Placental transcriptomic signatures of spontaneous preterm birth

Alison G. Paquette, PhD; James MacDonald, MS; Theo Bammler, PhD; Drew B. Day, PhD; Christine T. Loftus, PhD; Erin Buth, MPH; W. Alex Mason, PhD; Nicole R. Bush, PhD; Kaja Z. Lewinn, ScD; Carmen Marsit, PhD; James A. Litch, MD, DTMH; Michael Gravett, MD; Daniel A. Enquobahrie, MD, PhD; Sheela Sathyanarayana, MD, MPH

BACKGROUND: Spontaneous preterm birth accounts for most preterm births and leads to significant morbidity in the newborn and childhood period. This subtype of preterm birth represents an increasing proportion of all preterm births when compared with medically indicated preterm birth, yet it is understudied in omics analyses. The placenta is a key regulator of fetal and newborn health, and the placental transcriptome can provide insight into pathologic changes that lead to spontaneous preterm birth.

OBJECTIVE: This analysis aimed to identify genes for which placental expression was associated with spontaneous preterm birth (including early preterm and late preterm birth).

STUDY DESIGN: The ECHO PATHWAYS consortium extracted RNA from placentas collected from the Conditions Affecting Neurocognitive Development and Learning in Early Childhood and the Global Alliance to Prevent Prematurity and Stillbirth studies. Placental transcriptomic data were obtained by RNA sequencing. Linear models were fit to estimate differences in placental gene expression between term birth and spontaneous preterm birth (including gestational age subgroups defined by the American College of Obstetricians and Gynecologists). Models were adjusted for numerous confounding variables, including labor status, cohort, and RNA sequencing batch. This analysis excluded patients with induced labor, chorioamnionitis, multifetal gestations, or medical indications for preterm birth. Our combined cohort contained gene expression data for 14,023 genes in 48 preterm and 540 term samples. Genes and pathways were considered statistically significantly different at false discovery rate—adjusted $P$ value of <.05.

RESULTS: In total, we identified 1728 genes for which placental expression was associated with spontaneous preterm birth with more differences in expression in early preterm samples than late preterm samples when compared with full-term samples. Of those, 9 genes were significantly decreased in both early and late spontaneous preterm birth, and the strongest associations involved placental expression of IL1B, ALPL, and CRLF1. In early and late preterm samples, we observed decreased expression of genes involved in immune signaling, signal transduction, and endocrine function.

CONCLUSION: This study provides a comprehensive assessment of the differences in the placental transcriptome associated with spontaneous preterm birth with robust adjustment for confounding. Results of this study are in alignment with the known etiology of spontaneous preterm birth, because we identified multiple genes and pathways for which the placental and chorioamniotic membrane expression was previously associated with prematurity, including IL1B. We identified decreased expression in key signaling pathways that are essential for placental growth and function, which may be related to the etiology of spontaneous preterm birth. We identified increased expression of genes within metabolic pathways associated exclusively with early preterm birth. These signaling and metabolic pathways may provide clinically targetable pathways and biomarkers. The findings presented here can be used to understand underlying pathologic changes in premature placetas, which can inform and improve clinical obstetrics practice.

Key words: ALPL, chemokine signaling, GABRP, IL1B, placenta, placental metabolism, signal transduction, spontaneous preterm birth, transcriptomics

Introduction

Preterm birth (PTB) is the leading cause of newborn death worldwide, it is responsible for most newborn morbidity, and it imparts a significant social and financial burden to caregivers and the healthcare system. The majority (60%) of PTBs are spontaneous, arising either from premature membrane rupture (25%) or spontaneous preterm labor (35%). The remaining 40% of PTBs occur as a consequence of multifetal pregnancies, fetal malformations, and medical indications such as pre eclampsia or chorioamnionitis. Characterizing the etiology of spontaneous preterm birth (sPTB) is paramount because neonatal and later life outcomes differ based on subtle shifts in delivery timing. Late-preterm infants (born at 34–37 weeks’ gestation) and early-term infants (born at 37–39 weeks’ gestation) are at increased risk for adverse childhood health outcomes when compared with full-term infants. Consequently, the American College of Obstetricians and Gynecologists (ACOG) recommended more precise definitions of gestational age for research and clinical studies to best predict and prevent adverse outcomes. In addition, sPTBs comprise an increasing proportion of all PTBs, whereas the rate of medically indicated PTB is decreasing. The placenta is an essential regulator of the in utero environment. It transports gas, nutrients, and waste, it provides immunologic surveillance for the developing fetus, and it produces neuropeptides, growth factors, and steroid hormones that are released into the maternal circulation to coordinate
Why was this study conducted?
This study aimed to identify differences in the placental transcriptome of placentas from preterm infants when compared with term infants.

Key findings
This study identified and characterized genes associated with prematurity that are involved in immune, endocrine, and signal transduction pathways. We noted a subset of pathways involved in metabolic processes containing genes with increased expression only in early-preterm infants when compared with full-term infants.

What does this add to what is known?
This study contributes to the field by providing a catalog of genes and pathways for which placental expression is associated with early and late spontaneous preterm birth (sPTB) based on investigations from a large, retrospective cohort with nuanced covariate measures and robust adjustment for confounding. This study confirms gene expression changes identified in previous case-control investigations examining placental and chorioamniotic membrane physiology and PTB and yielded novel insight into shared and distinct immune signaling mechanisms and metabolic perturbations related to early and late sPTB.

