

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

Adi L. Tarca^{1,2,3*}, Bálint Ármin Pataki⁴, Roberto Romero^{1,5-9*}, Marina Sirota^{10,11}, Yuanfang Guan¹², Rintu Kutum¹³, Nardhy Gomez-Lopez^{1,2,14}, Bogdan Done^{1,2}, Gaurav Bhatti^{1,2}, Thomas Yu¹⁵, Gaia Andreoletti^{10,11}, Tinnakorn Chaiworapongsa², The DREAM Preterm Birth Prediction Challenge Consortium, Sonia S. Hassan^{2,16,17}, Chaur-Dong Hsu^{1,2,17}, Nima Aghaepour¹⁸, Gustavo Stolovitzky^{19,20*}, Istvan Csabai⁴, James C. Costello^{21*}

¹Perinatology 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, U. S. Department of Health and Human Services, Bethesda, Maryland 20892, and Detroit, Michigan 48201, USA;

²Department of Obstetrics and Gynecology, Wayne State University School of Medicine, Detroit, Michigan 48201 USA;

³Department of Computer Science, Wayne State University College of Engineering, Detroit, Michigan 48202, USA;

⁴Department of Physics of Complex Systems, ELTE Eötvös Loránd University, Budapest, Hungary;

⁵Department of Obstetrics and Gynecology, University of Michigan, Ann Arbor, Michigan 48109, USA;

⁶Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, Michigan 48824, USA;

⁷Center for Molecular Medicine and Genetics, Wayne State University, Detroit, Michigan 48201, USA;

⁸Detroit Medical Center, Detroit, Michigan 48201 USA;

⁹Department of Obstetrics and Gynecology, Florida International University, Miami, Florida 33199, USA;

¹⁰Department of Pediatrics, University of California, San Francisco, California, 94143, USA;

¹¹Bakar Computational Health Sciences Institute, University of California, San Francisco, California 94158, USA;

¹²Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan 48109, USA;

¹³Informatics and Big Data Unit, CSIR-Institute of Genomics and Integrative Biology, New Delhi, India;

¹⁴Department of Biochemistry, Microbiology and Immunology, Wayne State University School of Medicine, Detroit, Michigan 48201 USA;

¹⁵Sage Bionetworks, Seattle, WA, USA;

¹⁶Office of Women's Health, Integrative Biosciences Center, Wayne State University, Detroit, Michigan 48202 USA;

¹⁷Department of Physiology, Wayne State University School of Medicine, Detroit, Michigan 48201, USA;

¹⁸Department of Anesthesiology, Perioperative and Pain Medicine, and Department of Pediatrics, and Department of Biomedical Data Sciences Stanford University School of Medicine, Stanford, California 94305, USA;

¹⁹Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA;

47 ²⁰IBM T.J. Watson Research Center, Yorktown Heights, New York 10598, USA.

48 ²¹Department of Pharmacology, University of Colorado Anschutz Medical Campus, Aurora,
49 Colorado 80045, USA.

50

* Corresponding authors:

Adi L Tarca: atarca@med.wayne.edu

Roberto Romero: prbchiefstaff@med.wayne.edu

Gustavo Stolovitzky: gustavo@us.ibm.com

James C Costello: james.costello@cuanschutz.edu

51 **Disclosure statement: The authors declare no conflicts of interest.**

52

Abstract

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. We found that whole blood gene expression predicts ultrasound-based gestational ages in normal and complicated pregnancies ($r=0.83$), as well as the delivery date in normal pregnancies ($r=0.86$), with an accuracy comparable to ultrasound. However, unlike the latter, transcriptomic data collected at <37 weeks of gestation predicted the delivery date of one third of spontaneous (sPTB) cases within 2 weeks of the actual date. Based on samples collected before 33 weeks in asymptomatic women we found expression changes preceding preterm prelabor rupture of the membranes that were consistent across time points and cohorts, involving, among others, leukocyte-mediated immunity. Plasma proteomic random forests predicted sPTB with higher accuracy and earlier in pregnancy than whole blood transcriptomic models (e.g. AUROC=0.76 vs. 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 therapeutic interventions. One of the “great obstetrical syndromes”¹, preterm birth, defined as birth prior to 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 America². In the United States, the rate of prematurity remained fundamentally unchanged in recent years³ and it has an annual societal economic burden of at least \$26.2 billion⁴. The high incidence of preterm birth is concerning: 29% of all neonatal deaths worldwide, approximately 1 million deaths in total, can be attributed to complications of prematurity⁵. 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 etiologies⁹, and its complexity makes accurate prediction by a single set of biomarkers difficult. While genetic risk factors for preterm birth have been reported¹⁰, 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.¹¹ 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 < 33 weeks of gestation by 45% and decreases the rate of neonatal complications, including neonatal respiratory distress syndrome¹²⁻¹⁴. The role of 17-alpha-

hydroxyprogesterone caproate for preventing preterm delivery in patients with an episode of preterm labor is controversial^{15,16}.

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 signatures¹⁷, with the latter eventually refined by single-cell genomics^{17,18}, and by expanding the types of data collected via multi-omics platforms^{10,19,20}. While molecular profiles have been shown to be strongly modulated by advancing gestation in the maternal blood proteome^{21,22}, transcriptome^{17,23}, and vaginal microbiome^{24,25}, the timing of delivery based on such molecular clocks of pregnancy is still challenging¹⁷. A recent meta-analysis²⁶ 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 predictive models to make inferences in asymptomatic women early in pregnancy has not been evaluated. This topic is important, since early identification is necessary to develop treatment strategies to reduce the impact of prematurity.

Therefore, we generated longitudinal whole blood transcriptomic and plasma proteomic data on 216 women and leveraged the Dialogue for Reverse Engineering Assessments and Methods (DREAM) crowdsourcing framework²⁷ to engage over 500 members of the computational biology community and 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 have also predicted delivery dates in women with spontaneous birth (**Fig. 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.²⁸ in a cohort in Calgary. The separate consideration of both spontaneous preterm birth phenotypes, i.e. spontaneous preterm delivery with intact membranes (sPTD) and premature prelabor rupture of membranes (PPROM), allowed us to pinpoint that previously reported leukocyte activation-related RNA changes preceding preterm birth are shared across cohorts only for the PPRM phenotype but not sPTD. Moreover, the evaluation of plasma proteomics and blood multi-omics data to determine the earliest stage in gestation when biomarkers have predictive value (**Fig. 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 herein, this work sets a benchmark for evaluation of longitudinal omics data in pregnancy research. The computational lessons and algorithms for risk prediction from longitudinal omics data derived herein can also impact 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,

NICHD/NIH/DHHS; the Detroit Medical Center; and the Wayne State University School of Medicine. The patient population included women with a normal pregnancy who delivered at term (≥ 37 weeks) (Controls, N=49), women who delivered before 37 completed weeks of gestation by spontaneous preterm delivery with intact (sPTD, N=34) or ruptured (PPROM, N=37) membranes, and women who experienced an indicated delivery before 34 weeks due to early preeclampsia (N=13) (Dataset Detroit_HTA **Fig. 2a**). After including data from 16 additional normal pregnancies from the same population²³ (Dataset GSE1139, 32 transcriptomes), the resulting set of 149 pregnancies (see demographic characteristics 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 (**Fig. 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 last menstrual period (LMP) and ultrasound fetal biometry, and to make predictions based only on gene expressions 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 for the test samples, and the best submission (smallest Root Mean Squared Error, RMSE) was retained for each unique team. We received 331 submissions for this sub-challenge from 87 participating teams, of which 37 teams provided the required details on the computational methods used to be qualified for the final team ranking in this sub-challenge (**Table S2**).

Team ranking robustness analysis (see **Methods**) suggested that the predictions of the first-ranked team were significantly better (Bayes factor > 3) than those of the second- and third-ranked teams

(Fig. S2). Among the top 20 teams, the most frequent methods used to select predictor genes included univariate gene ranking and meta-gene building via principal components 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 the first-ranked team in sub-challenge 1 (Team 1) predicted the test set's gestational ages at blood draw with an RMSE of 4.5 weeks (Pearson correlation between actual and predicted values, $r=0.83$, $p<0.001$) (Fig. 2b). This prediction model (M_GA_Team1) was based on ridge regression, and the predictors were meta-genes derived by using principal components from expression data of 6,106 genes. As shown in Fig. 2b, the gestational age predictions showed little bias in second trimester (14-28 weeks) samples (mean error 0.6 weeks); however, gestational ages predicted 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, in general, was robust to obstetrical disease- and parturition-related perturbations in gene expression data (Fig. S3).

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 >6000 predictor genes, we combined linear mixed effects modeling for longitudinal data to prioritize gene

expressions and then used these as input in a LASSO regression model. The resulting 249 gene regression model (M_GA_Core) (Fig. S4, Table S3) had an RMSE=5.1 weeks (Pearson correlation between actual and predicted values, $r=0.80$) and involved two tightly connected modules related to immune response, leukocyte activation, inflammation- and development-related Gene Ontology biological processes (Fig. 2c, Table S4). We previously reported that several member genes of these networks (e.g. MMP8, CECAM8, and DEFA4) were most highly modulated in the normal pregnancy group used herein²⁹, and others have shown the same to be true at a cell-free RNA level in a Danish cohort¹⁷. In addition, these data are consistent with the concept that pregnancy is characterized by a systemic cellular inflammatory response³⁰⁻³⁴. 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 (Table S1).

Comparison of gene expression models and clinical standard in predicting time to delivery 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 profiling¹⁷, we used the same methods as described above for model M_GA_Team1, except for the use of a time variable defined backward from delivery and, hence, independent of LMP and ultrasound estimations [Time to delivery, TTD=Date at sample – Date at delivery (weeks)], as response. As in the study by Ngo et al¹⁷, only those patients with spontaneous term delivery were included in this analysis, thus omitting even the subset of normal pregnancies that had been truncated by elective cesarean delivery. The prediction performance on the test set of the resulting model (M_sTD_TTD) was compared to the LMP and ultrasound fetal biometry,

with the latter predicting delivery at 40 weeks of gestation. As shown in **Fig. 3a**, the gene expression model significantly predicted time to delivery 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 test set accuracy of the gene expression model, defined as predicting delivery within one week of the actual date³⁵, 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%). Of note, 45% accuracy was also reported by Ngo et al¹⁷ using cell free RNA based on second and third trimester samples.

When data from all pregnancies with spontaneous term delivery were used to train a transcriptomic model of time to delivery, and apply it to data from women with spontaneous preterm birth, the prediction was found to be statistically significant. However, the error increased (RMSE=5.6) (**Fig. 3b**) relative to the estimate (RMSE=4.5) for prediction of time to delivery in women with spontaneous term delivery (**Fig. 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 (**Fig. 3b**). The overestimation (mean prediction error) was 2.3 weeks compared to the five-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. Based on data of 1 to 4 samples collected at 24-37 weeks of gestation for each woman with preterm birth, the M_sTD_TTD model identified 33% of pregnancies to be at risk of delivery within 2 weeks or already past due, with one-half of these, 16% (11/70), predicted to deliver ± 1 week from the actual

date. This finding contrasts with the LMP-and-ultrasound method, which predicted no cases delivering within ± 1 or ± 2 weeks of the actual delivery date. This evidence further suggests that the M_sTD_TTD model captured both gene expression changes related to immune- and development-related processes establishing the age of pregnancy as well as the effects of the common pathway of parturition^{36,37}. Hence, it 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)

In this study, we demonstrated that a transcriptomic model of maternal blood derived from data of women with spontaneous term delivery (M_sTD_TTD) can predict spontaneous preterm birth when the data has been collected early in pregnancy as well as near or at the time of preterm parturition up to less than 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, the maximum number of predictor genes that participating teams could use as predictors in this sub-challenge was limited to a total of 100 genes for prediction of both phenotypes of preterm birth: sPTD and PPRM.

After the first phase of sub-challenge 2 in which predictions were optimistically biased, given the confounding effects of the gestational ages at sampling (see **Methods**), we drew from the Detroit

cohort longitudinal study (**Fig. 4a**) only samples collected at specific gestational-age intervals while women were asymptomatic (i.e., prior to an eventual diagnosis of spontaneous preterm delivery or PPRM). Two scenarios of prediction of preterm birth were devised to include cases and controls with available samples collected at the 17-23 and 27-33 weeks (**Fig. 4a**); a third scenario involved patients with available samples collected at three gestational age intervals (17-22, 22-27, and 27-33 weeks) (**Fig. 4b**). 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.²⁸ 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 technique^{38,39} in samples collected at 17-23 and 27-33 weeks of gestation from 66 women prior to the diagnosis of preterm birth (62 sPTD and 4 PPRM). These samples were profiled in the same experimental batch with samples from 39 normal pregnancies that we previously described^{22,40}, which herein served as controls (**Fig. 4c**). The characteristics of pregnancies with available proteomics profiles are shown in **Table S1**.

The prediction algorithms generated by 13 teams who participated in the second phase of 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 (**Table 1**). 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 (**Table 1**). In all cases, there were no differences in terms of number of samples and gestational age at sampling between the cases and controls (**Fig. 4**).

Scenario	Platform	Sample GA (weeks)	Train set	Test set	Number of datasets	Outcomes predicted
----------	----------	-------------------	-----------	----------	--------------------	--------------------

C2	Transcriptomics	17-23 & 27-33	Calgary	Calgary	10*	sPTD, PPROM
D2	Transcriptomics	17-23 & 27-33	Detroit	Detroit	10*	sPTD, PPROM
D3	Transcriptomics	17-22 & 22-27 & 27-33	Detroit	Detroit	10*	sPTD, PPROM
CtoD	Transcriptomics	17-23 & 27-33	Calgary	Detroit	10 ⁺	sPTD, PPROM
DtoC	Transcriptomics	17-23 & 27-33	Detroit	Calgary	10 ⁺	sPTD, PPROM
CDtoD	Transcriptomics	17-23 & 27-33	Calgary + Detroit	Detroit	10**	sPTD, PPROM
DP2	Proteomics	17-23 & 27-33	Detroit	Detroit	10*	sPTD, sPTB

Table 1: 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 replacement from the original cohort to create different versions of the training set, and the trained model was then applied to the original test cohort. sPTB, spontaneous preterm birth; sPTD, spontaneous preterm delivery with intact membranes; PPROM, preterm prelabor rupture of membranes.

To assess the prediction performance of the test set 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 for imbalanced datasets, e.g., the proteomics set that features more cases than controls (**Fig. 4c**).

Figure 5 depicts the resulting 28 prediction performance scores (7 scenarios x 2 outcomes x 2 metrics) for each team after the conversion of AUROC and AUPRC metrics into Z scores. Final team rankings were obtained by aggregating the ranks over all prediction performance scores that were significant, according to at least one team, after multiple testing correction (**Table S5**). A rank robustness analysis (**Fig. S5**) determined that the first-ranked team outperformed the second-ranked team and that the second- and third-ranked teams outperformed the fifth-ranked team (Bayes factor > 3). For all scenarios (**Table 1**), the models of Team 1 involved data from 50 genes collected at the last available measurement (closest to delivery), while Team 2 used data collected at the last two available time points for 50 genes 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 **Methods**). Of note, the top two ranked teams were the same in both sub-challenges 1 and 2.

