Heart Disease. JAMA 315, 2532–2541.

29. Williams, S.A., Kivimaki, M., Langenberg, C., Hingorani, A.D., Casas, J.P., Bouchard, C., Jonasson, C., Sarzynski, M.A., Shipley, M.J., Alexander, L., et al. (2019). Plasma protein patterns as comprehensive indicators of health. Nat. Med. 25, 1851–1857.

30. Stolovitzky, G., Monroe, D., and Califano, A. (2007). Dialogue on reverse-engineering assessment and methods: the DREAM of high-throughput pathway inference. Ann. N Y Acad. Sci. 1115, 1–22.

31. Heng, Y.J., Pennell, C.E., McDonald, S.W., Vinturache, A.E., Xu, J., Lee, M.W., Briollais, L., Lyon, A.W., Slater, D.M., Bocking, A.D., et al. (2016). Maternal Whole Blood Gene Expression at 18 and 28 Weeks of Gestation Associated with Spontaneous Preterm Birth in Asymptomatic Women. PLoS ONE 11, e0155191.

32. Tin, A., Yu, B., Ma, J., Masushita, K., Daya, N., Hoogeveen, R.C., Ballantyne, C.M., Couper, D., Rebholz, C.M., Grams, M.E., et al. (2019). Reproducibility and Variability of Protein Analytes Measured Using a Multiplexed Modified Aptamer Assay. J. Appl. Lab. Med. 4, 30–39.

33. Nazarov, P.V., Muller, A., Kaoma, T., Nicot, N., Maximo, C., Birembaut, P., Tran, N.L., Dittmar, G., and Vallar, L. (2017). RNA sequencing and transcriptome arrays analyses show opposing results for alternative splicing in patient derived samples. BMC Genomics 18, 443.

34. Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. 67. https://doi.org/10.18637/jss.v067.i01.

35. Gomez-Lopez, N., Romero, R., Hassan, S.S., Bhatti, G., Berry, S.M., Kusanovic, J.P., Pacora, P., and Tarca, A.L. (2019). The Cellular Transcriptome in the Maternal Circulation During Normal Pregnancy: A Longitudinal Study. Front. Immunol. 10, 2863.

36. Sacks, G.P., Studena, K., Sargent, K., and Redman, C.W. (1998). Normal pregnancy and preeclampsia both produce inflammatory changes in peripheral blood leukocytes akin to those of sepsis. Am. J. Obstet. Gynecol. 179, 80–86.

37. Naccasha, N., Gervasi, M.T., Chaiworapongsa, T., Berman, S., Yoon, B.H., Maymon, E., and Romero, R. (2001). Phenotypic and metabolic characteristics of monocytes and granulocytes in normal pregnancy and maternal infection. Am. J. Obstet. Gynecol. 185, 1118–1123.

38. Germain, S.J., Sacks, G.P., Sooranna, S.R., Sargent, I.L., and Redman, C.W. (2007). Systemic inflammatory priming in normal pregnancy and preeclampsia: the role of circulating syncytiotrophoblast microparticles. J. Immunol. 178, 5949–5956.

39. Gomez-Lopez, N., Tanaka, S., Zaeem, Z., Metz, G.A., and Olson, D.M. (2013). Maternal circulating leukocytes display early chemotactic responsiveness during late gestation. BMC Pregnancy Childbirth 13 (Suppl 1), S8.

40. Aghaeepour, N., Ganio, E.A., Mcilwain, D., Tsai, A.S., Tingle, M., Van Gassen, S., Gaudilliere, D.K., Baca, Q., McNeil, L., Okada, R., et al. (2017). An immune clock of human pregnancy. Sci. Immunol. 2, eaan2946.

41. Savitz, D.A., Terry, J.W., Jr., Dole, N., Thorp, J.M., Jr., Siega-Riz, A.M., and Herring, A.H. (2002). Comparison of pregnancy dating by last menstrual period, ultrasound scanning, and their combination. Am. J. Obstet. Gynecol. 187, 1660–1666.

42. Romero, R., Espinoza, J., Kusanovic, J.P., Gotsch, F., Hassan, S., Erez, O., Chaiworapongsa, T., and Mazor, M. (2006). The preterm parturition syndrome. BJOG 113 (Suppl 3), 17–42.

43. Norwitz, E.R., Robinson, J.N., and Challis, J.R. (1999). The control of labor. N. Engl. J. Med. *341*, 660–666.

44. Davies, D.R., Gelinas, A.D., Zhang, C., Rohloff, J.C., Carter, J.D., O'Connell, D., Waugh, S.M., Wolk, S.K., Mayfield, W.S., Burgin, A.B., et al. (2012). Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. Proc. Natl. Acad. Sci. USA *109*, 19971–19976.

