Large-scale proteomics in the first trimester of pregnancy predict psychopathology and temperament in preschool children: an exploratory study

Jessica L. Buthmann,1† Jonas G. Miller,2† Nima Aghaeepour,3,4,5 Lucy S. King,1 David K. Stevenson,3 Gary M. Shaw,3 Ronald J. Wong,3 and Ian H. Gotlib1

1Department of Psychology, Stanford University, Stanford, CA, USA; 2Department of Psychological Sciences, University of Connecticut, Storrs, CT, USA; 3Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA; 4Department of Anesthesiology, Perioperative, and Pain Medicine, Stanford University School of Medicine, Stanford, CA, USA; 5Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, USA

Background: Understanding the prenatal origins of children's psychopathology is a fundamental goal in developmental and clinical science. Recent research suggests that inflammation during pregnancy can trigger a cascade of fetal programming changes that contribute to vulnerability for the emergence of psychopathology. Most studies, however, have focused on a handful of proinflammatory cytokines and have not explored a range of prenatal biological pathways that may be involved in increasing postnatal risk for emotional and behavioral difficulties.

Methods: Using extreme gradient boosted machine learning models, we explored large-scale proteomics, considering over 1,000 proteins from first trimester blood samples, to predict behavior in early childhood. Mothers reported on their 3- to 5-year-old children's (N = 89, 51% female) temperament (Child Behavior Questionnaire) and psychopathology (Child Behavior Checklist).

Results: We found that machine learning models of prenatal proteomics predict 5%–10% of the variance in children's sadness, perceptual sensitivity, attention problems, and emotional reactivity. Enrichment analyses identified immune function, nervous system development, and cell signaling pathways as being particularly important in predicting children's outcomes.

Conclusions: Our findings, though exploratory, suggest processes in early pregnancy that are related to functioning in early childhood. Predictive features included far more proteins than have been considered in prior work. Specifically, proteins implicated in inflammation, in the development of the central nervous system, and in key cell-signaling pathways were enriched in relation to child temperament and psychopathology measures.

Keywords: Biomarkers; child development; machine learning; prenatal; prediction.

Pregnancy is a critical developmental period when 'programming' of offspring psychobiology is hypothesized to occur (Wadhwa, Buss, Entringer, & Swanson, 2009). According to the developmental origins of health and diseases hypothesis, the biological signals received by the fetus in utero may disrupt development in a manner that increases risk for problematic functioning postnatally (Hantsoo, Kornfield, Anguera, & Epperson, 2019). Given epidemiological data indicating that infection during pregnancy is associated with maladaptive sequelae in offspring (Reisinger et al., 2015), scientists have identified prenatal inflammation and glucocorticoids as possible contributors to child mental health difficulties (Hantsoo et al., 2019). While important, most prior studies have considered only a small subset of cytokines measured during pregnancy in relation to offspring psychological functioning. Furthermore, relative to inflammation and glucocorticoids, we know far less about the possible roles of other prenatal biological pathways in the context of child mental health. Here, we used large-scale proteomics profiling of plasma obtained in the first trimester of pregnancy to investigate whether prenatal proteins and pathways can predict temperament and symptoms of psychopathology during the preschool-age period.

Investigators examining prenatal inflammation have documented associations between increased maternal proinflammatory cytokine levels during pregnancy and child difficulties (Hantsoo et al., 2019; Monk, Lugo-Candelas, & Trumpff, 2019). For example, higher levels of maternal inflammation (i.e., interleukin-6 [IL-6], tumor necrosis factor alpha [TNF-α], and monocyte chemoattractant protein-1 [MCP-1]) during the third trimester have been found to be associated with greater negative affect in 6-month-old infants and with higher levels of attention-deficit hyperactivity disorder symptoms at 4–6 years of age (Gustafsson et al., 2018, 2020). Higher levels of maternal IL-6 during pregnancy have been associated with poorer executive function in 2-year-old children (Rudolph et al., 2018), and levels of C-reactive protein and
glycoprotein acetyl in maternal plasma across pregnancy have been associated with increased risk for neurodevelopmental delay in 7- to 10-year-old children (Girchenko et al., 2020).

