Multi-omic signals associated with maternal epidemiological factors contributing to preterm birth

June 8 2023, by Thamarasee Jeewandara

A pregnant person checks the pocket watch that tracks the chronicity of their pregnancy while many factors that impact pregnancy progression surround them.

Credit: *Science Advances* (2023). DOI: 10.1126/sciadv.ade7692

In low- and middle-income countries, and in general, the prospect of preterm birth is the leading cause of mortality in children under the age of five, although comprehensive studies are hindered to the phenomenon due to its complex etiology. Researchers have previously characterized the epidemiological associations between preterm birth and maternal characteristics.

In a new report now featured on the cover page of *Science Advances*, Camilo A. Espinosa and a team of scientists at Stanford Medicine, Stanford University, the Aga Khan University, Pakistan, and several other international research institutes used multi-omics profiling to investigate characteristic biological signatures.

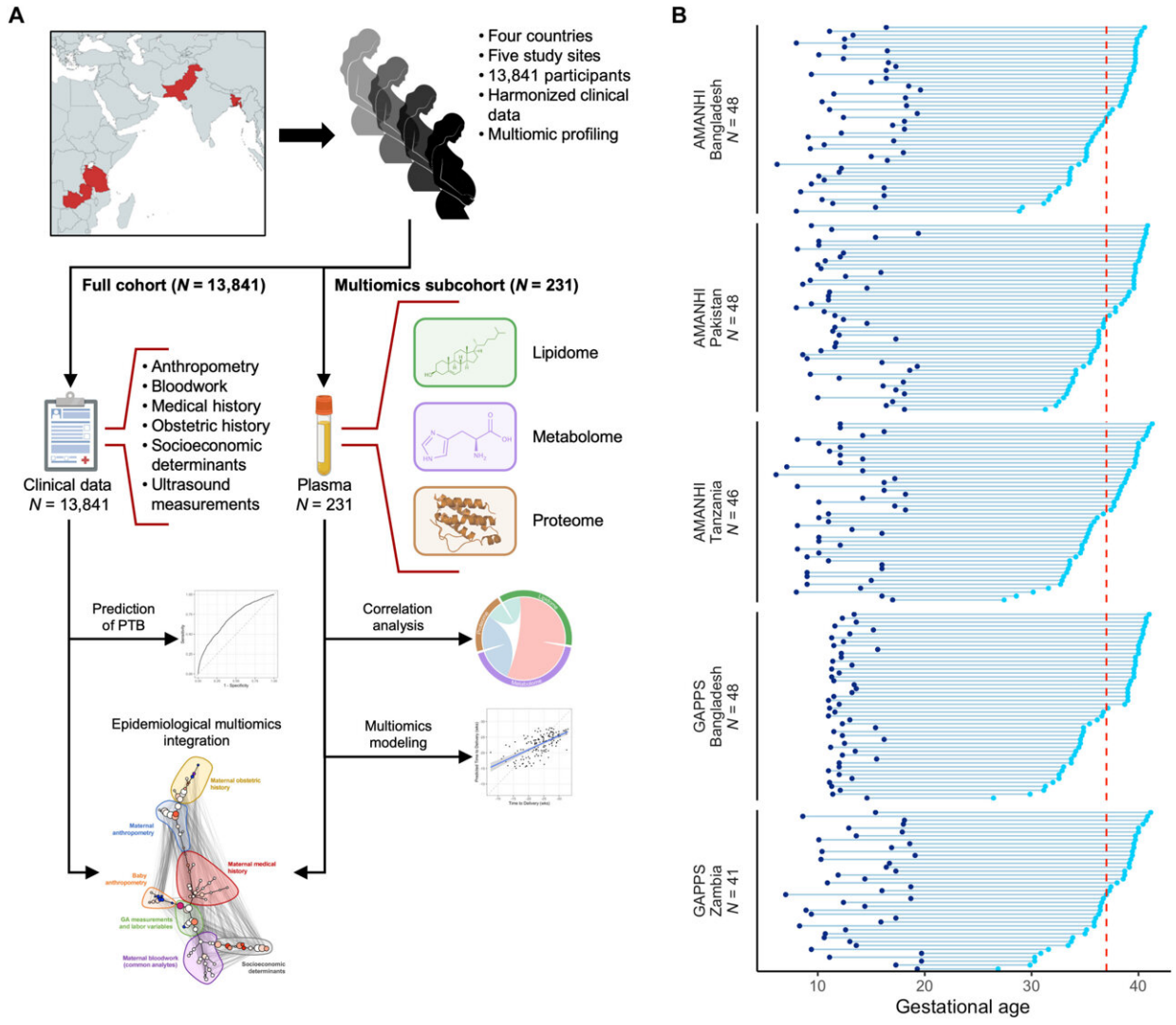
The researchers collected maternal covariates during pregnancy, from 13,481 pregnant women and analyzed plasma samples from patient subsets to generate proteomic, metabolomic and lipidomic (i.e., multi-omic) datasets. The scientists used machine learning models with robust performance to predict preterm birth, time-of-delivery, maternal age and the body-mass-index among a range of parameters.