As depicted in **Fig. 6a**, with the approach of Team 1, one transcriptomic profile at 27-33 weeks of gestation from asymptomatic women predicted PPROM across the cohorts and microarray platforms [AUROC=0.6 (0.56-0.64)] when training the model on the Calgary cohort and testing on the Detroit cohort. Nearly the same performance [AUROC=0.61 (0.56-0.66)] was observed when one-fifth of the Detroit set was added to the Calgary cohort so that microarray platform and cohort effects can be accounted for during the data preprocessing and model training. Prediction of PPROM when training on the Detroit cohort and testing on the Calgary cohort was shy of significance [AUROC=0.54 (0.5-0.57)] based on one sample collected at the 27-33 weeks interval but increased to an AUROC of 0.6 (0.56–0.63); if, in a post-challenge analysis, borrowing from the approach of Team 2, for the same predictor genes, the data from both time points (17-22 and 27-33 weeks of gestation) were used as independent predictors in the model of Team 1 (**Fig S6**). 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 an individual patient data meta-analysis, which identified 402 differentially expressed genes after adjusting for cohort and time point (moderated t-test q-values <0.1) (**Fig. 6b, Table S6**). A highly connected protein-protein interaction sub-network corresponding to genes significant in this meta-analysis is shown in **Fig. 6c**, illustrating some of the Gene Ontology biological processes significantly enriched in PPROM and included vesicle-mediated transport and leukocyte- (myeloid and lymphocyte) mediated immunity, among others (**Table S7**). 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 labor⁴¹⁻⁴⁴ and PPRM⁴⁵.

Unlike PPRM, changes in the maternal whole blood transcriptome preceding spontaneous preterm delivery with intact membranes were not shared across cohorts at either the 17-23 or 27-33 week intervals. Within the Detroit cohort, prediction of spontaneous preterm delivery was best performed by Team 2, whose approach leveraged information from the last-available two time points; the prediction performance was significant [AUROC=0.66 (0.6-0.73)] if the data collected at 22-27 and 27-33 weeks of gestation were used, and it was not significant if the data collected at 17-22 and 27-33 weeks of gestation were used instead (**Fig. S7**). This finding is in agreement with previous observations that that closer the sampling to the clinical diagnosis, the higher the predictive value of the biomarkers^{40,46,47}. Of interest, 20 of the 50 most highly expressed genes in the Detroit cohort, chosen as predictors by Team 2, showed a significant correlation with gestational age at delivery at both time points simultaneously, suggesting a within-cohort-across-time-point consistency in gene expression changes with gestational age at delivery ($q < 0.1$) (**Table S8**).

Although participating teams in this sub-challenge did not have access to the longitudinal plasma proteomics data in preterm birth included herein when they developed prediction algorithms, their algorithms, when applied to the plasma proteomics set (**Fig. 4c**), resulted in models with test prediction performances that surpassed those obtained using transcriptomic data (**Fig. 5** and **Table S5** and **Fig. 7a**). Prediction of spontaneous preterm delivery by the approach of Team 1 involved 50 plasma proteins selected by random forest model importance from the panel of 1,125 available

proteins. The test set accuracy was the highest when using data collected at 27-33 weeks of gestation [AUROC=0.76 (0.72-0.8)] (**Fig. 7A, Table S9**). However, importantly, even one proteome profile at 17-22 weeks of gestation predicted significantly spontaneous preterm delivery [AUROC=0.62 (0.58-0.67)] (**Table S10**), suggesting that this approach has value in the early identification of women at risk. The addition of four cases with PPRM to those with spontaneous preterm delivery did not impact the prediction performance of the proteomics models of Team 1, suggesting that this approach could generalize to both preterm birth phenotypes. Indeed, the increase in plasma protein abundance of PDE11A and ITGA2B preceded the diagnosis of both spontaneous preterm delivery and PPRM in the Detroit cohort at 27-33 weeks of gestation (**Fig. 7b,c**). The tightly interconnected network of proteins built from among those with differential profiles with spontaneous preterm delivery in asymptomatic women at 27-33 weeks of gestation ($q < 0.1$, **Fig. 7b** and **Table S9**) included not only several previously known markers of preterm delivery (IL6, ANGPT1) but also MMP7 and ITGA2B, which we previously described as dysregulated in women with preeclampsia⁴⁶. Member proteins of this network perturbed prior to a diagnosis of spontaneous preterm delivery are annotated to biological processes such as regulation of cell adhesion, response to stimulus, and development (**Fig. 7d**).

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 spontaneous preterm delivery cases for which both types of data originated in 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 (**Fig. 8**), hence confirming the superior value of proteomics relative to transcriptomics for this endpoint. Of note, for a fixed number of 50 predictors allowed, a stacked generalization⁴⁸ approach combining predictions from individual platform models via a LASSO logistic regression led to 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 (**Fig. 8**).

To extract further insight into 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 aspects of its method except for the temporal information considered among the following: a) first point (T1), b) change in expression between T2 and T1 (slope), or c) 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 spontaneous preterm delivery). As shown in **Fig. S8**, none of these approaches would have improved prediction performance relative to the baseline approach that only considered 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: a) to start with only the top half of the most highly abundant features on each platform, b) to convert the binary classification (preterm versus term) into a regression of gestational age at delivery, and c) given the selected 50 predictor genes based on the correlation of T2 expression values with the outcome, to add the expression of those gene 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 unique genes, slightly outperformed the approach of Team 1's overall prediction scenarios (**Fig S8**) and led to the consistent prediction of PPRM in all cross-study analyses (see improvement in prediction from **Fig. 6a** to **Fig. S6**). Interestingly, simply doubling the number of genes measured 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 genes at T2 (**Fig. S8**). This finding suggests that for preterm birth prediction, it is more important to measure the right markers at one additional time point than to double the number of markers at the most recent time point.

Discussion

In this study, we have evaluated maternal blood omics data to predict gestational age and the risk of preterm birth. Although the main interest herein 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 framework²⁷ 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 tasks⁴⁹⁻⁵³. Moreover, the results gained from these Challenges define community standards and advancements in many scientific fields^{54,55}.

Collectively, sub-challenge 1 and the additional post-challenge analyses demonstrated that models based on the maternal whole blood transcriptome i) significantly predict LMP and ultrasound-defined gestational age at venipuncture in both normal and complicated pregnancies (RMSE=4.5), ii) predict a delivery date within ± 1 week in women with spontaneous term delivery with an accuracy (45%) comparable to the clinical standard (55%), and iii) outperform the latter approach for timing of delivery in women who are destined to experience spontaneous preterm birth (33% predicted within ± 2 weeks of delivery versus 0% for the LMP and ultrasound method). Of interest, the accuracy of dating gestation in women with spontaneous term delivery was similar to the report by Ngo et al.¹⁷ 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 herein. However, in the study by Ngo et al.¹⁷, the time-to-delivery transcriptomic model derived from samples of women with normal pregnancy failed to predict delivery dates on independent cohorts of women with preterm birth, while 33% of preterm birth cases herein were accurately identified as being at risk of delivery within ± 2 weeks. 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 preterm birth by omics data collected up to less than 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^{10,20,56,57}.

445 In the context of sub-challenge 2 of the DREAM Preterm Birth Prediction Challenge, we have
 446 tackled the issue of predicting preterm birth from samples collected while women were
 447 asymptomatic prior to 33 weeks of gestation. Overall, although prediction performance was low
 448 (AUROC=0.6 at 27-33 weeks of gestation), the sub-challenge and post-challenge analyses provide
 449 evidence of changes in maternal whole blood gene expression that precede a diagnosis of PPROM
 450 and that are shared across gestational-age time points and cohorts/microarray platforms. However,
 451 although some correlations of gene expression to gestational age at delivery were consistent across
 452 time points within the Detroit cohort (**Table S8**), the cross-cohort changes preceding spontaneous
 453 preterm delivery with intact membranes were not consistent. Moreover, the Detroit within-cohort
 454 proteomic results, obtained with methods developed without any tuning to the proteomic data,
 455 represent evidence of superior performance by plasma proteomics (AUROC=0.76 at 27-33 weeks)
 456 compared to whole blood transcriptomics in the prediction of spontaneous preterm delivery.
 457 Although we and other investigators reported the value of aptamer-based SomaLogic assays to
 458 predict early⁴⁶ and late preeclampsia⁴⁰, this is the first study conducted to evaluate this platform to
 459 predict preterm birth, and we found it to be of superior value as compared to whole blood
 460 transcriptomics platform in predicting spontaneous preterm delivery. Two possible limitations to
 461 the comparison between platforms are the lower sample size utilized to analyze the same blood
 462 draws and the much larger number of transcriptomic than proteomic features, which made the
 463 “needle in the haystack” problem more difficult for the transcriptomic platform. This curse of
 464 dimensionality was noted when transcriptomic and proteomic features were combined, resulting
 465 in a lower performance estimate for the multi-omics model obtained with the approach of Team 1,
 466 than for proteomics data alone. Although herein the remedy to this issue was to combine the
 467 predictions of each platform into a meta-model (stacked generalization) (**Fig. 8**), alternative

approaches focus on biologically plausible sets of features derived by single-cell genomics. This latter category of methods was demonstrated to predict preeclampsia^{47,58} and to distinguish between women with spontaneous preterm labor and the gestational age-matched controls^{43,44}.

The use of crowdsourcing to evaluate computational approaches and longitudinal multi-omics data to predict preterm birth is a major strength of this study. The primary advantage of importance is that many perspectives of this problem were implemented by the machine learning 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 crowd-sourcing initiatives, e.g., sbv IMPROVER^{23,59,60}, CAGI⁶¹⁻⁶⁴ and DREAM^{50,65,66}. The second 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. This robust evaluation of prediction performance, combined with a separate consideration of preterm birth phenotypes (spontaneous preterm delivery 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. Finally, the work herein has resulted in computational algorithms with implementations 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.

Methods

Study design

Women who provided blood samples included in the transcriptomic and proteomic 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.

Clinical definitions

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 ≥ 140 mm Hg and/or ≥ 90 mm Hg, respectively, measured on at least two occasions, 4 hours to 1 week apart) and proteinuria (≥ 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)⁶⁷. 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 gestation⁶⁸. The diagnosis of PPROM was determined by a sterile speculum examination with documentation of either vaginal

pooling or a positive nitrazine or ferning test⁶⁹. 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 gestation⁷⁰.

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 we previously described²⁹. 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). 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)⁷¹ implemented in the *oligo* package⁷² 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*⁷³ package in *Bioconductor*⁷⁴. 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.

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^{38,39}. Proteomic profiling services were provided by SomaLogic, Inc. (Boulder, CO, USA). The plasma samples were diluted and then incubated with the respective SOMAmer mixes, and after following a suite of steps

described elsewhere^{38,39}, the signal from the SOMAmer reagents was measured using microarrays. 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^{22,40}. 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 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 (**Fig. 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 1 team rank stability analysis

To determine whether differences in gestational-age prediction accuracy between the different teams were substantial, 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.

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 regression⁷⁵ 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 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 at 17-36

585 weeks of gestation from women with a normal pregnancy and from cases of preterm birth
586 (spontaneous preterm delivery and PPRM) illustrated in **Fig. 2**. The training set was composed
587 of data from the Calgary cohort and a fraction of the Detroit cohort (**Fig. 2**), while the test set
588 comprised the remainder of the Detroit cohort. Teams were requested to submit a risk value
589 (probability) for all samples when classifying test samples as sPTD versus Control, and as PPRM
590 versus Control. The AUROC and AUPRC were calculated separately for each prediction task and
591 the ranks for each of the resulting four performance measures were calculated for each team and
592 aggregated by summation. Two predictions per team were allowed and performance results on the
593 test set were posted to a live leaderboard from August 15, 2019, to December 5, 2019. Because
594 the prediction models developed in this phase of the Challenge could have captured eventual
595 differences between the cases and controls in terms of the timing and number of samples, a second
596 phase of the Challenge was organized (December 5, 2019 to January 3, 2020) for which teams
597 were asked to provide prediction algorithms (computer code) instead of predictions of a given test
598 dataset. The algorithms were applied as implemented by the participants without any tuning to the
599 70 pairs of training and test datasets described in **Table 1**. In each of the 7 scenarios in **Table 1**,
600 there were 2 outcomes predicted (sPTD vs Control; and PPRM vs Control), except for proteomic
601 data (scenario DP2), where the feasible comparisons were sPTD versus control and PTB versus
602 control; the PTB group was defined as the union of sPTD and PPRM cases. As in the first phase
603 of the challenge, the AUROC and AUPRC were used to assess predictions for each outcome. The
604 resulting 28 prediction performance scores (7 scenarios x 2 outcomes x 2 metrics) for each team
605 were converted into Z-scores by subtracting the mean and dividing by the standard deviation of
606 these metrics obtained from 1,000 random predictions (random uniform posterior probabilities).
607 Further, only the combinations of scenarios and outcomes resulting in a significant prediction

performance (False Discovery Rate-adjusted p-value, $q < 0.05$) for at least one of the 13 teams, were considered further for team ranking, resulting in 20 performance criteria for each team. Teams were ranked by each of the 20 prediction performance criteria (columns in **Fig. 8**), and a final rank was generated based on the sum of the ranks over all criteria (**Table S5**).

Sub-challenge 2 team rank robustness analysis

To assess the significance of the differences in prediction performance of preterm birth among the teams based on omics data, we have 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 correspond to a bootstrap sample of the 10 train/test instances pertaining to each scenario (**Table 1**) and, at the same time, taking a bootstrap sample of the prediction criteria (columns in the **Fig. 8**). 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.

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 PPRM 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 regressors 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* package⁷⁶ in R.

Post-challenge 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* package⁷⁷ 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* package⁷⁸ to determine the over-representation of Gene Ontology⁷⁹ biological processes among the significant genes. The background list in the enrichment analyses featured all genes profiled on the microarray platform. For proteomic-based enrichment analyses, protein-to-gene annotations from the manufacturer (SomaLogic) were used as input in the *stringApp* version (1.5.0)⁸⁰ in *Cytoscape* (version 3.7.2)⁸¹ using the whole genome as background. A false discovery rate adjusted $q < 0.05$ was used in 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.

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 described²⁹. 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 (λ) was determined by cross-validation, leading to 249 genes with non-zero coefficients in the model

(**Table S2**). 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* package⁸² in R.

Data availability

The transcriptomic and proteomic data from the Detroit cohort described herein is available as a 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>).

Code availability

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 (www.synapse.org/pretermbirth). Direct links to method write-up and computer implementations for prediction of gestational age and preterm birth are also available in **Tables S2** and **Table S5**, respectively. Moreover, R code vignettes demonstrating the use of participant methods and key post-challenge analyses were also provided at www.synapse.org/pretermbirth.

Acknowledgements

This research was supported, in part, by the 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, U.S. Department of Health and Human Services (NICHD/NIH/DHHS) and, in part, with federal funds from NICHD/NIH/DHHS under Contract No. HHSN275201300006C.

A.L.T. and N.G.-L. were also supported by the Wayne State University Perinatal Initiative in Maternal, Perinatal and Child Health.

