

Title

Pattern of variation in DNA methylation during pregnancy among mothers who delivered preterm in the GARBH-Ini cohort

Author names:

Jagyashila Das¹, Indranil Bagchi¹, Shekhar Ghosh¹, Nitya Wadhwa², Uma Chadramouli Nachu^{2,4}, Ramachandran Thiruvengadam², Pallavi Kshetrapal², Shinjini Bhatnagar^{2*}, Partha P. Majumder^{1,3*}, Arindam Maitra^{1*}, GARBH-Ini Team^{1,2,3,4}

Author affiliations:

- 1: National Institute of Biomedical Genomics, Kalyani, India
- 2: Translational Health Science and Technology Institute, Faridabad, India
- 3: Indian Statistical Institute, Kolkata, India
- 4: St. John's Research Institute, Bengaluru

ORCID iD: Jagyashila Das¹(<https://orcid.org/0000-0001-7776-2378>)
Arindam Maitra^{1*}(<https://orcid.org/0000-0002-0249-8856>)

*Author for correspondence:

Arindam Maitra^{1*}, am1@nibmg.ac.in
Partha Pratim Majumder, ppm1@nibmg.ac.in
Shinjini Bhatnagar, shinjini.bhatnagar@thsti.res.in

Author Contributions: PPM, SB, AM conceptualized the study. SB, NW, UCN, RT, PK conducted the participant recruitment, clinical data and biospecimen collection, storage and DNA extraction. DNA methylation data generation and analysis was undertaken by JD, IB, SG, AM. Statistical analysis was undertaken by PPM, JD. Manuscript was written by JD, AM and reviewed and contributed by PPM, SB, NW, UCN, RT, PK.

Acknowledgements: Professor M. K. Bhan will always be remembered reverently for his critical scientific and technical feedback. The authors recognize the efforts of the personnel of the GARBH-Ini Cohort and laboratory support from the Core Technologies Research Initiative (CoTeRI). We would like to thank Sahana Ghosh of NIBMG for assistance on submission of the data in GEO. This work was funded by the Department of Biotechnology, Ministry of Science and Technology, Govt. of India (BT/PR9983/MED/97/194/2013) and for some components of the biorepository by the Grand Challenges India – All Children Thriving Program (supported by the Programme Management Unit), Biotechnology Industry Research Assistance Council (BIRAC/GCI/0114/03/14–ACT). PPM acknowledges support of his J. C. Bose Fellowship and National Science Chair. JD is supported by the Research Fellowship (NET) of the University Grants Commission (UGC), India. The authors are thankful to Mr. Sukhdev Sinha and Dr. Anamika Gambhir for their technical feedback and support.

Financial disclosure: This work was funded by the Department of Biotechnology, Ministry of Science and Technology, Govt. of India (BT/PR9983/MED/97/194/2013) and for some components of the biorepository by the Grand Challenges India – All Children Thriving Program (supported by the Programme Management Unit), Biotechnology Industry Research Assistance Council (BIRAC/GCI/0114/03/14–ACT). PPM acknowledges support of his J. C.

Bose Fellowship. JD is supported by the Research Fellowship (NET) of the University Grants Commission (UGC), India.

Information pertaining to writing assistance: None.

Ethical disclosure: The authors state that they have obtained appropriate institutional review board approval for all human experimental investigations. In addition, informed consent has been obtained from the participants involved in the study.

Peer Review: This article has been submitted to Epigenomics.

Abstract

Background: DNA methylation (DNAm) may play an important role in birth outcomes.

Material and Methods: Genome wide DNAm was analysed in peripheral blood DNA of women at multiple time points during gestation. A novel empirical method was used to identify CpG sites with high temporal variance in methylation associating with preterm birth.

Results: High variability at 1296 CpG sites from the promoter regions of 1197 genes significantly associated with PTB. These genes belonged to pathways involved in signalling by platelet derived growth factor, platelet homeostasis, collagen degradation, extracellular matrix and circadian clock.

Conclusions: The findings provide novel information which might help in development of predictive biomarkers of preterm birth outcome.

Key words

Epigenetic, DNA Methylation, Temporal, Preterm Birth, Pregnancy

Introduction

Preterm birth (PTB) is any live birth before 37 completed weeks of gestation [1]. It is the leading cause of neonatal morbidity. Individuals born preterm face greater risk of delayed development of cognitive and motor skills, sensory deficits, learning disabilities and have higher risk of cerebral

palsy, respiratory illness, immunodeficiency and other late onset diseases, etc. [2,3]. In India, 3.6 million of 27 million babies are born preterm every year [4]. PTB is a syndrome attributable to multiple etiologies [5,6]. Epidemiological studies suggest increased risk of PTB can be attributed to nutritional deficiencies, stress, and other factors such as smoking, maternal infection (bacterial vaginosis/intra-amniotic infection), and race [7]. These factors can modulate cellular functioning primarily via DNA methylation [8–10].

Biochemical and molecular signatures in peripheral blood provide a unique potential window into the health and phenotype of an individual. Longitudinal epigenomic profiling of peripheral blood of pregnant women across all three trimesters of pregnancy and at delivery is expected to provide an improved understanding of the physiological changes that take place during gestation, in relation to birth outcome, including PTB [11]. Currently, limited data are available on the role of maternal DNA methylation alterations in pregnancy outcomes; available in only African-American population [12,13]. Blood samples for these studies were collected either at labor or shortly after delivery and hence do not capture the longitudinal changes that take place during the course of pregnancy. A detailed review on existing studies of maternal DNAm during pregnancy discusses the existing limitations in the field [14].

A multidisciplinary Group for Advanced Research on Birth-outcomes - DBT India Initiative (GARBH-Ini) program, was established in 2014 to identify clinical, epidemiologic, genomic, epigenomic, proteomic, metabolic and microbial correlates of PTB [15]. A cohort (named the GARBH-Ini cohort) of pregnant women was established as part of this programme with the primary objective of generating a risk-prediction algorithm for preterm birth based on multidimensional risk factors assessed during pregnancy, using a time series approach. We have undertaken the present study, on the GARBH-Ini platform, to test the hypothesis that increased epigenomic variation in maternal blood DNA during pregnancy results in preterm birth. In particular, we hypothesized that instability in the extent of promoter DNA methylation during pregnancy of key genes that regulate

the period of gestation will be associated with preterm birth. The instability in methylation can be measured by the variance in the level of methylation during the course of pregnancy.

Thus, we posit that women who delivered preterm will exhibit greater temporal variance in DNA methylation during gestation, including at time of delivery, compared to those who delivered at term.

Methods

Study Participants

GARBH-Ini cohort was established at the district hospital in Gurugram, Haryana, India. Women are enrolled within 20 weeks of conception as assessed by ultrasonography and are followed up at least once during each trimester during the antenatal period until delivery and once within 6 months of post-partum. Detailed information is collected from the participants on their socio-demography, environmental and clinical details, nutritional parameters. Serial ultrasound scans are done across pregnancy for uterine, fetal and placental morphology and biospecimens are collected from each woman during the course of her pregnancy at defined intervals. The profile, detailed objectives and methodology of this cohort, have been published earlier [15].