These association studies implicate prenatal inflammation as a potential biomarker of risk for difficulties in offspring that might inform both the early identification of vulnerable children and the design of perinatal interventions to prevent the development of psychopathology. This would require, however, that prenatal biology can inform an approach that allows intervention and treatment at the individual level (Bzdok & Meyer-Lindenberg, 2018). In this context, a growing body of research in prenatal biology is using machine learning approaches to predict individuals’ outcomes. Machine learning allows researchers to use an approach that is data-driven, rather than hypothesis-driven, can handle a large number of variables, and can lead to the identification of novel biomarkers of risk. Much of the work in this area has focused on pregnancy outcomes, such as preterm birth and preeclampsia (Aghaeepour et al., 2018; D'Silva, Hyett, & Coorssen, 2018; Tarca et al., 2019). For example, D'Silva et al. found that women who experienced a spontaneous preterm birth had first trimester levels of proteins related to various pathways (regulation of the complement cascade and coagulation pathways, immune modulation, metabolic processes, cell signaling) that were up- or downregulated compared with women in a full-term control group (D'Silva et al., 2018). The proteome of women who developed preeclampsia has also been found to differ from controls, with protein levels from the second trimester predicting more than two-thirds of preeclampsia cases (Tarca et al., 2019). To date, however, investigators have not conducted large-scale proteomics profiling during pregnancy to predict children’s subsequent symptoms of psychopathology or their temperament, which reflects emergent risk for psychopathology. Data from such studies have the potential to identify novel biomarkers of risk and fetal developmental processes that, if disrupted, may have long-term effects on well-being.

In addition, most previous studies of the prenatal origins of offspring psychological functioning have focused on specific proinflammatory cytokines, such as IL-6, C-reactive protein, and TNF-α (Gustafsson et al., 2020; Rudolph et al., 2018; Spann, Monk, Scheinost, & Peterson, 2018). It is important to recognize that prenatal environments that present risk for maladaptive outcomes in offspring may include a broader number of inflammatory components as well as an altered enrichment of non-inflammatory proteins and pathways (Gyllenhammer et al., 2022; Györfi et al., 2016). In this context, therefore, data-driven approaches, such as machine learning, are useful in integrating a more comprehensive spectrum of prenatal proteins than have been examined in previous studies. Moreover, adopting a data-driven approach can enable the discovery of otherwise undetected novel biomarkers of risk when using a hypothesis-driven approach that focuses on a few select cytokines and can guide the generation of hypotheses in future work.

In the present study, we assessed prenatal biology using large-scale proteomics analysis of blood samples collected from women during their first trimester of pregnancy. We used a machine learning approach to explore whether comprehensive prenatal proteomics data are associated with subsequent symptoms of psychopathology and temperament in the women’s preschool-age children. We conducted follow-up analyses to identify specific proteins that are related to child outcomes. Finally, we explored enrichment of biological pathways during pregnancy that are overrepresented in the prediction of child outcomes.

**Methods**

**Participants**

Pregnant women who were at least 18 years old and in their first trimester of a singleton pregnancy participated in a study sponsored by the March of Dimes Prematurity Research Center at Stanford University to examine factors associated with birth outcomes. Women’s blood was drawn into EDTA tubes through antecubital venipuncture during the first trimester (7–14 weeks). English-speaking mothers who experienced normal births (i.e., >36 weeks' gestation, without complications in labor/delivery) were invited to participate in an in-person assessment when their child was 3–5 years of age. This follow-up assessment included 97 mothers and their children. Mothers reported on their children’s psychopathology symptoms and temperament. The final sample for analyses included 89 mothers and children with data from both the prenatal and the preschool-age assessment. A majority of mothers reported having a graduate degree, 69% were White, and 91% were married; 51% of the children were assigned female at birth (see Table 1 for detailed demographic characteristics of the sample). Children with congenital abnormalities, developmental disabilities, and severe medical conditions were excluded from the study. Compared to women who did not participate in the follow-up assessment with their children, women who participated in the follow-up assessment were more likely to be older at baseline (Mean full = 30.15 years (SD = 5.61), Mollow-up = 31.84 (SD = 4.37), p = .001), more likely to have older partners at baseline (Meanfull = 31.70 years (SD = 7.18), Mollow-up = 33.61 (SD = 4.90), p = .002), less likely to be of Hispanic ethnicity (full sample 71.2% not Hispanic, follow-up sample 86% not Hispanic, p = .014), and more likely to work outside the home (full sample 55% worked outside the home, follow-up sample 74% worked outside the home, p=.001). Mothers signed consent forms and were compensated for their time. This study protocol was approved by the Stanford University Institutional Review Board (Protocol #21956).