45. Erez, O., Romero, R., Maymon, E., Chaemsaithong, P., Done, B., Pacora, P., Panaitescu, B., Chaiworapongsa, T., Hassan, S.S., and Tarca, A.L. (2017). The prediction of late-onset preeclampsia: results from a longitudinal proteomics study. PLoS ONE *12*, e0181468.

46. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425.

47. Gervasi, M.T., Chaiworapongsa, T., Naccasha, N., Blackwell, S., Yoon, B.H., Maymon, E., and Romero, R. (2001). Phenotypic and metabolic characteristics of maternal monocytes and granulocytes in preterm labor with intact membranes. Am. J. Obstet. Gynecol. *185*, 1124–1129.

48. Zhang, J., Shynlova, O., Sabra, S., Bang, A., Briollais, L., and Lye, S.J. (2017). Immunophenotyping and activation status of maternal peripheral blood leukocytes during pregnancy and labour, both term and preterm. J. Cell. Mol. Med. *21*, 2386–2402.

49. Paquette, A.G., Shynlova, O., Kibschull, M., Price, N.D., and Lye, S.J.; Global Alliance to Prevent Prematurity and Stillbirth Systems Biology of Preterm Birth Team (2018). Comparative analysis of gene expression in maternal peripheral blood and monocytes during spontaneous preterm labor. Am. J. Obstet. Gynecol. *218*, 345.e1–345.e30.

50. Pique-Regi, R., Romero, R., Tarca, A.L., Sendler, E.D., Xu, Y., Garcia-Flores, V., Leng, Y., Luca, F., Hassan, S.S., and Gomez-Lopez, N. (2019). Single cell transcriptional signatures of the human placenta in term and preterm parturition. eLife *8*, e52004.

51. Gervasi, M.T., Chaiworapongsa, T., Naccasha, N., Pacora, P., Berman, S., Maymon, E., Kim, J.C., Kim, Y.M., Yoshimatsu, J., Espinoza, J., and Romero, R. (2002). Maternal intravascular inflammation in preterm premature rupture of membranes. J. Matern. Fetal Neonatal Med. *11*, 171–175.

52. Tarca, A.L., Romero, R., Benshalom-Tirosh, N., Than, N.G., Gudicha, D.W., Done, B., Pacora, P., Chaiworapongsa, T., Panaitescu, B., Tirosh, D., et al. (2019). The prediction of early preeclampsia: results from a longitudinal proteomics study. PLoS ONE *14*, e0217273.

53. Wolpert, D.H. (1992). Stacked generalization. Neural Netw. *2*, 241–259.

54. Küffner, R., Zach, N., Norel, R., Hawe, J., Schoenfeld, D., Wang, L., Li, G., Fang, L., Mackey, L., Hardiman, O., et al. (2015). Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression. Nat. Biotechnol. *33*, 51–57.

55. Keller, A., Gerkin, R.C., Guan, Y., Dhurandhar, A., Turu, G., Szalai, B., Mainland, J.D., Ihara, Y., Yu, C.W., Wolfinger, R., et al.; DREAM Olfaction Prediction Consortium (2017). Predicting human olfactory perception from chemical features of odor molecules. Science *355*, 820–826.

56. Salcedo, A., Tarabichi, M., Espiritu, S.M.G., Deshwar, A.G., David, M., Wilson, N.M., Dentro, S., Wintersinger, J.A., Liu, L.Y., Ko, M., et al.; DREAM SMC-Het Participants (2020). A community effort to create standards for evaluating tumor subclonal reconstruction. Nat. Biotechnol. *38*, 97–107.

57. Schaffter, T., Buist, D.S.M., Lee, C.I., Nikulin, Y., Ribli, D., Guan, Y., Lotter, W., Jie, Z., Du, H., Wang, S., et al.; DM DREAM Consortium (2020). Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. JAMA Netw. Open *3*, e200265.

58. Costello, J.C., Heiser, L.M., Georgii, E., Gönen, M., Menden, M.P., Wang, N.J., Bansal, M., Ammad-ud-din, M., Hintsanen, P., Khan, S.A., et al.; NCI DREAM Community (2014). A community effort to assess and improve drug sensitivity prediction algorithms. Nat. Biotechnol. *32*, 1202–1212.

59. Guinney, J., and Saez-Rodriguez, J. (2018). Alternative models for sharing confidential biomedical data. Nat. Biotechnol. *36*, 391–392.

60. Saez-Rodriguez, J., Costello, J.C., Friend, S.H., Kellen, M.R., Mangravite, L., Meyer, P., Norman, T., and Stolovitzky, G. (2016). Crowdsourcing biomedical research: leveraging communities as innovation engines. Nat. Rev. Genet. *17*, 470–486.