This study was designed as case-control study nested in GARBH-Ini cohort. Approvals from Institutional Ethics Committees of Gurugram Civil Hospital, Translational Health Science and Technology Institute, National Institute of Biomedical Genomics and the collaborating hospitals were obtained. Each woman recruited into this study provided prior informed voluntary consent for participation. The study population from which the women who delivered preterm -- a live birth less than 37 completed weeks of gestation -- and at term were sampled included those who gave birth to singleton baby by spontaneous delivery without congenital abnormalities or medical complications during pregnancy and had visited the clinical site <14 weeks, 18-20, 26-28 and at delivery to contribute, with informed consent, biospecimens for analysis. Time of sample collection for each

study participant is provided in supplementary table 2. Mothers with concomitant complexities like intrauterine growth restriction (IUGR), uterine infection, preeclampsia, any medical or surgical illness, e.g., HIV, malignancy, immunodeficiency or a history of long-term intake of immunomodulators and steroids were excluded. A total of 17 pairs of mothers, who delivered at term and preterm, matched for month of delivery and parity, were selected for this study based on the criteria described above.

Sample collection and DNA methylation assay

Maternal peripheral blood samples were collected in EDTA vacutainers at 11-14 weeks, 18-20 weeks, 26-28 weeks of gestation and at the time of delivery for DNA extraction from each study participant. DNA was isolated using QIAamp blood midi kit (Qiagen) and quantified in Qubit fluorimeter (Thermo). Bisulphite conversion of genomic DNA was carried out using the EZ DNA methylation kit (Zymo Research) and used for Infinium MethylationEPIC BeadChip (Illumina) according to the manufacturer's guidelines. The processed chips were scanned using iScan reader (Illumina) and raw data were analyzed using the methylation module under the GenomeStudio software (Illumina) version 2011.1.

DNA methylation levels for each probe was converted from fluorescence intensities of methylated (M) and unmethylated (U) alleles to methylation level, ranging from 0 (no methylation) to 1 (completely methylated), given by $\beta = M / (M + U + 100)$ [16].

Probes with <3 beads in at least 5% of samples, along with 2991 non-CpG probes, 81493 CpG sites which were identified by Illumina as SNPs, 49 multi-hit probes along with 2991 non-cpG probes, and 17844 probes on the sex chromosomes were removed using ChAMP package version 2.8.3 [17–19] in R version 3.6.3 (2020-02-29) [20].

Illumina EPIC methylation array uses two types of probes which have different dynamic ranges as well as a type 2 probe bias. Beta Mixture Quantile dilation (BMIQ) was used to normalise for the two types of probes used in the array [21]. Quantile normalisation (QN) was performed to equalize

the distribution of probe intensities from each array [22,23]. Cell type proportion using the algorithm described previously for CD4+ T cells, CD8+ T cells, natural killer (NK) cells, B cells, monocytes, and granulocytes were estimated using champ.refbase command from the ChAMP analysis pipeline to account for the potential differences in DNA methylation (DNAm) that arise due to cell composition variability in whole blood [19,24].

We used Illumina Infinium EPIC manifest file in which annotation is based on hg19 reference sequence. We considered probes in promoter specific regions of the genome, i.e. from 200 to 1500 bases upstream of transcription start sites and 5' untranslated regions (TSS200, TSS1500, 5'UTR) for analysis. We have used probes with β value ranging from 0.2 to 0.8 (corresponding M value ranging from -2 to 2) for subsequent analysis [25]. We proceeded with 26,107 probes for subsequent analysis.

Statistical Method

We performed unpaired two-tailed t-tests to compare equality of mean values of the period of gestation (POG), maternal age at conception, and infant birth weight between the two groups of mothers. We tested whether sex of the child is associated with the status of the birth outcome using a chi-squared test.

To test the equality of temporal variances between the two groups over the four time points during gestation and at delivery at which biospecimens were collected, we adopted the following empirically-driven statistical approach.

For each mother and each promoter probe, we calculated V_{ij} , variance of $M_{ij}^{(1)}, M_{ij}^{(2)}, M_{ij}^{(3)}, M_{ij}^{(4)}$, where $M_{ij}^{(i)}$ denotes the M value of the j^{th} probe ($j=1,2,\dots,103517$) for the i^{th} mother (preterm group= 1,2,...17; term group= 18,19,...34). We then computed $\bar{V}_{(1)j} = \frac{1}{17} \sum_{i=1}^{17} V_{ij}$ and $\bar{V}_{(2)j} = \frac{1}{17} \sum_{i=18}^{34} V_{ij}$, the mean values of temporal variances across mothers in each of the two groups, for

the j^{th} probe; and, $d_j = (\bar{V}_{(1)j} - \bar{V}_{(2)j})/S_j$, where $S_j^2 = \text{Var}(V_{1j}, \dots, V_{17j}, V_{18j}, \dots, V_{34j})$. It may be noted that d_j is analogous to the t-statistic for testing equality of mean values.

In other words, d_j is the statistic to compare equality of temporal variances in methylation for the j^{th} probe. To test the statistical significance of the observed d_j , we calculated an empirical sampling distribution of the d_j -statistic by re-sampling M values with replacement for the j^{th} probe for the 34 mothers, using the spirit of bootstrapping. An observed value of d_j was declared to be significant at the 1% level if the value of the d_j belonged to the upper 1% tail of the empirical distribution.

We used the list of promoter probes for which the temporal variance between mothers who delivered at term and preterm were significantly different. We mapped these probes to promoter regions of genes. A brief overview of the workflow is represented in Supplementary figure 1 and supplementary table 1.

Pathway Enrichment Analysis

The 1197 genes identified as mentioned above were analysed for enrichment of biological pathways. REACTOME [26] and KEGG [27] data resources were used in this analysis. ClueGO V2.5.1 plugins [28] and CluePedia V1.5.1 [29] and of Cytoscape (v. 3.7.0) [30] were used to analyse interrelations of terms and pathways in biological networks and visualise known interactions of genes in pathways from the data matrix, respectively. For the enrichment, a right-sided hyper-geometric test was used. The Kappa score threshold value was set to ≥ 0.4 . Criteria to identify a pathway as significant were that the pathway (a) should have minimum of 10 genes and (b) the p-value corrected for multiple-testing using the Benjamini-Hochberg method should be < 0.05 .

