Epigenetic biomarkers and preterm birth

Bongsoo Park, Rasheda Khanam, Vinesh Vinayachandran, Abdullah H. Baqui, Stephanie J. London, and Shyam Biswal

1Department of Environmental Health and Engineering, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA; 2Department of International Health, Johns Hopkins Bloomberg School of Public Health, International Center for Maternal and Newborn Health, Baltimore, MD 21205, USA; 3School of Medicine, Cardiovascular Research Institute, Case Western Reserve University, Cleveland, OH 44106, USA; and 4Department of Health and Human Services, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA

*Correspondence address. E-mail: sbiswal1@jhu.edu
Managing Editor: Dana Dolinoy

Abstract

Preterm birth (PTB) is a major public health challenge, and novel, sensitive approaches to predict PTB are still evolving. Epigenomic markers are being explored as biomarkers of PTB because of their molecular stability compared to gene expression. This approach is also relatively new compared to gene-based diagnostics, which relies on mutations or single nucleotide polymorphisms. The fundamental principle of epigenome diagnostics is that epigenetic reprogramming in the target tissue (e.g. placental tissue) might be captured by more accessible surrogate tissue (e.g. blood) using biochemical epigenome assays on circulating DNA that incorporate methylation, histone modifications, nucleosome positioning, and/or chromatin accessibility. Epigenomic-based biomarkers may hold great potential for early identification of the majority of PTBs that are not associated with genetic variants or mutations. In this review, we discuss recent advances made in the development of epigenome assays focusing on its potential exploration for association and prediction of PTB. We also summarize population-level cohort studies conducted in the USA and globally that provide opportunities for genetic and epigenetic marker development for PTB. In addition, we summarize publicly available epigenome resources and published PTB studies. We particularly focus on ongoing genome-wide DNA methylation and epigenome-wide association studies. Finally, we review the limitations of current research, the importance of establishing a comprehensive biobank, and possible directions for future studies in identifying effective epigenome biomarkers to enhance health outcomes for pregnant women at risk of PTB and their infants.

Key words: preterm birth (PTB); epigenome; diagnostics; methylation; chromatin

Introduction

Preterm birth (PTB, birth <37 completed weeks of gestation) is a major global public health problem because of its high incidence rate and associated high morbidity, mortality, and long-term disability [1]. Each year ~15 million (11.1% of all live births) worldwide occur preterm [2]. The incidence ranges from ~5% in many European countries to 18% in certain South Asian and sub-Saharan African countries [3]. More than 60% of all PTBs occur in South Asia and sub-Saharan Africa [3]. High-income countries are also affected, and the USA is one of the top 10 countries with the largest number of PTBs [3]. In the USA, there
are significant and persistent racial, ethnic, and socioeconomic differences in the rates of PTB. In 2018, the rate of PTB in various subpopulations was as follows: African American women (13.6%), American Indian/Alaska Native (11.3%), Hispanic (9.4%), White women (9.0%), and Asian/Pacific Islanders (8.7%) [4]. Preterm infants experience greatly increased mortality. In the USA, 34.3% of all infant deaths were classified as PTB associated, and 95% of PTB-associated deaths occurred in infants born at <32 weeks of gestation and with weight <1500 g [5]. Globally, PTB complications account for ~1 million deaths or 15.4% of all deaths in <5-year-old children, making them the leading cause of death for children in this age group [6]. PTB is also associated with greater lifetime disability, including reduced cognitive function, and visual and hearing impairment, as well as associated with a greater risk of future health conditions such as neurodevelopmental disorders and cardiovascular diseases [7, 8]. The economic burden and lifetime costs of PTB of a child to the caregiver are substantial [9].

The etiology of PTB is not completely understood because of the complex interplay of genetic, environmental, and host factors. There has been a significant effort to understand the disease mechanisms of PTB with several risk factors identified. Rubens et al. [10] has summarized the current state of knowledge and potential etiologic pathways involved in PTB. PTB risk analysis from a Danish family cohort showed more significant maternal genetic effects than paternal effects using historic family medical records [11]. Large cohort genome-wide association studies (GWAS) have described potential causal variants associated with PTB [12]. In addition, environmental exposure (e.g., cigarette smoke, heavy metals) is known risk factors for PTB [12]. Placental dysfunction is another important risk factor for PTB, implicated in major complications during pregnancy such as preeclampsia, fetal growth restriction, and stillbirth [13]. Maternal cardiometabolic disease also contributes to PTB risk. For example, obesity or obstetric factors, such as gestational diabetes and hypertension can increase the risk of PTB [14, 15]. PTB may be associated with enhanced oxidative stress as indicated by increased levels of lipid peroxidation biomarker 8-iso-PGF2α metabolite in urine samples [16]. Malnutrition of pregnant women can also increase the risk for PTB [17]. For example, deficiency of iron, folic acid, or vitamin D during pregnancy may increase the risk for PTB [17]. The alteration of maternal micronutrients (folic acid, vitamin B-12) has been found to increase homocysteine and oxidative stress levels, which may in turn cause epigenetic modifications that contribute to PTB [18].

In high-income countries, substantial reductions in PTB-related mortality have been achieved through improved management and care of women with premature labor and PTB, despite increasing rates of PTB [19]. However, the “Born Too Soon” partnership estimated that 50% of preterm deaths in low- to middle-income countries (LMICs) could be prevented by improving the delivery of feasible interventions, such as the use of antenatal corticosteroids in women with preterm labor [20], antibiotics usage for preterm–premature rupture of membranes [21], management of respiratory distress syndrome, and use of Kangaroo Mother Care [22]. The effectiveness of potential interventions requires early identification of increased risk for PTB. Screening for PTB in resource-limited settings is based on clinical grounds with limited predictive ability and not through ultrasonic or biochemical assessments [13]. Obstetric premature delivery screening approaches include cervical length and fetal fibronectin measurements, though their ability to predict early PTB is limited [23]. Determining PTB-related epigenome markers would help to improve understanding of the underlying molecular mechanisms and could serve as a supplementary diagnostic tool utilizing more accessible surrogate bio-samples.

**Table 1** summarizes some recent studies exploring epigenome biomarkers potentially related to PTB, with a focus on investigating blood tissue. We noticed that only DNA methylation has been studied as a primary epigenome biomarker, although other types of epigenome profiles are available. In this review, we will outline recent progress made in the development of epigenome assays and discuss the challenges and opportunities for the potential of epigenomic biomarkers for predicting PTB. The purpose of this review is to describe the currently available technologies and available literature on their application to the study of PTB.