**Proteomics assay**

Immediately after collection, maternal blood was placed on ice, centrifuged within 60 min, and plasma was stored at −80°C until analysis. Briefly, samples were first analyzed at the Humane Immune Monitoring Center at Stanford University
The prenatal proteome predicts child psychopathology

**Child psychopathology**

Mothers completed the Child Behavior Checklist, Preschool Form to assess symptoms of psychopathology (CBCL; Achenbach, 1999). This 100-item questionnaire asks parents to rate how true statements about their child’s functioning are on a scale from 0 (not true) to 2 (very or often true). This measure is used for children aged 1.5–5 years of age. These questions comprise the following seven subscales, with an example of an item making up the scale in parentheses: emotional reactivity (rapid shifts between sadness and excitement); anxious/depressed (too fearful or anxious); somatic complaints (aches or pains); withdrawn (does not want to go out of home); sleep problems (talks or cries out in sleep); attention problems (quickly shifts from one activity to another); and aggressive behavior (cruel to animals). Means, standard deviations, and internal reliability (Cronbach’s alpha) for all subscales are presented in Table 2.

**Predictive modeling of child psychopathology and temperament**

We used extreme gradient boosted models to predict each child outcome measure from 1,196 proteins derived from prenatal plasma samples. These models are robust to collinearity and outliers in independent variables; there were no outliers (±3 SDs) in any of the dependent variables. No data were missing in the CBCL data. In the CBQ data, 59% had complete data, 20% were missing one item, 5% were missing two items, 7% outliers in independent variables; there were no outliers (±3 SDs) in any of the dependent variables. No data were missing

© 2024 Association for Child and Adolescent Mental Health.
were missing three items, and 9% were missing four or more items. Missing data were imputed with the mean value for that item. All proteins were included as predictors in the models. For each outcome, the gradient boosting algorithm built a prediction model sequentially using additive weak learners (Friedman, 2001). Data were randomly split in half 20 times; each time, one half of the data were used for training the algorithm and the other half for testing the algorithm. Results were aggregated and reported only from the test-set. This analysis was conducted using the xgboost package (Chen & He, 2023) in R (R Core Team, 2021) with default loss function and calibration parameters. We evaluated prediction performance using Spearman correlations between predicted and observed outcome values. The source code is provided in the Supporting Information.

Protein enrichment analysis

As described in Aghaeepour et al. (2018), we used gene ontology analysis to infer enrichment of biological pathways (i.e. an overrepresentation of proteins implicated in certain biological processes) among proteins included in the predictive models. Enrichment analysis was performed using WikiPathways with all measured proteins used as the background (Slenter et al., 2018; Yu, Wang, Han, & He, 2012). We used a $p = .05$, without correction for multiple comparisons, as a significance threshold for these exploratory analyses. We explored correlations between individual proteins and behavioral outcomes to identify whether proteins were up- or downregulated and to examine additional biological pathways (i.e. other than the top pathway identified by WikiPathways) with which the proteins may be associated.

Results

Spearman correlations between model-predicted and observed outcome values indicated that four behavioral measures were predicted more accurately than would be expected by chance ($p < .05$): the CBQ-SF sadness ($r = .32, p = .002$) and perceptual sensitivity ($r = .23, p = .033$) subscales, and the CBCL attention problems ($r = .27, p = .011$) and emotional reactivity ($r = .25, p = .017$) subscales. The prediction performance for these measures is presented as scatterplots in Figure 1 and shows the relations between model-predicted and observed outcomes. The models predicted 5%-10% of the variance in these four outcomes. Thus, prenatal proteomics accounted for a modest amount of variance in preschool-age children's sadness, perceptual sensitivity, emotional reactivity, and attention problems. The prediction performance for all behavioral outcomes is presented in Table S1. As a sensitivity analysis, we randomly permuted the assignment of outcomes among participants and re-ran the extreme gradient boosted machine learning models. Only one of the permuted outcome measures (CBQ Fearfulness) was significantly predicted by the model (see Table S2). In addition, models based on demographic and delivery information (i.e. variables listed in Table 1, maternal medical history, delivery complications, birthweight) did not accurately predict any behavioral outcome measure (see Table S3).

Enrichment analyses indicated further that several pathways had a significant number of proteins that were up- or downregulated in relation to sadness, attention problems, and emotional reactivity, but not perceptual sensitivity. Table 3 presents each pathway that was overrepresented in relation to each behavioral measure. Biological pathways related to inflammation, nervous system development, and cellular signaling were associated with sadness (all $p < .043$), and pathways related to nervous system development and cell signaling were associated with attention problems (all $p < .027$) and emotional reactivity (all $p < .049$).