Results

Participant details

The summary of demographic and other information of the participants are provided in Table 1, and Supplementary Figure 2. For the 17 pairs of mothers, who delivered preterm (mean POG 33.1 ± 2.8 weeks) and at term (mean POG 39.2 ± 1.2 weeks), there were no significant difference in mean maternal age ($t = -1.5765$, $df = 32$, $p = 0.1248$) or any association of sex of the child with status of birth ($\chi^2 = 1.91$, $df = 1$, $p = 0.1671$). As expected, there was a significant difference in mean weight of a child (mean weight PTB = 1.92kg, mean weight term = 2.77kg) at birth ($t = 5.0066$, $df = 32$, $p = 1.949 \times 10^{-05}$).

DNA methylation analysis

Of the 219,163 probes located in the promoter regions of genes, CpG sites with M values $\leq +2$ and ≥ -2 were not considered in all samples. This is to reduce the number of non-variable sites to further reduce the statistical burden of multiple testing in the subsequent analyses, as used in previous studies [31,32]. Data on the 26,107 CpG sites that remained were analyzed further. We identified 1296 probes from 1197 genes with standardized difference in mean values of temporal variance of methylation between the two groups of mothers above the 99th percentile of the distribution of empirical standardized differences. Mean and SD values are provided in supplementary Table 3.

The ten probes with the largest d-statistics belonged to *UST*, *OSBPL3*, *CGNLI*, *TBC1D24*, *NFYC*, *ZNF165*, *BMP5*, *CKM*, *C5orf47*, *TNRC6B* genes (Figure 1). The probe with the largest effect size of 1.76 in our analysis was cg03366867 which belonged to the promoter of *UST*.

The four CpG sites from *UST* (cg03366867), *OSBPL3* (cg00891520), *CGNLI* (cg26185468), *TBC1D24* (cg08190844) with the highest temporal variation in PTB group are depicted in Figure 2.

Pathway enrichment

Pathway analysis with the significant probes (1296 promoter probes from 1197 genes) that showed the highest standardised mean differences in temporal variance of DNA methylation between the two groups of mothers resulted in identification of twenty enriched pathways (BH corrected P <0.05). List of probes (1296) along with their corresponding d-statistics are provided in supplementary table 4. The relative abundance of the most significant pathways based on the BH corrected p-values are provided in Figure 3. The most significantly enriched pathway was signaling by platelet derived growth factor (*COL2A1*, *COL9A2*, *COL9A3*, *FURIN*, *GRB2*, *PLAT*, *RAPGEF1*, *SRC*, *STAT1*, *STAT3*, *STAT5B*, *THBS2*), followed by collagen degradation (*COL12A1*, *COL1A2*, *COL26A1*, *COL2A1*, *COL8A2*, *COL9A2*, *COL9A3*, *CTSB*, *CTSK*, *FURIN*, *MMP19*), platelet homeostasis (*ATP2B2*, *ATP2B4*, *FGR*, *GNAS*, *GNB1*, *GNG12*, *GNG4*, *GNG7*, *KCNMB1*, *ORAI1*, *PDE5A*, *PPP2R5A*, *PRKG2*, *SLC8A1*, *TRPC6*), circadian clock pathways (*AVP*, *CPT1A*, *CRY2*, *F7*, *NCOA1*, *NCOA2*, *NCOR1*, *NR3C1*, *PPPICA*, *RAI1*, *UBB*). Extracellular matrix (ECM) proteoglycans and ECM-receptor interaction genes also exhibited higher temporal variation in PTB delivering mothers compared to those who delivered at term. The list of the 14 enriched pathways are provided in Table 2.

Discussion

Pregnancy is a synchronized cascade of multiple physiological events triggered by changes in gene expression [32]. In view of the emerging perception that adverse pregnancy outcomes like PTB might result from dysregulation of this cascade, we hypothesized that dysregulation in DNA methylation changes during pregnancy may be a good predictor of PTB. Some studies have been initiated on maternal blood transcriptomic alterations during healthy pregnancy and adverse events associated with pregnancy [33]. However, the inherent problems associated with the transient nature of the alterations in mRNA levels in whole blood and the impact on stability of these mRNA by non-physiological factors related to extraction and storage, might limit the ability of such

investigations. On the other hand, DNA methylation, used in our study, is expected to be unaffected by the above problems [34,35].

In our report, we have used a novel statistical method of analyzing temporal variation in DNA methylation between two groups of mothers with contrasting birth outcomes which allows for robust identification of large effects even in modest sample sizes. Methods to compare epigenome wide DNA methylation levels between these groups at individual time points may not have revealed the temporal differences between these groups of mothers during the course of pregnancy. We have identified those loci which exhibit larger variance over pregnancy in mothers delivering preterm birth giving compared to those who delivered at term. We view our findings as exploratory and recognize the importance of validating these in an independent set of samples. Nutrition and environmental exposures are important factors that may change the epigenetic pattern and it would be interesting to evaluate whether these findings hold true for women with different nutritional status as assessed by BMI and with variations in the environmental exposures like ambient air pollution.

Changes in DNA methylation of the promoter CpG sites might lead to altered gene expression[36,37]. We identified promoter CpG sites in 1197 genes with the highest difference in mean levels of temporal variance of methylation between the two groups of mothers, amongst which the one with the largest effect belonged to the gene coding Uronyl 2-Sulfotransferase (*UST*). A previous study on gene expression analysis of peripheral blood in preterm delivering women compared those delivering at term reported a 2.05 fold higher expression of *UST* [38]. Our results probably indicate that the longitudinal variability of promoter DNA methylation of this gene might be the underlying cause of this observation.

Placenta mediates rapid physiological changes in pregnancy. During later part of pregnancy, certain genes which are expressed in placenta during mid-gestation, remain highly expressed until term. One such gene is *BMP5* (Bone morphogenetic protein 5), which showed higher temporal variance in preterm delivering women. It has been observed earlier that *BMP5* gene expression in placenta

peaks in the second trimester (58.89 fold) compared to the first trimester, followed by a sharp decline at term in healthy term deliveries [39]. Studies have associated non-significant differential expression of *BMP5* gene in pregnancy related complications like preeclampsia, gestational diabetes, including small for gestational age as well as large for gestational age in the mid-gestation placenta. *BMP5* is implicated in bone and cartilage related disorders such as rheumatoid arthritis and osteoarthritis, also known to contribute to fetal development and growth. Probably, the variability in DNA methylation pattern of *BMP5* detected in our study might serve to be a PTB specific signature in maternal blood. However, strictly to maternal peripheral blood specific signature for PTB.