Materials and Methods
Study population
This retrospective cohort study aimed to analyze the relationship between the placental transcriptome and sPTB. ECHO PATHWAYS consortium harmonized data from 2 preexisting cohorts, the Conditions Affecting Neurocognitive Development and Learning in Early Childhood (CANDLE) study, which enrolled 1504 women from Shelby County, Tennessee and the Global Alliance to Prevent Prematurity and Stillbirth (GAPPS) study, which enrolled >600 pregnant women and children from Seattle, Washington, and Yakima, Washington. Each cohort had its own inclusion criteria. Inclusion criteria for the CANDLE study included individuals between 16 and 40 years of age with uncomplicated singleton pregnancies who planned to deliver at a participating study hospital. Inclusion criteria for the GAPPS study included a participant age of ≥18 years or medically emancipated and confirmed to be pregnant by self-test or by physician’s medical testing. The ECHO PATHWAYS consortium sequenced a subset
of placentas from each cohort (n=289 from GAPPs and n=794 from CANDLe). Our study-specific exclusion criteria excluded patients with multifetal pregnancies, preeclampsia, placental abruption, chorioamnionitis (based on individual cohort collection information explained in the clinical definitions section), recorded cervical insufficiency, and induced labor (Figure 1, A). Data analyzed here represent a subset of each population with complete RNA sequencing data and complete covariate data but without the study-specific exclusion criteria (Figure 1, A). All ECHO PATHWAYS research activities pertaining to these samples were approved by the University of Washington Institutional Review Board (IRB). All research activities for the CANDLe cohort were approved by the University of Tennessee Health Sciences Center IRB. The Seattle Children’s Research Institute IRB approved all research activities for the GAPPs cohort.

**Clinical definitions**

PTB and gestational age subgroups were defined using the most recent ACOG guidelines (preterm birth, <37 weeks’ gestation; early-preterm birth, <34 weeks’ gestation; late-preterm birth, 34–37 weeks’ gestation; early-term birth, 37–39 weeks’ gestation; full-term birth, 39–41 weeks’ gestation; and late-term birth, >41 weeks). In the CANDLe study, gestational age was calculated from medical record abstraction and self-report. About 57.3% of CANDLe study participants reported ultrasound dating as the primary method of calculating gestational age. In GAPPs, gestational age was derived from clinical entered dates, which captured both the date of the last menstrual period and the expected date based on the first or second trimester ultrasound in alignment with ACOG criteria. Clinical information about the exclusion criteria was also gathered and defined by the retrospective cohorts based on medical record abstraction. From the CANDLe study, we excluded individuals with preeclampsia based on an indication in the medical records of preeclampsia or hypertension and individuals with chorioamnionitis based on medical record abstraction of confirmed clinical chorioamnionitis. No data on cervical insufficiency was available for CANDLe study participants. From the GAPPs cohort, we also excluded individuals with preeclampsia, cervical insufficiency, and chorioamnionitis based on medical record abstraction data. Premature preterm rupture of membranes (PPROM) was defined as the rupture of membranes before labor onset in infants born before
37 weeks’ gestation based on medical record abstraction. For both cohorts, phenotype data were harmonized by applying a uniform data structure and consistent coding rules for phenotype variables.

**Placental collection**

For placentas collected from the CANDLE study, 1 rectangular piece of placental villous tissue (approximate dimensions of 2 cm × 0.5 cm × 0.5 cm) was dissected from the middle of the parenchyma within 15 minutes of delivery as previously described. The tissue cubes were placed in a 50 mL tube with 20 mL RNAlater and refrigerated at 4°C overnight before being stored in RNAlater at −80°C. This tissue was manually dissected and cleared of maternal decidua. For placentas collected from the GAPPS biorepository, four 8-mm vertical tissue punches from the placental disc were collected from 2 sites separated by 7 cm on the placental disc and placed in 5 mL tubes containing approximately 3 mL of RNAlater within 30 minutes of delivery and stored at −20°C as previously described. Specimens were then shipped to the GAPPS facility and stored at −80°C. The fetal villous tissue was manually dissected and cleared of maternal decidua using standard protocols developed by the GAPPS placental biorepository.

**RNA sequencing processing and analysis**

RNA isolation and RNA sequencing quantification methods have been described previously for the CANDLE study placental samples, and the same protocols were used for GAPPS study. Briefly, approximately 30 mg of fetal villous placental tissue was homogenized using a TissueLyser LT instrument (Qiagen, Germantown, MD). RNA was isolated using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen). RNA integrity was determined using a Bioanalyzer 2100 with RNA 6000 NanoChips (Agilent, Santa Clara, CA), and only RNA samples with an RNA integrity number > 7 were sequenced. RNA sequencing was performed at the University of Washington Northwest Genomics Center. The total RNA was poly-A enriched, and complementary DNA libraries were prepared using the TruSeq Stranded mRNA kit (Illumina, San Diego, CA). Each library was sequenced to an approximate depth of 30 million reads on an Illumina HiSeq 4000 instrument. RNA sequencing quality control was performed using both the FASTX-tool (version 0.0.13) and FastQC (version 0.11.2) toolkits. Transcript abundances were estimated by aligning to the GRCh38 transcriptome (Gencode version 33) using Kallisto, then collapsed to the gene level using the Bioconductor tximport package, scaling to the average transcript length. Only protein-coding genes and long noncoding RNAs were included in this analysis. The transcriptomic results of the 2 cohorts were pooled together and then filtered to remove those genes that were unexpressed or expressed at an unreliably low level, defined as a mean log counts/million (logCPM) of < 0. After filtering, 14,023 genes remained.