61. Bukowski, R., Sadovsky, Y., Goodarzi, H., Zhang, H., Biggio, J.R., Varner, M., Parry, S., Xiao, F., Esplin, S.M., Andrews, W., et al. (2017). Onset of human preterm and term birth is related to unique inflammatory transcriptome profiles at the maternal fetal interface. PeerJ *5*, e3685.

62. Heng, Y.J., Pennell, C.E., Chua, H.N., Perkins, J.E., and Lye, S.J. (2014). Whole blood gene expression profile associated with spontaneous preterm birth in women with threatened preterm labor. PLoS ONE *9*, e96901.

63. Liong, S., and Lappas, M. (2014). Endoplasmic reticulum stress is increased after spontaneous labor in human fetal membranes and myometrium where it regulates the expression of prolabor mediators. Biol. Reprod. *91*, 70.

64. Kyathanahalli, C., Organ, K., Moreci, R.S., Anamthathmakula, P., Hassan, S.S., Caritis, S.N., Jeyasuria, P., and Condon, J.C. (2015). Uterine endoplasmic reticulum stress-unfolded protein response regulation of gestational length is caspase-3 and -7-dependent. Proc. Natl. Acad. Sci. USA *112*, 14090–14095.

65. Andraweera, P.H., Dekker, G.A., and Roberts, C.T. (2012). The vascular endothelial growth factor family in adverse pregnancy outcomes. Hum. Reprod. Update *18*, 436–457.

66. Matthews, S.G., Gibb, W., and Lye, S.J. (2000). Endocrine and paracrine regulation of birth at term and preterm. Endocr. Rev. *21*, 514–550.

67. Gomez-Lopez, N., Olson, D.M., and Robertson, S.A. (2016). Interleukin-6 controls uterine Th9 cells and CD8(+) T regulatory cells to accelerate parturition in mice. Immunol. Cell Biol. *94*, 79–89.

68. Prins, J.R., Gomez-Lopez, N., and Robertson, S.A. (2012). Interleukin-6 in pregnancy and gestational disorders. J. Reprod. Immunol. *95*, 1–14.

69. Romero, R., Avila, C., Santhanam, U., and Sehgal, P.B. (1990). Amniotic fluid interleukin 6 in preterm labor. Association with infection. J. Clin. Invest. *85*, 1392–1400.

70. Jehan, F., Sazawal, S., Baqui, A.H., Nisar, M.I., Dhingra, U., Khanam, R., Ilyas, M., Dutta, A., Mitra, D.K., Mehmood, U., et al.; Alliance for Maternal and Newborn Health Improvement, the Global Alliance to Prevent Prematurity and Stillbirth, and the Prematurity Research Center at Stanford University (2020). Multiomics Characterization of Preterm Birth in Low- and Middle-Income Countries. JAMA Netw. Open *3*, e2029655.

71. Tarca, A.L., Gong, X., Romero, R., Yang, W., Duan, Z., Yang, H., Zhang, C., and Wang, P. (2018). Human blood gene signature as a marker for smoking exposure: computational approaches of the top ranked teams in the sbv IMPROVER Systems Toxicology challenge. Comput. Toxicol. *5*, 31–37.

72. Tarca, A.L., Lauria, M., Unger, M., Bilal, E., Boue, S., Kumar Dey, K., Hoeng, J., Koeppl, H., Martin, F., Meyer, P., et al.; IMPROVER DSC Collaborators (2013). Strengths and limitations of microarray-based phenotype prediction: lessons learned from the IMPROVER Diagnostic Signature Challenge. Bioinformatics *29*, 2892–2899.

73. Hoskins, R.A., Repo, S., Barsky, D., Andreoletti, G., Moult, J., and Brenner, S.E. (2017). Reports from CAGI: The Critical Assessment of Genome Interpretation. Hum. Mutat. *38*, 1039–1041.

74. Andreoletti, G., Pal, L.R., Moult, J., and Brenner, S.E. (2019). Reports from the fifth edition of CAGI: The Critical Assessment of Genome Interpretation. Hum. Mutat. *40*, 1197–1201.

75. Pejaver, V., Mooney, S.D., and Radivojac, P. (2017). Missense variant pathogenicity predictors generalize well across a range of function-specific prediction challenges. Hum. Mutat. *38*, 1092–1108.

76. Clark, W.T., Kasak, L., Bakolitsa, C., Hu, Z., Andreoletti, G., Babbi, G., Bromberg, Y., Casadio, R., Dunbrack, R., Folkman, L., et al. (2019). Assessment of predicted enzymatic activity of α-N-acetylglucosaminidase variants of unknown significance for CAGI 2016. Hum. Mutat. *40*, 1519–1529.

77. Menden, M.P., Wang, D., Mason, M.J., Szalai, B., Bulusu, K.C., Guan, Y., Yu, T., Kang, J., Jeon, M., Wolfinger, R., et al.; AstraZeneca-Sanger Drug Combination DREAM Consortium (2019). Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. Nat. Commun. *10*, 2674.