A substantial proportion of the genes (20.69%, corresponding to 12 genes) which were found to exhibit temporal variance in DNAm in preterm birth delivering mothers in this study to the Platelet-derived growth factor (PDGF) signaling pathway. PDGF is a major mitogen for fibroblasts, smooth muscle cells, and other cells [40]. Both PDGF and PDGF receptor (PDGFR) expression patterns are spatio-temporally regulated during development and physiological processes [41,42]. Mesenchymal expression of PDGFRs is low in vivo. Several factors induce PDGFR expression, including TGF- β , estrogen (probably linked to hypertrophic smooth muscle responses in the pregnant uterus), interleukin-1 α (IL-1 α), basic fibroblast growth factor-2 (FGF-2), tumor necrosis factor- α , and lipopolysaccharide [43]. Furthermore, increased expression of PDGF-A is observed in human uterine smooth muscle cells during the physiological hypertrophy of pregnancy [44]. Platelet turnover is usually within normal range during pregnancy, except during the third trimester when gestational thrombocytopenia sets in, which is associated with higher rate of cesarean delivery compared to women who do not have this condition[45,46]. Increased platelet aggregation has been observed during pregnancy [47]. Platelet activation and simultaneous release of vasoactive substances occur especially during repair of the utero-placental unit during obstetric complications [48]. Variability in promoter DNA methylation of PDGF signaling genes could lead to variable

expression which in turn might result in adverse outcomes like PTB by hindering downstream recruitment of vasoactive utero-placental repair.

Successful pregnancy is established by a series of physiological events including breakdown and remodelling of the extracellular matrix (ECM) in the endometrium. ECM proteoglycans play key roles in maintaining the connective tissue matrix and tensile strength of human fetal membranes and have been previously linked to pPROM (preterm premature rupture of membranes) [49]. ECM proteoglycans like decorin and biglycan act as biomarkers for increased risk of PPRM risk in maternal blood [49]. Excessive deposition of ECM in the connective tissue has also been associated with preeclampsia [50]. PPRM and preterm labor are discrete pathologic processes that can lead to spontaneous PTB. There is much overlap in the etiologic factors that predispose to these conditions [51], as well as exhibit much overlap in their clinical presentations. Alterations of DNA methylation leading of the ECM proteoglycans pathway genes leading to changes in their expression may therefore potentially explain their involvement in PTB outcome.

The primary component of ECM is collagen, which helps in endothelial cell proliferation, migration and differentiation, in order to form new vascular networks. Extensive collagen metabolism facilitates cervical remodelling. Cervical softening, shortening and dilation caused by early cervical remodelling can lead to spontaneous preterm birth(sPTB) [52]. The abundance of collagen in the fetal membranes decreases with gestational age [53,54]. Collagen degradation is initially brought by collagenase which is synthesized by cervical fibroblasts shortly before term delivery and activated during labor. Collagen degradation plays an essential role in prostaglandin (PG)-induced cervical ripening and the physiological maturation of the cervix in late pregnancy and at term [55,56].

Activity of collagen at the maternal fetal interface is therefore dynamic during the establishment of pregnancy. Interestingly, some of the genes which we identified belonged to the collagen degradation pathway. We hypothesize that temporal variance in the expression of these genes, as evidenced from the DNA methylation changes, might lead to preterm delivery.

Pregnancy is considered as a period susceptible to chronodisruption as pregnant women often undergo physiological, social and emotional changes, which may alter their usual lifestyle.

Chronodisruption can occur either due to polymorphisms in the core clock genes [57], or pregnant women can experience environmental factors that force them to sleep or wake at times different than their biological clock. Furthermore, nutrition, physical activity, and stress can also contribute to chronodisruption[58,59]. Such chronodisruptive factors can increase the risk of adverse pregnancy outcomes like miscarriage, low birth weight, and PTB [60–62]. Chronodisruption may affect pregnancy outcomes such as PTB, however, the interaction of maternal, fetal, and placental circadian systems to regulate gestational length is yet unclear. A study in a Canadian cohort have shown that lower levels of transcript of the circadian clock gene Cryptochrome 2 (*CRY2*) in second trimester peripheral blood from mothers delivering preterm compared to term delivering mothers, might be associated with an increased risk of sPTB [63]. The same study also identified degradation of the ECM as one of the most significant pathways that could potentially lead to sPTB. Our study shows similar results with the DNAm signatures in the Indian cohort, thereby particularly establishing the importance and relevance of maternal component towards preterm delivery.

The importance of the impact of maternal circadian rhythm on pregnancy and fetal development is only now being recognized. It has been proposed that the development of fetal suprachiasmatic nuclei (SCN) and fetal organs are entrained by maternal circadian oscillators which results in temporal order during fetal growth. This subsequently allows for postnatal integration of an adult-like circadian clock in the fetal system[64]. Further investigations in-depth are needed to fully comprehend the effect of chronodisruption on risk of preterm birth and its effect on infant health.

To our knowledge this is the first study that focuses on the associations between temporal DNAm levels and PTB in specifically an Indian cohort. Although novel, our study has certain limitations. This is a discovery study and demands subsequent validation of the findings in an independent group of mothers. Additionally, the mechanistic follow-up of our findings need to be undertaken, which could help in providing deep understanding of the biological implications.

This study has several limitations that should be considered when interpreting the findings. First, the lack of pre-pregnancy DNAm levels, which would allow us to examine the role of the change in these levels during the course of pregnancy in PTB. Second, our sample size was relatively small, thus future well-powered studies are needed in order to perform validation analyses which could shed more light on the DNAm variability that exist in the preterm delivering mothers compared to term delivery. To our knowledge this is the first study that focuses on the associations between temporal DNAm levels and PTB in specifically an Indian cohort.

Conclusion

We have tested our hypothesis that large variations in levels of genome-wide methylation correlate with preterm birth, using a novel method of analysis of longitudinal genome wide DNA methylation data in groups of individuals with contrasting pregnancy outcomes. Our study identified 1197 genes which exhibited significantly higher temporal variance in mothers delivering preterm as compared to those delivering at term. The results of our study, which is the first of its kind on maternal peripheral blood, has the potential to contribute towards improved understanding of PTB and might lead to development of a reduced invasive method of predicting risk of preterm delivery. We plan to validate our results in a larger set of mothers enrolled in GARBH-Ini, in order to to identify DNAm biomarkers which can be used for detection of PTB earlier during pregnancy.