The maternal age negatively correlated with collagen, gravidity with endothelial nitrous oxide, inflammatory chemokine, and the body-mass index was influenced by leptin and some structural proteins. The outcomes provided a full-view of epidemiological factors underlying preterm birth, while identifying biological signatures of related clinical covariates.

Maternal factors underlying preterm-birth

Birth prior to 37 weeks of gestation is deemed preterm and is a leading cause of mortality in children under the age of five around the globe. The occurrence has high prevalence in low and middle-income countries. Child-birth at preterm can increase the risk of life-threatening short-term complications, including long-term neurological, cardiovascular and metabolic morbidities.

This phenomenon affects the life of the mother, child and family, posing a high burden to global public health. The etiology of preterm birth is complex and carries a diverse range of effects to varying degrees.



Study overview. (A) Maternal clinical data and plasma samples were collected from a cohort of 13,841 pregnant women across five sites in four low- and middle-income countries (LMICs). Plasma samples taken during early and mid pregnancy from a subcohort of 231 of these women were further analyzed to generate targeted lipidomic, untargeted metabolomic, and targeted proteomic datasets. Clinical data from the full cohort were used for the prediction of preterm birth (PTB). Multiomic profiling data from the multiomics subcohort were used for interomic correlation analysis and for the prediction of maternal clinical covariates. Clinical data from the full cohort and multiomic data from the multiomics subcohort were used for the epidemiological multiomic integration. (B) Raster plot depicting the gestational age (GA) at sampling for each woman in the multiomics subcohort stratified by site of origin, where each line represents an individual woman and the dark blue circles and light blue circles represent sampling dates and delivery dates, respectively. The dashed red line at 37 weeks of GA indicates the boundary between preterm (PTB) and term births. Credit: *Science Advances* (2023). DOI: 10.1126/sciadv.ade7692