78. Gönen, M., Weir, B.A., Cowley, G.S., Vazquez, F., Guan, Y., Jaiswal, A., Karasuyama, M., Uzunangelov, V., Wang, T., Tsherniak, A., et al.; Broad-DREAM Community (2017). A Community Challenge for Inferring Genetic Predictors of Gene Essentialities through Analysis of a Functional Screen of Cancer Cell Lines. Cell Syst. *5*, 485–497.e3.

79. Tarca, A.L., Romero, R., Erez, O., Gudicha, D.W., Than, N.G., Benshalom-Tirosh, N., Pacora, P., Hsu, C.D., Chaiworapongsa, T., Hassan, S.S., and Gomez-Lopez, N. (2020). Maternal whole blood mRNA signatures identify women at risk of early preeclampsia: a longitudinal study. J. Matern. Fetal Neonatal Med. https://doi.org/10.1080/14767058.2019.1685964.

80. Kistka, Z.A., Palomar, L., Lee, K.A., Boslaugh, S.E., Wangler, M.F., Cole, F.S., DeBaun, M.R., and Muglia, L.J. (2007). Racial disparity in the frequency of recurrence of preterm birth. Am. J. Obstet. Gynecol. *196*, 131.e1–131.e6.

81. Ghaemi, M.S., Tarca, A.L., Romero, R., Stanley, N., Fallahzadeh, R., Tanada, A., Culos, A., Ando, K., Han, X., Blumenfeld, Y.J., et al. (2021). Proteomic signatures predict preeclampsia in individual cohorts but not across cohorts - implications for clinical biomarker studies. J. Matern. Fetal Neonatal Med. https://doi.org/10.1080/14767058.2021.1888915.

82. Tsang, J.C.H., Vong, J.S.L., Ji, L., Poon, L.C.Y., Jiang, P., Lui, K.O., Ni, Y.B., To, K.F., Cheng, Y.K.Y., Chiu, R.W.K., and Lo, Y.M.D. (2017). Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. Proc. Natl. Acad. Sci. USA *114*, E7786–E7795.

83. Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA sequencing: the teenage years. Nat. Rev. Genet. *20*, 631–656.

84. Carvalho, B.S., and Irizarry, R.A. (2010). A framework for oligonucleotide microarray preprocessing. Bioinformatics *26*, 2363–2367.

85. Smyth, G.K. (2012). Limma: linear models for microarray data. In Bioinformatics and Computational Biology Solutions Using R and Bioconductor, R. Gentleman, V.J. Carey, W. Huber, R.A. Irizarry, and S. Dudoit, eds. (Springer), pp. 397–420.

86. Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. J. Stat. Softw. *33*, 1–22.

87. Otasek, D., Morris, J.H., Bouças, J., Pico, A.R., and Demchak, B. (2019). Cytoscape Automation: empowering workflow-based network analysis. Genome Biol. *20*, 185.

88. Falcon, S., and Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. Bioinformatics *23*, 257–258.

89. ACOG Committee on Practice Bulletins–Obstetrics (2002). ACOG practice bulletin. Diagnosis and management of preeclampsia and eclampsia. Obstet. Gynecol. *99*, 159–167.

90. von Dadelszen, P., Magee, L.A., and Roberts, J.M. (2003). Subclassification of preeclampsia. Hypertens. Pregnancy *22*, 143–148.

91. Gomez, R., Romero, R., Edwin, S.S., and David, C. (1997). Pathogenesis of preterm labor and preterm premature rupture of membranes associated with intraamniotic infection. Infect. Dis. Clin. North Am. *11*, 135–176.

92. Guinn, D.A., Goldenberg, R.L., Hauth, J.C., Andrews, W.W., Thom, E., and Romero, R. (1995). Risk factors for the development of preterm premature rupture of the membranes after arrest of preterm labor. Am. J. Obstet. Gynecol. *173*, 1310–1315.

93. Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4*, 249–264.

94. Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80.

95. Hoerl, A.E., and Kennard, R.W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics *12*, 55–67.

96. Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (2019). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. https://rdrr.io/rforge/e1071/.

97. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

98. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.; The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29.

99. Schaefer, C.F., Anthony, K., Krupa, S., Buchoff, J., Day, M., Hannay, T., and Buetow, K.H. (2009). PID: the Pathway Interaction Database. Nucleic Acids Res. *37*, D674–D679.

100. Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. *27*, 29–34.

101. Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., et al. (2020). The reactome pathway knowledgebase. Nucleic Acids Res. *48* (*D1*), D498–D503.