**Genetic Susceptibility to PTB**

Genetic susceptibility is one of the key factors to determine the risk of PTB. Candidate gene and genome-wide analysis have been widely utilized to understand the role of the genetic background in environmental epidemiology field. For example, candidate genetic markers for PTB or low birth weights, such as ADCY5, CDKAL1, HHEX-IDE, and GCK, have been reported [30]. While several candidate genes have been associated with PTB, lack of replication in different cohorts is a major problem [31]. Efforts to find genetic risk and genetic variants associated with PTB continue, and GWAS have identified candidate regulatory loci associated with PTB. A recent GWAS revealed additional markers (EBF1, EEFSEC, AGT K2) substantially related to gestational length and as well as PTB [32]. By performing integrated analysis using curated genes and GWAS, Uzun et al. [33] reported 15 significant pathways using pathway-based PTB risk analysis (e.g., breast cancer estrogen signaling and oxidoreductase activity). Large cohort samples in GWAS have provided a foundational catalog of genetic variations related to PTB. However, PTB is also highly associated with environmental factors. Incorporating environmental interaction into GWAS can identify novel loci that would be missed in analyses of genetic variants alone [34].

**Environmental Exposures and PTB**

Several causes of spontaneous PTB have been identified, including maternal smoking during pregnancy, maternal stress, infectious disease, inflammation, and uterine distension [35, 36]. In addition to smoking, various environmental exposures have been identified as potential risk factors for PTB [37]. Several studies in the USA have identified ambient air pollution, including particulate matter, as a risk factor for PTB [38]. Meteorological conditions have also been implicated [39], raising the possibility that climate change could lead to increased PTB in the future. Martinez-Zamudio and Ha [40] summarized the evidence of increased PTB risk associated with exposure to environmental metals (chromium, arsenic, nickel, and cadmium). For example, arsenic exposure from drinking water can contribute to epigenetic reprogramming, resulting in the formation of inflammatory cells and more serious conditions like cancers [41]. Also, arsenic reported to affect birth weight through reduced gestational age and reduced maternal weight gain during pregnancy [42].

Many studies have investigated the mechanisms of one toxicant at a time, though such studies do not accurately represent the complex and real-life exposures to complex mixtures. Animal models serve as an informative approach to potentially
identify such complex and interactive molecular mechanisms. Ongoing research by the NEHS TaRGET II Epigenome Consortium will provide a unique opportunity to examine the effect of various toxicants in the prenatal-exposed mouse model [43]. Maternal exposure can leave epigenetic signatures that share the properties across all tissue type, and this meta-stable epialleles are another target we can investigate for developing effective epigenome biomarker [44–47].

**Epigenome-Wide Association Studies**

DNA methylation has been investigated as a major epigenetic mechanism for PTB; in particular, it has been reported that prenatal level of arsenic leads to changes in DNA methylation in placental tissue [48]. In addition, placental DNA methylation has been associated with cadmium [49]. Recent work from epigenome-wide association studies (EWAS) has shown that many of the prenatal exposures implicated in PTB are also related to changes in DNA methylation in cord blood. This finding has been most consistently demonstrated for maternal smoking in pregnancy, where there is widespread alteration in methylation in newborns [50]. EWAS have also shown reproducible associations between blood DNA methylation in newborns and maternal folate levels [51]. Exposure to traffic-related air pollution [52] and particulate matter [53] during pregnancy has been linked to cord blood methylation. EWAS have found exposure to metals, including lead and mercury, during pregnancy to be related to methylation in cord blood [54, 55]. For example, genome-wide analyses have shown that exposure to arsenic in utero has led to DNA methylation changes in multiple loci such as PLA2G2C, loci-SQSTM1, SLC4A4, and IGH [56]. Chromium found in airborne particles [57] or contaminated water causes various epigenetic modifications including DNA methylation [58] and histone modifications [40]. While a number of studies have suggested that the exposure-related methylation differences might mediate the effects of exposures on birth outcomes, the causal epigenomic mechanism remains unclear. For example, it has been shown that maternal smoking [59] complicates the use of mediation analysis to identify causal relationships [60, 61].

**Multi-Omics Resource for PTB Studies**

We checked for available multi-omics datasets in Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/ (1 February 2020, date last accessed) using the search terms “preterm birth” and “homo sapiens,” and found 53 studies and 3050 datasets (Table 2). For gene expression datasets, Affymetrix or Illumina Bead Array methods have been used generating 1756 transcriptome datasets. In addition, more recently, high-throughput sequencing methods have been employed generating 681 datasets. A total of 2437 transcriptome datasets (mRNA and miRNA) were found to be available for PTB meta-analysis in the NIH GEO repository. For methylation datasets, Illumina 450K arrays have been used in the larger population studies. We found 613 DNA methylation samples from 11 studies. GWAS datasets have been utilized to construct a comprehensive database of single nucleotide polymorphisms (SNPs) related to PTB called dbPTB [62]. Recently, the GWAS Catalog reported 63 associations from 15 studies [63]. There is no available PTB open chromatin analysis (ATAC-seq).

We summarized examples of research on genome-wide methylation in newborns to PTB or gestational age (Table 3), and the literature suggests that there are widespread epigenetic effects throughout the genome in PTB. The placenta has been investigated as an associated tissue for PTB, and particular immune cell types (e.g. monocytes and CD4 T cells) have also been studied for potential biomarkers in PTB diagnostics. Although the genetic variation and epigenome changes are not completely understood at the mechanistic level, and further investigations are needed to provide effective treatment and health management.

**Epigenetic Reprogramming as a Major Diagnostic Source for PTB**

The epigenome encompasses DNA methylation, histone modifications, high-level chromatin structure, and chromatin accessibility [81]. The most well characterized and highly studied epigenome is DNA methylation at cytosine residues [81]. DNA methylation has useful features for biomarker development. DNA methylation is covalent methylation that is biochemically stable. The most commonly assayed epigenetic marker in current epidemiological research at the population level is DNA methylation [82]. DNA methylation can be captured by a relatively low cost of Illumina DNA methylation arrays (e.g. 450K and EPIC) and sequencing-based methods (e.g. whole-genome bisulfite sequencing).

Histone markers, the post-translational modifications are the second epigenome marks on the N-terminal-exposed tails of histones [83]. Histones (H2A, H2B, H3, and H4) are highly conserved and positively charged proteins that are packaged with the DNA to form nucleosomes [84]. A total of 147 bp of DNA wraps around two copies of each histone molecule [85]. Packaging of the DNA not only serves to make the genome more compact but also to regulate the accessibility of cis- and trans-regulators for the DNA-templated processes such as replication, transcription, recombination, and repair. Covalent modifications in the DNA, N-terminal tails of histones, and to some extent, heritable non-coding RNA, are termed as epigenetic modifications [81, 83]. Most of these covalent modifications change the nucleosome composition, which in turn alters gene expression without any change in the underlying DNA.