**Identification of differentially expressed genes**

Differentially expressed genes (DEGs) were identified by fitting a weighted linear regression using the limma-voom pipeline from the Bioconductor limma package, which converts gene counts into log counts/million counts (logCPM) and then uses observation-level weights to account for the mean variance relationship of the logCPM values. Comparisons between groups were made using empirical-Bayes adjusted t statistics implemented in the Bioconductor limma package. We selected potential confounders to be included in our analysis a priori through a review of the literature on PTB and gestational age and then constructed a directed acyclic diagram (Figure 1, B). In all models, we adjusted for RNA sequencing batch, fetal sex, cohort, study site, labor status, PPROM status, and maternal age, body mass index, self-reported smoking status, self-reported alcohol consumption, income, education, race, and ethnicity (Figure 1, B). We identified unmeasured confounding using surrogate variable analysis (SVA) and adjusted for these artifacts in our model. For all analyses, genes were considered statistically significant with a false discovery rate (FDR)–adjusted P < 0.05 using the Benjamini-Hochberg method. Lists of DEGs were compared using the “UpSetR” R package (R Studio, Vienna, Austria). Pathway enrichment analysis

We identified differences in gene expression within the Kyoto Encyclopedia of Genes and Genomes (KEGG) biologic pathways using the roast function in the Bioconductor limma package, which compares the average t statistic for each gene set to a null distribution estimated from random rotations of the residuals of the underlying model (10,000 rotations). We present results from the directional test, which first infers the direction of change for the gene set based on the preponderance of up- or down-regulated genes and then computes the average t statistic based on the subset of genes in the set with the inferred direction of change. We included all KEGG pathways except disease pathways (KEGG release 98.1). Pathways were considered statistically significant at an FDR-adjusted P < 0.05.

**Results**

**Clinical characteristics**

We identified differences in expression from placentas collected from 461 individuals from the CANDLE study and from 127 individuals from the GAPPS cohort enrolled in the ECHO PATHWAYS consortium study with complete RNA sequencing and covariate data (Table). Our combined cohort was 42.7% Black, and 44.7% of participants had a high school education or lower. This combined study included 48 preterm infants with 38.7% of those cases presenting with PPROM (n = 24 based on medical record abstraction). A total of 91.7% of participants who delivered prematurely went into labor. Our study also included 540 participants who were delivered at term, and 74.3% of these participants went into labor. Our exclusion criteria eliminated all preterm and term infants with medically...
indicated causes of delivery based on medical record abstraction, including preeclampsia, multi-fetal gestation, chorioamnionitis or cervical insufficiency, and we adjusted for labor status and PPROM as confounders (Figure 1), leaving us with a transcriptional signature attributable to sPTB. Gestational length did not differ between the CANDLE (mean, 38.89 weeks) and GAPPS (mean, 38.92 weeks) cohorts ($P = .9$, $t$ test).

**Global assessment of placental transcriptomic signatures of prematurity**

We analyzed differences in placental gene expression by first using a binary grouping of births at $< 37$ weeks’ gestation (preterm, $n = 48$) vs births at $\geq 37$ weeks’ gestation (term, $n = 540$). We observed broad differences in the preterm transcriptomes with 1062 genes exhibiting statistically significant differences in expression based on a statistical threshold of FDR adjusted $P < 0.05$ (ie, DEGs) (Figure 2). We performed a separate analysis for each gestational age subgroup, as defined by ACOG guidelines (early preterm, late preterm, early term, full term, and late term), with the reference group defined as full-term infants ($39e41$ weeks’ gestation). In this analysis, we observed more DEGs in early-preterm infants (1359 DEGs) than in late-preterm infants (27 DEGs) when compared with full-term infants (Figure 2) based on a statistical threshold of FDR adjusted $P < 0.05$. There were no statistically significant differences in placental expression of early-term or late-term infants when compared with full-term infants. Within the group of preterm infants alone ($n = 48$), we performed stratified analyses of transcriptomic differences associated with labor ($n = 44$) vs no labor ($n = 4$), as well as PPROM ($n = 24$) vs no PPROM ($n = 38$); however, no genes were statistically significant for either analysis (FDR–adjusted $P < .05$). We also performed an analysis of fetal sex as a potential modifier of the relationship