Many of the proteins that were significantly correlated with the behavioral measures of child temperament and psychopathology have been implicated in a range of biological processes, including inflammation and immune function, neural development, glucocorticoid regulation, and cell signaling pathways. Children's sadness was significantly associated with 56 proteins, perceptual sensitivity with 157 proteins, attention problems with 315 proteins, and emotional reactivity with 186 proteins (see Tables S4–S7 for full list by behavioral outcome). Whereas there were an equal number of up- and downregulated proteins associated with sadness, 92% of the predictors of perceptual sensitivity were upregulated. Conversely, for models predicting symptoms of psychopathology, 95% of the proteins associated with attention problems and 75% of the proteins associated with emotional reactivity were downregulated. Of the significant proteins, 131 were associated with two or more of the four behavioral scales shown in Table S8; the functions of these proteins are diverse, with 22 involved in immune system function, 22 involved in protein metabolism, 17 involved in apoptosis, 15 involved in nervous system development, 15 involved in protein transport, and 10 involved in glucocorticoid regulation.

Most of the individual proteins that were significantly associated with children's temperament did not survive correction for multiple comparisons; however, we explore them here to generate hypotheses about the mechanisms that might contribute to child outcomes. Many proteins that have been implicated in the upregulation of proinflammatory responses were associated with sadness, although they varied with respect to whether they were positively or negatively associated with sadness. For example, IL-1 receptor-like 2 (IL1RL2) and IL-32 were positively and negatively associated with sadness, respectively (IL1RL2: $r = .24, p = .01$; IL-32 $r = -.23, p = .04$). Children's perceptual sensitivity was the other dimension of temperament that was significantly associated with our proteomics model before correction for multiple comparisons. Perceptual sensitivity was most strongly and positively associated with several individual proteins that are involved in inflammation and immune function, such as IL-8, P-selecting (SELP), platelet and...
endothelial cell adhesion molecule 1 (PECAM1), CD244 molecule (CD244), and TNF superfamily member 14 (TNFSF14; see Table S7). It is noteworthy that the majority of the proteins involved in inflammation and immune function that were related to sadness and perceptual sensitivity have not been considered in previous studies of child psychosocial development.

With respect to individual proteins that were associated with child symptoms of psychopathology, many were negatively related to both emotional reactivity and attentional problems (see Tables S4 and S5). For example, IL-1 receptor antagonist type 2 (IL-1RT2) was negatively associated with emotional reactivity and attentional problems. It is noteworthy, however, that compared to the profiles of individual proteins that were related to child sadness and perceptual sensitivity, fewer proteins implicated in inflammation and immune function were significantly related to emotional reactivity and attentional problems. Emotional reactivity was related to a number of nervous system development-related proteins, including Cadherin 1 (CDH1), BOC Cell Adhesion Associated (BOC), and Fibroblast Growth Factor 2 (FGF2), as well as cell signaling-related proteins, including Growth Differentiation Factor 2 (GDF2), Erb-B2 Receptor Tyrosine Kinase 2 (ERBB2), and Latent Transforming Growth Factor Beta Binding Protein 2 (LTPB2). Attentional problems, too, were associated with a number of nervous system development-related proteins, including Protagenin (PRTG) and Fibronectin Leucine-Rich Transmembrane Protein 2 (FLRT2), cell signaling-related proteins including Cellular Communication Network Factor 4 (WISP-1) and Cadherin 6 (CDH6), and immune-related proteins including TNF Superfamily Member 13 (TNFSF13) and Inducible T-Cell Costimulator Ligand (ICOSLG).

**Figure 1** Correlations of observed values and proteomics-based predicted values. The machine learning algorithm successfully predicted four child behavioral outcomes from first trimester maternal proteome data. In panels A–D, the x-axis presents the observed (parent-report) values and the y-axis presents the proteomics-based predicted values. $p_v = p$-value for significance testing. Panels A–B show the Spearman correlations for Child Behavior Checklist psychopathy measures of Attentional Problems and Emotional Reactivity, respectively. Panels C–D show the Spearman correlations for Child Behavior Questionnaire temperament measures of sadness and perceptual sensitivity, respectively.

Discussion
In this exploratory study, we used extreme gradient boosted machine learning models of over 1,000 proteins from first trimester maternal plasma samples to predict parent-reported measures of behaviors in their 3- to 5-year-old children. The algorithm successfully predicted scores on four behavioral measures: sadness and perceptual sensitivity (temperament), and attention problems and emotional reactivity (psychopathology risk). Enrichment analyses and correlations with individual proteins suggested that pathways involved in neural development, inflammation, glucocorticoid regulation, and cell signaling are important prenatal processes...
for the development of child temperament and emerging psychopathology in early childhood.