R.R. has contributed to this work as part of his official duties as an employee of the United States Federal Government.

The authors acknowledge Alison Paquette for insightful discussions about available maternal omics datasets in preterm birth and Maureen McGerty (Wayne State University) and Corina Ghita for proofreading and copyediting this manuscript.

N.A. was supported by the Bill and Melinda Gates Foundation (OPP1112382 and OPP1113682), the March of Dimes Prematurity Research Center at Stanford University, the Burroughs Wellcome Fund, the National Center for Advancing Translational Sciences, and the Robertson Family Foundation (KL2TR003143).

I.C and B.A.P. were supported by the National Research, Development and Innovation Fund of Hungary (Project No. FIEK_16-1-2016-0005).

Y.G. was supported by grants from the National Institutes of Health (NIH/NIGMS R35GM133346-01) and National Science Foundation (NSF/DBI #1452656).

R.K. was supported by a Senior Research Fellowship Award from the Council of Scientific and Industrial Research of India (HCP00013).

G.A. and M.S. were supported by the March of Dimes.

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 Consortium provided predictions for sub-challenge 1, as well as method implementations and descriptions for sub-challenge 1 and 2. A.L.T. and B.D. applied method implementations to datasets for sub-challenge 2 and designed and applied hybrid methods between approaches of top two teams. A.L.T., B.D., B.P.A., G.B., G.A. and J.C.C. performed 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., J.C.C. wrote the manuscript. A.L.T., R.R., J.C.C. and G.S. supervised the project.

The DREAM Preterm Birth Prediction Challenge Consortium

Nima Aghaeepour¹⁸, Gaia Andreoletti^{10,11}, Benan Bardak²², Madhuchhanda Bhattacharjee²³, Gaurav Bhatti^{1,2}, Michael Blair²⁴, Huiyuan Chen²⁵, Feng Cheng²⁶, Tinnakorn Chiaworapongsa², Changje Cho²⁷, Junseok Choe²⁸, Mohit Choudhary²⁹, James C. Costello²¹, Istvan Csabai⁴, Yang Dai³⁰, Ophilia Daniel³¹, Bikram K. Das³², Francisco de Abreu e Lima³³, Anjali Dhall³⁴, Bogdan Done^{1,2}, İşıksu Ekşioğlu²², Bogdan N. Gavrilovic³⁵, Nardhy Gomez-Lopez^{1,2,14}, Yuanfang Guan¹², Akshay Gupta²⁹, Romeharsh Gupta²⁹, Rohan Gurve²⁹, Dániel Györfy^{36,37}, Sonia S. Hassan^{2,16,17}, Eric D. Hill³⁸, Chaur-Dong Hsu^{1,2,17}, Jinseub Hwang³⁹, Yuguang F. Ipsen⁴⁰, Rıza Işık²², Priyansh Jain⁴¹, Pratheepa Jeganathan⁴², Sujae Jeong⁴³, Chan-Seok Jeong⁴⁴, Anshul Jha²⁹, JinZhu Jia⁴⁵, Jaewoo Kang⁴⁶, Hyojin Kang⁴⁴, Gaurang A. Karwande⁴⁷, Harpreet Kaur⁴⁸, Hannah Kim⁴⁹, Keonwoo Kim²⁸, Sunkyu Kim²⁸, Dohyang Kim²⁷, Junseok Kim²⁷, Jongtae Kim³⁹, Min-Jeong Kim⁵⁰, Amrit Koirala³², Adriana N. König⁵¹, Prachi Kothiyal⁵², Vladimir B. Kovacevic³⁵, Aleksandra V. Kovacevic⁵³, Shiu Kumar⁵⁴, Chandrani Kumari⁵⁵, Christoph F. Kurz^{51,56}, Rintu Kutum¹³, Taeyong Kwon²⁷, Thuc D. Le³¹, Kyeongjun Lee³⁹, Hyungyu Lee⁵⁷, Dawoon Leem⁵⁷, Shuya Li⁵⁸, Weng Khong Lim^{59,60}, Xinyue Liu⁶¹, Yunan Luo⁶², Bahattin C. Maral²², Suyash Mishra⁶³, Yeongeun Nam²⁷, Leelavati Narlikar⁶⁴, Thin Nguyen⁶⁵, Zoran Obradovic⁶⁶, Hyeju Oh²⁷, Kousuke Onoue⁶⁷, Hyojung Paik⁴⁴, Wenchu Pan⁶⁸, Bogyu Park⁵⁷, Balint Armin Pataki⁴, Sumeet Patiyal³⁴, Jian Peng⁶², Dimitri Perrin²⁴, Kaike Ping⁶⁹, Alidivinas Prusokas⁷⁰, Augustinas Prusokas⁷¹, Peng Qiu⁷², Gajendra P.S. Raghava³⁴, Derek Reiman³⁰, Renata Retkute⁷³, Roberto Romero^{1,5-9}, Nay Min Min Thaw Saw³⁸, Neelam Sharma³⁴, Alok Sharma⁷⁴⁻⁷⁷, Ronesh Sharma⁵⁴, Rahul Siddharthan⁵⁵, Musalula Sinkala^{78,79}, Marina Sirota^{10,11}, Alex Soupir³², Marija Stanojevic⁶⁶, Gustavo Stolovitzky^{19,20}, Yufeng Su⁶², Alexander M. Sutherland⁸⁰, András Szilágyi³⁷, Mehmet Tan²², Adi L. Tarca^{1,2,3}, Nandor G. Than^{37,81}, Buu Truong³¹, Edwin Vans^{54,77}, Fangping Wan⁵⁸, Rohan B.H. Williams⁸², Wendy S.W. Wong⁸³, Jeong Woong⁵⁷, Li Xiaomei³¹, Dongchan Yang⁴³, Sanghoo Yoon³⁹, Dakota York³², James Young³², Thomas Yu¹⁵, Wei Zhu⁸⁴

²²Department of Computer Engineering, TOBB University of Economics and Technology, Ankara, Turkey

²³School of Mathematics and Statistics, University of Hyderabad, Hyderabad, India

761 ²⁴School of Computer Science, Queensland University of Technology, Brisbane, Australia
762 ²⁵Departments of Computer and Data Sciences, Case Western Reserve University, Cleveland, OH, USA
763 ²⁶Department of Pharmaceutical Science, College of Pharmacy, University of South Florida, Tampa, FL, USA
764 ²⁷Department of Statistics, Daegu University, Gyeongsan, Republic of Korea
765 ²⁸Department of Computer Science and Engineering, Korea University, Seoul, Republic of Korea
766 ²⁹Department of Computer Science and Engineering, Indraprastha Institute of Information Technology, New Delhi,
767 India
768 ³⁰Department of Bioengineering, University of Illinois at Chicago, Chicago, USA
769 ³¹UniSA STEM, University of South Australia, South Australia, Australia
770 ³²Department of Biology and Microbiology, College of Natural Science, South Dakota State University, Brookings,
771 SD, USA
772 ³³Max-Planck Institute of Molecular Plant Physiology, Potsdam, Germany
773 ³⁴Department of Computational Biology, Indraprastha Institute of Information Technology, New Delhi, India
774 ³⁵School of Electrical Engineering, University in Belgrade, Belgrade, Serbia
775 ³⁶Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest, Hungary
776 ³⁷Institute of Enzymology, Research Centre for Natural Sciences, Budapest, Hungary
777 ³⁸Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore
778 ³⁹Division of Mathematics and Big Data Science, Daegu University, Gyeongsan, Republic of Korea
779 ⁴⁰Research School of Finance, Actuarial Studies & Statistics, Australian National University, Canberra, Australia
780 ⁴¹ZS Associates India Private Ltd., Bangalore, India
781 ⁴²Department of Statistics, Stanford University, Stanford, CA, USA
782 ⁴³Department of Bio and Brain engineering, Korea Advanced Institute of Science and Technology, Daejeon,
783 Republic of Korea
784 ⁴⁴Korea Institute of Science and Technology Information, Center for Supercomputing Application, Daejeon,
785 Republic of Korea
786 ⁴⁵Health Science Center, Peking University, Beijing, China
787 ⁴⁶Department of Computer Science and Engineering, Korea University, Interdisciplinary Graduate Program in
788 Bioinformatics, Korea University, Seoul, Republic of Korea
789 ⁴⁷Department of Electrical Engineering, Veermata Jijabai Technological Institute, Mumbai, India
790 ⁴⁸Bioinformatics Centre, CSIR-Institute of Microbial Technology, Chandigarh, India
791 ⁴⁹Bioinformatics Program, Department of Biology, Temple University, Philadelphia, PA, USA
792 ⁵⁰Patient-Centered Clinical Research Coordinating Center, National Evidence-based Healthcare Collaborating
793 Agency, Seoul, Republic of Korea
794 ⁵¹Munich School of Management and Munich Center of Health Sciences, Ludwig-Maximilians-University Munich,
795 Germany, Munich, Germany
796 ⁵²SymbioSeq LLC, Ashburn, VA, USA
797 ⁵³Faculty of Pharmacy, University of Belgrade, Belgrade, Serbia
798 ⁵⁴School of Electrical and Electronic Engineering, Fiji National University, Suva, Fiji
799 ⁵⁵The Institute of Mathematical Sciences (HBNI), Chennai, India
800 ⁵⁶Institute of Health Economics and Health Care Management, Helmholtz Zentrum München, Munich, Germany
801 ⁵⁷DoAI Inc., Republic of Korea
802 ⁵⁸Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China
803 ⁵⁹Cancer & Stem Cell Biology Program, Duke-NUS Medical School, Singapore
804 ⁶⁰SingHealth Duke-NUS Institute of Precision Medicine, Singapore
805 ⁶¹GeneDx, Gaithersburg, MD, USA
806 ⁶²Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA
807 ⁶³ZS Associates International, Inc., London, UK
808 ⁶⁴CSIR National Chemical Laboratory, Pune, India
809 ⁶⁵Applied Artificial Intelligence Institute (A2I2), Deakin University, Victoria, Australia
810 ⁶⁶Center for Data Analytics and Biomedical Informatics, Computer and Information Sciences Department, Temple
811 University, Philadelphia, PA, USA
812 ⁶⁷NEC Corporation, Tokyo, Japan
813 ⁶⁸Department of Statistics, Peking University, Beijing, China
814 ⁶⁹Guizhou Center for Disease Control and Prevention, Guiyang, Guizhou, China
815 ⁷⁰Plant and Microbial Sciences, School of Natural and Environmental Sciences, University of Newcastle, Newcastle,
816 UK

817 ⁷¹School of Life Sciences, Imperial College London, London, UK
818 ⁷²Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA,
819 USA
820 ⁷³Epidemiology and modelling group, Department of Plant Sciences, University of Cambridge, Cambridge, UK
821 ⁷⁴Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University,
822 Tokyo, Japan
823 ⁷⁵Institute for Integrative and Intelligent Systems, Griffith University, Brisbane, Australia
824 ⁷⁶Laboratory for Medical Science Mathematics, RIKEN, Yokohama, Japan
825 ⁷⁷School of Engineering and Physics, University of the South Pacific, Suva, Fiji
826 ⁷⁸University of Cape Town, Faculty of Health Sciences, Department of Integrative Biomedical Sciences,
827 Computational Biology Division, Cape Town, South Africa
828 ⁷⁹University of Zambia, School of Health Sciences, Department of Biomedical Sciences, Lusaka, Zambia
829 ⁸⁰San Francisco, CA, USA
830 ⁸¹Maternity Clinic of Obstetrics and Gynecology, Budapest, Hungary
831 ⁸²Singapore Centre for Environmental Life Sciences Engineering, National University of Singapore, Singapore
832 ⁸³Coherent Logic Limited, McLean, VA, USA
833 ⁸⁴CodexSage, LLC, Gaithersburg, MD, USA
834

835

References

1. Brosens, I., Pijnenborg, R., Vercruysse, L. & Romero, R. The "Great Obstetrical Syndromes" are associated with disorders of deep placentation. *American journal of obstetrics and gynecology* **204**, 193-201 (2011).
2. Blencowe, H., *et al.* 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 (2012).
3. Martin, J.A., Hamilton, B.E. & Osterman, M.J.K. Births in the United States, 2018. *NCHS Data Brief*, 1-8 (2019).
4. in *Preterm Birth: Causes, Consequences, and Prevention* (eds. Behrman, R.E. & Butler, A.S.) (Washington (DC), 2007).
5. Lawn, J.E., Kerber, K., Enweronu-Laryea, C. & Cousens, S. 3.6 million neonatal deaths--what is progressing and what is not? *Semin Perinatol* **34**, 371-386 (2010).
6. Mwaniki, M.K., Atieno, M., Lawn, J.E. & Newton, C.R. Long-term neurodevelopmental outcomes after intrauterine and neonatal insults: a systematic review. *Lancet* **379**, 445-452 (2012).
7. Spittle, A., Orton, J., Anderson, P.J., Boyd, R. & Doyle, L.W. Early developmental intervention programmes provided post hospital discharge to prevent motor and cognitive impairment in preterm infants. *Cochrane Database Syst Rev*, CD005495 (2015).
8. Goldenberg, R.L., Culhane, J.F., Iams, J.D. & Romero, R. Epidemiology and causes of preterm birth. *Lancet* **371**, 75-84 (2008).
9. Romero, R., Dey, S.K. & Fisher, S.J. Preterm labor: one syndrome, many causes. *Science* **345**, 760-765 (2014).
10. Knijnenburg, T.A., *et al.* Genomic and molecular characterization of preterm birth. *Proc Natl Acad Sci U S A* **116**, 5819-5827 (2019).
11. Romero, R., *et al.* A blueprint for the prevention of preterm birth: vaginal progesterone in women with a short cervix. *Journal of perinatal medicine* **41**, 27-44 (2013).
12. Romero, R., *et al.* 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. *American journal of obstetrics and gynecology* **206**, 124 e121-119 (2012).
13. Conde-Agudelo, A., *et al.* 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. *American journal of obstetrics and gynecology* **219**, 10-25 (2018).
14. Hassan, S.S., *et al.* Vaginal progesterone reduces the rate of preterm birth in women with a sonographic short cervix: a multicenter, randomized, double-blind, placebo-controlled trial. *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology* **38**, 18-31 (2011).
15. Rubin, R. Confirmatory Trial for Drug to Prevent Preterm Birth Finds No Benefit, So Why Is It Still Prescribed? *JAMA* (2020).