<table>
<thead>
<tr>
<th>TABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant information</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cohort, $n$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CANDLE</td>
</tr>
<tr>
<td>GAPPS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gestational length subgroup, $n$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early preterm ($&lt; 34$ wk)</td>
</tr>
<tr>
<td>Late preterm ($34$–$37$ wk)</td>
</tr>
<tr>
<td>Early term ($37$–$39$ wk)</td>
</tr>
<tr>
<td>Full term ($39$–$41$ wk)</td>
</tr>
<tr>
<td>Late term ($41$–$42$ wk)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Preterm premature rupture of membranes, $n$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fetal sex, $n$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Male</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Labor status, $n$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor</td>
</tr>
<tr>
<td>No labor</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Delivery method, $n$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cesarean delivery</td>
</tr>
<tr>
<td>Vaginal</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maternal ethnicity, $n$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Latino</td>
</tr>
<tr>
<td>Latino</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maternal race, $n$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian</td>
</tr>
<tr>
<td>Black</td>
</tr>
<tr>
<td>Multiple race</td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>White</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maternal education, $n$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High school or less</td>
</tr>
<tr>
<td>College or higher</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>History of preterm birth, $n$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
</tr>
</tbody>
</table>

between placental gene expression and sPTB but found no statistically significant instances of an interaction (FDR-adjusted \( P < 0.05 \)). Across all analyses, there were more DEGs with decreased expression than increased expression in preterm infants when compared with term infants (Figures 2 and 3, A).

**Assessment of confounding by gestational age using placental maturation signatures**

Because the placental tissues are not collected from the same time point in our preterm and term infants, our analysis is intrinsically confounded by developmental changes that are related to gestational age, but it is not logistically feasible or ethical to collect placental samples from nonpathologic placentas at the same gestational time point to act as a control group. To investigate how the gene expression changes we identified in association with prematurity might be attributable to normal developmental processes (i.e., placental maturation) related to gestational age, we created a curated list of 1420 placental maturation signatures collected from maternal blood (\( n = 522 \) genes), cell-free RNA throughout pregnancy (\( n = 9 \) genes), as well as genes for which placental protein expression was significantly different between the second trimester and term gestations (\( n = 954 \) genes) (Supplemental Figure 1, A; Supplemental Table 1). A similar approach was used by Lien et al to disentangle gestational age-related differences in their transcriptomic study of sPTB. Overall, 101 of the DEGs associated with sPTB (vs term birth),

---

**TABLE**

**Participant information (continued)**

<table>
<thead>
<tr>
<th>Participant information</th>
<th>Preterm (n=48)</th>
<th>Term (n=540)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal smoking status during pregnancy, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>46 (95.8)</td>
<td>500 (92.6)</td>
</tr>
<tr>
<td>Yes</td>
<td>2 (4.2)</td>
<td>40 (7.4)</td>
</tr>
<tr>
<td>Maternal alcohol use during pregnancy, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>45 (93.8)</td>
<td>474 (87.8)</td>
</tr>
<tr>
<td>Yes</td>
<td>3 (6.3)</td>
<td>66 (12.2)</td>
</tr>
</tbody>
</table>

CANDELE, Conditions Affecting Neurocognitive Development and Learning in Early Childhood study; GAPPS, Global Alliance to Prevent Prematurity and Stillbirth study.


---

**FIGURE 2**

Volcano plots depicting significance and strength across analysis

**A** Preterm vs. Term Analysis  
**B** Early Preterm vs. Full Term Analysis  
**C** Late Preterm vs. Full Term Analysis

Volcano plots our analyses of (A) preterm vs term infants and (B) early and (C) late preterm infants vs. full term infants. Purple Xs represent placental maturation signatures, and the top 5 genes based on log fold change are labeled.

ACOG, American College of Obstetricians and Gynecologists.


---

73.e6 American Journal of Obstetrics & Gynecology JANUARY 2023
115 of the DEGs associated with early-preterm birth (vs full-term birth), and 5 of the DEGs associated with late-preterm birth (vs full-term birth) were genes attributable to placental maturity based on this list (Supplemental Figure 1, B). The results that were statistically significant may have been confounded by gestational age based on these maturation signatures that are presented in Supplemental Table 2 and these were removed from all downstream analyses. The remaining genes that were statistically significant may have been confounded by gestational age based on these maturation signatures that are presented in Supplemental Table 3.

Shared and distinct changes to the placental transcriptome between prematurity subgroups

We present shared and distinct DEGs between each PTB analysis subgroup (Figure 3, A and B). Two genes were significantly increased in all 3 analyses (GABRP, KIF26B), and 9 genes were significantly decreased across all 3 analyses (with 1L1B expression the most decreased overall) (Supplemental Figure S2). A total of 468 DEGs were significantly associated with early-preterm (compared with full-term infants) and all preterm infants (vs term infants). Nine DEGs were significantly associated with late-preterm (compared with full-term infants) and all preterm infants (vs term infants). Among the top 5 genes from each subgroup (Figure 3, B), we observed increased expression of multiple chemokines (CXCL8, IL1B) and decreased expression of other immune signaling genes (CRLF1, SAA1, PTX3).