---

**Table 2: Selected studies on epigenome biomarkers and PTB**

| Epigenome | Brief summary | Number of sites | Sample size |
|-----------|--------------|----------------|-------------|
| DNAm      | Fetal leukocyte DNA | 29 CpG sites | 50 [24] |
| DNAm      | Multi-omics profiles (RNA-seq, WGS, DNA methylation) using maternal blood | 72 genes | 791 [25] |
| DNAm      | Comparison between cord tissue and blood | 10 CpG sites | 1019 [26] |
| DNAm      | SLC9B1 methylation, fetal intolerance | 4 CpG sites | 57 [27] |
| DNAm      | Prenatal smoke exposure, epigenome marker | 15 CpG sites | 754 [28] |
| DNAm      | Birth weight, epigenome marker | 34 CpG sites | 711 [29] |
The chromatin immunoprecipitation with parallel DNA sequencing (ChIP-seq) is used to profile histone modifications and transcription factor-binding sites. It requires a large amount of input material and is therefore difficult to deploy in large cohort studies. In addition, there are substantial sample-to-sample variations in both transcription factor-binding sites [86] and histone modifications [87]. Due to this reason, there are very few examples of large-scale population-level studies using ChIP-seq. The recent advancement in molecular assay enables low input ChIP-seq to be used for particular biological assays involving a small number of cells [88]. The insulator protein CTCF is cell-type-specific; therefore, it has been used for cancer cell profiling. Interestingly, the plasticity in CTCF-binding occupancy is also highly correlated with changes in DNA methylation [89].

It is challenging to use the classical assay for chromatin accessibility, such as DNAse hypersensitivity. This issue has been largely overcome due to the recently developed ATAC-seq assay for chromatin accessibility, which has low input requirements and an easy-to-use protocol [90]. Since its introduction, this assay has been used in several population-level studies [91]. Finally, the study of the 3D structure of cells and chromatin has gained increased attention in the past years. The 4D Nucleome Project [92] is dedicated to exploring the structure of cells and chromatin. High-level chromatin structure may be a source to elucidate the most important epigenetic markers once the technology is matured for commercialization of a more cost-effective method for large-scale population-level studies. This emerging molecular technology is in the early stage, and the analytical method is being developed.

Applying Epigenome Biomarkers to Identify Prenatal Complications

DNA methylation can occur in diverse regulatory regions: promoters, enhancers, and insulators. A recent study has shown the significance of DNA methylation in regulating the transposable elements, including LTRs, LINEs, SINEs in response to environmental exposure [93]. In general, gene expression is negatively correlated with hypermethylation in the promoter. DNA methylation in CTCF-binding sites (four CpG sites in the motif) can disturb regulation of surrounding genes, causing epigenetic transcription reprogramming. As noted above, DNA methylation at CpG sites is a stable and sensitive biomarker for early

Table 2: Available transcriptome and epigenome datasets for PTB studies in GEO datasets

| Methods               | Descriptions and relevant assays | Number of studies | Number of samples |
|-----------------------|----------------------------------|-------------------|------------------|
| Gene expression       |                                  |                   |                  |
| Microarray            | Affymetrix, Illumina Beadarray   | 32                | 1789             |
| mRNA                  | High-throughput sequencing       | 9                 | 516              |
| Non-coding RNA        |                                  |                   |                  |
| miRNA                 | High-throughput sequencing or array | 9              | 379              |
| DNA methylation       |                                  |                   |                  |
| Illumina arrays       | 450K, EPIC arrays                | 12                | 692              |
| MeDIP-seq             | High-throughput sequencing       | 1                 | 30               |
| RRBS or WGBS         | High-throughput sequencing        | 1                 | 52               |
| Open chromatin        |                                  |                   |                  |
| ATAC-seq              | High-throughput sequencing        | None              | None             |

aWGBS datasets for prenatal exposure with smoke [64] are available in European Genome-phenome Archive (EGAS00001000455). The resources for WGBS are limited. The above number of the datasets was extracted in February 2020.

Table 3: Examples of DNA methylation studies using multiple tissue types

| Tissue type                  | Sample size | Brief study description                                           |
|------------------------------|-------------|-------------------------------------------------------------------|
| Placenta                     | 480         | PTB, birth weight and gestational age [65]                        |
| Placenta                     | 206         | Maternal tobacco smoke and gestational age [66]                  |
| Placenta                     | 158         | Locus-specific, MTR gene [67]                                    |
| Placenta, cord blood         | 83          | Preeclampsia, circadian clock [68]                                |
| Cord blood                   | 1753        | Gestational age [69]                                             |
| Cord blood                   | 397         | PTB, birth weight [70]                                            |
| Cord blood                   | 181         | PTB and infection [71]                                            |
| Cord blood                   | 141         | PTB and gestational age [72]                                      |
| Cord blood                   | 83          | Very PTB [73]                                                    |
| Cord blood                   | 73          | PTB and infection [71]                                            |
| Cord blood                   | 50          | PTB, African American [74]                                        |
| Cord blood                   | 50          | PTB, gestational age [24]                                         |
| Amnion                       | 121         | PTB and labor [75]                                                |
| Blood spots at birth         | 24          | PTB, effect of gestational age and long-term legacy [76]          |
| Maternal leukocytes          | 24          | PTB [77]                                                          |
| Blood                        | 144         | PTB, twin study [78]                                              |
| Maternal, cord blood         | 300         | PTB, African American [79]                                        |
| Buccal cell DNA              | 536         | Neurobehavioral variation, preterm infants (29 CpG) [80]          |

PTB, preterm birth.

The chromatin immunoprecipitation with parallel DNA sequencing (ChIP-seq) is used to profile histone modifications and transcription factor-binding sites. It requires a large amount of input material and is therefore difficult to deploy in large cohort studies. In addition, there are substantial sample-to-sample variations in both transcription factor-binding sites [86] and histone modifications [87]. Due to this reason, there are very few examples of large-scale population-level studies using ChIP-seq. The recent advancement in molecular assay enables low input ChIP-seq to be used for particular biological assays involving a small number of cells [88]. The insulator protein CTCF is cell-type-specific; therefore, it has been used for cancer cell profiling. Interestingly, the plasticity in CTCF-binding occupancy is also highly correlated with changes in DNA methylation [89].

It is challenging to use the classical assay for chromatin accessibility, such as DNAse hypersensitivity. This issue has been largely overcome due to the recently developed ATAC-seq assay for chromatin accessibility, which has low input requirements and an easy-to-use protocol [90]. Since its introduction, this assay has been used in several population-level studies [91]. Finally, the study of the 3D structure of cells and chromatin has gained increased attention in the past years. The 4D Nucleome Project [92] is dedicated to exploring the structure of cells and chromatin. High-level chromatin structure may be a source to elucidate the most important epigenetic markers once the technology is matured for commercialization of a more cost-effective method for large-scale population-level studies. This emerging molecular technology is in the early stage, and the analytical method is being developed.