16. Godlewski, B.J., Sobolik, L.I., King, V.J. & Harrod, C.S. Accelerated Approval of 17alpha-Hydroxyprogesterone Caproate: A Cautionary Tale. *Obstetrics and gynecology* **135**, 1207-1213 (2020).
17. Ngo, T.T.M., *et al.* Noninvasive blood tests for fetal development predict gestational age and preterm delivery. *Science* **360**, 1133-1136 (2018).
18. Pavlicev, M., *et al.* Single-cell transcriptomics of the human placenta: inferring the cell communication network of the maternal-fetal interface. *Genome Res* **27**, 349-361 (2017).
19. Ghaemi, M.S., *et al.* Multiomics modeling of the immunome, transcriptome, microbiome, proteome and metabolome adaptations during human pregnancy. *Bioinformatics* **35**, 95-103 (2019).
20. Romero, R., *et al.* The use of high-dimensional biology (genomics, transcriptomics, proteomics, and metabolomics) to understand the preterm parturition syndrome. *BJOG* **113 Suppl 3**, 118-135 (2006).
21. Aghaepour, N., *et al.* A proteomic clock of human pregnancy. *American journal of obstetrics and gynecology* **218**, 347.e341-347.e314 (2018).
22. Romero, R., *et al.* The maternal plasma proteome changes as a function of gestational age in normal pregnancy: a longitudinal study. *American journal of obstetrics and gynecology* **217**, 67.e61-67.e21 (2017).
23. Tarca, A.L., *et al.* 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 (2019).
24. Romero, R., *et al.* 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 (2014).
25. DiGiulio, D.B., *et al.* Temporal and spatial variation of the human microbiota during pregnancy. *Proc Natl Acad Sci U S A* **112**, 11060-11065 (2015).
26. Vora, B., *et al.* Meta-Analysis of Maternal and Fetal Transcriptomic Data Elucidates the Role of Adaptive and Innate Immunity in Preterm Birth. *Frontiers in immunology* **9**, 993 (2018).
27. Stolovitzky, G., Monroe, D. & Califano, A. Dialogue on reverse-engineering assessment and methods: the DREAM of high-throughput pathway inference. *Ann N Y Acad Sci* **1115**, 1-22 (2007).
28. Heng, Y.J., *et al.* Maternal Whole Blood Gene Expression at 18 and 28 Weeks of Gestation Associated with Spontaneous Preterm Birth in Asymptomatic Women. *PLoS One* **11**, e0155191 (2016).
29. Gomez-Lopez, N., *et al.* The Cellular Transcriptome in the Maternal Circulation During Normal Pregnancy: A Longitudinal Study. *Frontiers in immunology* **10**, 2863 (2019).
30. Sacks, G.P., Studena, K., Sargent, K. & Redman, C.W. Normal pregnancy and preeclampsia both produce inflammatory changes in peripheral blood leukocytes akin to those of sepsis. *American journal of obstetrics and gynecology* **179**, 80-86 (1998).
31. Naccasha, N., *et al.* Phenotypic and metabolic characteristics of monocytes and granulocytes in normal pregnancy and maternal infection. *American journal of obstetrics and gynecology* **185**, 1118-1123 (2001).

32. Germain, S.J., Sacks, G.P., Sooranna, S.R., Sargent, I.L. & Redman, C.W. Systemic inflammatory priming in normal pregnancy and preeclampsia: the role of circulating syncytiotrophoblast microparticles. *J Immunol* **178**, 5949-5956 (2007).
33. Gomez-Lopez, N., Tanaka, S., Zaeem, Z., Metz, G.A. & Olson, D.M. Maternal circulating leukocytes display early chemotactic responsiveness during late gestation. *BMC Pregnancy Childbirth* **13 Suppl 1**, S8 (2013).
34. Aghaeepour, N., *et al.* An immune clock of human pregnancy. *Sci Immunol* **2**(2017).
35. Savitz, D.A., *et al.* Comparison of pregnancy dating by last menstrual period, ultrasound scanning, and their combination. *American journal of obstetrics and gynecology* **187**, 1660-1666 (2002).
36. Romero, R., *et al.* The preterm parturition syndrome. *BJOG* **113 Suppl 3**, 17-42 (2006).
37. Norwitz, E.R., Robinson, J.N. & Challis, J.R. The control of labor. *N Engl J Med* **341**, 660-666 (1999).
38. Davies, D.R., *et al.* Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. *Proc Natl Acad Sci U S A* **109**, 19971-19976 (2012).
39. Gold, L., *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* **5**, e15004 (2010).
40. Erez, O., *et al.* The prediction of late-onset preeclampsia: Results from a longitudinal proteomics study. *PLoS One* **12**, e0181468 (2017).
41. Gervasi, M.T., *et al.* Phenotypic and metabolic characteristics of maternal monocytes and granulocytes in preterm labor with intact membranes. *American journal of obstetrics and gynecology* **185**, 1124-1129 (2001).
42. Zhang, J., *et al.* Immunophenotyping and activation status of maternal peripheral blood leukocytes during pregnancy and labour, both term and preterm. *J Cell Mol Med* **21**, 2386-2402 (2017).
43. Paquette, A.G., *et al.* Comparative analysis of gene expression in maternal peripheral blood and monocytes during spontaneous preterm labor. *American journal of obstetrics and gynecology* **218**, 345 e341-345 e330 (2018).
44. Pique-Regi, R., *et al.* Single cell transcriptional signatures of the human placenta in term and preterm parturition. *Elife* **8**(2019).
45. Gervasi, M.T., *et al.* Maternal intravascular inflammation in preterm premature rupture of membranes. *J Matern Fetal Neonatal Med* **11**, 171-175 (2002).
46. Tarca, A.L., *et al.* The prediction of early preeclampsia: Results from a longitudinal proteomics study. *PLoS One* **14**, e0217273 (2019).
47. Tarca, A.L., *et al.* Maternal whole blood mRNA signatures identify women at risk of early preeclampsia: a longitudinal study. *J Matern Fetal Neonatal Med*, 1-12 (2020).
48. Wolpert, D.H. Stacked generalization. *Neural Networks* **2**, 241-259 (1992).
49. Kuffner, R., *et al.* Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression. *Nat Biotechnol* **33**, 51-57 (2015).
50. Keller, A., *et al.* Predicting human olfactory perception from chemical features of odor molecules. *Science* **355**, 820-826 (2017).
51. Salcedo, A., *et al.* A community effort to create standards for evaluating tumor subclonal reconstruction. *Nat Biotechnol* **38**, 97-107 (2020).

965 52. Schaffter, T., *et al.* Evaluation of Combined Artificial Intelligence and Radiologist
966 Assessment to Interpret Screening Mammograms. *JAMA Netw Open* **3**, e200265 (2020).
967 53. Costello, J.C., *et al.* A community effort to assess and improve drug sensitivity prediction
968 algorithms. *Nat Biotechnol* **32**, 1202-1212 (2014).
969 54. Guinney, J. & Saez-Rodriguez, J. Alternative models for sharing confidential biomedical
970 data. *Nat Biotechnol* **36**, 391-392 (2018).
971 55. Saez-Rodriguez, J., *et al.* Crowdsourcing biomedical research: leveraging communities as
972 innovation engines. *Nat Rev Genet* **17**, 470-486 (2016).
973 56. Bukowski, R., *et al.* Onset of human preterm and term birth is related to unique
974 inflammatory transcriptome profiles at the maternal fetal interface. *PeerJ* **5**, e3685
975 (2017).
976 57. Heng, Y.J., Pennell, C.E., Chua, H.N., Perkins, J.E. & Lye, S.J. Whole blood gene expression
977 profile associated with spontaneous preterm birth in women with threatened preterm
978 labor. *PLoS One* **9**, e96901 (2014).
979 58. Tsang, J.C.H., *et al.* Integrative single-cell and cell-free plasma RNA transcriptomics
980 elucidates placental cellular dynamics. *Proc Natl Acad Sci U S A* **114**, E7786-E7795 (2017).
981 59. Tarca, A.L., *et al.* Human blood gene signature as a marker for smoking exposure:
982 computational approaches of the top ranked teams in the sbv IMPROVER Systems
983 Toxicology challenge. *Comput Toxicol* **5**, 31-37 (2018).
984 60. Tarca, A.L., *et al.* Strengths and limitations of microarray-based phenotype prediction:
985 lessons learned from the IMPROVER Diagnostic Signature Challenge. *Bioinformatics* **29**,
986 2892-2899 (2013).
987 61. Hoskins, R.A., *et al.* Reports from CAGI: The Critical Assessment of Genome Interpretation.
988 *Hum Mutat* **38**, 1039-1041 (2017).
989 62. Andreoletti, G., Pal, L.R., Moul, J. & Brenner, S.E. Reports from the fifth edition of CAGI:
990 The Critical Assessment of Genome Interpretation. *Hum Mutat* **40**, 1197-1201 (2019).
991 63. Pejaver, V., Mooney, S.D. & Radivojac, P. Missense variant pathogenicity predictors
992 generalize well across a range of function-specific prediction challenges. *Hum Mutat* **38**,
993 1092-1108 (2017).
994 64. Clark, W.T., *et al.* Assessment of predicted enzymatic activity of alpha-N-
995 acetylglucosaminidase variants of unknown significance for CAGI 2016. *Hum Mutat* **40**,
996 1519-1529 (2019).
997 65. Menden, M.P., *et al.* Community assessment to advance computational prediction of
998 cancer drug combinations in a pharmacogenomic screen. *Nat Commun* **10**, 2674 (2019).
999 66. Gonen, M., *et al.* A Community Challenge for Inferring Genetic Predictors of Gene
1000 Essentialities through Analysis of a Functional Screen of Cancer Cell Lines. *Cell Syst* **5**, 485-
1001 497 e483 (2017).
1002 67. ACOG practice bulletin. Diagnosis and management of preeclampsia and eclampsia.
1003 Number 33, January 2002. *Obstetrics and gynecology* **99**, 159-167 (2002).
1004 68. von Dadelszen, P., Magee, L.A. & Roberts, J.M. Subclassification of preeclampsia.
1005 *Hypertension in pregnancy* **22**, 143-148 (2003).
1006 69. Gomez, R., Romero, R., Edwin, S.S. & David, C. Pathogenesis of preterm labor and preterm
1007 premature rupture of membranes associated with intraamniotic infection. *Infectious*
1008 *disease clinics of North America* **11**, 135-176 (1997).

70. Guinn, D.A., *et al.* Risk factors for the development of preterm premature rupture of the membranes after arrest of preterm labor. *American journal of obstetrics and gynecology* **173**, 1310-1315 (1995).
71. Irizarry, R.A., *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. **4**, 249-264 (2003).
72. Carvalho, B.S. & Irizarry, R.A. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**, 2363-2367 (2010).
73. Smyth, G.K. Limma: linear models for microarray data. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (eds. Gentleman, R., Carey, V.J., Huber, W., Irizarry, R.A. & Dudoit, S.) 397-420 (Springer, 2012).
74. Gentleman, R.C., *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80 (2004).
75. Hoerl, A.E. & Kennard, R.W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **12**, 55-67 (1970).
76. Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. & Leisch, F. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package (2019).
77. Ritchie, M.E., *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
78. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257-258 (2007).
79. Ashburner, M., *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).
80. Doncheva, N.T., Morris, J.H., Gorodkin, J. & Jensen, L.J. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J Proteome Res* **18**, 623-632 (2019).
81. Otasek, D., Morris, J.H., Boucas, J., Pico, A.R. & Demchak, B. Cytoscape Automation: empowering workflow-based network analysis. *Genome Biol* **20**, 185 (2019).
82. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).

Figure Legends

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:** Whole blood transcriptomic data were generated from samples collected in normal pregnancies and those complicated by spontaneous preterm birth with intact (sPTD) or ruptured membranes (PPROM), or preeclampsia. Participating teams were provided gene expression data to develop their own 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 **Fig. S1** for training/test partition). In a post challenge analysis, the approach of the top team in sub-challenge 1 (smallest test set Root Mean Squared Error) was applied to predict time to delivery. **Sub-challenge 2:** Participating teams submitted risk prediction algorithms designed to use as input omics data at two or more time points from individual women coupled with outcome information (control, sPTD or PPRM) 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 **Table 1**) to assess within- and across-cohort predictions of preterm birth by whole blood 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 **Methods**).

Figure 2. Prediction of gestational age by whole blood transcriptomics. A) Detroit cohort whole blood transcriptomics study design. Each line corresponds to one patient and each dot represents a sample. Gestational ages at delivery are marked by a triangle. B) Test set prediction of gestational age by the model of the top-ranked team (M_GA_Team1). 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.

Figure 3. Prediction of time to delivery by whole blood transcriptomics. A) The top panel shows the test set time-to-delivery (TTD) estimates from the M_sTD_TTD model plotted against actual values. 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 al¹². T1: <12 weeks; T2 = 12-24 weeks, and T3 = 24-37 weeks of gestation. B) Prediction of time-to-delivery in women with spontaneous preterm birth by a gene expression model established in women with spontaneous term delivery (M_sTD_TTD).

Figure 4. Prediction of preterm birth in asymptomatic women by Detroit omics datasets. From the transcriptomic study in **Fig. 2A**, only controls and preterm birth groups were included. A) Cases were selected if they were asymptomatic at 17-23-week and 27-33-week intervals or B) at 17-22-week, 22-27-week, and 27-33-week intervals. C) Plasma proteomic samples in cases and controls at 17-23-week and 27-33-week intervals.

Figure 5. Prediction performance and team ranking in sub-challenge 2. Prediction performance for preterm birth of the approaches of 13 teams under 7 scenarios (Table 1). AUROC and AUPRC metrics were converted into Z-scores and shown as a heatmap.

Figure 6. Prediction of preterm prelabor rupture of the membranes from samples collected in asymptomatic women. A) Receiver operating characteristic curve (ROC) representing prediction of preterm prelabor rupture of the membranes (PPROM) by 50 genes across the cohorts and microarray platforms using the Team 1 approach. B) Heatmap of 402 genes differentially expressed in PPRM 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 PPRM. Significantly enriched biological processes are highlighted.

Figure 7. Prediction of spontaneous preterm delivery by plasma proteomic data. A) Receiver operating characteristic (ROC) curve represents the predictions of spontaneous preterm delivery (sPTD) and spontaneous preterm birth (sPTB) [which includes sPTD and preterm prelabor rupture of the membranes (PPROM)] for Team 1. B) Plasma protein abundance for all proteins deemed as significant according to a moderated t-test ($q\text{-value} < 0.1$), of which those selected by the top teams as predictors in their models are marked on the left side of the heatmap. C) Overlap of protein changes in those with sPTD at an earlier time point (17-22 weeks of gestation) and with PPRM at 27-33 weeks of gestation. D) Network of proteins among those shown in panel B: each protein node is annotated to biological processes based on corresponding gene ontology.

Figure 8. Comparison of prediction performance of spontaneous preterm delivery between platforms. Receiver operating characteristic curve (ROC) for prediction of spontaneous preterm delivery by models obtained with the approach of Team 1 based on a subset of samples for which data from both platforms were available. The multi-omics model was obtained 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 logistic regression.