Comparative analysis of prematurity signatures from single-cell sequencing analysis

Complementary work by Pique Regi et al27 used single-cell RNA sequencing data collected from women who underwent term and preterm labor (n=3/group) to characterize transcriptomic differences within individual cell types related to prematurity. Thirty-four of these signatures overlapped with our transcriptomic signatures of early-preterm vs full-term infants or preterm infants.
COMPARATIVE ANALYSIS OF PREMATURE SUFFICIENT SIGNED DIFFERENCES DERIVED FROM CHORIOAMNOITIC MEMBRANES

Because we did not generate transcriptome data on chorioamnion membranes in this study, we compared the changes in expression we noted in the placenta in a previously published case-control study that identified 270 DEGs for which expression in chorioamnion membranes was associated with prematurity using similar exclusion and inclusion criteria and study design as our study, including adjustment for PROM as a confounder. A total of 62 of these 270 DEGs were also associated with sPTB in our analysis (Supplemental Figure S4). Two genes (ALPL and ILL1B) were significantly decreased in the placenta across all 3 of our analyses but were increased in the chorioamnionic membranes in PTB cases. Pathway enrichment of these DEGs revealed that these 62 overlapping DEGs were significantly overrepresented in 11 KEGG pathways (FDR adjusted P < 0.05). The top pathways based on P value included chemokine signaling, cytokine-cytokine receptor interactions, and osteoclast differentiation.

IDENTIFICATION OF SHARED AND DISTINCT BIOLOGIC PATHWAYS

Pathway analysis was performed using self-contained gene set testing and pathways were considered statistically significant based on a threshold of FDR adjusted P < 0.05. In the comparison of sPTB vs term birth, we identified 133 down-regulated KEGG pathways and 16 up-regulated pathways. In early-preterm infants, 120 pathways were down-regulated, and 34 pathways were up-regulated when compared with full-term infants. In late-preterm infants, 72 pathways were down-regulated when compared with full-term infants. Complete results of the pathway enrichment analysis are presented in Supplemental Table 4. A total of 65 pathways (36.3% of all significant pathways) contained genes with decreased expression in all 3 analyses (Figure 4, A). These pathways were members of KEGG subgroups involving the immune system, signal transduction, signaling molecules, organismal systems, metabolism, and cellular processes (Figure 4, B). The immune signaling KEGG subgroup was the largest (17 pathways) and contained the chemokine signaling pathway, which was the most significant pathway overall. The second largest subgroup was signal transduction (14 pathways), and the most significant pathway here was the phospholipase D signaling. The third largest subgroup was the endocrine system, a component of organismal systems (11 pathways), and the most significant pathway in this group was parathyroid hormone synthesis, secretion, and action.

No pathways were increased across all 3 PTB subgroups, but we noted that several pathways had a higher proportion of genes that were increased only among late-preterm infants (3 pathways; top pathway basal transcription factors) or in early-preterm infants (21 pathways; top pathway: spliceosome). Of note, two-thirds (14/21) of pathways only increased among early-preterm infants and involved metabolism, including carbohydrates (butanoate and glyoxylate), amino acids (alanine, aspartate glutamine, tryptophan selenocompounds), lipids, and cofactor and vitamin metabolism (porphyrin, vitamin B6, retinol and ubiquinone).

RESULTS IN CONTEXT

The findings from our study are in alignment with the known etiology of sPTB and previous transcriptomic analyses but expand on these by studying the placenta, which is a core regulator of the in utero environment, and by using well-phenotyped samples and adjusting for confounding variables. We observed a higher number and more significant differences in placental gene expression associated with early PTB (vs full-term birth) than with late sPTB (vs full-term birth), suggesting broader differences in placental function in early sPTB. Other studies have identified unique molecular signatures in early-preterm infants (both spontaneous and medically indicated). A study involving biomarkers collected during the first trimester screening identified a unique subset of biomarkers associated with early PTB. Another multi-omics assessment of maternal whole blood identified unique differences in the methylome and transcriptome associated with very-early preterm labor (<28 weeks’ gestation), but that were not significantly associated with preterm vs term labor or other preterm phenotypes. We noted that 15 of the DEGs in this study were also significantly associated with differences in expression in monocytes and whole blood in relation to spontaneous preterm labor in our previous study, including the top DEG from the previous study, ADAMTS2, as well as several interleukin signaling genes that were demonstrated to be regulated by microRNAs in the context of preterm labor in a follow-up study. Our study adds to this by providing a
Rotational gene set testing revealed shared and distinct pathways for which placental expression was associated with prematurity

A. Shared and distinct pathways associated with early and late preterm birth and with binary preterm birth analysis

B. 65 Pathways containing genes whose placental expression was associated with early and late preterm birth

A, Venn diagram of shared and distinct pathways between analysis groups. B, 65 pathways associated with all 3 analyses. The size of the dots is scaled to the log-adjusted P value and ordered on the X axis based on the proportion of genes that are up- or down-regulated. Pathways are grouped into KEGG subgroups.

KEGG: Kyoto Encyclopedia of Genes and Genomes.
clinical implications

The topmost DEGs in all 3 analyses provide high-confidence results that may be useful for future biomarker discovery, particularly for infant comorbidities related to sPTB. These signatures could also potentially serve as candidates for predictive markers of sPTB, which would need to be tested in maternal blood throughout pregnancy. Such analyses are beyond the scope of this retrospective analysis. Clinical implications and invasion in vitro by stimulating metalloproteinases such as MMP9. IL1B mediates the onset of premature labor in the context of infection, and increased levels of IL1B in the cervical, myometrium, and fetal membranes in the presence or absence of infection are associated with PTB; increased levels of IL1B were identified in the decidual and placenta in association with term labor. IL1B has been proposed to be a potential therapeutic target for PTB, but preclinical studies have shown reduced efficacy of IL1B antagonists to prevent infection-induced PTB, as reviewed by Nadeau-Vallee et al. Our study demonstrates decreased expression of IL1B in the placenta in preterm infants when compared with term infants, which is contrary to findings in other prenatal tissues. This decreased expression could reflect compensatory mechanisms or alternative signaling mechanisms occurring within the placenta. More work is needed to confirm these findings in an independent cohort and to understand IL1B expression across multiple pregnancy tissues during preterm labor.