Applying Epigenome Biomarkers to Identify Prenatal Complications

DNA methylation can occur in diverse regulatory regions: promoters, enhancers, and insulators. A recent study has shown the significance of DNA methylation in regulating the transposable elements, including LTRs, LINEs, SINEs in response to environmental exposure [93]. In general, gene expression is negatively correlated with hypermethylation in the promoter. DNA methylation in CTCF-binding sites (four CpG sites in the motif) can disturb regulation of surrounding genes, causing epigenetic transcription reprogramming. As noted above, DNA methylation at CpG sites is a stable and sensitive biomarker for early
detection of the impact of some environmental toxins. As the human embryo and fetus are susceptible to toxins in the pregnant mother’s environment [94], epigenome markers can be investigated to understand diseases induced by environmental exposures (phenotype–epigenome association).

Understanding the molecular mechanisms, including epigenetic reprogramming, is a critical step to gain insight into complex prenatal disease etiology. Toxins, such as arsenic, bisphenol A, lead, tetrachlorodibenzo-p-dioxin, diethylhexyl phthalate, and particulate matter (particles <2.5 nm), are being investigated in a consortium effort by the Toxicant Exposures and Responses by Genomic and Epigenomic Regulators of Transcription (TaRGET) program funded by NIEHS [43]. In efforts to identify reliable epigenome markers for PTB, the consortium and other researchers have been particularly investigating blood-driven biomarkers. For example, researchers have used neonatal blood to successfully identify CpG sites related to PTB [76, 95]. Interestingly, many of these CpG sites are located in genes related to the Notch signaling pathway. The TaRGET program supports research on multiple toxins and associated disease outcomes by applying the state-of-the-art multi-omics technology such as Whole-genome Bisulfite Sequencing (WGBS), ATAC-seq, ChIP-seq, and RNA-sequencing.

**Potential Advantages of WGBS Over Microarray-Based Methods**

Environmental scientists and epidemiologists have often utilized microarray-based DNA methylation assays (e.g., HumenMethylation27, HumenMethylation450, and Illumina EPIC array) due to relatively lower cost compared with whole-genome sequencing methods. While WGBS is a standard method for checking global methylation patterns in the genome, it remains expensive, and thus, it is not practical to use WGBS for population-level studies. In addition, because of the high cost, it is tempting to use low coverage in larger population studies creating possibilities for false-positive results given the paucity of studies with sequencing data for replication. In contrast, the Illumina EPIC array is affordable, and as a consequence, it is widely used in population-level studies enabling meta-analysis and replication.

WGBS can provide much higher resolution compared to the EPIC array, providing ~28M CpG sites in the human genome. WGBS can detect methylation levels at every single CpG site, which is important to the field of environmental epigenetics because methylation differences associated with environmental exposures are not uniformly distributed across the genome [96]. While this reflects the need for larger sample sizes and better assessment of exposures to reduce measurement error for exposures that are less prevalent or have smaller effects, it is possible that these exposures impact CpGs not covered by existing platforms. WGBS is also an unbiased assay compared with reduced representation bisulfite sequencing (RRBS) [97] or eRRBS [98], which are C/G restriction enzyme-based assays. Restriction enzyme digestion can create biased sequencing fragments, which can produce difficulties to examine reproducibility and to detect consistently differential methylation regions. Furthermore, WGBS provides more information regarding regulatory regions (Table 3).

Nonetheless, WGBS does have some limitations. For example, WGBS alone cannot distinguish between methylation (5 mC) and hydroxymethylation (5 hmC) sites. But the primary limitation of WGBS is the greater difficulty of analysis of sequencing data compared to that of microarray-based technologies.

As sequencing technologies continue to advance, sequencing costs are expected to decline, and expertise in analyzing sequencing will continue to increase. Using the recently released S4 flow cell of Illumina’s NovaSeq 6000, it is possible to process more than 30 WGBS samples in one lane, generating considerable savings. Environmental epigenetic research aims to sequence 30× coverage for WGBS. However, it is suggested to be a cost-effective approach to sequence 5–15× average coverage using more biological replicates (increasing number of samples) [99]. Portable, lower-cost sequencing devices such as Oxford Nanopore GridION could be effective for future diagnostics [100]. However, the accuracy level of these portable devices is poor compared to the Illumina high-throughput sequencing machines. We nonetheless anticipate that the accuracy of portable devices will improve in the future. A portable or small-size sequencer combined with effective biomarkers could be useful for quick diagnostics.

**Chromatin Accessibility and Population-Level ATAC-seq**

To perform comprehensive diagnostics using epigenome markers, it is also possible to examine open chromatin signatures (or accessibility) using blood samples. ATAC-seq is widely used for determining open chromatin regions in the genome, incorporating hyperactive Tn5 transposase into accessible regions of chromatin [101]. The open chromatin regions are important candidates as transcription-binding sites (e.g., sequence-specific or general transcription factors) and potential regulatory regions. ATAC-seq is specifically designed to effectively sequence samples consisting of a relatively small number of cells; for instance, blood ATAC-seq requires only 5000 cells [102]. Thus, it has become a popular tool in epigenomics research and can be used in population-level epigenome studies [103]. Omni-ATAC is an improved ATAC-seq protocol that allows chromatin accessibility profiling [104]. This advanced protocol significantly enhances signal-to-background ratios; therefore, the information contained under the peaks (bioinformatically predicted open chromatin regions) is higher. Moreover, this protocol can be applied to archival frozen blood and tissue samples, making it especially suitable for epigenetic epidemiological studies.

TaRGET consortium has developed a robust quality control matrix for high-quality ATAC-seq assays, and open-source scripts are available to the scientific community (http://www.targetepigenomics.org/). In particular, the bioinformatics group has been currently generating visualizations of sequenced ATAC-seq reads in a genome browser for validation and analysis of the binding of reads at different functional regions, including enhancers, transcription starts sites, gene body, and transcription end sites of genes (active or inactive), or at intergenic regions. The profiles of open chromatin could be classified using diverse genomic features as described in Table 4. In epigenetic diagnostics, it would be important to keep recordings, software version, and the reference datasets used for accurate validation.

**Cell-free DNA and Nucleosome Signatures**

Histones (H2A, H2B, H3, and H4) proteins are packaged with DNA to form nucleosomes [84], positioned in the specific
Table 4: Summary of reference features widely used in epigenome analysis

| Genomic features      | Descriptions and relevant assays                                                                 |
|-----------------------|--------------------------------------------------------------------------------------------------|
| Transcription start sites | Bivalent/poised TSS, flanking TSS, CAGE analysis [105]                                           |
| Promoter              | 1 kb from TSS, enriched marks: H3K4me1, H3K4me3                                                   |
| Enhancer              | H3K4me1, H3K27ac, p300 ChIP-seq                                                                  |
| Insulator             | CTCF-binding sites, Hi-C, Hi-ChIP                                                                |
| CpG island, sites     | 28 million CpG sites (human), 21 million CpG sites (mouse)                                       |
| CpG shore             | Surrounding area (2 kb) from CpG islands                                                          |
| Exon                  | miRNA, lincRNA, GENCODE prediction, RNA-seq                                                        |
| Intron                | Transcript, alternative splicing, GENCODE prediction, RNA-seq                                     |
| Repressed marks       | PolyComb, heterochromatin, H3K27me3 ChIP-seq                                                      |
| Activated genes       | Transcriptome, PolII, H3K36me3 ChIP-seq                                                          |
| TFBS                  | CTCF-binding sites, sequence-specific factor bindings (CG rich)                                   |
| MicroRNA              | Non-coding RNA, NIH epigenome roadmap [106]                                                       |

distance from each other. Nucleosome positioning indicates that nucleosomes are located in the genomic DNA sequence, and it varies between cell types and disease status. Recently, cell-free DNA (cfDNA), which is typically a short fragment of 200 bp, [107], has been isolated from circulating blood plasma and utilized for detecting signatures of nucleosome occupancy in cancer cells [108]. The size distribution of cfDNA is 147 bp (nucleosome size), and sometimes, it includes the size of a linker histone (167 bp) [109, 110].