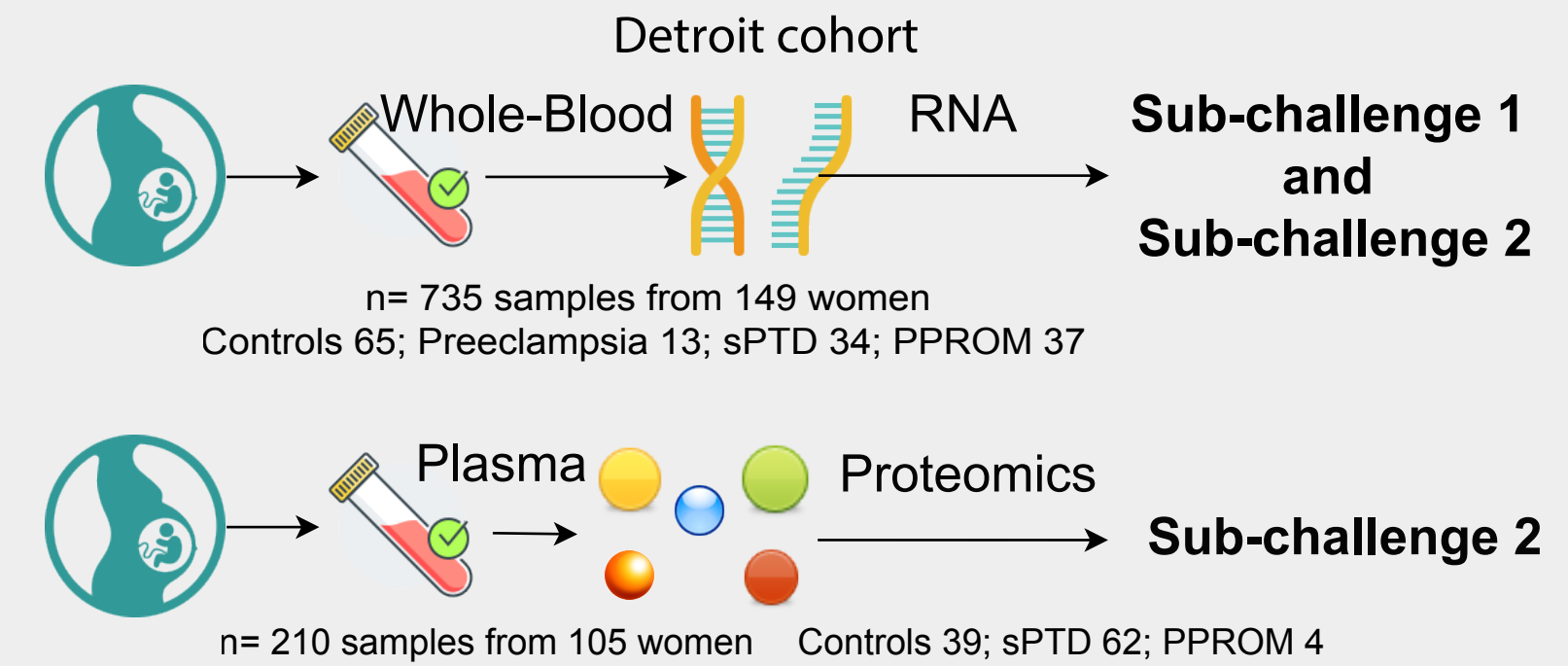
Challenge



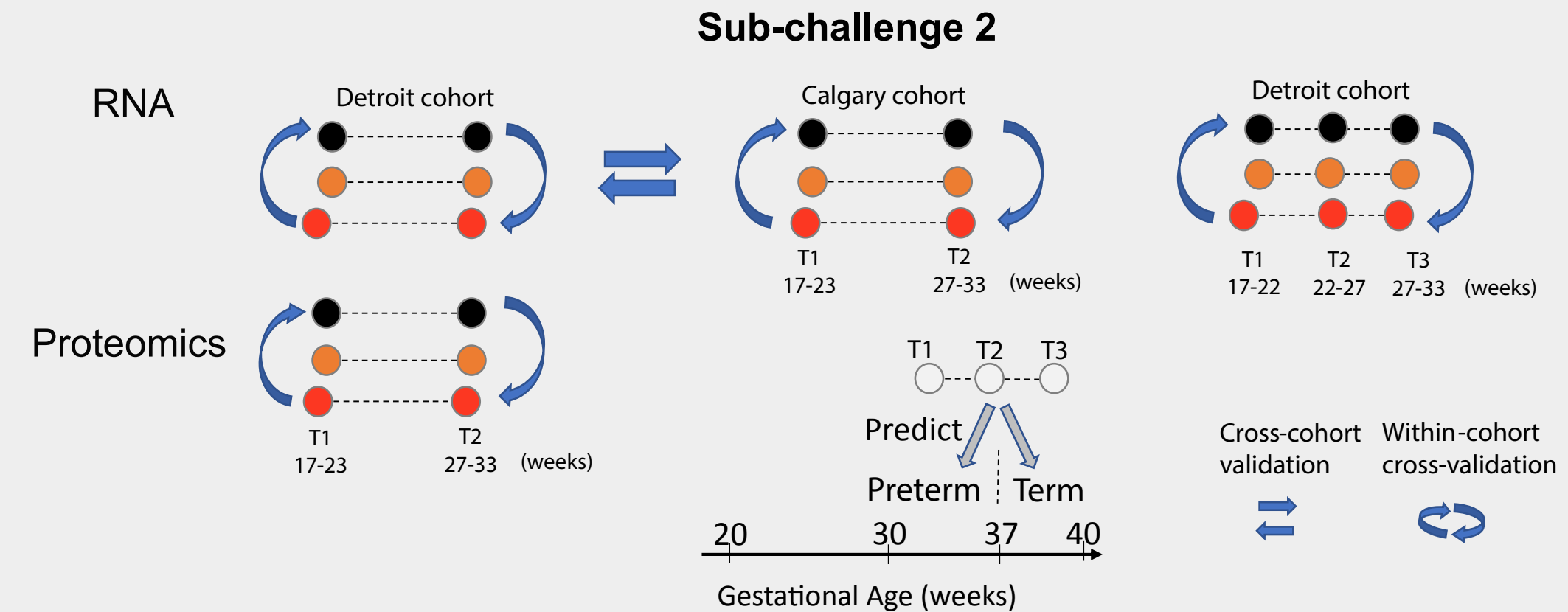
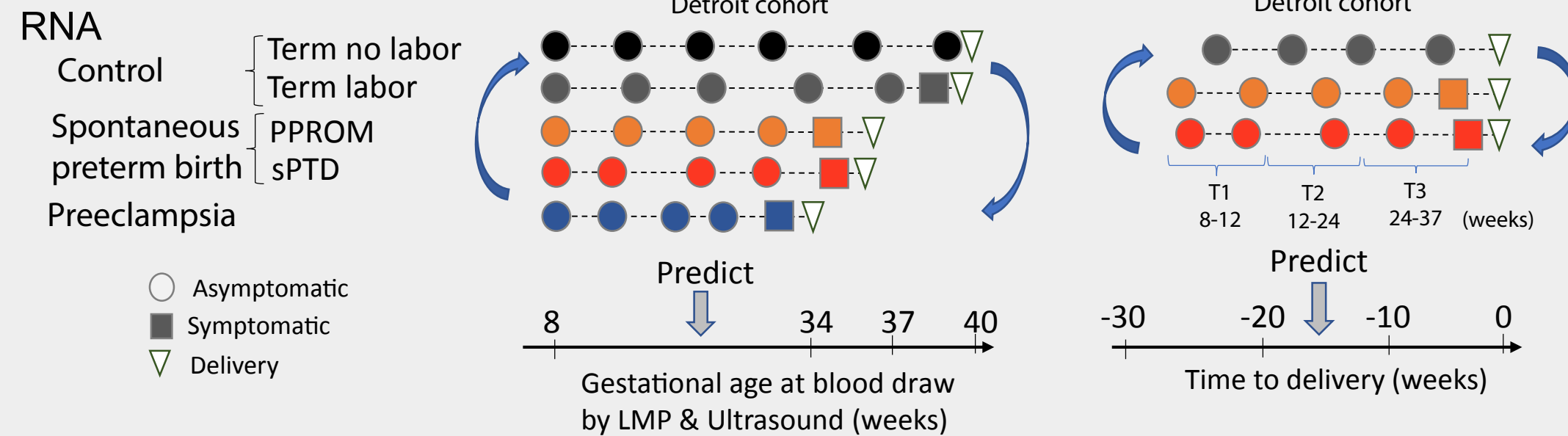
Sub-challenge 1: Prediction of gestational age.

Sub-challenge 2: Prediction of spontaneous preterm birth

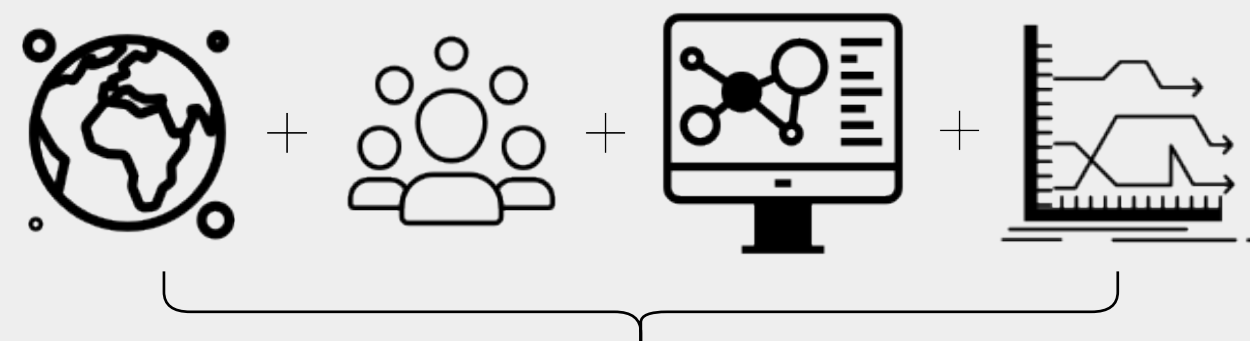
Data



Design



Assessment



DREAM
CHALLENGES
powered by Saae Bionetworks

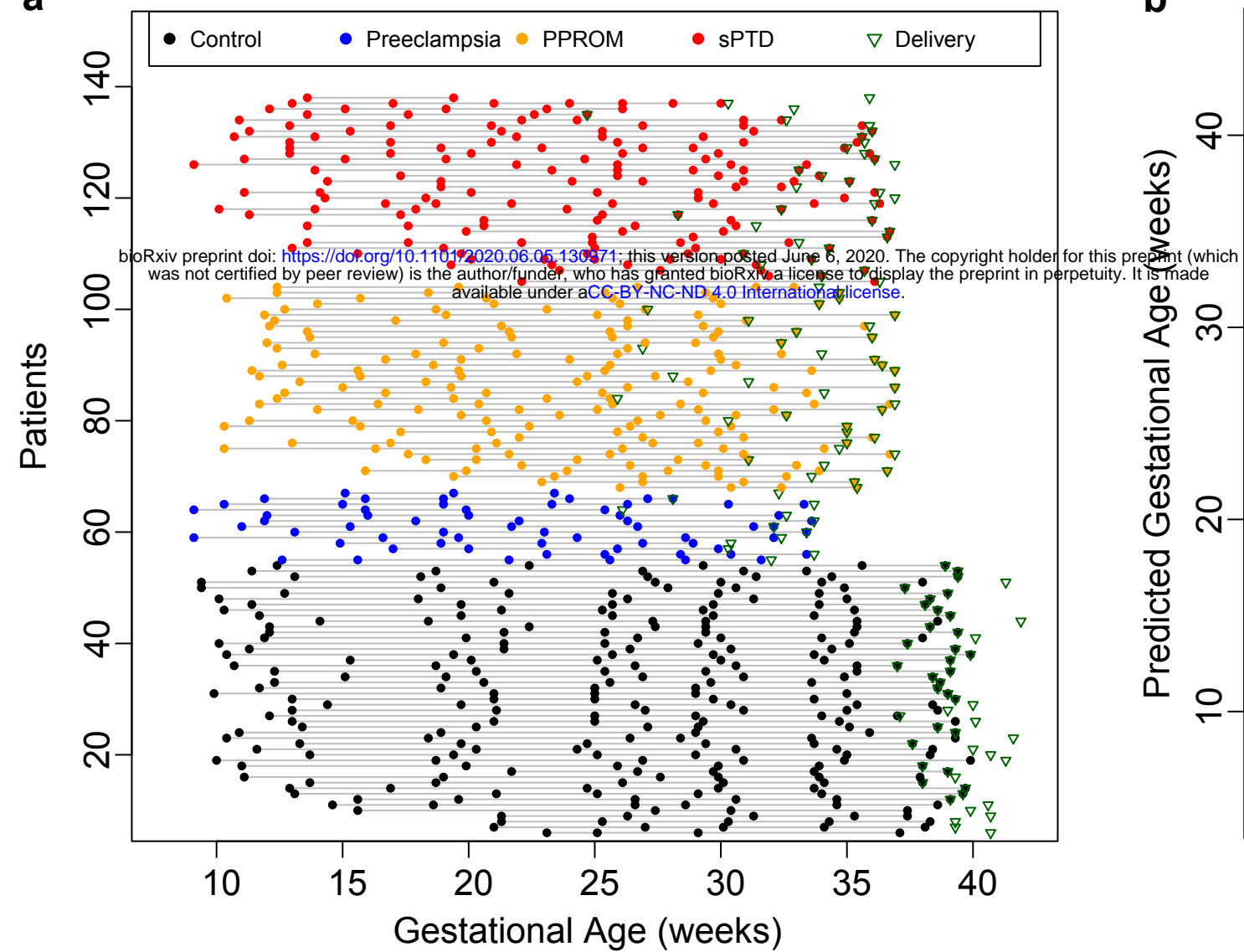
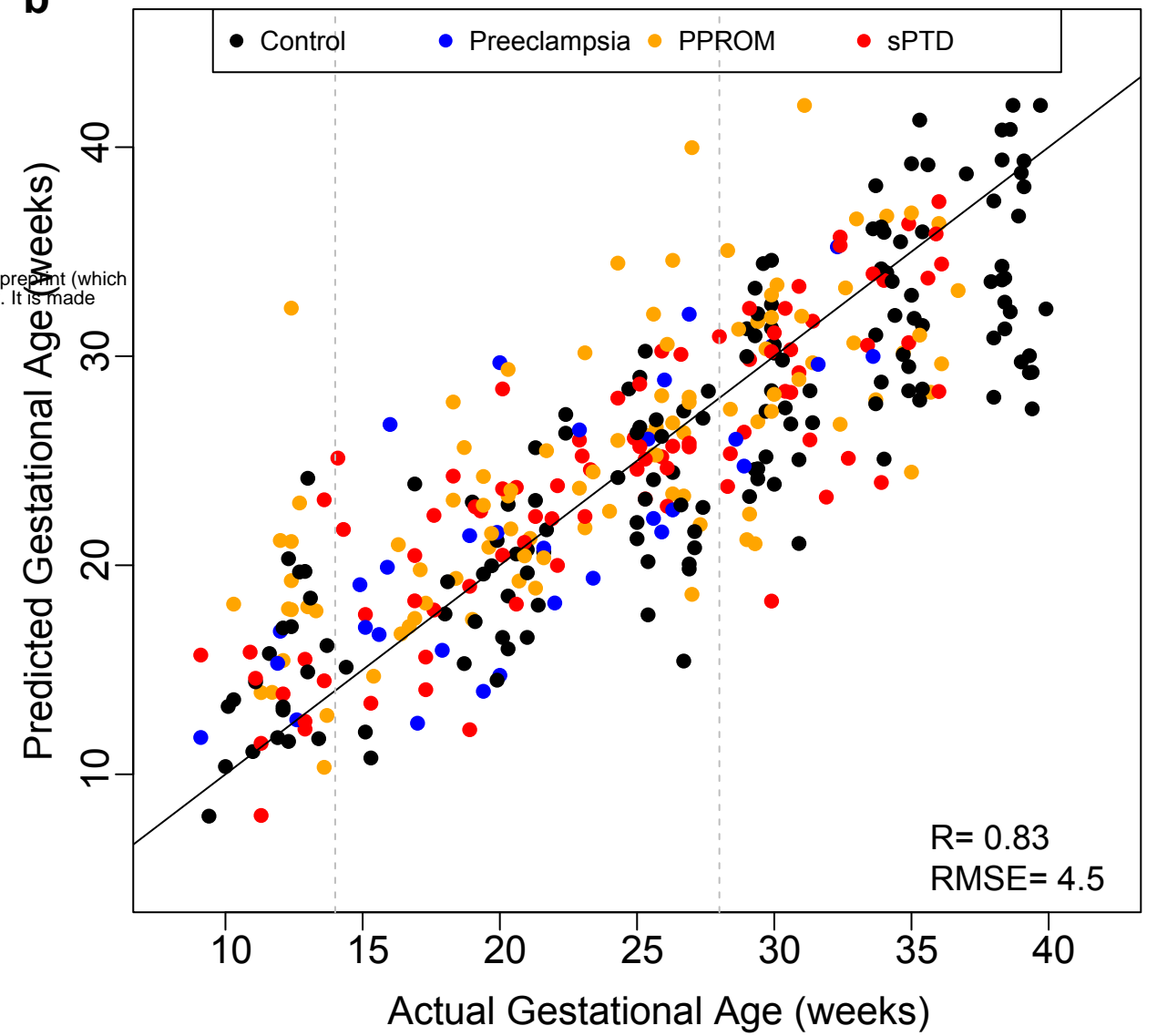
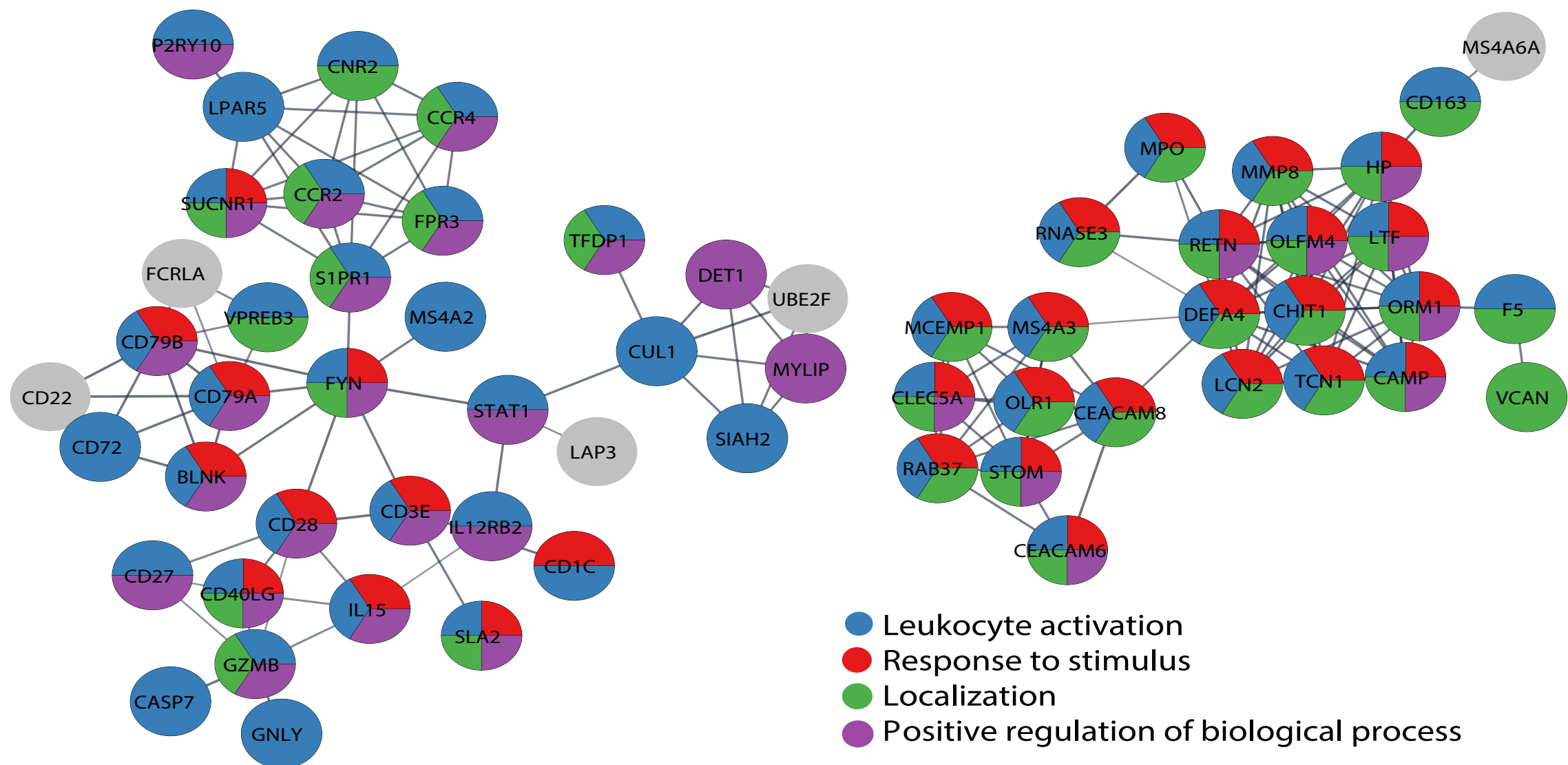
Sub-challenge 1:
May 4 - Aug 15, 2019