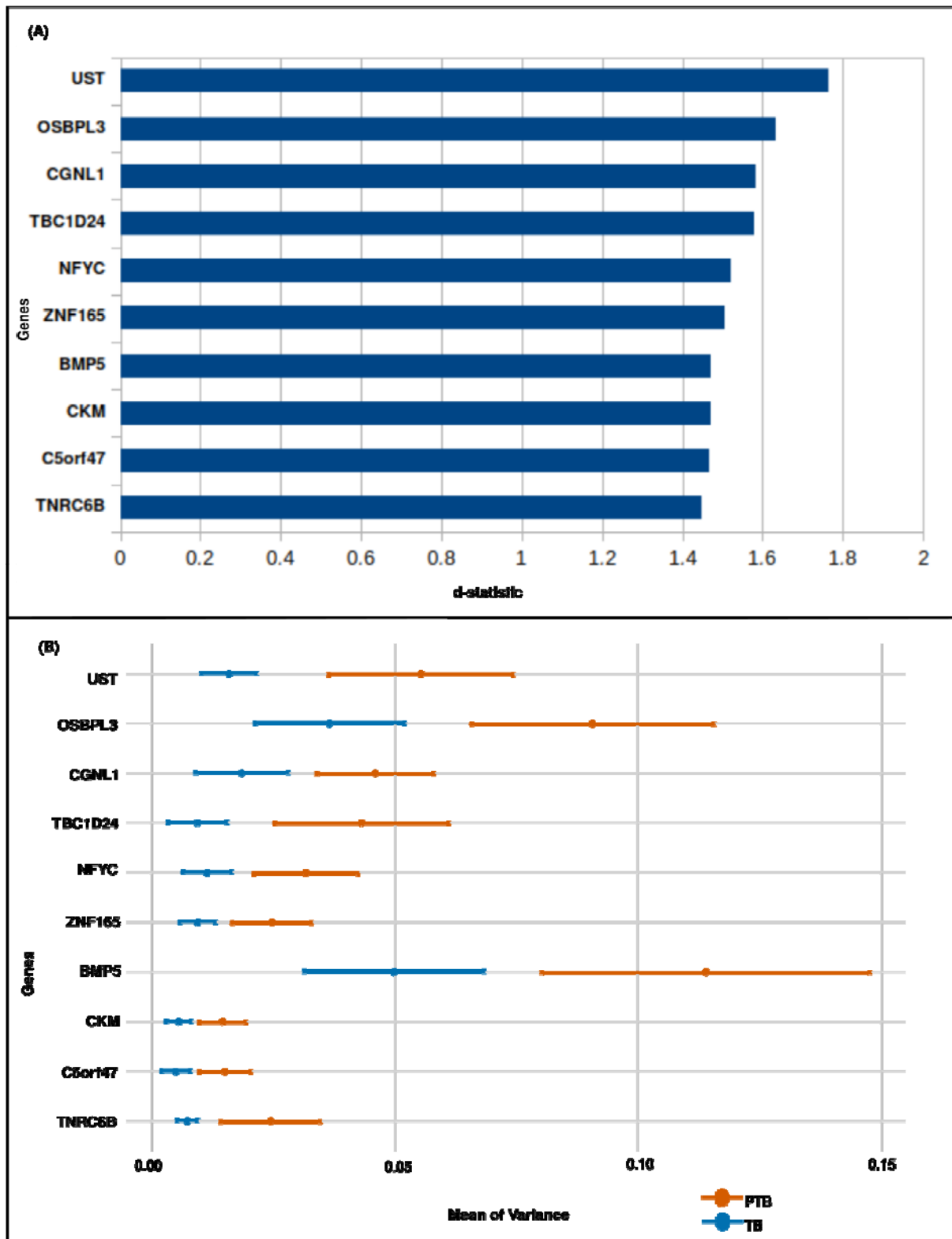


Fig 1: (a) Genes arranged in descending order of value of the d-statistic (b) Ten genes with the largest standardized mean difference in temporal variance of methylation between the mothers who delivered at term and preterm plotted with the mean of variance and confidence interval.

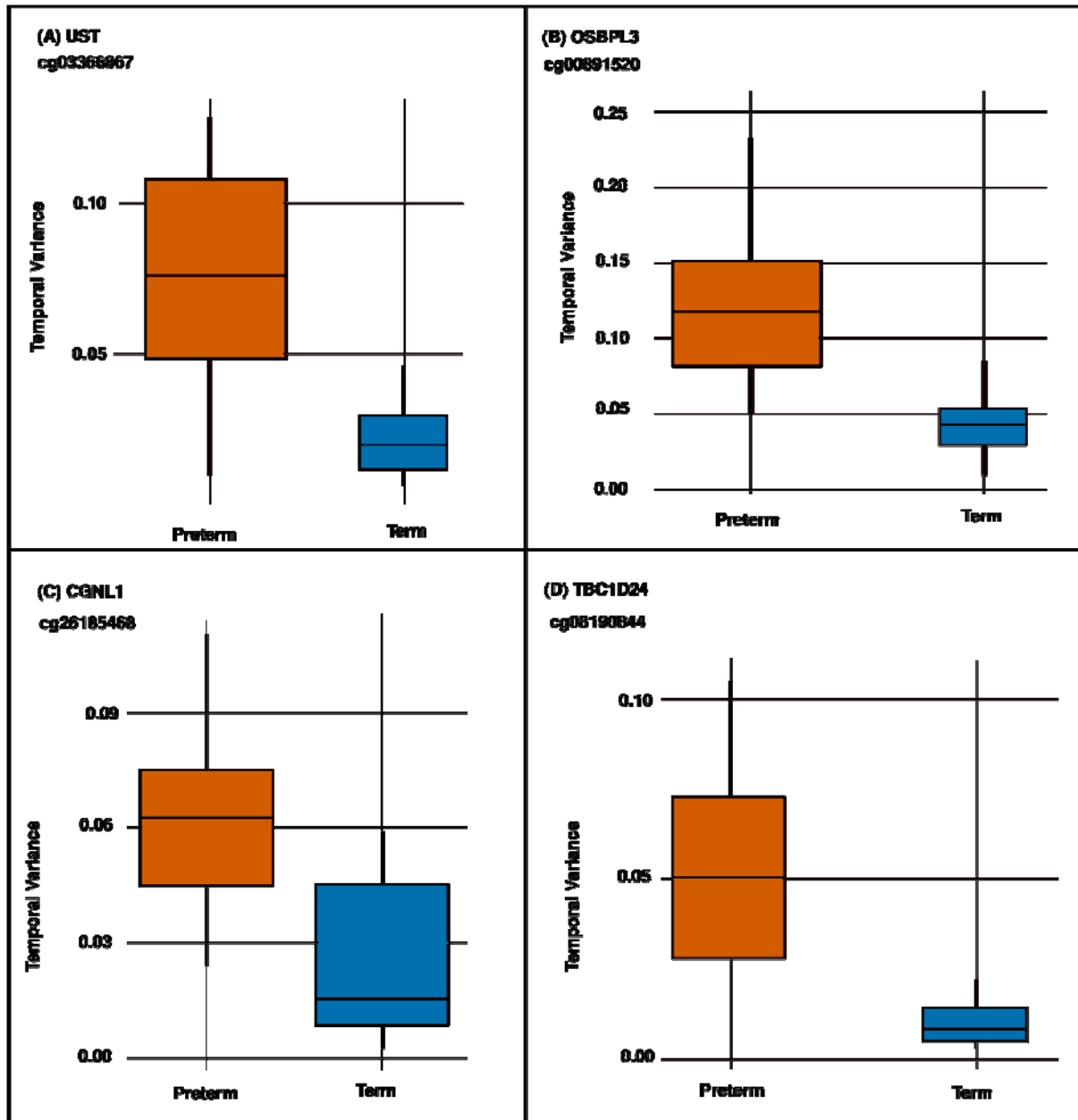


Fig 2: (A-D) Temporal variance of the probes with largest d-statistic in the x axis and status in the Y-axis.