102. Martens, M., Ammar, A., Riutta, A., Waagmeester, A., Slenter, D.N., Hanspers, K., A Miller, R., Digles, D., Lopes, E.N., Ehrhart, F., et al. (2021). WikiPathways: connecting communities. Nucleic Acids Res. *49* (*D1*), D613–D621.

103. Doncheva, N.T., Morris, J.H., Gorodkin, J., and Jensen, L.J. (2019). Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. J. Proteome Res. *18*, 623–632.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human whole blood and plasma samples	Perinatology Research Branch, an intramural program of the Eunice Kennedy Shriver NICHD, NIH, DHHS, Wayne State University (Detroit, MI, USA), and the Detroit Medical Center (Detroit, MI, USA)	N/A
Critical commercial assays		
GeneChip WT Pico Reagent Kit	Affymetrix (Thermo Fisher Scientific)	P/N 703262 Rev. 1
Human Transcriptome Arrays (HTA 2.0)	Affymetrix (Thermo Fisher Scientific)	P/N 902162
SOMAmer proteomic assays and profiling services (1,125 proteins)	SomaLogic, Inc.	Gene Expression Omnibus: GPL28509
Deposited data		
Raw and preprocessed transcriptomics data	This paper	Gene Expression Omnibus:: GSE149440
Raw and preprocessed proteomics data	This paper	Gene Expression Omnibus:: GSE150167
Software and algorithms		
oligo	Carvalho and Irizarry[84]	https://www.bioconductor.org/packages/release/bioc/html/oligo.html
limma	Smyth[85]	https://www.bioconductor.org/packages/release/bioc/html/limma.html
lme4	Bates et al.[34]	https://cran.r-project.org/web/packages/lme4/index.html
glmnet	Friedman et al.[86]	https://cran.r-project.org/web/packages/glmnet/index.html
Cytoscape	Otasek et al.[87]	https://cytoscape.org/
GOstats	Falcon and Gentleman[88]	https://bioconductor.org/packages/release/bioc/html/GOstats.html
Predictive modeling; Sub-challenge 1, Team 1	This paper	https://www.synapse.org/#!Synapse:syn20684755
Predictive modeling; Sub-challenge 2, Team 1	This paper	https://www.synapse.org/#!Synapse:syn21443858
MSigDB curated gene sets	Liberzon et al.[46]	http://www.gsea-msigdb.org/gsea/msigdb/collections.jsp#C2
Other		
Calgary cohort transcriptomics data	Heng et al.[31]	GEO: GSE59491
Resource website for the DREAM Preterm Birth Prediction Challenge, including data, software code, and vignettes	This paper	https://www.synapse.org/pretermbirth

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Adi L. Tarca (atarca@med.wayne.edu).

Materials availability
This study did not generate new unique reagents.

Data and code availability

The accession number for the transcriptomic and proteomic data from the Detroit cohort described herein are Gene Expression Omnibus super-series GSE149440 and GSE150167, respectively. They were also submitted to the March of Dimes repository (https://www.immport.org/shared/study/SDY1636).

Analysis scripts for transcriptomic data preprocessing and for building prediction models based on the approaches of the participating teams in sub-challenges 1 and 2 are available from the Challenge website (https://www.synapse.org/pretermbirth). Direct links to method write-ups and computer code for prediction of gestational age and preterm birth are also available in Tables S2 and S6, respectively. Moreover, R code vignettes demonstrating the use of participant methods and key post-challenge analyses were also provided at https://www.synapse.org/pretermbirth.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human subjects, clinical specimens, and definitions

Women who provided blood samples included in the transcriptomic (n = 149) and proteomic (n = 105) studies described in the Results section were enrolled in a prospective longitudinal study at the Center for Advanced Obstetrical Care and Research of the Perinatology Research Branch, NICHD/NIH/DHHS; the Detroit Medical Center; and the Wayne State University School of Medicine. Blood samples were collected at the time of prenatal visits, scheduled at four-week intervals from the first or early second trimester until delivery, during the following gestational-age intervals: 8- < 16 weeks, 16- < 24 weeks, 24- < 28 weeks, 28- < 32 weeks, 32- < 37 weeks, and > 37 weeks. Collection of biological specimens and the ultrasound and clinical data was approved by the Institutional Review Boards of Wayne State University (WSU IRB#110605MP2F) and NICHD (OH97-CH-N067) under the protocol entitled "Biological Markers of Disease in the Prediction of Preterm Delivery, Preeclampsia and Intra-Uterine Growth Restriction: A Longitudinal Study." Cases and controls were selected retrospectively and sample size was determined based on sample availability and cost of experiments.