Interestingly, cell-free fetal DNA has been detected in the plasma of a pregnant woman [110]. Dugoff et al. [111] assessed the connection between the first and second trimesters of cfDNA and confirmed that enhanced concentrations of fetal fractions during gestation (14.1–20.0 weeks) were substantially associated with enhanced incidence of PTB. van Boeckel et al. [112] investigated cell-free fetal DNA and spontaneous PTBs, and they found potentially pro-inflammatory properties of cell-free fetal DNA. Snyder et al. [108] hypothesized that cfDNA is the detritus of cell death, and the boundaries of cfDNA could reflect patients’ nucleosome signatures in target tissues, making cfDNA an important epigenetic surrogate material. Nucleosomes are favorably positioned near regulatory regions such as transcription start sites [113] and exon–intron boundaries [114, 115]. By creating nucleosome maps for healthy and diseased target and surrogate cells in experimental studies, potential disease risks can be categorized.

**The Challenges of Cell Heterogeneity for Developing Epigenomic Biomarkers**

Like most tissues, blood is a heterogeneous mixture of different cell types, and human blood cells are constantly generated in the bone marrow [116–118]. Hematopoietic stem cells are differentiated and divided into common myeloid progenitors and common lymphoid progenitors [116]. Later, common myeloid progenitors differentiate into megakaryocytes, erythrocytes, mast cells, and myeloblasts [116]. Due to the complex mixture of the resulting blood tissues, researchers sort the cells into the specific immune cell (e.g., monocyte) and examine the cell-type-specific DNA methylation profiles. Therefore, it is necessary to create an integrative deconvolution epigenomic analysis pipeline. Most studies of epigenetic changes and PTB to date (Table 3) have used whole blood, and thus statistical methods were used to adjust for cell composition, but to date, reference databases are small, and methods are evolving. In addition, some of these studies were done prior to the availability of newborn reference panels. In addition to the usual issues of estimating cell-type composition, studies of gestational age in relation to newborn blood DNA are complicated by the presence of nucleated red cells that decrease with longer gestation. Reference-free methods do not require prior cell-type-specific knowledge [119, 120] but run the risk of removing some of the true variations under study. An intermediate approach is a mixed model such as a semi-reference-free method [121]. In contrast, reference-based methods require a database of cell-type-specific epigenome profiles [122], and these are limited for newborns. Larger reference panels would advance the field.

**Emerging Single-Cell Technologies and Computational Methods**

Researchers have developed various single-cell-based epigenome assays including single-cell methylation [123], single-cell ATAC-seq [124], and single-cell RNA-seq [125]. These assays effectively address the cell heterogeneity issue, providing comprehensive epigenome or transcriptome profiles. Currently, these cutting-edge technologies have not been sufficiently developed for diagnostics usage, and these are not cost-effective, but in the future, due to advancement in this field, these technologies could be applied to many clinical and biomedical applications. The single-cell level epigenome is still under development, and it is not practical to apply it to the population-based epigenome biomarker development. In fact, effective biomarkers may not be the actual epigenetic mediators or modifiers involved in the molecular mechanisms of the disease. To find reliable epigenome biomarkers, we need large cohort studies to generate robust epigenetic profiles. Once such a large cohort omics dataset is produced, effective computational methods, and models could be derived. Based on the information on the genetic variation, SNP, and family history using electronic medical record, we can incorporate the additional epigenome profiles such as DNA methylation, histone modifications, chromatin accessibility, and cell-free nucleosome profiles. Combining large datasets requires computationally intensive analysis pipeline and sophisticated analytic methods. Emerging machine learning tools and artificial intelligence can facilitate the robustness of the prediction power for PTB.

**Biobank for Exploring Epigenomic Marks of PTB**

Around the world, there have been several cohorts on pregnancy and childhood outcomes. The Global Pregnancy and Childhood Epigenetics (PACE) Consortium, which connects over 39 studies in different countries [126], has examined associations between DNA methylation, primarily using the Illumina 450K array, and various prenatal exposures including smoking, air pollution, maternal body mass index, and alcohol intake [50, 52, 53, 127] and health outcomes including birth weight, maternal hypertension, and childhood asthma [128–131]. Combining data from many studies using meta-analysis has...
the advantage of generating large sample sizes and also provides built-in replication, reducing false-positive findings. PACE publications to date have used DNA methylation data primarily from newborns and children. Some PACE studies have also DNA methylation data from pregnant women, but no PACE meta-analyses have yet focused on these data. PACE cohorts are predominantly based in high-income countries. Although the burden of PTB is much higher in LMIC, there is hardly any epigenetic research to identify risks and markers of PTB from these countries. Efforts to pursue research in LMIC have been hindered due to a lack of functional biobanks in those settings. Recently, a number of biorepositories in pregnancy cohorts have been established in LMICs which will facilitate discovery of epigenetic markers of PTB [132, 133]. These biobanks can create new opportunities to conduct research to identify markers of PTB for LMICs.

Conclusions

The development of effective biomarkers that can predict increased risk of delivering PTB during pregnancy is essential. Genetic variation, while useful, will not be sufficient to explain the risks of PTB due to environmental factors. Epigenome signatures are dynamically changed as a response to environmental exposure. Thus, in addition to understanding the genetic variations, we need to develop epigenome biomarkers that can capture genome-wide epigenome signatures such as DNA methylation and chromatin accessibility. Multi-omics approaches to PTB will also be useful as the availability of multi-dimensional datasets increases. These datasets include transcriptome, epigenome, proteome, and metabolomics. In addition to multi-omics datasets, we need to continue developing the standardized and effective way to measure maternal stress, socioeconomic status, ethnic background that contribute the increase risk of PTB. In fact, the mechanisms of PTB are complex and poorly understood. Developing highly predictive biomarkers will enable public health interventions and also increase understanding of the underlying mechanisms. For developing such biomarkers, a large number of samples from prospective cohorts will be needed. Identification of omics markers in target tissue such as placental samples may be easier but not sufficient. It will be important to identify pregnant women at risk of PTB early in pregnancy so that they can be preferentially managed. Therefore, any biomarkers identified in placenta will have to be identified on surrogate tissue samples e.g. blood or even in samples that can be collected non-invasively (e.g. urine, saliva). PTB remains a challenge in the USA and globally as it contributes to health disparities. We encourage the translational bioinformatics and genomics community to contribute to deciphering the complex signatures of epigenome biomarkers and understanding their role in PTB.