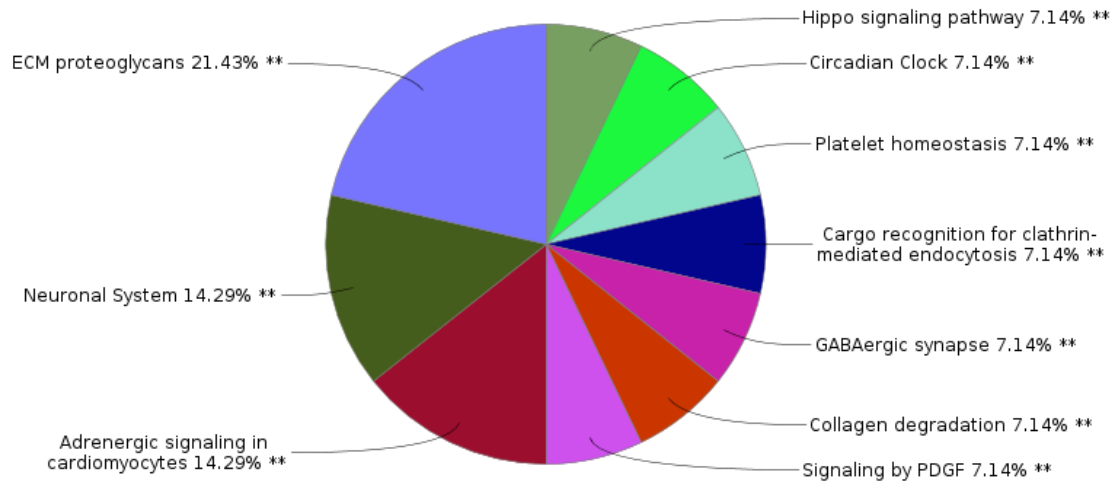


Fig 3: Percentage of terms per group, most significant pathways with BH corrected p-value (<0.05)

Variable	Statistic	Mothers delivering		
		Preterm (n=17)	Term (n=17)	
Maternal age (years)	Minimum	19	19	
	Maximum	27	29	
	Mean	21.3	22.8	
	SD	2.4	3.2	
Period of gestation* (weeks)	Minimum	27.1	37	
	Maximum	35.8	41.6	
	Mean	33.1	39.2	
	SD	2.8	1.2	
Weight of child at birth* (kg)	Minimum	1.04	2.29	
	Maximum	3.22	3.34	
	Mean	1.92	2.77	
	SD	0.62	0.32	
Proportion of female:male children at birth		5:12	10:7	
BMI category	Underweight	7 (41.18%)	3 (17.65%)	
	Normal	9 (52.94%)	11 (64.71%)	
	Overweight	1 (5.88%)	2 (11.76%)	
	Obese	-	1 (5.88%)	
History of vaginal discharge	Present	1 (5.88%)	1 (5.88%)	
	Absent	16 (94.12%)	16 (94.12%)	
History of bleeding per vagina	Present	2 (11.76%)	-	
	Absent	15 (88.24%)	17 (100%)	
Socioeconomic status	Upper Middle Class	3 (17.65%)	3 (17.65%)	
	Lower Middle Class	10 (58.82%)	3 (17.65%)	
	Upper Lower Middle Class	4 (23.53%)	10 (58.82%)	
	Lower Class	-	1 (5.88%)	
Toilet usage	Flush/ pour flush toilet	16 (94.12%)	17 (100.00%)	
	Bucket latrine	1 (5.88%)	-	
Laboratory characteristics	POG at sampling	Enrolment	11w3d (8w1d,12w6d)	10w3d (8w1d,12w4d)
		Visit 2	18w4d (18w4d,19w4d)	18w4d (18w2d,19w5d)
		Visit 3	26w2d (26w1d,26w3d)	26w2d (26w1d,26w3d)

Table 1: Summary of information of mothers included in the study.

*Difference is significant ($p < 0.05$) between term and preterm delivering mothers.

GO Term	Term PValue BH-Corrected	% Associated Genes	No.of genes	Associated Genes
Signaling by PDGF	0.01	20.69	12.00	COL2A1, COL9A2, COL9A3, FURIN, GRB2, PLAT, RAPGEF1, SRC, STAT1, STAT3, STAT5B, THBS2
Collagen degradation	0.03	17.19	11.00	COL12A1, COL1A2, COL26A1, COL2A1, COL8A2, COL9A2, COL9A3, CTSB, CTSK, FURIN, MMP19
Platelet homeostasis	0.01	17.05	15.00	ATP2B2, ATP2B4, FGR, GNAS, GNB1, GNG12, GNG4, GNG7, KCNMB1, ORAI1, PDE5A, PPP2R5A, PRKG2, SLC8A1, TRPC6
Circadian Clock	0.05	15.71	11.00	AVP, CPT1A, CRY2, F7, NCOA1, NCOA2, NCOR1, NR3C1, PPP1CA, RAI1, UBB
ECM proteoglycans	0.05	14.47	11.00	ACAN, COL1A2, COL2A1, COL9A2, COL9A3, DCN, ITGA2B, ITGA8, LRP4, MATN4, TNXB
Potassium Channels	0.04	14.14	14.00	GNB1, GNG12, GNG4, GNG7, KCNA3, KCNA4,

				KCNA6, KCNAB2, KCNH2, KCNJ10, KCNJ8, KCNK3, KCNMB1, KCNQ3
ECM-receptor interaction	0.05	13.64	12.00	COL1A2, COL2A1, COL9A2, COL9A3, GP1BA, GP6, ITGA2B, ITGA8, LAMC3, SV2B, THBS2, TNXB
GABAergic synapse	0.05	13.48	12.00	GABRG2, GNB1, GNG12, GNG4, GNG7, PRKACB, PRKCA, SLC38A1, SLC6A1, SLC6A12, SRC, TRAK2
Cargo recognition for clathrin-mediated endocytosis	0.05	13.21	14.00	AP2B1, AP2M1, AVP, CD3D, CHRM2, EPGN, EPN2, EPS15, GRB2, NECAP2, SLC2A8, STON1- GTF2A1L, SYT2, UBB
Dopaminergic synapse	0.05	12.12	16.00	CREB3L1, CREB5, GNAL, GNAS, GNB1, GNG12, GNG4, GNG7, MAPK10, MAPK12, PPP1CA, PPP2R2B, PPP2R5A, PRKACB, PRKCA, SCN1A
Hippo signaling pathway	0.04	11.69	18.00	BMP5, BMP7, BMP8B, BMPR1A, BMPR1B, CTNNA2, FRMD1,

				FRMD6, PARD6G, PPP1CA, PPP2R2B, TCF7, TEAD1, TEAD3, TEAD4, TP53BP2, TP73, YWHAG
Adrenergic signaling in cardiomyocytes	0.04	11.41	17.00	ATP2B2, ATP2B4, CACNB2, CACNG4, CREB3L1, CREB5, GNAS, KCNE1, MAPK12, PPP1CA, PPP2R2B, PPP2R5A, PRKACB, PRKCA, RAPGEF4, SLC8A1, TNNT2
Focal adhesion	0.05	10.55	21.00	COL1A2, COL2A1, COL9A2, COL9A3, GRB2, ITGA2B, ITGA8, LAMC3, MAPK10, MYLK, PAK5, PAK6, PARVG, PGF, PPP1CA, PRKCA, PXN, RAPGEF1, SRC, THBS2, TNXB
Neuronal System	0.01	10.24	42.00	AP2B1, AP2M1, APBA2, ARHGEF7, BCHE, CACNB2, CACNG4, CHRNA1, DLGAP1, GABRG2, GNAL, GNB1, GNG12, GNG4, GNG7, GRIK4, KCNA3, KCNA4, KCNA6,

It is made available under a [CC-BY-NC-ND 4.0 International license](#) .