The first ultrasound scan during pregnancy was used to establish gestational age if this estimate was more than 7 days from the LMP-based gestational age. The first ultrasound scan was obtained before 14 weeks of gestation for 70% of the women, and 95% of the women underwent the first ultrasound before 20 weeks of gestation. Preeclampsia was defined as new-onset hypertension that developed after 20 weeks of gestation (systolic or diastolic blood pressure $\geq$ 140 mm Hg and/or $\geq$ 90 mm Hg, respectively, measured on at least two occasions, 4 hours to 1 week apart) and proteinuria ($\geq$ 300 mg in a 24-hour urine collection, or two random urine specimens obtained 4 hours to 1 week apart containing $\geq$ 1+ by dipstick or one dipstick demonstrating $\geq$ 2+ protein).[89] Early preeclampsia was defined as preeclampsia diagnosed before 34 weeks of gestation, and late preeclampsia was defined by diagnosis at or after 34 weeks of gestation.[90] The diagnosis of PPROM was determined by a sterile speculum examination with documentation of either vaginal pooling or a positive nitrazine or ferning test.[91] Spontaneous preterm labor and delivery was defined as the spontaneous onset of labor with intact membranes and delivery occurring prior to the 37th week of gestation.[92] Demographic characteristics of the study population are summarized in Table S1, and they are available for each individual patient in the GEO datasets (see Data and code availability).

METHOD DETAILS

Maternal whole blood transcriptomics

RNA was isolated from PAXgene® Blood RNA collection tubes (BD Biosciences, San Jose, CA; Catalog #762165) and hybridized to GeneChip Human Transcriptome Arrays (HTA) 2.0 (P/N 902162), as previously described.[35] Microarray experiments were carried out at the University of Michigan Advanced Genomics Core, a part of the Biomedical Research Core Facilities, Office of Research (Ann Arbor, MI, USA).

Maternal plasma proteomics

Maternal plasma protein abundance was determined by using the SOMAmer (Slow Off-rate Modified Aptamer) platform and reagents to profile 1,125 proteins.[27,44] Proteomic profiling services were provided by SomaLogic, Inc. (Boulder, CO, USA). As we previously described,[21] plasma samples were diluted and then incubated with SOMAmer mixes pre-immobilized onto streptavidin-coated beads. The beads were washed to remove non-specifically bound proteins and other matrix constituents. Proteins that remained bound to their cognate SOMAmer reagents were tagged using an NHS-biotin reagent. After the labeling reaction, the beads were exposed to an anionic competitor solution that prevents non-specific interactions from reforming after disruption. Pure cognate-SOMAmer complexes and unbound (free) SOMAmer reagents are released from the streptavidin beads using ultraviolet light that cleaves the photo-cleavable linker used to quantitate protein. The photo-cleavage eluate, which contains all SOMAmer reagents (some bound to a biotin-labeled protein and some free), was separated from the beads and then incubated with a second streptavidin-coated bead that binds the biotin-labeled proteins and the biotin-labeled protein-SOMAmer complexes. The free SOMAmer reagents were then removed by washing. In the last elution step, protein-bound SOMAmer reagents were released from their cognate proteins using denaturing conditions. SOMAmer reagents were then quantified by hybridization to custom DNA microarrays. The Cyanine-3 signal from the SOMAmer reagent was detected on microarrays.[27,44]

Sub-challenge 1 organization

For sub-challenge 1, aimed at predicting gestational age at sampling from whole blood transcriptomic data in normal and complicated pregnancies, a training set and a test set were generated (Figure S1; Video S1). Transcriptomic gene expression data were made available to participants for both the training and test sets. Gestational age was provided for the training set and participants were required to submit predicted gestational-age values for the test set, which were compared in real time against the gold standard; the RMSE was posted to a leaderboard that was live from May 22, 2019, to August 15, 2019. Up to five submissions per team were allowed, and they were ranked by the RMSE, and the smallest value was retained as entry for each unique team (Table S1). Only the teams who described their approach and provided the analysis code were retained in the final team rankings.

Sub-challenge 2 organization

In the first phase of sub-challenge 2, participants were invited to develop preterm birth prediction algorithms using gene expression data from longitudinal transcriptomic data collected from 17 to less than 37 weeks of gestation from women with a normal pregnancy and from cases of preterm birth (sPTD and PPROM) illustrated in Figure 2. The training set was composed of data from the Calgary cohort and a fraction of the Detroit cohort (Figure 2), while the test set comprised the remainder of the Detroit cohort. Teams were requested to submit a risk value (probability) for all samples when classifying test samples as sPTD versus Control, and as PPROM versus Control. The AUROC and AUPRC were calculated separately for each prediction task and the ranks for each of the resulting four performance measures were calculated for each team and aggregated by summation. Two predictions per team were allowed and performance results on the test set were posted to a live leaderboard from August 15, 2019, to December 5, 2019.

Because the prediction models developed in the first phase of sub-challenge 2 could have captured eventual differences between the cases and controls in terms of the timing and number of samples, a second phase of the sub-challenge 2 was organized (December 5, 2019 to January 3, 2020) for which teams were asked to provide prediction algorithms (computer code) instead of predictions of a given test dataset.