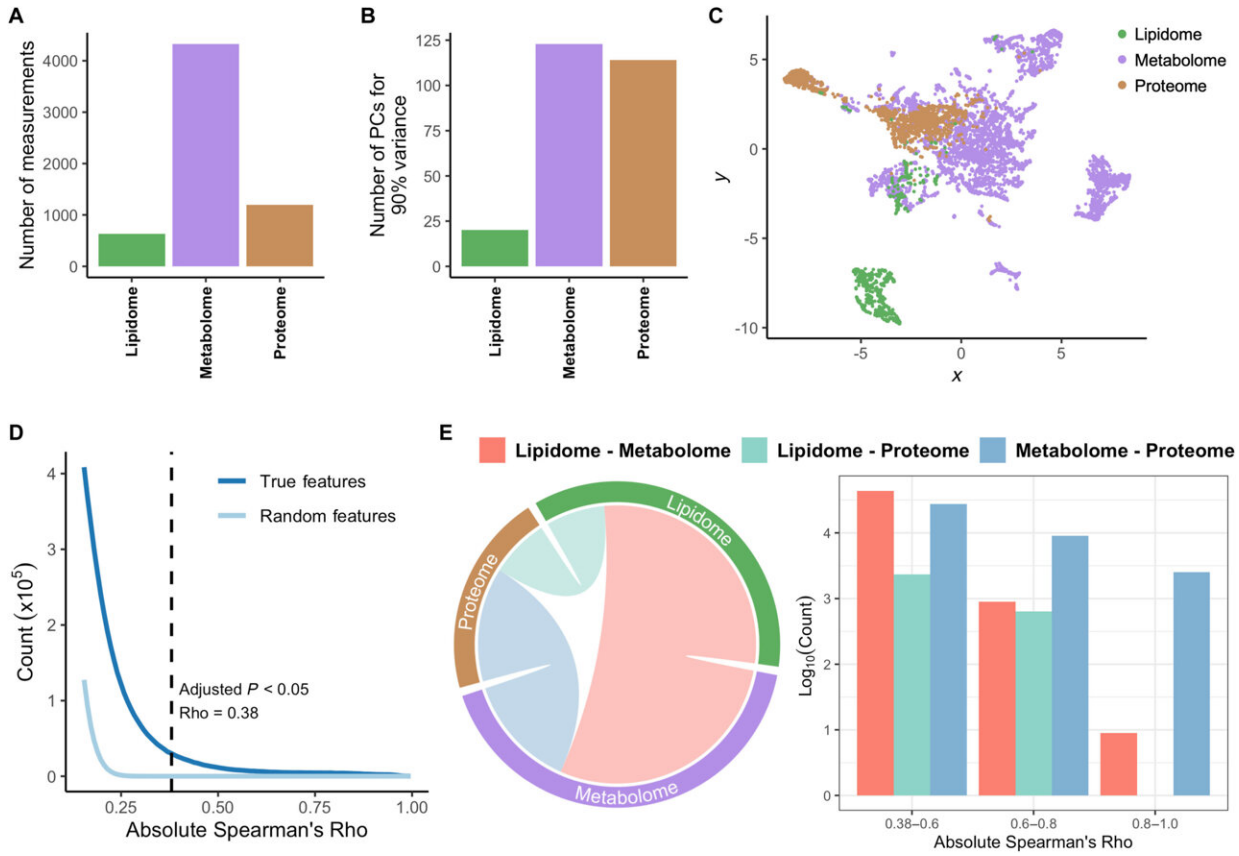
While researchers have described the epidemiological associations between preterm birth and maternal clinical history, the outcomes are limited by a lack of biological data. Socioeconomic determinants of health are key players that affect the outcome of gestation, while environmental exposures can influence the health of a pregnancy; further complicating the risks of preterm birth.

During this work, the team used a multinational cohort of pregnant women data to understand epidemiologically derived factors associated with the phenomenon. The outcomes allowed Espinosa et al. to propose innovative concepts to study key connections between population-level risk factors and the underlying pathological mechanisms of disease.

Participant data and study design

The research team collected maternal covariates and plasma samples of patients across different international communities. Of the pregnancies on record, a few resulted in preterm births. The researchers obtained details, demographics and characteristics of the pregnancy and measured the maternal proteome, metabolome and lipidome during early and mid-pregnancy.

They visualized the combined multi-omics datasets and examined the distribution of significant correlations within intra- and inter-omics during pregnancy to reveal the rigorous orchestration of biological processes during gestation. The resultant pregnancy interactome outlined a strict regulation of biological systems and uncovered their crosstalk during gestation.



Multiomic characterization of early and mid pregnancy. Plasma samples taken during early and mid pregnancy from a subcohort of 231 women were analyzed to generate targeted lipidomic, untargeted metabolomic, and targeted proteomic datasets. (A) Quantification of the number of measurements (features) of each different omic analyzed. (B) Estimation of the modularity—i.e., the degree of internal correlation between features in a given omic dataset—using the number of principal components needed to explain 90% of the variance. (C) A two-dimensional representation of the multiomic correlation space was generated by first calculating the correlation matrix of the feature space and then using the UMAP dimensionality reduction algorithm for visualization, where green, purple, and brown circles represent lipidomic, metabolomic, and proteomic features, respectively. (D and E) Interomic and intraomic Spearman correlations were quantified and assessed for significance using a cutoff value of Bonferroni-adjusted Spearman P 0.38. (D) Distribution of all correlations by absolute strength of association for the multiomic dataset (dark blue) and a random dataset (light blue), where each simulated feature was generated through bootstrapping of the true feature distribution. (E) Visualization of significant interomic correlations. Left: Chord diagram showing the relative distribution of significant interomic correlations, where the outer ring depicts each individual omic and links correspond to interomic correlations, with colors assigned as shown in the legend. The size of the links is proportional to the total number of significant interactions normalized to the total number of possible correlations between each omic pair. Right: Quantification of the number of significant weak (0.38 to 0.6), moderate (0.6 to 0.8), and strong (0.8 to 1.0) absolute correlations between each omic pair. Credit: *Science Advances* (2023). DOI: 10.1126/sciadv.ade7692

This work also highlighted epidemiological characteristics underlying preterm birth, which they cross-validated with a multivariable approach by using a gradient-boosted tree model. The performance of the model showed the existence of a potential risk in the preterm birth population.