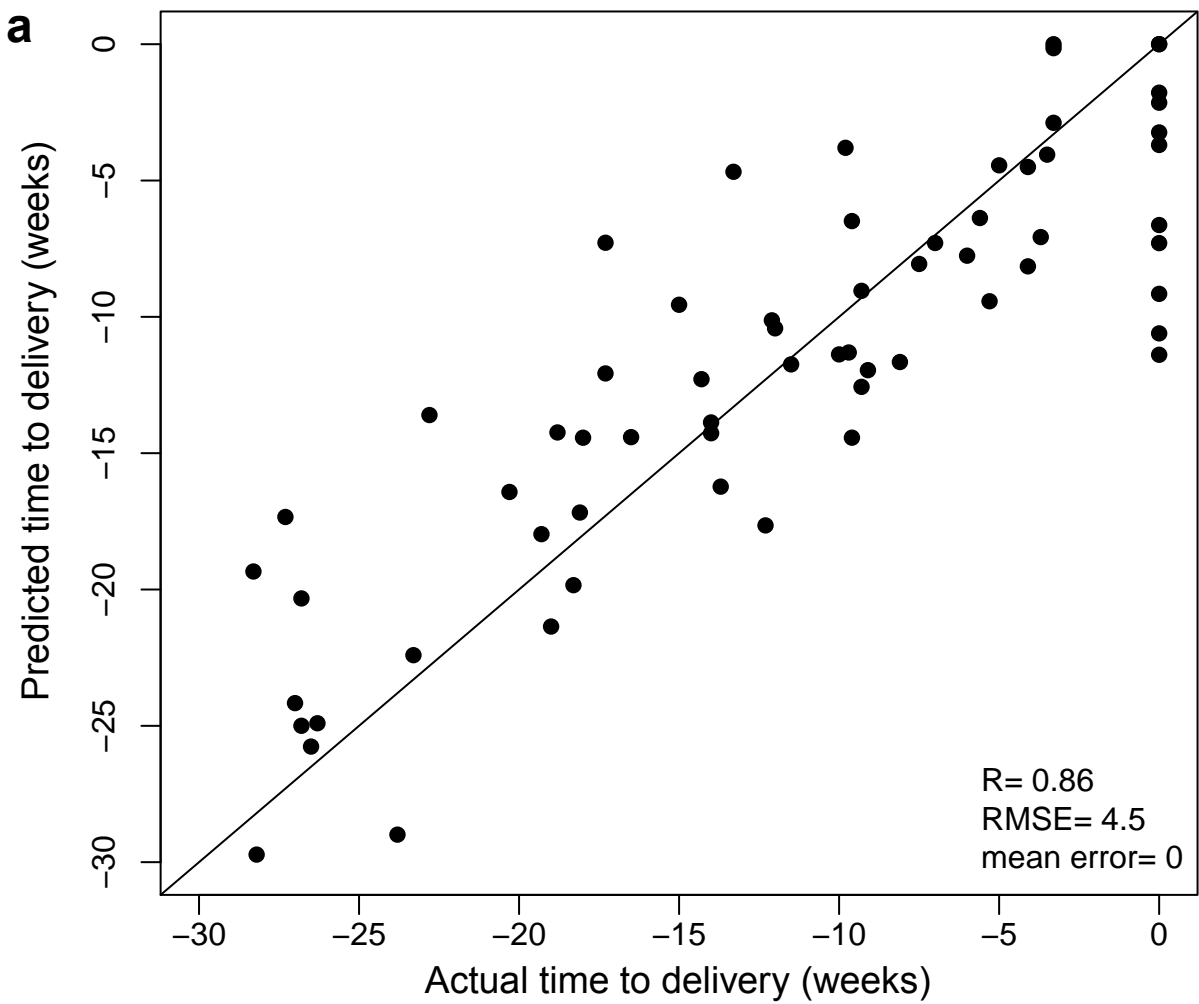
Sub-challenge 2:
Aug 15, 2019 - January 3, 2020

533 registered participants

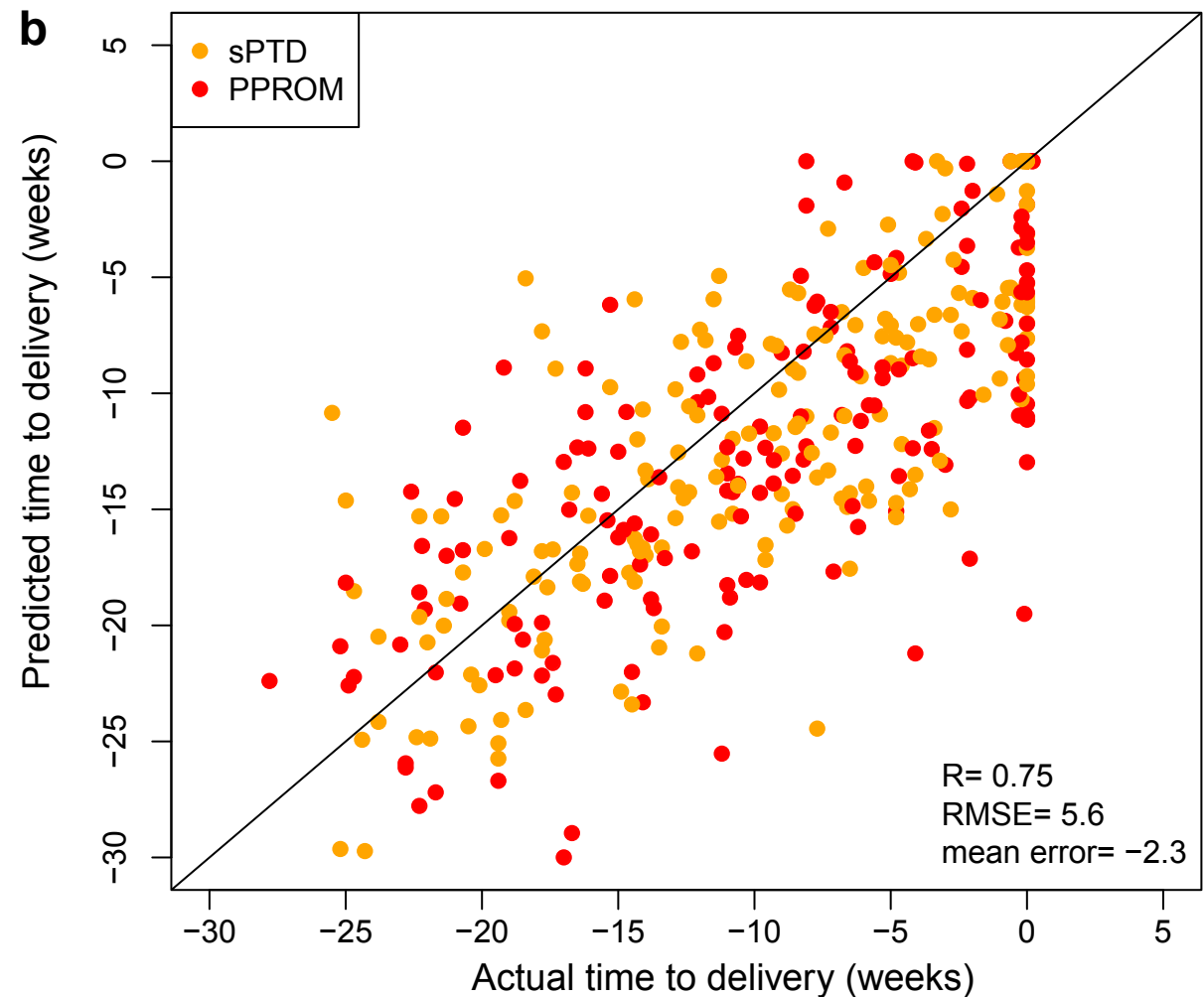
87 unique teams submitted predictions of gestational ages of samples in the test set. Predictions were assessed by Root Mean Squared Error

13 teams submitted computer algorithms that were applied by challenge organizers to 70 pairs of training/test to generate and test models. Predictions were assessed by Area Under the Receiver Operating Characteristic and Area Under Precision Recall curves.