We found that prenatal proteomic models predict 5%–10% of the variance in specific aspects of preschool-age children’s temperament and symptoms of psychopathology. Although exploratory, these results suggest that as early as the first trimester of pregnancy, large-scale proteomics can identify biomarkers that forecast offspring psychosocial development. Although the proteomic

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
<th>GeneRatio</th>
<th>BgRatio</th>
<th>p-value</th>
<th>q-value</th>
<th>geneID</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP2858</td>
<td>Ectoderm differentiation</td>
<td>7/99</td>
<td>16/595</td>
<td>0.0090</td>
<td>0.4438</td>
<td>91653/4772/8675/50937/10253/6781/1004</td>
</tr>
<tr>
<td>WP2018</td>
<td>RANKL/RANK signaling pathway</td>
<td>5/99</td>
<td>10/595</td>
<td>0.0145</td>
<td>0.4438</td>
<td>3383/6714/8792/4772/7186</td>
</tr>
<tr>
<td>WP2848</td>
<td>Pluripotent stem cell differentiation pathway</td>
<td>5/99</td>
<td>10/595</td>
<td>0.0145</td>
<td>0.4438</td>
<td>4851/7018/2247/3082/2056</td>
</tr>
<tr>
<td>WP4806</td>
<td>EGFR tyrosine kinase inhibitor resistance</td>
<td>7/99</td>
<td>19/595</td>
<td>0.0257</td>
<td>0.4521</td>
<td>2621/6714/1956/2247/6198/2064/3082</td>
</tr>
<tr>
<td>WP4540</td>
<td>Hippo signaling regulation pathways</td>
<td>7/99</td>
<td>20/595</td>
<td>0.0342</td>
<td>0.4521</td>
<td>999/1956/1015/1001/4915/1004/1969</td>
</tr>
<tr>
<td>WP2064</td>
<td>Neural crest differentiation</td>
<td>5/99</td>
<td>12/595</td>
<td>0.0346</td>
<td>0.4521</td>
<td>999/4851/329/2064</td>
</tr>
<tr>
<td>WP3303</td>
<td>RAC1/PAK1/p38/MMP2 pathway</td>
<td>5/99</td>
<td>12/595</td>
<td>0.0346</td>
<td>0.4521</td>
<td>7075/6714/1956/9068/2064</td>
</tr>
<tr>
<td>WP4541</td>
<td>Hippo-Merlin signaling dysregulation</td>
<td>1/11</td>
<td>30/595</td>
<td>0.0456</td>
<td>0.4974</td>
<td>999/1956/3694/1015/1001/4915/1004/1969/2321</td>
</tr>
<tr>
<td>WP2380</td>
<td>Brain-derived neurotrophic factor (BDNF) signaling pathway</td>
<td>5/99</td>
<td>13/595</td>
<td>0.0489</td>
<td>0.4974</td>
<td>6714/5777/6198/4772/5777/1004/1969/2321/28514</td>
</tr>
<tr>
<td>WP4262</td>
<td>Breast cancer pathway</td>
<td>5/99</td>
<td>14/595</td>
<td>0.0662</td>
<td>0.5167</td>
<td>4851/1956/2247/6198/2064</td>
</tr>
<tr>
<td>WP4155</td>
<td>Endometrial cancer</td>
<td>4/99</td>
<td>10/595</td>
<td>0.0677</td>
<td>0.5167</td>
<td>999/1956/2247/2064</td>
</tr>
<tr>
<td>WP4674</td>
<td>Head and neck squamous cell carcinoma</td>
<td>4/99</td>
<td>10/595</td>
<td>0.0677</td>
<td>0.5167</td>
<td>4851/1956/329/2064</td>
</tr>
<tr>
<td>WP3888</td>
<td>VEGFA-VEGFR2 signaling pathway</td>
<td>13/99</td>
<td>53/595</td>
<td>0.0816</td>
<td>0.5332</td>
<td>3383/2022/3486/7148/6714/1666/4772/5777/10963/6198/3685/2817/1969</td>
</tr>
<tr>
<td>WP2037</td>
<td>Prolactin signaling pathway</td>
<td>4/99</td>
<td>11/595</td>
<td>0.0932</td>
<td>0.5332</td>
<td>6714/5777/6198/2064</td>
</tr>
<tr>