Our pathway analysis revealed decreased placental expression of genes within a variety of signal transduction, endocrine, and immune pathways that are clinically relevant because these pathways may provide modifiable targets for pharmacologic interventions. For example, many of the signal transduction pathways in which we observed decreased expression are well studied in the context of cancer with well-documented pharmaceutical treatments, including the MAP-kinase, Ras, and JAK-Stat pathways. These pathways are essential for placental functions, as discussed below, and so future studies could explore these pathways as targets for clinical intervention.

We also observed increased placental expression of genes in multiple metabolic pathways unique to sPTB placentas, which may provide clinically modifiable targets through dietary intervention. These included pathways involved in metabolism of retinol, vitamin B6, pyruvate, and tryptophan. Of note, we observed increased expression of genes within the selenocompound metabolism pathway, which is responsible for converting selenocysteine to methylselenomethionine, which is involved in core biological processes regulated by the placenta, such as inflammatory processes and redox regulation. Other studies have identified selenium deficiency as a modifiable risk factor for premature birth, particularly in extremely-preterm infants. Genetic polymorphisms in the loci encoding selenocysteine transfer RNA—specific eukaryotic elongation factor (participating in the incorporation of selenocysteine into selenoproteins) have been implicated in both gestational age and in sPTB. The significant positive association between the selenocompound metabolism pathway and early sPTB suggests a role for maternal selenium micronutrient supplementation in the prevention of sPTB, which supports recently launched clinical trials in this field. This study...
suggests that metabolism of these key micronutrients may be perturbed in preterm placentas based on differences in gene expression and it contributes to a body of evidence that suggest these may be modifiable risk factors for the prevention of prematurity. Future studies will be needed to validate these findings in an independent cohort.

**Research implications**

Several of the signaling pathways that were different in preterm placentas may be related to altered placental growth and invasion. We observed decreased expression of genes within the MAPK or JAK-STAT signaling pathways, which are involved in the invasion and syncytialization of placental trophoblast cells. We also observed decreased expression of VEGF and HIF-1 signaling, which is essential for placental angiogenesis. We also noted decreased expression of genes in endocrine pathways (including the relaxin, estrogen, oxytocin, and cortisol pathways) that are produced by the placenta and are involved in key endocrine processes important to pregnancy. Our research suggests that decreased expression of genes within these pathways may lead to perturbed molecular mechanisms and altered placental function and maternal-fetal signaling, ultimately culminating in sPTB. Both our gene- and pathway-level findings highlight the importance of immune signaling in the context of sPTB. We noted 17 different immune pathways with decreased expression in preterm vs term samples, as well as a number of individual genes, including decreased expression of IL1B, CRLF1, CXCL8, and PTX3. Although there are numerous studies on the role of the fetal membranes and maternal blood in the context of premature birth, this study highlights the role of immune signaling within the placenta itself.

Our global findings also generated new hypotheses at both the gene and pathway level. For example, the top increased DEG across all analyses was GABRP, which encodes the inhibitory neurotransmitter gamma-aminobutyric acid (GABA) type A receptor subunit π. This form of the GABA neurotransmitter has been found in a variety of reproductive tissues, including the placenta, where it alters the sensitivity of GABA receptors to pregnanolone. In extravillous trophoblast cells, GABRP promotes apoptosis, which ultimately leads to decreased invasion. Placental expression of GABRP was also positively associated with preeclampsia. Our work suggests that placental expression of GABRP may be altered in the context of prematurity, but more experimental work is needed to elucidate the biologic basis behind changes in expression of these genes.

**Strengths and limitations**

Our results should be interpreted in light of the inherent limitations in RNA sequencing analysis. We capture the placental transcriptome at birth, which is a snapshot of a highly regulated temporal process. We cannot collect samples matched for gestational age in preterm and term samples, and so the differences we observe in preterm placentas may be reflective of gestational age-related differences instead of pathologic changes related to prematurity. We address this in part by using placental maturation signatures. Expression was quantified using bulk RNA sequencing data, and so our findings may be confounded by different cell types collected within each sample and discrepancies in placental sampling protocols, which is a well-established challenge in this field; however, we address this in part by adjusting for cellular heterogeneity using SVA. This study did not deploy a spatial transcriptomics approach to investigate multiple sites within the same placenta, which is not feasible in such a large retrospective cohort study but is an important factor, because normal variation in placental tissue has been noted. We do not account for all risk factors and phenotypic subgroups of premature infants, such as extra-uterine infections, cervical shortening or insufficiency, or maternal trauma because of the low prevalence and inherent challenges in measuring these factors, however, we accounted for as many feasible factors in our study design as possible. Our analysis did not include pregnancies with other pathologies (e.g., preeclampsia, gestational diabetes, etc.) because our sample collection strategy was not designed specifically for these complications. Because this was a retrospective cohort study, we were unable to attain any histopathologic data on these placental specimens, which limits our ability to understand molecular differences. Future analyses should consider integration of histopathology data with transcriptomic data, which was successfully used to provide molecular endotypes of preeclampsia and fetal growth restriction in other studies. Future studies may also consider other subtypes of sPTB caused by factors such as chorioamnionitis, PPROM, or preeclampsia, which we were unable to fully profile with the clinical data available. Integration of other omics data including proteomics data from matched samples may also generate a more robust molecular signature.