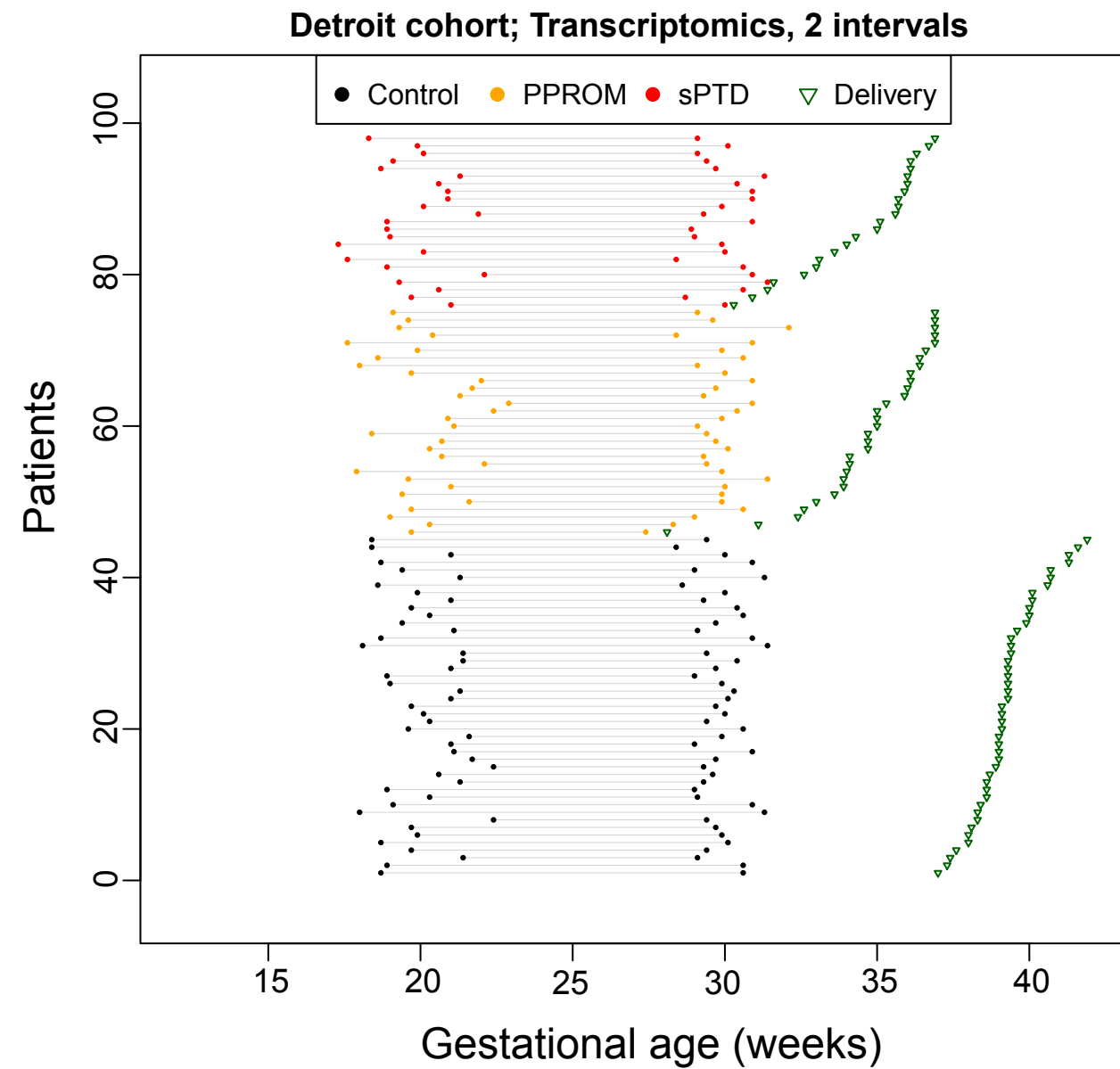
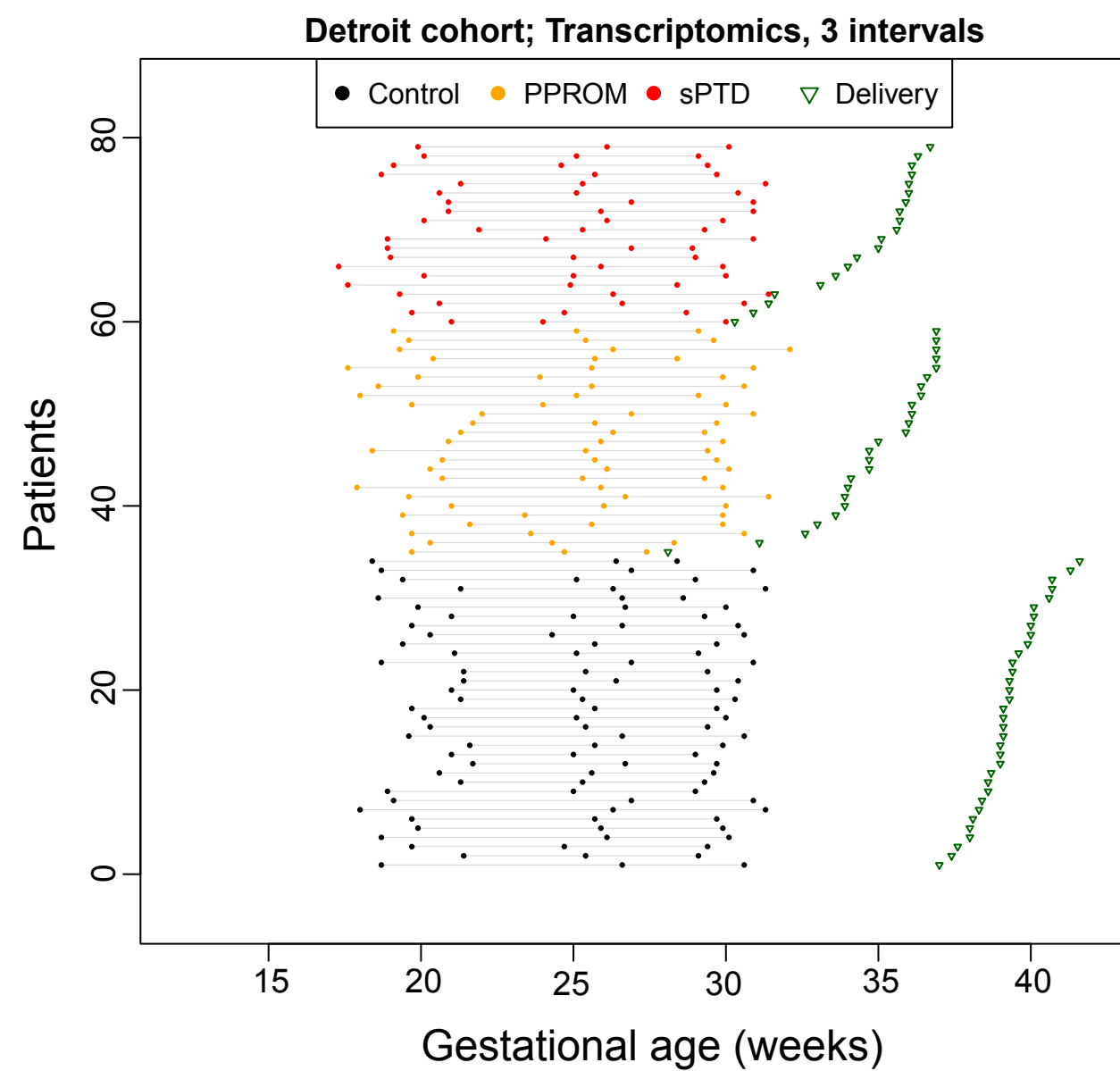
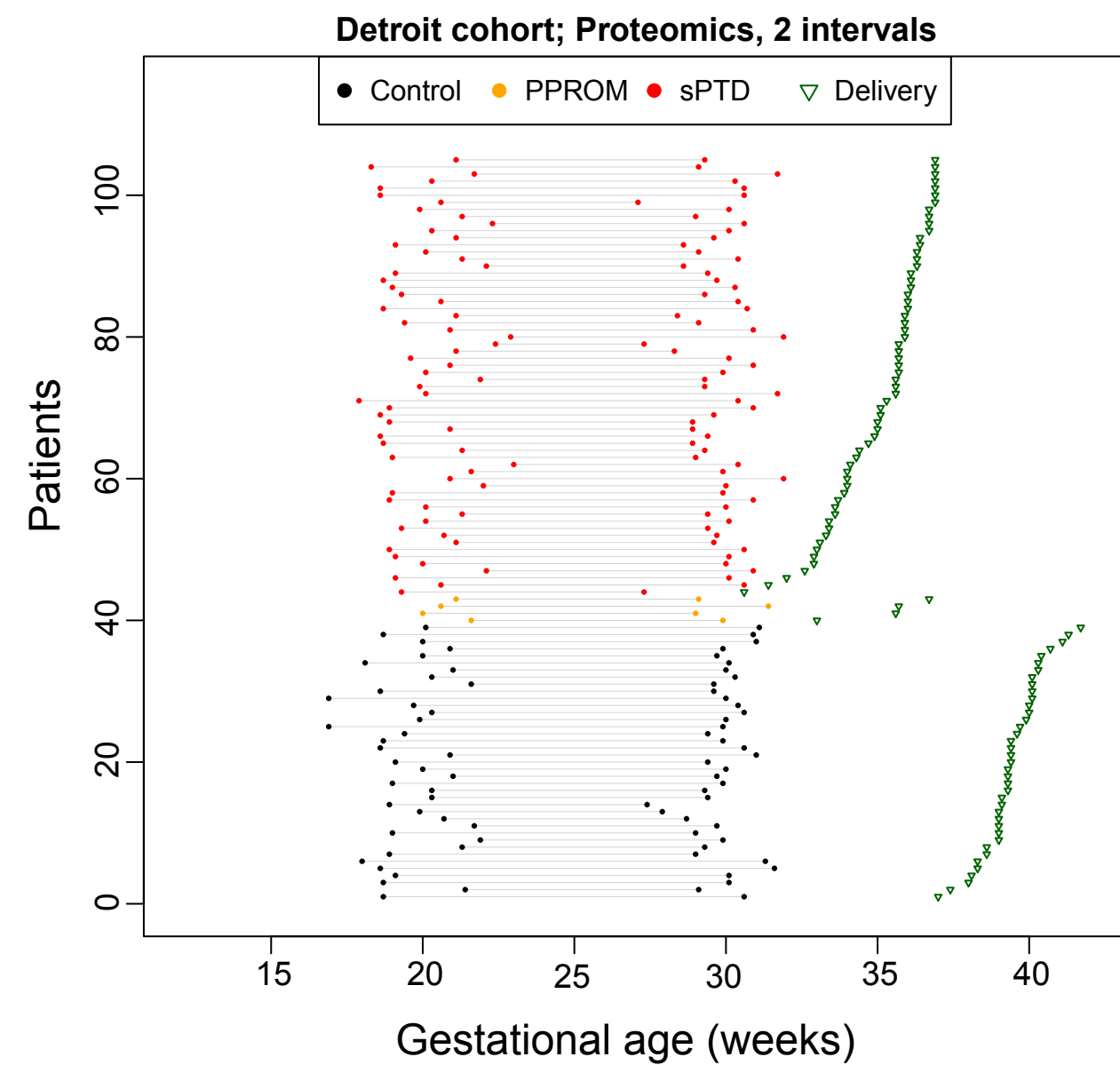
a**b****c**

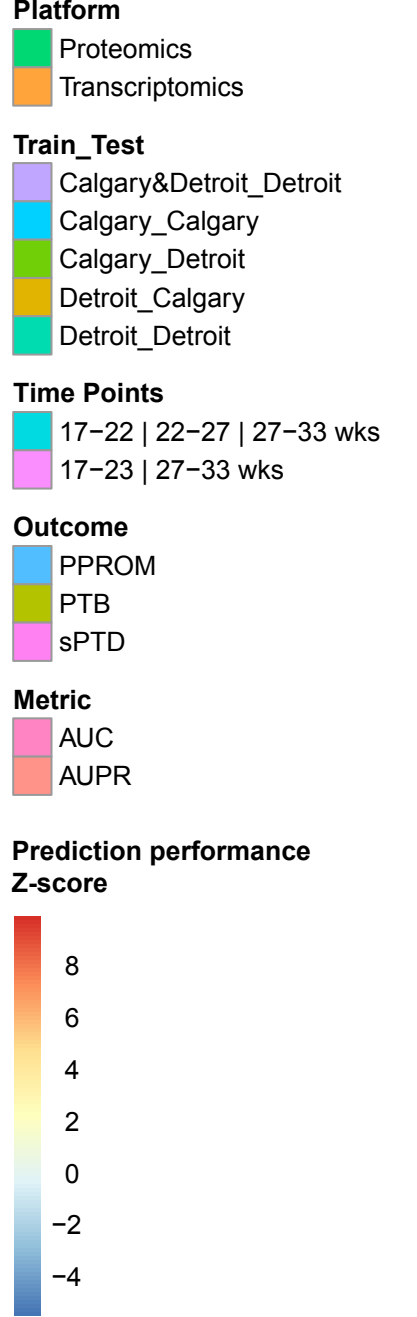
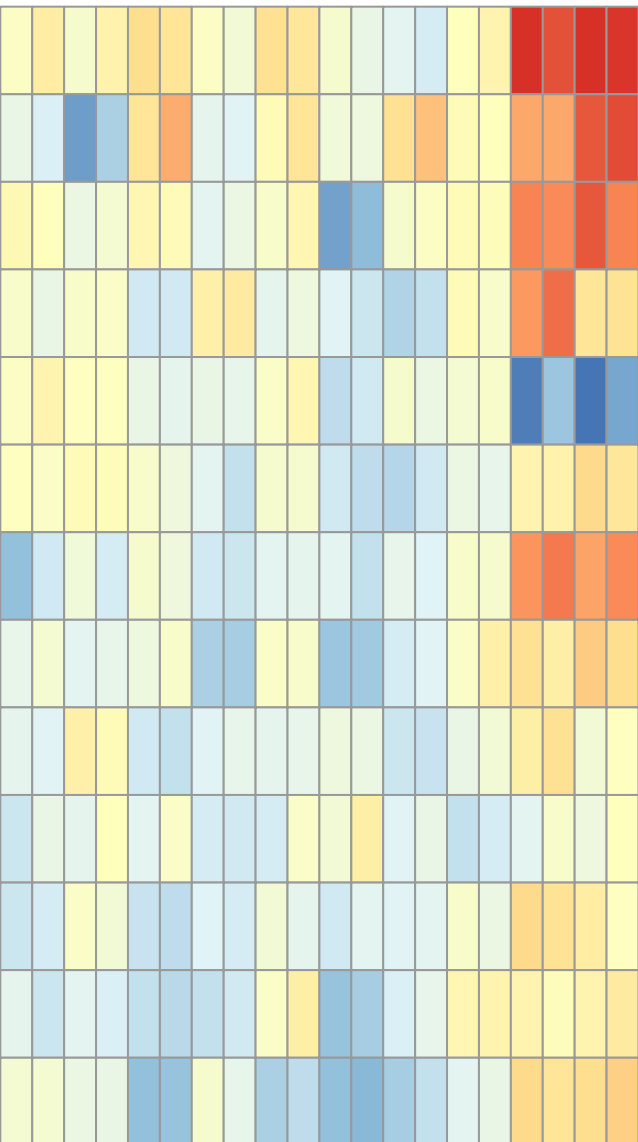
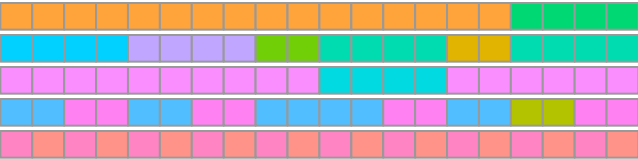