<td>WP4657</td>
<td>22q11.2 copy number variation syndrome</td>
<td>4/99</td>
<td>11/595</td>
<td>0.0932</td>
<td>0.5332</td>
<td>1956/3934/65078/91179</td>
</tr>
<tr>
<td>WP5180</td>
<td>DYSK1A</td>
<td>4/99</td>
<td>11/595</td>
<td>0.0932</td>
<td>0.5332</td>
<td>6714/5777/6198/2064</td>
</tr>
<tr>
<td>WP2064</td>
<td>Neural crest differentiation</td>
<td>1/11</td>
<td>30/595</td>
<td>0.0456</td>
<td>0.4974</td>
<td>999/1956/3694/1015/1001/4915/1004/1969/2321/28514</td>
</tr>
<tr>
<td>WP5144</td>
<td>NRP1-triggered signaling pathways in pancreatic cancer</td>
<td>8/142</td>
<td>12/595</td>
<td>0.0018</td>
<td>0.0838</td>
<td>4233/7049/1003/1956/9158/3082/7422/2321</td>
</tr>
<tr>
<td>WP4540</td>
<td>Hippo signaling regulation pathways</td>
<td>10/142</td>
<td>20/595</td>
<td>0.0088</td>
<td>0.2714</td>
<td>999/4233/1003/1956/1000/2263/4915/1004/1969/2321</td>
</tr>
<tr>
<td>WP4806</td>
<td>EGFR tyrosine kinase inhibitor resistance</td>
<td>9/142</td>
<td>19/595</td>
<td>0.0197</td>
<td>0.4561</td>
<td>4233/2621/558/1956/3082/2263/7422/2064/2065</td>
</tr>
<tr>
<td>WP4239</td>
<td>Epithelial to mesenchymal transition in colorectal cancer</td>
<td>6/142</td>
<td>11/595</td>
<td>0.0264</td>
<td>0.4886</td>
<td>999/4851/1282/1000/8828/28514</td>
</tr>
<tr>
<td>WP4674</td>
<td>Head and neck squamous cell carcinoma</td>
<td>5/142</td>
<td>10/595</td>
<td>0.0639</td>
<td>0.9871</td>
<td>4851/1956/2263/7422/2064</td>
</tr>
<tr>
<td>WP4541</td>
<td>Hippo-Merlin signaling dysregulation</td>
<td>11/142</td>
<td>30/595</td>
<td>0.0752</td>
<td>0.9939</td>
<td>999/4233/3689/1003/1956/1000/2263/4915/1004/1969/2321</td>
</tr>
<tr>
<td>WP5095</td>
<td>Overview of proinflammatory and profibrotic mediators</td>
<td>5/17</td>
<td>52/595</td>
<td>0.0116</td>
<td>0.2200</td>
<td>6362/2919/4283/6372/6372</td>
</tr>
<tr>
<td>WP5115</td>
<td>Network map of SARS-CoV-2 signaling pathway</td>
<td>5/17</td>
<td>53/595</td>
<td>0.0126</td>
<td>0.2200</td>
<td>2919/4283/6372</td>
</tr>
<tr>
<td>WP5039</td>
<td>SARS-CoV-2 innate immunity evasion and cell-specific immune response</td>
<td>3/17</td>
<td>22/595</td>
<td>0.0213</td>
<td>0.2200</td>
<td>2919/4283/6372</td>
</tr>
<tr>
<td>WP2064</td>
<td>Neural crest differentiation</td>
<td>2/17</td>
<td>12/595</td>
<td>0.0429</td>
<td>0.2200</td>
<td>3688/2247</td>
</tr>
<tr>
<td>WP2118</td>
<td>Arrhythmogenic right ventricular cardiomyopathy</td>
<td>2/17</td>
<td>12/595</td>
<td>0.0429</td>
<td>0.2200</td>
<td>3688/81</td>
</tr>
<tr>
<td>WP2572</td>
<td>Primary focal segmental glomerulosclerosis (FSGS)</td>
<td>2/17</td>
<td>12/595</td>
<td>0.0429</td>
<td>0.2200</td>
<td>3688/81</td>
</tr>
<tr>
<td>WP5088</td>
<td>Prostaglandin signaling</td>
<td>2/17</td>
<td>12/595</td>
<td>0.0429</td>
<td>0.2200</td>
<td>2919/4283</td>
</tr>
<tr>
<td>WP51</td>
<td>Regulation of actin cytoskeleton</td>
<td>2/17</td>
<td>12/595</td>
<td>0.0429</td>
<td>0.2200</td>
<td>2247/7430</td>
</tr>
</tbody>
</table>