Funding

S.B. is partly supported by NIEHS (U01ES026721). R.K. and A.H.B. are part of Bangladesh AMANHI study funded by the Bill and Melinda Gates Foundation and coordinated by the World Health Organization. R.K. and A.H.B. are also part of Multi-omics for Mothers and Infants (MOMI) consortium funded by the Bill and Melinda Gates Foundation. S.J.L. is supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (NIH Z01 ES49019).

Conflict of interest statement. None declared.

References

1. Behrman RE, Butler AS. In: RE Behrman and AS Butler (eds), Preterm Birth: Causes, Consequences, and Prevention, Washington (DC): National Academies Press (US), 2007.
2. Harrison MS, Goldenberg RL. Global burden of prematurity. Semin Fetal Neonatal Med 2016;21:74–9.
3. Blencowe H, Cousins S, Oestergaard MZ, Chou D, Moller A-B, Narwal R, Adler A, Vera Garcia C, Rohde S, Say L, Lawn JE. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. Lancet 2012;379:2162–72.
4. Talih M, Huang DT. Measuring progress toward target attainment and the elimination of health disparities in Healthy People 2020. Healthy People Statistical Notes, no 27. Hyattsville, MD: National Center for Health Statistics. 2016.
5. Callaghan WM, MacDorman MF, Rasmussen SA, Qin C, Lackritz EM. The contribution of preterm birth to infant mortality rates in the United States. Pediatrics 2006;118:1566–73.
6. Liu L, Oza S, Hogan D, Perin J, Rudan I, Lawn JE, Cousens S, Mathers C, Black RE. Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis. Lancet 2015;385:430–40.
7. Soleimani F, Zaheri F, Abdi F. Long-term neurodevelopmental outcomes after preterm birth. Iran Red Crescent Med J 2014;16:e17965.
8. Tanz IJ, Stuart JJ, Williams PL, Rimm EB, Missmer SA, Rexrode KM, Mukamal KJ, Rich-Edwards JW. Preterm delivery and maternal cardiovascular disease in young and middle-aged adult women. Circulation 2017;135:578–89.
9. Frey HA, Klebanoff MA. The epidemiology, etiology, and costs of preterm birth. Semin Fetal Neonatal Med 2016;21:68–73.
10. Rubens CE, Sadovsky Y, Muglia L, Gravett MG, Lackritz E, Gravett C. Prevention of preterm birth: harnessing science to address the global epidemic. Sci Transl Med 2014;6:262sr5.
11. Boyd HA, Poulsen G, Wohlffahrt J, Murray JC, Feenstra B, Melbye M. Maternal contributions to preterm delivery. Am J Epidemiol 2009;170:1358–64.
12. Bezold KY, Karjalainen MK, Hallman M, Teramo K, Muglia LJ. The genomics of preterm birth: from animal models to human studies. Genome Med 2013;5:34.
13. Sovio U, White IR, Dacey A, Pasupathy D, Smith GCS. Screening for fetal growth restriction with universal third trimester ultrasonography in nulliparous women in the Pregnancy Outcome Prediction (POP) study: a prospective cohort study. Lancet 2015;386:2089–97.
14. Leddy MA, Power ML, Schulkin J. The impact of maternal obesity on maternal and fetal health. Rev Obstet Gynecol 2008;1:170–8.
15. Kramer MR, Hogue CR. What causes racial disparities in very preterm birth? A biosocial perspective. Epidemol Rev 2009;31:84–98.
16. Rosen EM, van’t Erve TJ, Boss J, Sathyanarayana S, Barrett ES, Nguyen RH, Bush NR, Milne GL, McElrath TF, Swan SH, Ferguson KK. Urinary oxidative stress biomarkers and accelerated time to spontaneous delivery. Free Radic Biol Med 2019;130:419–25.
17. Gernand AD, Schulze KJ, Stewart CP, West KP, Christian P. Micronutrient deficiencies in pregnancy worldwide: health effects and prevention. Nat Rev Endocrinol 2016;12:274–89.

18. Dhole M, Joshi S. Altered maternal micronutrients (folic acid, vitamin B(12)) and omega 3 fatty acids through oxidative stress may reduce neurotrophic factors in preterm pregnancy. J Matern Fetal Neonatal Med 2012;25:317–23.

19. Chang HH, Larson J, Blencowe H, Spong CY, Howson CP, Cairns-Smith S, Lackritz EM, Lee SK, Mason E, Serazin AC, Walani S, Simpson JL, Lawn JE. Preventing preterm births: analysis of trends and potential reductions with interventions in 39 countries with very high human development index. Lancet 2013;381:223–34.

20. Vidaeff AC, Doyle NM, Gilstrap LC 3rd. Antenatal corticosteroids for fetal maturation in women at risk for preterm delivery. Clin Perinatol 2003;30:825–40, vii.

21. Kenyon S, Boulvain M, Neilson JP. Antibiotics for preterm rupture of membranes. Cochrane Database Syst Rev 2010;8:CD001058.

22. March of Dimes, TPFM, N & Child Health, Save the Children, World Health Organization. Born Too Soon: The Global Action Report on Preterm Birth. WHO Hinari. Geneva Switzerland, 2012.

23. Magro-Malosso ER, Saccone G, Di Tommaso M, Roman A, Berghella V. Exercise during pregnancy and risk of gestational hypertensive disorders: a systematic review and meta-analysis. Acta Obstet Gynecol Scand 2017;96:921–31.

24. Paretas SE, Conneely KN, Kilaru V, Fortunato SJ, Syed TA, Saade G, Smith AK, Menon R. Fetal DNA methylation associates with early spontaneous preterm birth and gestational age. PLoS One 2013;8:e67489.

25. Knijnenburg TA, Vockley JG, Chambwe N, Gibbs DL, Leinonen K, Linthorst J, Baveja R, Baker R, Solomon BD, Eley DE, Miller M, Slagel J, Elasady S, Roach JC, Kramer R, Durante-Dessau M, Langdon R, Relton CL, Davey Smith G, Melbye M, Teramo K, Nohr EA, Davey Smith G, Melbye M, Teramo K, Nohr EA, Davey Smith G, Melbye M, Jacobsson B, Muglia LJ. Genetic associations with gestational duration and spontaneous preterm birth. N Engl J Med 2017;377:1156–67.