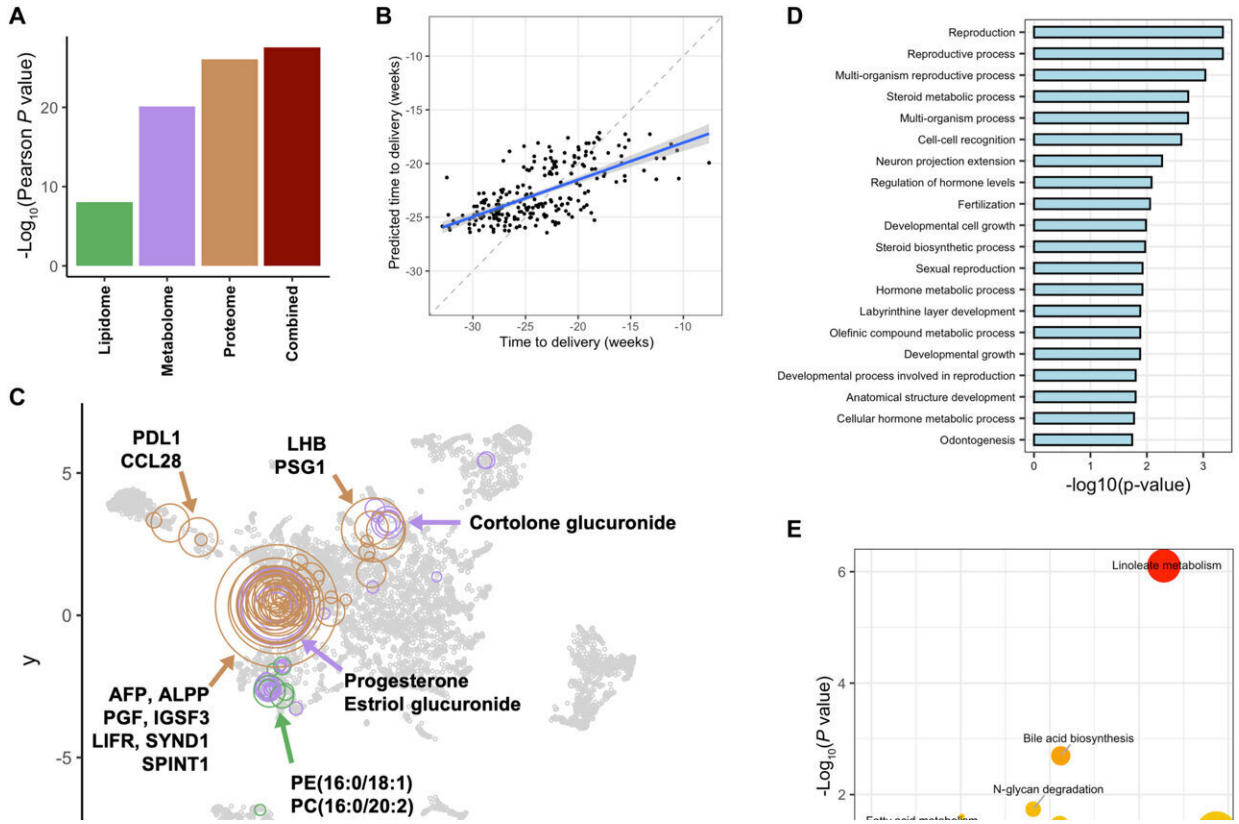
They explored several clinical categories to predict preterm birth based

on epidemiological features. The analysis described multivariate risk factors associated with the phenomenon to reveal predictive measurements.

Multi-omics modeling to predict premature delivery

Espinosa et al. conducted multi-omics modeling of the maternal proteome, metabolome and lipidome to predict the time of delivery. The process of pregnancy and the onset of labor are highly coordinated biological events that can be measured via multi-omic profiling. To determine the presence of biological signatures representing the time-frame of delivery, the scientists built a cross-validated XGBoost model.