ID identifies the pathway of biological processes to which relevant proteins are overrepresented; gene ratio refers to the total differentially expressed proteins in the given gene ontology term; BgRatio refers to the size of the geneset relative to the size of the unique genes in the collection of genesets; q-value represents the p-value that is corrected for multiple testing; geneID identifies the genes from which relevant proteins were encoded.
prediction accuracies were not high, the amount of variance explained by our models is comparable to the performance of other biology-based predictive models of psychological outcomes, such as polygenic prediction of cognitive function (Lee et al., 2018). Complex traits in early childhood, including dimensions of temperament and psychopathology, may be associated with many prenatal proteins that individually have small effects but collectively exhibit larger associations. Nonetheless, the proteomic models do not explain the majority of variance in these early childhood outcomes. One interpretation of these findings is that large-scale proteomic modeling in the first trimester of pregnancy has stronger implications for social than for predictive developmental and clinical science. From a basic science perspective that is focused on building explanatory models, our results further highlight the importance of prenatal inflammation and immune function in postnatal development. Our results also point to the potential importance of other proteins and pathways that have not been considered in prior work. Thus, the study of large-scale proteomics may reveal important gaps in current hypotheses regarding the prenatal origins of child well-being. From a prediction perspective, however, machine learning algorithms using prenatal proteomics do not produce child temperament and psychopathology scores that strongly resemble mother-reported scores. This finding raises the possibility that the study of large-scale proteomics during pregnancy, by itself, is not of practical value for predicting and preventing mental health difficulties at the level of individual children. We further discuss below the prospect that accurately predicting complex psychosocial outcomes requires taking into account considerably more information about children’s unique characteristics.

Prior studies have focused primarily on a subset of proinflammatory cytokines in relation to offspring psychological outcomes (Girchenko et al., 2020; Gustafsson et al., 2018, 2020; Rudolph et al., 2018). Using an exploratory data-driven approach with large-scale proteomics, we identified novel biomarkers of immune function in pregnant women that proteins associated with offspring development, particularly young children’s proneness to sadness. Dispositional sadness underlies some internalizing problems and has frequently been associated with the development of psychopathology (Eggum et al., 2012; Eisenberg et al., 2009). We found that prenatal proteins involved in upregulating inflammation, such as sortilin-1 (SORT1), are among the most strongly, positively associated with children’s sadness. Research with mice has implicated SORT1 in the regulation of proinflammatory cytokines including IL-6 and interferon gamma (Talbot et al., 2019). This protein, however, has not been the focus of prior studies with humans aimed at uncovering prenatal processes implicated in offspring risk. Interestingly, many of the immune system proteins were downregulated in mothers who were higher in sadness, such as IL-32 and C-X-C Motif Chemokine Ligands (CXCL) 1, 6, and 9. Proteins implicated in inflammatory processes were also related to emotional reactivity and attentional problems, including CD93 (Woodbury-Smith, Paterson, Szatmari, & Scherer, 2020) and IL-1RT2 (Byrne et al., 2022), which have also been associated with risk for psychopathology in human models. We also identified inflammatory-related proteins Cathepsin S (CTSS; Smith et al., 2017) and EGF Like Repeats and Discoidin Domains 3 (EDIL3; Richetto et al., 2017) as associated with emotional reactivity and attentional problems; these proteins have been associated with alterations in brain development in rodent models. Although several of the associations we found do not survive correction for multiple comparisons, our exploratory, hypothesis-generating approach yielded a list of proteins that could guide future studies in considering a wider range of prenatal predictors of child development. In future work, researchers should experimentally validate the roles of these proteins in larger, more heterogeneous samples to confirm these findings.

Consistent with the perspective that maternal immune and glucocorticoid processes contribute to prenatal programming of offspring risk (Hantsoo et al., 2019), the enrichment analyses further highlighted the association between proinflammatory and profibrotic mediators pathway and child sadness. Thus, our findings provide support for evidence from prior studies that have identified associations between prenatal inflammation and altered behavioral development using hypothesis-driven approaches focused on candidate proteins (Gustafsson et al., 2018; Hantsoo et al., 2019; Rudolph et al., 2018). The enrichment analyses also provided insight into prenatal biological pathways implicated in offspring psychosocial development that have not previously been considered. Proteins associated with nervous system development pathways and Hippo signaling regulation pathways were overrepresented in relation to children’s emotional reactivity and attentional problems. For example, several erythropoietin-producing human hepatocellular (Eph) receptors, which are part of the nervous system development pathway and mediate processes such as cell migration and adhesion, were significantly associated with both emotional reactivity and attentional problems. The Hippo signaling pathway is involved in cell proliferation, differentiation, and survival, and is highly active during embryonic development. Emotional reactivity and attentional problems were both associated with members of this pathway, such as tyrosine kinases like FGFR2 and FLT1 and tyrosine kinase receptors like MET and NTRK2,
which has also been found to be associated with infant temperament from the placental transcriptome (Buthmann et al., 2022).