Our findings stand apart as a large transcriptome-wide assessment of sPTB—related differences in the placenta, quantified using RNA sequencing. We collected detailed covariate data, which was harmonized across both cohorts, allowing us to exclude PTBs with other pathologies and to adjust for potential confounding variables that were not addressed in our previous analyses of PTB, including key variables such as labor status. In addition, our study is more rigorous because we performed a stratified analysis using gestational age subgroups as defined by ACOG. This is clinically significant because most premature births are late preterm, necessitating better understanding of the molecular and pathologic differences between these subgroups.

**Conclusion**

Overall, this work supports the role of placental omics analyses in understanding the relationship between placental physiology and premature birth. The differences in gene expression presented here can be used to understand underlying pathologic changes in premature placentas, which can inform and improve clinical obstetrics practice.
GLOSSARY

**American College of Obstetricians and Gynecologists (ACOG):** a professional organization of physicians specializing in gynecology and obstetrics.

**Benjamini-Hochberg P value:** the Benjamini-Hochberg approach is applied to a list of P values to reduce the false discovery rate. The Benjamini-Hochberg method ranks the P values and calculates a new P value based on the gene’s rank divided by the total number of genes.

**Conditions Affecting Neurocognitive Development and Learning in Early childhood (CANDLE):** an observational, longitudinal cohort study that enrolled pregnant individuals from Shelby County Tennessee between 2006 and 2011.

**Con Founding variable:** a variable that may influence both the dependent variable (in this analysis our dependent variable was placental gene expression) and the independent variable (in this analysis our independent variable was preterm birth and preterm birth subgroups), causing spurious associations.

**Differentially expressed gene (DEG):** a gene with expression levels that are significantly different between 2 conditions based on a predefined statistical cutoff. In this manuscript, we used a cutoff of a false discovery rate—adjusted P < .05.

**Early term birth:** birth occurring between 28 and 34 weeks of gestation.

**Empirical analysis of digital gene expression data in R (EdgeR):** This is a popular R package distributed through Bioconductor that contains functions to perform differential expression analysis of RNA sequencing data. This package implements methodologies based on negative binomial distributions of the data.

**False discovery rate (FDR):** the rate of rejected null hypothesis that are false. In studies that make a large number of multiple comparisons (including transcriptomics analyses), it is essential to control for the false discovery rate because of the increased chances of rejecting a null hypothesis as a consequence of the large number of tests.

**Full-term birth:** birth occurring between 39 and 41 weeks of gestation, as defined by ACOG committee opinion 579.6

**Global Alliance to Prevent Prematurity and Stillbirth (GAPPS) biorepository:** a placental biorepository with a collection of standardized, high-quality samples and questionnaire data from pregnant individuals. This study uses GAPPS biorepository samples collected from Seattle and Yakima, Washington that are part of the ECHO PATHWAYS Consortium.

**Kallisto:** a program to quantify transcript abundances from RNA sequencing files using pseudoalignments, which align the reads to a transcriptome index.34

**Kyoto Encyclopedia of Genes and Genomes (KEGG):** KEGG is a database for understanding high-level functions and utilities of biologic systems from molecular-level information, especially large-scale molecular data sets generated by genome sequencing and other high-throughput experimental technologies. KEGG pathways are manually curated from existing biologic literature and are continuously updated.

**Limma-Voom:** the limma voom pipeline is a common statistical approach to analyze RNA sequencing data. VOOM (variance modeling at the observational level) calculates precision weights that estimate the mean-variance relationships of the log count data. This is then used in the LIMMA pipeline, which uses empirical bayes analysis to estimate the relationship between gene expression and covariate data.

**Late preterm birth:** birth occurring between 34 and 37 weeks of gestation

**Late term birth:** birth occurring between 41 and 42 weeks of gestation, as defined by ACOG committee opinion 579.6

**Log counts per million (LogCPM):** the log-scaled counts that are scaled for the library count sum and multiplied by a million. This metric corrects for sequencing depth and is a popular unit among differential expression analysis methods.

**Placental maturation signatures:** we created a curated list of 1420 gestational length signatures collected from maternal blood (n=522 genes)13 cell-free RNA across pregnancy (n=9 genes),15 as well as another list of genes for which the placental protein expression was significantly different between the second trimester and term gestations (n=954 genes)18

**Premature preterm rupture of membranes (PPROM):** rupture of the amniotic sac before the onset of labor and occurring before 37 weeks of gestation. In this analysis, this was characterized via medical record abstraction.