				KCNAB2, KCNH2, KCNJ10, KCNJ8, KCNK3, KCNMB1, KCNQ3, NRXN2, PPFIA4, PPFIBP1, PPFIBP2, PRKACB, PRKAR1B, PRKCA, RPS6KA1, SHANK1, SHANK2, SLC38A1, SLC6A1, SLC6A12, SRC, SYT2, TUBB3
--	--	--	--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Table 2: Details of twenty significantly enriched pathways identified based on differential DNA methylation results.

References

1. Collins JL, Lehnherr J, Posner SF, Toomey KE. Ties that bind: Maternal and child health and chronic disease prevention at the centers for disease control and prevention. *Preventing chronic disease* 6(1) (2009).
2. Butler AS, Behrman RE, others. *Preterm birth: Causes, consequences, and prevention*. (Editions). National Academies Press (2007).
3. Moore ML. Preterm birth: A continuing challenge. *The Journal of perinatal education* 11(4), 37 (2002).
4. Sharma P, Khan S, Ghule M *et al*. Rationale & design of the PROMISES study: a prospective assessment and validation study of salivary progesterone as a test for preterm birth in pregnant women from rural India. *Reprod. Health* 15 (2018).
5. Chawanpaiboon S, Vogel JP, Moller A-B *et al*. Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. *Lancet Global Health* 7(1), e37–e46 (2019).
6. Romero R, Espinoza J, Kusanovic JP *et al*. The preterm parturition syndrome. *BJOG: An International Journal of Obstetrics & Gynaecology* 113, 17–42 (2006).
7. Nieuwenhuijsen MJ, Dadvand P, Grellier J, Martinez D, Vrijheid M. Environmental risk factors of pregnancy outcomes: A summary of recent meta-analyses of epidemiological studies. *Environmental health* 12(1), 6 (2013).
8. Menon R, Conneely KN, Smith AK. DNA methylation: An epigenetic risk factor in preterm birth. *Reproductive Sciences* 19(1), 6–13 (2012).
9. Hoyo C, Murphy SK, Jirtle RL. Imprint regulatory elements as epigenetic biosensors of exposure in epidemiological studies. *J. Epidemiol. Community Health* 63(9), 683–684 (2009).
10. Relton CL, Smith GD. Epigenetic epidemiology of common complex disease: Prospects for prediction, prevention, and treatment. *PLoS medicine* 7(10), e1000356 (2010).

11. Koh W, Pan W, Gawad C *et al.* Noninvasive in vivo monitoring of tissue-specific global gene expression in humans. *Proceedings of the National Academy of Sciences* 111(20), 7361–7366 (2014).
12. Hong X, Sherwood B, Ladd-Acosta C *et al.* Genome-wide DNA methylation associations with spontaneous preterm birth in US blacks: findings in maternal and cord blood samples. *Epigenetics* 13(2), 163–172 (2018).
13. Parets SE, Conneely KN, Kilaru V, Menon R, Smith AK. DNA methylation provides insight into intergenerational risk for preterm birth in african americans. *Epigenetics* 10(9), 784–792 (2015).
14. Das J, Maitra A. Maternal DNA Methylation During Pregnancy: a Review. *Reprod. Sci.* (2021).
15. Bhatnagar S, Majumder PP, Salunke DM, Advanced Research on Birth Outcomes—DBT India Initiative (GARBH-Ini) IG for. A pregnancy cohort to study multidimensional correlates of preterm birth in india: Study design, implementation, and baseline characteristics of the participants. *American journal of epidemiology* 188(4), 621–631 (2019).
16. Bibikova M, Barnes B, Tsan C *et al.* High density dna methylation array with single cpg site resolution. *Genomics* 98(4), 288–295 (2011).
17. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of infinium dna methylation beadchip probes. *Nucleic acids research* 45(4), e22–e22 (2017).
18. Nordlund J, Backlin CL, Wahlberg P *et al.* Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol.* 14(9), r105 (2013).
19. Tian Y, Morris TJ, Webster AP *et al.* ChAMP: Updated methylation analysis pipeline for illumina beadchips. *Bioinformatics* 33(24), 3982–3984 (2017).
20. R Core Team. *R: A language and environment for statistical computing. (Editions)*. R Foundation for Statistical Computing, Vienna, Austria (2020).
21. Teschendorff AE, Marabita F, Lechner M *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* 29(2), 189–196 (2013).

22. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19(2), 185–193 (2003).
23. Marabita F, Almgren M, Lindholm ME *et al.* An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. *Epigenetics* 8(3), 333–346 (2013).
24. Houseman EA, Accomando WP, Koestler DC *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC bioinformatics* 13(1), 86 (2012).
25. Du P, Zhang X, Huang C-C *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinf.* 11(1), 1–9 (2010).
26. Jassal B, Matthews L, Viteri G *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* 48(D1), 498–503 (2020).
27. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic acids research* 27(1), 29–34 (1999).
28. Bindea G, Mlecnik B, Hackl H *et al.* ClueGO: A cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25(8), 1091–1093 (2009).
29. Bindea G, Galon J, Mlecnik B. CluePedia cytoscape plugin: Pathway insights using integrated experimental and in silico data. *Bioinformatics* 29(5), 661–663 (2013).
30. Shannon P, Markiel A, Ozier O *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome research* 13(11), 2498–2504 (2003).
31. Wu P, Farrell WE, Haworth KE *et al.* Maternal genome-wide dna methylation profiling in gestational diabetes shows distinctive disease-associated changes relative to matched healthy pregnancies. *Epigenetics* 13(2), 122–128 (2018).
32. Fryer AA, Emes RD, Ismail KMK *et al.* Quantitative, high-resolution epigenetic profiling of CpG loci identifies associations with cord blood plasma homocysteine and birth weight in humans. *Epigenetics* 6(1), 86 (2011).