Several pathways that were significantly related to the behavioral measures are related to cancer and other diseases. It is important to note that the proteins in these pathways are also affiliated with the development of the central and peripheral nervous systems, cell signaling, apoptosis, and immune function. Curry and Glasgow (Curry & Glasgow, 2021) recently noted that proteins that are prevalent during normative neurodevelopmental processes in utero are also active in disease processes; for example, cell differentiation and angiogenesis occur during the embryonic stage of gestation, as well as during tumor formation. Furthermore, the sadness scale was related to two SARS-CoV-2 pathways; the proteins in these pathways are also involved in immune responses to other pathogens and stress responses.

We should note that no biological pathways remained significantly enriched in relation to sadness, attention problems, or emotional reactivity after adjusting for multiple comparisons. Larger samples may be necessary to elucidate more robust and precise relations between prenatal proteomics and offspring outcomes in early childhood. These efforts will be more financially feasible as mass cytometry technology becomes less expensive. In addition, the extreme gradient boosted algorithm significantly predicted only four of 23 behavioral scales and, of those four, the algorithm predicted a relatively small proportion of variance (5%–10%). Whether this is considered appreciable accuracy for making clinical decisions at the individual level is debatable. The low prediction accuracy fits well with a perspective emphasizing equi- and multi-finality in developmental psychopathology (Cicchetti & Rogosch, 1996). There are likely many paths to the emergence of psychopathology, not all of which are specific to prenatal processes. In addition, it is unlikely that the upregulation of prenatal biomarkers of risk will always lead to offspring difficulties. Thus, from these perspectives, prenatal proteomic models should not be expected to account for large percentages of variance in highly complex outcomes, such as child psychosocial development. Nonetheless, it is possible that large-scale proteomics during pregnancy represent one piece of a precision health approach that considers many factors in formulating diagnosis and generating unique prevention strategies for individual families.

We should note several limitations of this exploratory study. First, the sample is relatively homogeneous with respect to racial composition and socioeconomic status. Compared to the general population, a higher proportion of people in this study were White, married, and have attended graduate school. Furthermore, we did not recruit for this study participants who delivered preterm in order to avoid the confounding effect of early gestational age. These sample characteristics limit the generalizability of our findings, which require replication in more diverse samples, and may have constrained our ability to identify associations between basic demographic factors and child outcomes. Second, we used parent-reported measures of child behavior in this study. Future studies might consider including other assessments of child behavior. Additionally, the internal reliability of the sadness scale was low, and that of the other three significant scales was only moderate (0.66–0.71). The low-moderate internal reliability of some of the subscales may have contributed to the differential ability of the proteome to make accurate predictions. Third, we recognize that in using first trimester maternal blood samples to measure the proteome, we cannot make strong conclusions about the full gestational environment. The maternal proteome may reveal as much about maternal characteristics as it does about fetal developmental processes and child outcomes. We should note, however, that given the constraints of human prenatal research, our approach is among the closest feasible approximation of the first trimester gestational environment. Fourth, we cannot account for the influence of the maternal genome, the environment, or any other factors that may have affected both maternal inflammation and child behavior. Finally, did not have an independent cohort with which to replicate these findings. We did use half of the sample to train and the other half to test the algorithm; nevertheless, given that the half used to test were from the same cohort, an independent sample should be used to test the model. Despite these limitations, however, we were able to use first trimester maternal proteome samples to identify associations with child behavior at 3–5 years of age. Integrating such data with additional -omics data and with data from electronic health records may improve our ability to predict child development from gestational biospecimens and indicators of maternal health (De Francesco et al., 2023; Marić et al., 2022). Continuing to examine associations between maternal proteomics and child functioning will advance our understanding of gestational development and inform the identification of targets for prevention and intervention for child psychopathology.

Relevance
This work is important in generating recommendations for further science and informing clinical practice. Based on our findings that first trimester maternal biospecimens meaningfully predict child behavior at 3–5 years of age, we recommend that future work integrates multiple source and types of -omics data from across gestation to improve the
The prenatal proteome predicts child psychopathology


Accepted for publication: 23 November 2023

© 2024 Association for Child and Adolescent Mental Health.