QUANTIFICATION AND STATISTICAL ANALYSIS

Transcriptomics data preprocessing

Raw intensity data (CEL files) were generated from array images using the Affymetrix AGCC software. CEL files from this study and those for the Calgary cohort were preprocessed separately for each platform. ENTREZID gene level expression summaries were obtained with Robust Multi-array Average (RMA)[93] implemented in the *oligo* package[84] using suitable chip definition files from http://brainarray.mbni.med.umich.edu. Since samples in the Detroit cohort were profiled in several batches, correction for potential batch effects was performed using the *removeBatchEffect* function of the *limma*[85] package in *Bioconductor*.[94] Cross-study/platform analyses were performed on a combined dataset after quantile normalizing data across all samples for the set of common genes, followed by platform effect-removal.

Proteomics data preprocessing

The protein abundance in relative fluorescence units was obtained by scanning the microarrays. A sample-by-sample adjustment in the overall signal within a single plate was performed in three steps per manufacturer's protocol, as we previously described.[21,45] Outlier values (larger than $2 \times$ the 98th percentile of all samples) were set to $2 \times$ the 98th percentile of all samples. Data was log$_2$ transformed before applying machine learning and differential abundance analyses.

Sub-challenge 1 robustness analysis of team ranks

To determine whether differences in gestational-age prediction accuracy between the different teams were robust, we have simulated the challenge by drawing 1000 bootstrap samples of the test set. RMSE values were calculated for each submission (1 to at most 5) for each team, and we retained the submission with the smallest RMSE. Team ranks were calculated and the Bayes factors were then calculated as the ratio between the number of iterations in which the team k performed better than the team ranked next ($k+1$) relative to the number of iterations when the reverse was true. A Bayes factor > 3 was considered a significant difference in ranking (see Figure S1B).

Sub-challenge 1 top two algorithms

Team 1: The first-ranked team in this sub-challenge (authors B.A.P. and I.C.) used gene-level expression data after filtering out samples considered as outliers, followed by the standardization of gene expression for each microarray experiment batch separately. Genes were ranked by using singular value decomposition, and those genes having higher dot products with singular vectors that correspond to large singular values across the training samples were assigned a higher score.

In the next step, ~6000 genes were selected based on the described ranking, which was based on cross-validation results on the training set using a ridge regression model. Ridge regression[95] models were fitted using the *Sklearn* package in *Python* (version 3).

Team 2: The second-ranked team in this sub-challenge (author Y.G.) applied quantile normalization to gene level expression data, followed by the modeling of the gestational-age values using Generalized Process Regression and Support Vector Regression.

Model tuning parameters were optimized using a grid search, and predictions by the two approaches were weighted equally. Models were fit using *Octave*.

Sub-challenge 2 team ranking

The algorithms submitted by participants in the final stage of sub-challenge 2 were applied as implemented by the participants without any tuning to the 70 pairs of training and test datasets described in Figure 4A, Table S5, and Video S1. In each of the 7 scenarios in Figure 4A, there were 2 outcomes predicted (sPTD versus Control; and PPROM versus Control), except for proteomic data (scenario DP2), where the feasible comparisons were sPTD versus control and sPTB versus control; the sPTB group was defined as the union of sPTD and PPROM cases. The AUROC and AUPRC were used to assess predictions for each outcome based on predictions on each of the 10 test sets for each scenario, and then were averaged over the 10 tests sets. The resulting 28 prediction performance averages (7 scenarios x 2 outcomes x 2 metrics) for each team were converted into Z-scores by subtracting the mean and dividing by the standard deviation of these metrics obtained from 1,000 random predictions (random uniform posterior probabilities). Further, only the combinations of scenarios and outcomes resulting in a significant prediction performance (False Discovery Rate-adjusted p value derived from Z-scores, $q < 0.05$) for at least one of the 13 teams, were considered for team ranking, resulting in 20 performance criteria for each team. Teams were ranked by each of the 20 prediction performance criteria, and a final rank was generated based on the sum of the ranks over all criteria (Table S6).

Sub-challenge 2 robustness analysis of team ranks

To assess the significance of the differences in prediction performance of preterm birth among the teams based on omics data, we used the same ranking procedure described above in more than 1,000 simulated iterations of the sub-challenge. At each iteration, the rankings were calculated by using prediction performance results that corresponded to a bootstrap sample of the 10 train/test pairs pertaining to each scenario and, at the same time, taking a bootstrap sample of the prediction criteria (columns in Table S6). Bayes factors were then calculated as the ratio between the number of iterations in which the team k performed better than the team ranked next $(k+1)$ relative to the number of iterations when the reverse was true. A Bayes factor > 3 was considered a significant difference among rankings (Figure S3B).