33. Maron JL, Johnson KL, Slonim D *et al.* Gene expression analysis in pregnant women and their infants identifies unique fetal biomarkers that circulate in maternal blood. *J. Clin. Invest.* 117(10), 3007 (2007).
34. Gomez-Lopez N, Romero R, Hassan SS *et al.* The Cellular Transcriptome in the Maternal Circulation During Normal Pregnancy: A Longitudinal Study. *Front. Immunol.* 10 (2019).
35. Knight M, Nair M, Tuffnell D, Shakespeare J, Kenyon S, Kurinczuk J. On behalf of mbrface-uk. Saving lives, improving mothers' care-lessons learned to inform maternity care from the uk and ireland confidential enquiries into maternal deaths and morbidity 2013–15. *National Perinatal Epidemiology Unit, University of Oxford, Oxford* 24–36 (2017).
36. Tarca AL, Romero R, Xu Z *et al.* Targeted expression profiling by rna-seq improves detection of cellular dynamics during pregnancy and identifies a role for t cells in term parturition. *Scientific reports* 9(1), 848 (2019).
37. Aran D, Sabato S, Hellman A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome biology* 14(3), R21 (2013).
38. Moarii M, Boeva V, Vert J-P, Reyat F. Changes in correlation between promoter methylation and gene expression in cancer. *BMC genomics* 16(1), 873 (2015).
39. Enquobahrie DA, Williams MA, Qiu C *et al.* Early pregnancy peripheral blood gene expression and risk of preterm delivery: a nested case control study. *BMC Pregnancy Childbirth* 9, 56 (2009).
40. Uusküla L, Männik J, Rull K *et al.* Mid-gestational gene expression profile in placenta and link to pregnancy complications. *PLoS One* 7(11), e49248. (2012).
41. Sporn MB, Roberts AB, Wakefield LM, Assoian RK. Transforming growth factor-beta: biological function and chemical structure. *Science* 233(4763), 532–534 (1986).
42. Basciani S, Mariani S, Spera G, Gnessi L. Role of Platelet-Derived Growth Factors in the Testis. *Endocr. Rev.* 31(6), 916–939 (2010).
43. Andrae J, Gallini R, Betsholtz C. Role of platelet-derived growth factors in physiology and medicine. *Genes Dev.* 22(10), 1276 (2008).

44. Heldin CH, Westermark B. Mechanism of action and in vivo role of platelet-derived growth factor. *Physiol. Rev.* 79(4), 1283–1316 (1999).
45. Mendoza AE, Young R, Orkin SH, Collins T. Increased platelet-derived growth factor A-chain expression in human uterine smooth muscle cells during the physiologic hypertrophy of pregnancy. *Proc. Natl. Acad. Sci. U.S.A.* 87(6), 2177–2181 (1990).
46. Aster RH. Gestational Thrombocytopenia. *N. Engl. J. Med.* 323(4), 264–266 (1990).
47. Lin Y-H, Lo L-M, Hsieh C-C, Chiu T-H, Hsieh T-T, Hung T-H. Perinatal outcome in normal pregnant women with incidental thrombocytopenia at delivery. *Taiwanese Journal of Obstetrics and Gynecology* 52(3), 347–350 (2013).
48. Louden KA, Pipkin FB, Heptinstall S, Fox SC, Mitchell JR, Symonds EM. A longitudinal study of platelet behaviour and thromboxane production in whole blood in normal pregnancy and the puerperium. *Br. J. Obstet. Gynaecol.* 97(12), 1108–1114 (1990).
49. Ayhan A, Akkök E, Urman B, Yarali H, Dünder S, Kirazli S. Beta-thromboglobulin and platelet factor 4 levels in pregnancy and preeclampsia. *Gynecol. Obstet. Invest.* 30(1), 12–14 (1990).
50. Underhill LA, Avalos N, Tucker R, Zhang Z, Messerlian G, Lechner B. Serum Decorin and Biglycan as Potential Biomarkers to Predict PPROM in Early Gestation. *Reprod. Sci.* 1933719119831790. (2019).
51. Ducray JF, Naicker T, Moodley J. Pilot study of comparative placental morphometry in pre-eclamptic and normotensive pregnancies suggests possible maladaptations of the fetal component of the placenta. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 156(1), 29–34 (2011).
52. Menon R, Fortunato SJ. The Role of Matrix Degrading Enzymes and Apoptosis in Repture of Membranes. *J. Soc. Gynecol. Investig.* 11(7), 427–437 (2004).
53. Shi J-W, Lai Z-Z, Yang H-L *et al.* Collagen at the maternal-fetal interface in human pregnancy. *Int. J. Biol. Sci.* 16(12), 2220 (2020).

54. Kumar D, Moore RM, Mercer BM, Mansour JM, Redline RW, Moore JJ. The physiology of fetal membrane weakening and rupture: Insights gained from the determination of physical properties revisited. *Placenta* 42, 59–73 (2016).
55. Skinner SJ, Campos GA, Liggins GC. Collagen content of human amniotic membranes: effect of gestation length and premature rupture. *Obstet. Gynecol.* 57(4), 487–489 (1981).
56. Huszar G, Naftolin F. The Myometrium and Uterine Cervix in Normal and Preterm Labor. *N. Engl. J. Med.* 311(9), 571–581 (1984).
57. Uldbjerg N, Ulmsten U, Ekman G. The ripening of the human uterine cervix in terms of connective tissue biochemistry. *Clin. Obstet. Gynecol.* 26(1), 14–26 (1983).
58. Reid KJ, Zee PC. Circadian rhythm disorders. *Semin. Neurol.* 29(4), 393–405 (2009).
59. Fuller PM, Lu J, Saper CB. Differential rescue of light- and food-entrainable circadian rhythms. *Science* 320(5879), 1074–1077 (2008).
60. Erren TC, Reiter RJ. Defining chronodisruption. *J. Pineal Res.* 46(3), 245–247 (2009).
61. Abeysena C, Jayawardana P, D. E. A. Seneviratne R. Maternal sleep deprivation is a risk factor for small for gestational age: a cohort study. *Aust. N. Z. J. Obstet. Gynaecol.* 49(4), 382–387 (2009).
62. Zhu JL, Hjollund NH, Andersen A-MN, Olsen J. Shift work, job stress, and late fetal loss: The National Birth Cohort in Denmark. *J. Occup. Environ. Med.* 46(11), 1144–1149 (2004).
63. Reschke L, McCarthy R, Herzog ED, Fay JC, Jungheim ES, England SK. Chronodisruption: An untimely cause of preterm birth? *Best Pract. Res. Clin. Obstet. Gynaecol.* 52, 60–67 (2018).
64. Zhou G, Duong TV, Kasten EP, Hoffmann HM. Low CLOCK and CRY2 in 2nd trimester human maternal blood and risk of preterm birth: A nested casecontrol study. *Biol. Reprod.* (2021).
65. Serón-Ferré M, Mendez N, Abarzua-Catalan L *et al.* Circadian rhythms in the fetus. *Mol. Cell. Endocrinol.* 349(1), 68–75 (2012).