26. Wu Y, Lin X, Lim IY, Chen L, Teh AL, MacIsaac JL, Tan KH, Gernand AD, Schulze KJ, Stewart CP, West KP, Christian P. Micronutrient deficiencies in pregnancy worldwide: health effects and prevention. Nat Rev Endocrinol 2016;12:274–89.

27. Wu Y, Lin X, Lim IY, Chen L, Teh AL, MacIsaac JL, Tan KH, Gernand AD, Schulze KJ, Stewart CP, West KP, Christian P. Micronutrient deficiencies in pregnancy worldwide: health effects and prevention. Nat Rev Endocrinol 2016;12:274–89.
46. Silver MJ, Kessler NJ, Hennig BJ, Dominguez-Salas P, Laritsky E, Baker MS, Coarfa C, Hernandez-Vargas H, Castelino JM, Routledge MN, Gong YY, Herceg Z, Lee YS, Lee K, Moore SE, Fulford AJ, Prentice AM, Waterland RA. Independent genomewide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environmental. Genome Biol 2015;16:118.

47. Gunasekara CJ, Scott CA, Laritsky E, Baker MS, MacKay H, Duryea JD, Kessler NJ, Hennell G, Wood AC, Hodges KR, Gandhi M, Hair AB, Silver MJ, Moore SE, Prentice AM, Li Y, Chen R, Coarfa C, Waterland RA. A genomic atlas of systemic interindividual epigenetic variation in humans. Genome Biol 2019;20:105.

48. Cardenas A, Houseman EA, Baccarelli AA, Quamruzzaman Q, Rahman M, Mostofa G, Wright RO, Christiani DC, Kile ML. In utero arsenic exposure and epigenome-wide associations in placenta, umbilical artery, and human umbilical vein endothelial cells. Epigenetics 2015;10:1054–63.

49. Everson TM, Funston T, Jackson BP, Hao K, Lambertini L, Chen J, Karagas MR, Marist C. Cadmium-associated differential methylation throughout the placental genome: epigenome-wide association study of two U.S. Environ Health Perspect 2018;126:017010.

50. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton CV, Xu C-J, Annesi-Maesano I, Anto´ JM, Joubert BR, den Dekker HT, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton CV, Xu C-J, Anto´ JM, Yousefi P, Relton C, Merid SK, Breton CV, Gao L, Volk HE, Feinberg JI, Ligthart S, Bohlin J, Ligthart S, Chen R, Coarfa C, Waterland RA. A genomic atlas of systemic interindividual epigenetic variation in humans. Genome Biol 2019;20:105.

51. Gruzieva O, Xu C-J, Breton CV, Annesi-Maesano I, Anto´ JM, Affrayment L, Ballester B, Bousquet J, Bastmante M, Charles M-A, de Kluizenaar Y, den Dekker HT, Duijts L, Felix JF, Gehring U, Guexens M, Jaddoe VVV, Koppelman GH, Melén E. Epigenome-wide meta-analysis of methylation in children related to prenatal NO2 air pollution exposure. Environ Health Perspect 2017;125:104–10.

52. Gruzieva O, Xu C-J, Yousefi P, Relton C, Merid SK, Breton CV, Gao L, Volk HE, Feinberg JI, Ligthart S, Bakulski K, Auffray C, Lemonnier N, Plusquin M, Chantout A, Herceg Z, Nauront TS, Pizziz C, Richardi L, Rusconi F, Vines P, Kovgorinovs I, Felix JF, Duijts L, den Dekker HT, Jaddoe VVV, Ruiz JL, Bastmante M, Anto´ JM, Suyner J, Vrijkheid M, Gutzkow K, Grazevilecien R, Hernandez-Ferrer C, Arnesi-Maesano I, Lepeule J, Bouquet J, Bergströmm A, Kull I, Söderhäll C, Kere J, Gehring U, Brunkefress K, Just AC, Wright RJ, Peng C, Gold DR, Klooog I, DeMeo DL, Pershagen G, Koppelman GH, London SJ, Baccarelli AA, Melén E. Prenatal particulate air pollution and DNA methylation in newborns: an epigenome-wide meta-analysis. Environ Health Perspect 2019;127:057012.

53. Wu S, Hivert M-F, Cardenas A, Zhong J, Rīfas-Shiman SL, Agha G, Colicino E, Just AC, Amarisirawardena C, Lin X, Litonjua AA, DeMeo DL, Gillman MW, Wright RO, Oken E, Baccarelli AA. Exposure to low levels of lead in utero and umbilical cord blood DNA methylation in project viva: an epigenome-wide association study. Environ Health Perspect 2017;125:087019.

54. Argos M, Chen L, Jasmine F, Tong L, Pierce BL, Roy S, Paul-Brutus R, Game VB, Harter KN, Farvez F, Rahman M, Rakibuz-Zaman M, Slavkovich V, Baron JA, Graziano JH, Kibriya MG, Ahsan H. Gene-specific differential DNA methylation and chronic arsenic exposure in an epigenome-wide association study of adults in Bangladesh. Environ Health Perspect 2015;123:64–71.

55. Salnikow K, Zhitkovich A. Genetic and epigenetic mechanisms in metal carcinogenesis and cocarcinogenesis: nickel, arsenic, and chromium. Chem Res Toxicol 2008;21:28–44.

56. Kondo K, Takahashi Y, Hirose Y, Nagao T, Tsuyuguchi M, Sharma S, Istrail S, Padbury JF. dbPTB: a database for putative transcriptional binding sites. Database (Oxford) 2012;2012:bar069.

57. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milan A, Morales J, Pendlington ZM, Welter D, Burtett T, Hindorff L, Flicek P, Cunningham F, Parkinson H. The new NHGRI-EBI Catalog of published epigenetic biomarkers.
Genome-wide association studies (GWAS Catalog). Nucleic Acids Res 2017;45:D896–901.

64. Bauer T, Trump S, Ishaque N, Thurmann L, Gu L, Bauer M, Bieg M, Gu Z, Weichenhan D, Mallin J-P, Röder S, Herberth G, Takada E, Mücke O, Winter M, Junge KM, Grützmann K, Rolle-Kampczyk U, Wang Q, Lawerenz C, Borte M, Polte T, Schlesner M, Schanne M, Wiemann S, Georg C, Stunnenberg HG, Plass C, Rippe K, Mizuguchi J, Herrmann C, Elia R, Lehmann I. Environment-induced epigenetic reprogramming in genomic regulatory elements in smoking mothers and their children. Mol Syst Biol 2016;12:e861.

65. Filiberto AC, Maccani MA, Koestler DC, Wilhelm-Benartzi C, Avissar-Whiting M, Banister CE, Gagne LA, Marsit CJ. Birth weight is associated with DNA promoter methylation of the glucocorticoid receptor in human placenta. Epigenetics 2011;6:566–72.

66. Maccani IZJ, Koestler DC, Houseman EA, Marsit CJ, Kelsey KT. Placental DNA methylation alterations associated with maternal tobacco smoking at the RUNX3 gene are also associated with gestational age. Epigenomics 2013;5:619–30.