Sub-challenge 2, the top three algorithms

Team 1: The algorithm of the first-ranked team in this sub-challenge (authors B.A.P. and I.C.) starts with standardizing the input omics data so that they have a zero mean and a standard deviation of 1 for each omics platform (if more than one in an input set, which was the case while training and testing across the platforms). A random forest classifier with 100 trees was fit to each prediction task (sPTD versus Control and PPROM versus Control). The top 50 features, ranked by importance metric derived from the random forest, were selected for each task separately and used to fit a final model on the training data. Random forest models were fitted using the *Sklearn* package in *Python* (version 3).

Team 2: The approach of the second-ranked team in this sub-challenge (author Y.G.) first centers the data of each feature around the mean for each platform (if more than one) in a given input set. Then, data is quantile normalized to make identical the distributions of feature data across the samples. Next, the top 50 features with the highest average over all samples are retained, and the feature values for the last-available two time points for each subject are used as predictors (100 predictors) in a Generalized Process Regression model, a Bayesian non-parametric regression technique. The two parameters of GPR regression were preset to an eye value of 0.75, which represents how much noise is assumed in the data, and a sigma of 10, a data normalization factor. Models were fitted using Octave.

Team 3: The approach of the third-ranked team in this sub-challenge (author R.K.) starts with the selection of the top 50 features ranked by statistical significance p value derived from a t test or Wilcoxon test, depending on the normality of the data, and determined by a Shapiro test. Then, using the selected features, linear, sigmoid and radial Support Vector Machines models are fitted and compared via 5-fold cross validation, and the predictions for the best method were averaged over the five trained models. Models were fit using the *e1071* package[96] in R.

Assessing significance of gestational age prediction

The correlation between gestational ages predicted by the transcriptomics model of Team 1 in sub-challenge 1 (M_GA_Team1) and actual gestational ages at blood draw was assessed using a naive Pearson correlation test, but also via linear mixed-effects modeling to account for repeated-measurements from patients in the test set. This latter analysis involved fitting a linear mixed-effects model in which the dependent variable was the transcriptomics predicted gestational age and the independent variable was the actual gestational age. The patient identifier was included as a random effect in this model. A likelihood ratio test implemented in the *lme4* package[34] was used to determine the significance of the linear relation between actual and transcriptomics-predicted gestational ages.

Identification of a core transcriptome predicting gestational age

To identify a core transcriptome that can predict gestational age in normal and complicated pregnancies, linear mixed-effects models with splines were applied to prioritize genes that change with gestational age while accounting for the possible non-linear relation and for the repeated observations from each individual, as we previously described.[35] Of note, participating teams could have not used

such an approach given that sample-to-patient annotations were not provided on the training data. Then, the genes that did not change in average expression by at least 10% over the 10-40-week span were filtered out, and the remaining genes were ranked by p values from the linear mixed-effects models. The top 300 genes were then used as input in a LASSO regression model (elastic net mixing parameter alpha = 0.01) for which the shrinkage coefficient (lambda) was determined by cross-validation, leading to 249 genes with non-zero coefficients in the model (Table S3). Of note, using more than 300 genes as input in the ridge regression model did not further reduce the RMSE on the test set. LASSO models were fit using the *glmnet* package[86] in R.

Differential expression and abundance analyses

Differences in gene expression or protein abundance between the cases and controls were assessed based on linear models implemented in the *limma* package[97] in Bioconductor. When data across time points and/or cohorts were combined, these factors were included as fixed effects in the linear models. Downstream analyses of the differentially expressed genes involved enrichment analysis via a hypergeometric test implemented in the *GOstats* package[88] to determine the over-representation of Gene Ontology[98] biological processes among the significant genes. Additional enrichment analyses for both transcriptomics and proteomics platforms in sub-challenge 2 were based on a hypergeometric test with pathway definitions extracted from the C2 collection of the MSigDB database. The C2 collection in MSigDB includes pathways from the Pathway Interaction Database,[99] Kyoto Encyclopedia of Genes and Genomes,[100] Reactome database,[101] and Wiki Pathways,[102] among other sources. The background list in the enrichment analyses featured all genes profiled on the microarray platform. For proteomic-based enrichment analyses for sub-challenge 1, protein-to-gene annotations from the manufacturer (SomaLogic) were used as input in the *stringApp* version (1.5.0)[103] in *Cytoscape* (version 3.7.2)[87] using the whole genome as background. A false discovery rate adjusted q < 0.1 was used throughout enrichment analyses to infer significance. Networks of high-confidence protein-protein interactions (STRING confidence score > 0.7) were constructed from the lists of significant genes/proteins using *stringApp* in Cytoscape. For visualization, the most interconnected sub-networks were displayed and nodes were annotated to significantly enriched biological processes.