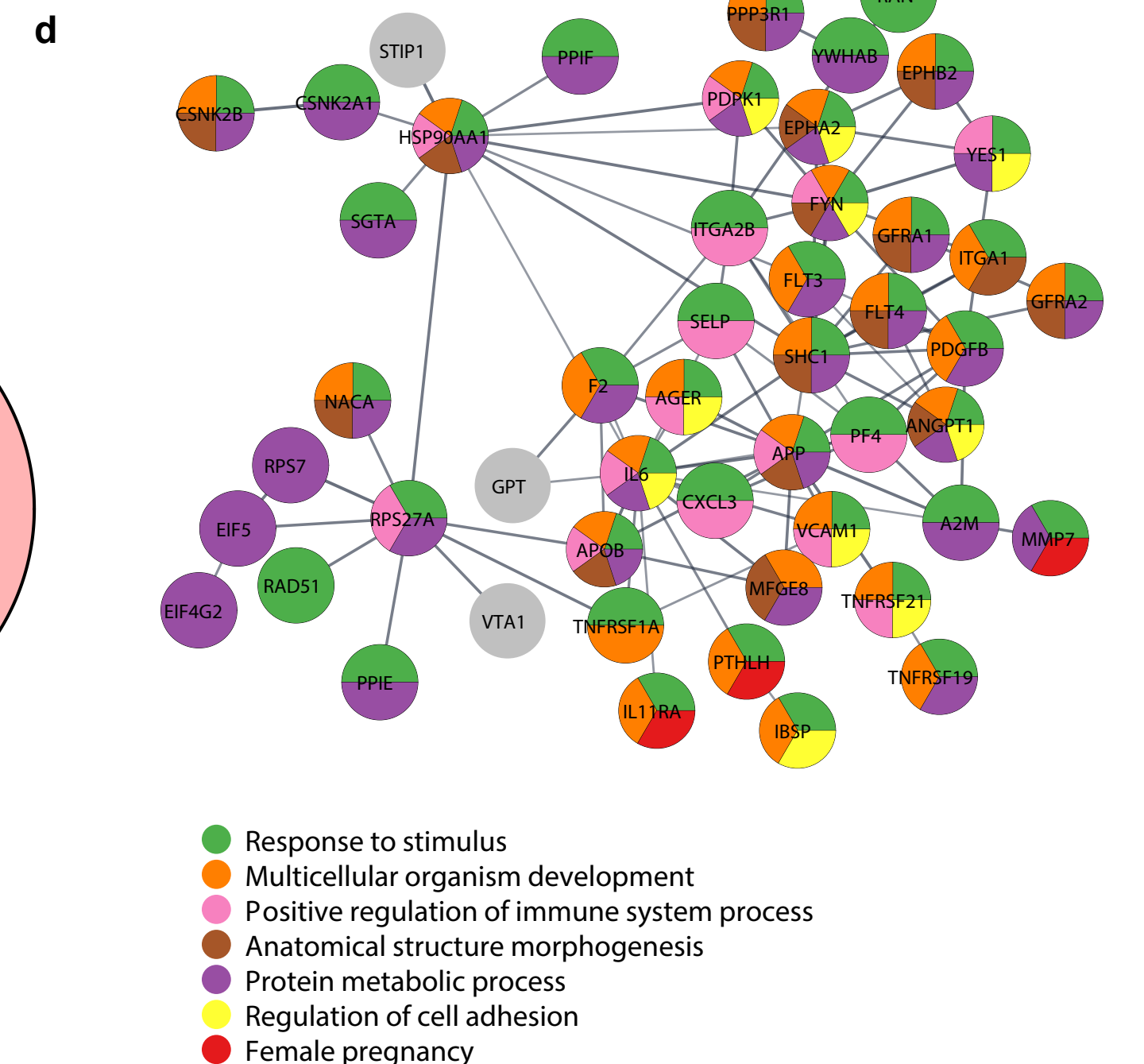
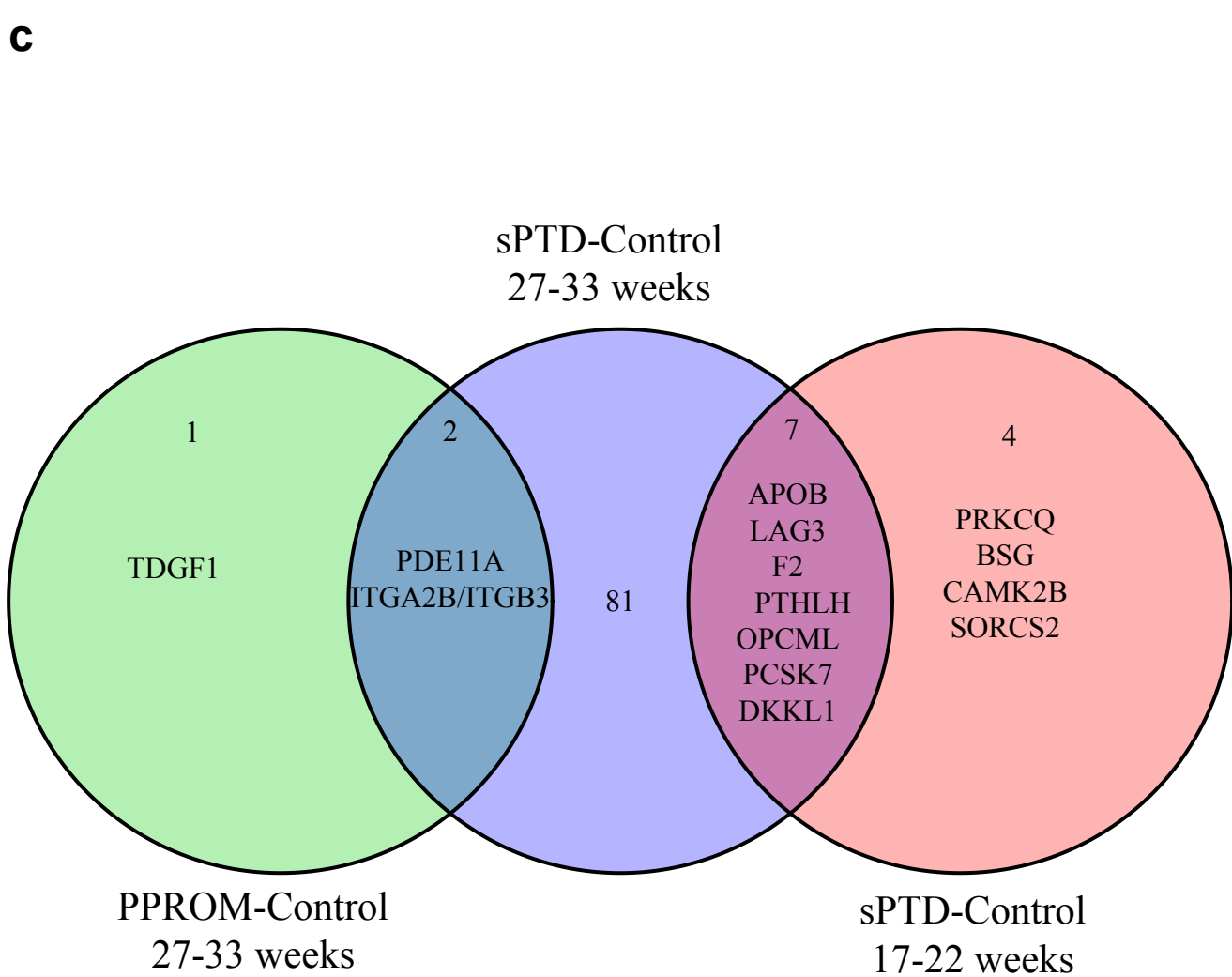
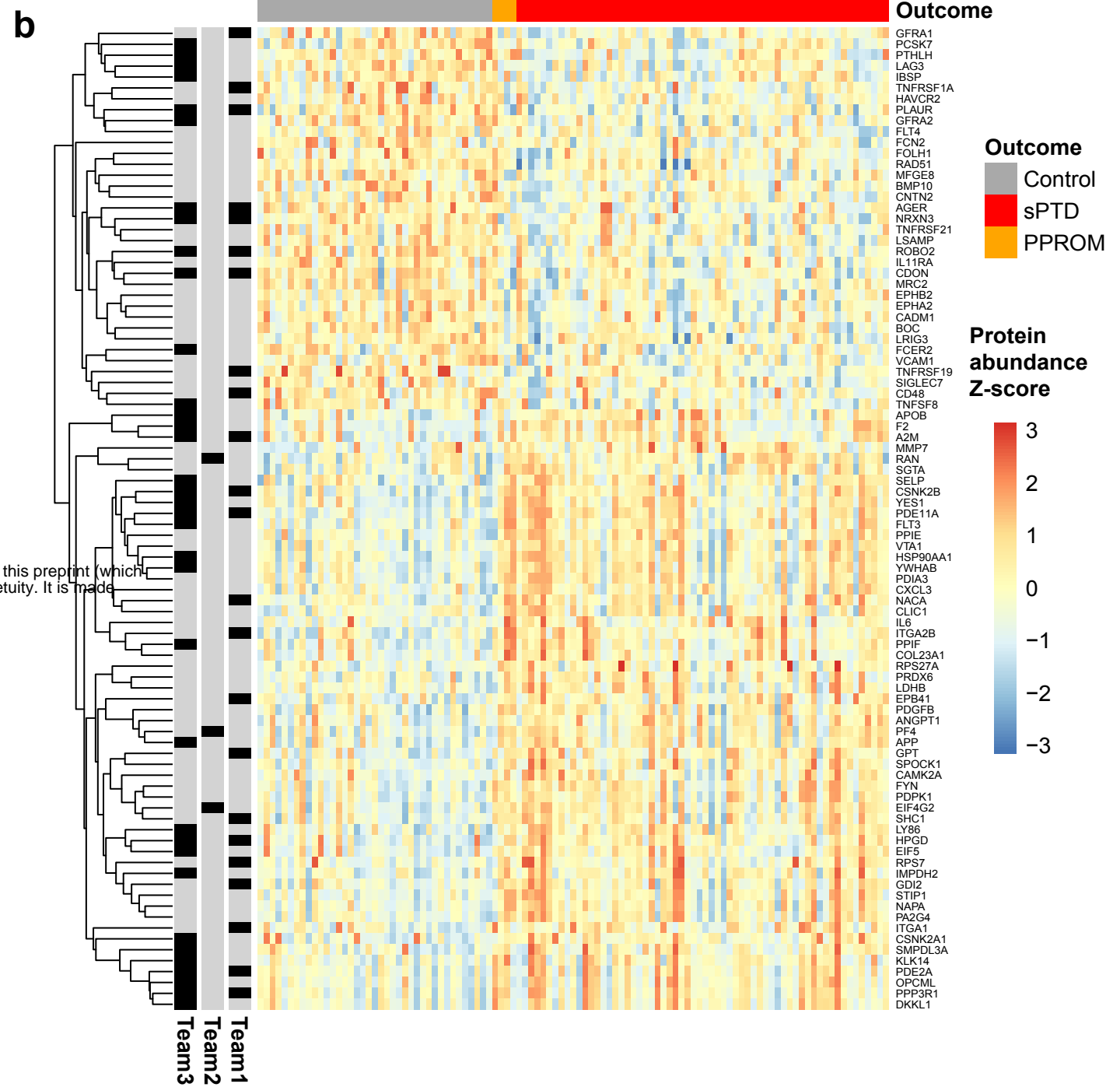
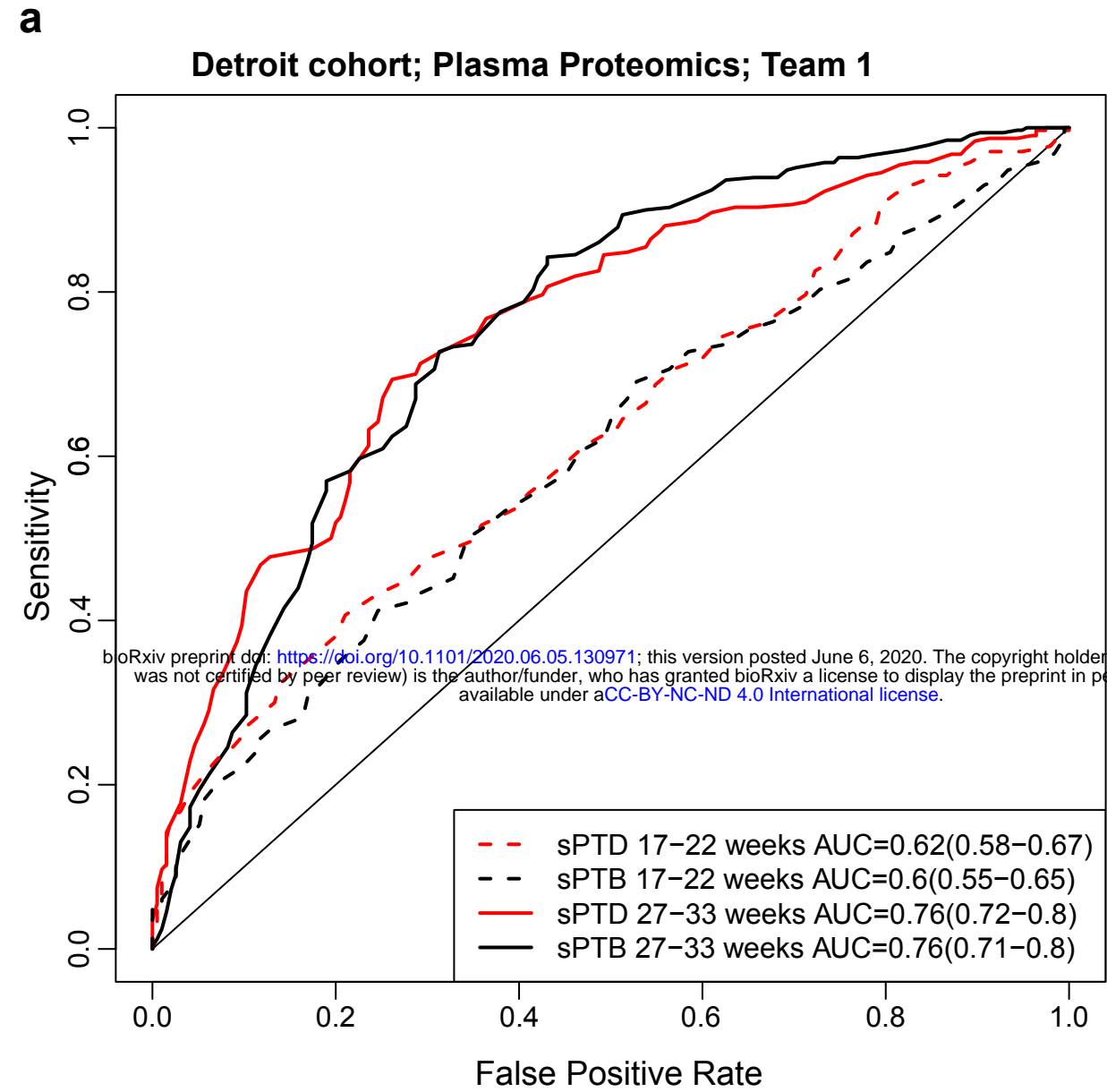
	Δ [observed – predicted delivery date (weeks)] (%)				
	< -2	-1 to -2	± 1	1 to 2	> 2
RNA T1 n= 3	0	33	0	0	67
RNA T2 n= 11	9	0	18	18	55
RNA T3 n= 11	18	9	45	9	18
LMP&Ultrasound n= 11	0	45	55	0	0



	Δ [observed – predicted delivery date (weeks)] (%)				
	< -2	-1 to -2	± 1	1 to 2	> 2
RNA T1 n= 17	29	6	18	0	47
RNA T2 n= 68	43	13	7	4	32
RNA T3 n= 70	67	6	16	6	6
LMP&Ultrasound n= 70	100	0	0	0	0

a**b****c**





sPTD prediction at 27–33 weeks