**Precision variable:** a variable that may influence the dependent variable (in this analysis our precision variable was placental gene expression) but is not related to our independent variable, which can lead to confounding.

**Rotational gene set testing (roast):** a pathway analysis tool that compares the average t statistic for each gene set with a null distribution estimated from random rotations of the residuals from the underlying model (10,000 rotations). We present results from the directional test, which first infers the direction of change for the gene set based on the preponderance of up- or down-regulated genes and then computes the average t statistic based on the subset of genes in the set with the inferred direction of change. The roast function was implemented in the Bioconductor limma package.41

**Spontaneous preterm birth (sPTB):** birth occurring before 37 weeks of gestation that is not medically indicated for conditions including preeclampsia, multifetal gestation, or placental abruption and is unrelated to cervical insufficiency or chorioamnionitis.

**Surrogate variable analysis (SVA):** a computational approach to overcome problems in the heterogeneity of data, including unmeasured confounding and cellular heterogeneity, by high dimensionality data reduction.39
Acknowledgments
The authors would like to thank the women who participated in the CANDLE and GAPPST study, as well as all the research staff and investigators. We would particularly like to thank Maureen Sorrells MPH, Cristina Urrutia BA, and Lisa Younglove MPH as well as the ECHO PATHWAYS data analysis team for their help with coordinating this.

References
24. Lapehn S, Paquette AG. The placental epi-genome as a molecular link Between prenatal exposures and fetal health outcomes Through the DOHaD hypothesis. Curr Environ Health Rep 2022 [Epub ahead of print].


Atlanta GA (Dr Marsit); Global Alliance to Prevent Preterm Birth and Stillbirth (GAPPS), Lynnwood, WA (Dr Litch); Department of Obstetrics and Gynecology, University of Washington, Seattle, WA (Gravett); and Department of Epidemiology, University of Washington, Seattle, WA (Drs Enquobahrie and Sathyanarayana).

Received March 19, 2022; revised July 1, 2022; accepted July 9, 2022.

The authors report no conflict of interest.

The ECHO PATHWAYS study was funded by the National Institutes of Health (NIH) under grants 1UG3OD023271-01 and 4UH3OD023271-03. The University of Washington Interdisciplinary Center for Exposures, Diseases, Genomics and Environment (EDGE Center) was funded by the NIH under grant P30ES007033. The Conditions Affecting Neuro-cognitive Development and Learning in Early Childhood (CANDLE) study was funded by the Urban Child Institute and the NIH under grant R01 HL109977. The Global Alliance to Prevent Prematurity and Stillbirth (GAPPS) Biorepository Program received financial support from numerous sources including the Seattle Children’s Hospital, the Bill and Melinda Gates Foundation, and the NIH. Work pertaining to this study was funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development under grant K99/ R00 HD096112.

This study was presented as an oral presentation at the 69th annual meeting of the Society for Reproductive Investigation, Denver, CO, March 18, 2022.

Corresponding author: Alison G. Paquette, PhD. Alison.Paquette@Seattlechildrens.org
**SUPPLEMENTAL FIGURE 1**

**A. Datasets used for curation of Placental Maturation Signatures**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Description</th>
<th>Date Source</th>
<th>Number of Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lien YC et al.</td>
<td>Human Placental Transcriptome Reveals Critical Alterations in Inflammation and Energy Metabolism with Fetal Sex Differences in Spontaneous Preterm Birth, Int J Mol Sci, 2021;22(15):7898, doi:10.3390/ijms22157898</td>
<td>Frozen placental samples underwent protein hydrolysis followed by LC-MS/MS on QExactive HF mass spectrometer. Differential proteins identified by TTest, FDR adjusted P&lt;0.05.</td>
<td>954</td>
</tr>
</tbody>
</table>


**B. Placental Maturation Signatures present our preterm birth analysis**


**SUPPLEMENTAL FIGURE 2**

**A. IL1B1 Expression across gestational age**

**B. IL1B1 expression in preterm subgroups**

A. Number of DEGs in each single cell cluster associated with PTL vs. TIL (Pique Regi et al, *elife*, 2019) also associated with sPTB in this analysis.

B. Heatmap of Log Fold Change Differences Reported

A. Overlap of Genes whose expression in fetal chorioamniotic membranes was associated with Preterm Birth (Pereyra et al, BMC Med Genomics, 2019) and this analysis

Late Preterm

Preterm vs. Term

Preterm Fetal Membrane Signatures
(Pereyra S, BMC Med Genomics, 2019)

208

0

0

7

0

32

21

9

447

414

9

2

758

B. Enrichment analysis of 62 DEGs whose expression in both the placenta and chorioamniotic membranes is associated with sPTB

Chemokine signaling pathway

Osteoclast differentiation

Fc gamma R-mediated phagocytosis

Parathyroid hormone synthesis, secretion and action

Cytokine–cytokine receptor interaction

NOD-like receptor signaling pathway

Neutrophil extracellular trap formation

Inflammatory mediator regulation of TRP channels

Viral protein interaction with cytokine and cytokine receptor

NF–kappa B signaling pathway

TNF signaling pathway

Number of DEGs

0

2

4

6

8

10

C. Characterization of 62 DEGs whose expression in both the placenta and chorioamniotic membranes is associated with sPTB