This model integrated the maternal proteome, metabolome and lipidome at the early stages and mid-pregnancy, while strongly predicting the time of delivery. Of the omics datasets, the proteome matched the highest predictive power of the integrated models, closely followed by the metabolome. The multi-omics profiling robustly captured biological processes underlying early and mid-pregnancy to predict time-of-delivery and show accelerated gestational clocks in women predicted to deliver prematurely.



Multiomic modeling of the maternal proteome, metabolome, and lipidome predicts time-to-delivery. A cross-validated gradient-boosted tree (XGBoost) model for the prediction of time from sampling to delivery was trained on each individual omic as well as the combined multiomic dataset. (A) Comparison of the cross-validated performance of each model for the prediction of time-to-delivery. (B) Time-to-delivery predictions of the combined multiomic model (Pearson's r = 0.65, 95% CI: 0.57 to 0.72, P = 2.6 × 10−28, RMSE = 3.7 weeks, MAE = 3.0 weeks, N = 226). (C) Two-dimensional UMAP visualization of the multiomic features significantly correlated with time-to-delivery. Each circle represents a feature, with sizes proportional to Spearman correlation with time-to-delivery, and colors representing modality. (D) Gene Ontology (GO) overrepresentation analysis performed on the plasma proteomic features significantly correlated with time-to-delivery as assessed using Fisher's exact test. (E) Metabolic pathway enrichment analysis performed on the metabolomic features significantly correlated with time-to-delivery. (F) Participants were randomly split into a training set (N = 159, 70%) and a test set (N = 67, 30%) to build a minimal XGBoost model for the prediction of time-to-delivery. Left:

Cross-validated model predictions in the training set (Pearson's r = 0.64, 95% CI: 0.54 to 0.73, P = 7.4 × 10−20, RMSE = 3.9 weeks, MAE = 3.1 weeks, N = 159). Right: Model predictions in the test set (Pearson's r = 0.68, 95% CI: 0.53 to 0.79, P = 1.7 × 10−10, RMSE = 3.4 weeks, MAE = 2.8 weeks, N = 67). (G) A cross-validated XGBoost model for the prediction of GA at sampling was trained on the combined multiomic dataset. Boxplot depicts the discrepancy between predicted GA and ultrasound GA stratified by PTB status. The blue lines and blue shadows in scatterplots represent the regression line and 95% CI of the predicted values versus the ground truth. Credit: *Science Advances* (2023). DOI: 10.1126/sciadv.ade7692

Age-independent markers of pregnancy memory

The details of a mother's age and obstetric history can impact her risk of preterm birth, although their independent risk contributions remain to be revealed. The multi-omics subset provided predictable variables where maternal age and gravidity correlated significantly, and type 9 collagen showed the strongest association with either covariate.

The scientists noted proteome features of a chemokine ligand (Chemokine ligand 13 CXCL13) and the endothelial nitric oxide synthase to have significant correlations with gravidity, although these biomolecular signatures were not associated with the age of the mother.

This observation suggests an age-independent effect on the maternal proteome, associated with maternal history. The findings highlighted the presence of previously unknown age-independent effects from prior pregnancies that can affect the maternal biology of future pregnancies.

Predictive modeling of maternal body mass index

Maternal body mass index is corelated to preterm birth and results in an

increased risk at both ends of the body-mass index spectrum. To examine the impact of this variable on maternal biology, the team-built, cross-validated model with integrated multi-omics data, further validated the strong signature of the body-mass index on the maternal proteome.

Significant features that correlated with this parameter further consisted of metabolism and biomolecule-storage-associated proteins and a cluster of intercorrelated lipid species. The outcomes provided an overview of the epidemiology-associated variables of preterm birth relative to their biological covariates.

Outlook

In this way, Camilo A. Espinosa and colleagues used a large, multinational cohort with widespread data collection to describe epidemiological factors associated with preterm birth. The integrated multi-omics model revealed key biological signatures underlying maternal biological adaptations to pregnancy and preterm birth.

Using a series of bioinformatics models and machine learning algorithms, the team provided fundamental insights to develop biological and socio-economic interventions to attenuate preterm birth across multiple populations.

More information: Camilo A. Espinosa et al, Multiomic signals associated with maternal epidemiological factors contributing to preterm birth in low- and middle-income countries, *Science Advances* (2023). DOI: 10.1126/sciadv.ade7692