67. Khot VV, Yadav DK, Shrestha S, Kaur L, Sundrani DP, Chavan-Gautam PM, Mehande SS, Chandak GR, Joshi SR. Hypermethylated CpG sites in the MTR gene promoter in preterm placenta. Epigenomics 2017;9:985–96.

68. van den Berg CB, Chaves I, Herzog EM, Willemsen SP, van der Horst CTJ, Steegers-Theunissen RPM. Early- and late-onset preeclampsia and the DNA methylation of circadian clock and clock-controlled genes in placental and newborn tissues. Chronobiol Int 2017;34:921–32.

69. Bohlin J, Häberl SE, Magnus P, Reese SE, Gjesing HK, Magnus MC, Parr CL, Page CM, London SJ, Nystad W. Prediction of gestational age based on genome-wide differentially methylated regions. Genome Biol 2016;17:207.

70. Vidal AC, Murphy SK, Murtha AP, Schildkraut JM, Soubry A, Wang X. Genome-wide DNA methylation associations with spontaneous preterm birth in US blacks: findings in maternal and cord blood samples. Epigenetics 2018;13:163–72.

71. Everson TM, Marsit CJ, Michael O’Shea T, Burt A, Hermetz K, Carter BS, Helderman J, Hofheimer JA, McGowan EC, Neal GR, Pastyrmak SL, Smith LM, Soliman A, DellaGrotta SA, Padbury JF, Lester BM. Epigenome-wide analysis identifies genes and pathways linked to neurobehavioral variation in preterm infants. Sci Rep 2019;9:6322.

72. Feinberg AP. The key role of epigenetics in human disease prevention and mitigation. N Engl J Med 2018;378:1523–34.

73. Fraser HB, Lami LL, Neumann SM, Kobor MS. Population-specificity of human DNA methylation. Genome Biol 2012;13:88.

74. Kasowski M, Grubert F, Heffelfinger C, Harirahan M, Asabere A, Waszak SM, Hablegger L, Rozowsky J, Shi M, Urban AE, Hong MY, Karczewski KJ, Huber W, Weissman SM, Gerstein MB, Korbel JO, Snyder M. Variation in transcription factor binding among humans. Science 2010;328:232–5.

75. Kornberg RD, Lorch Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell 1999;98:285–94.

76. Lugner K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 1997;389:251–60.

77. Kasowski M, Grubert F, Heffelfinger C, Harirahan M, Asabere A, Waszak SM, Hablegger L, Rozowsky J, Shi M, Urban AE, Hong MY, Karczewski KJ, Huber W, Weissman SM, Gerstein MB, Korbel JO, Snyder M. Variation in transcription factor binding among humans. Science 2010;328:232–5.

78. Dahl JA, Gilfillan GD. How low can you go? Pushing the limits of low-input ChIP-seq. Brief Funct Genomics 2018;17:89–95.

79. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M, Sandstrom R, Thurman RE, Kaul R, Daniels RM, Stamatoyanopoulos JA. Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res 2012;22:1680–8.

80. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 2013;10:1213–8.

81. Ucar D, Martínez-Espada M, Marches R, Rossi RJ, Uyar A, Wu T-C, George J, Stitziel NO, Palucka AK, Kuchel GA, Banchereau J. The chromatin accessibility signature of human immune aging stems from CD8(+) T cells. J Exp Med 2017;214:3123–44.

82. Dekker J, Belmont AS, Guttman M, Lesky VO, Lis JT, Lonvadars S, Mirny LA, O’Shea CC, Park PJ, Ren B, Politz JCR,
Shendure J, Zhong S. The 4D Nucleome Project. Nature 2017; 549:219–26.

93. Wang H, Lou D, Wang Z. Crosstalk of genetic variants. Front Genet 2018;9:969.

94. Perera F, Herbstman J. Prenatal environmental exposures, epigenetics, and disease. Reprod Toxicol 2011;31:363–73.

95. Liyanage V, Jarmasz J, Murugeshan N, Del Bigio M, Rastegar M, Davie J. DNA modifications: function and applications in normal and disease States. Biology (Basel) 2014;3:670–723.

96. Breton CV, Marsit CJ, Faustman E, Nadeau K, Goodrich JM, Dolinoy DC, Herbstman J, Holland N, LaSaille JM, Schmidt R, Yousefi P, Perera F, Joubert BR, Wiemels J, Taylor M, Yang IV, Chen R, Hew KM, Freeland DMH, Miller R, Murphy SK. Small-magnitude effect sizes in epigenetic end points are important in children’s environmental health studies: the children’s environmental health and disease prevention research center’s epigenetics working group. Environ Health Perspect 2017;125:511–26.

97. Meissner A. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res 2005;33:5868–77.

98. Garrett-Bakelman FE et al. Enhanced reduced representation bisulfite sequencing for assessment of DNA methylation at base pair resolution. J Vis Exp 2015;96:e52246.

99. Abugessaisa I, Mendez M, Hasegawa A, Noma S, Djebali S, Alam T, Medvedeva YA, Testa AC, Lipovich L, Yip C-W, Kojima L, Lizio M, Kawaji H, Kasukawa T, Itoh M, Burroughs AM, OJL, Gough J, Denisenko E, Schmeier S, Poulsen TM, Severin J, Lassmann T, Heutink P, Babina M, Wells CA, Kojima L, Gough J, Carninci P, Forrest ARR. An atlas of human T-cell chromatin accessibility genome-wide. Nat Methods 2017;14:407–10.

100. Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Hayashizaki Y, Carninci P, Forrest ARR. An improved ATAC-seq approach reveals lineage-specific re-opening of the open chromatin landscape via cooperation between BAF and p63. Nature 2010;464:219–26.

101. Orkin SH. Diversification of hematopoietic stem cells to some positioning and DNA methylation. Sci Immunol 2017;2:eaao1929.

102. Dugoff L et al. Cell-free DNA fetal fraction and preterm birth. Am J Obstet Gynecol 2016;215:231.e1–7.

103. Pedersen JS, Valen E, Velazquez AMV, Parker BJ, Rasmussen M, Lindgreen S, Lilje B, Tobin DJ, Kelly TK, Vang S, Andersson R, Jones PA, Hoover CA, Tikhonov A, Prokhortchouk E, Rubin EM, Sandelin A, Gilbert MTP, Krogh A, Willerslev E, Orlando L. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. Genome Res 2014;24:454–66.

104. Andersson R, Enroth S, Rada-Iglesias A, Wadelius C, Komorowski J. Nucleosomes are well positioned in exons and carry characteristic histone modifications. Genome Res 2009;19:1732–41.

105. Orkin SH. Development of the hematopoietic system. Curr Opin Genet Dev 1995;5:359–64.

106. Orkin SH. Development of the hematopoietic system. Curr Opin Genet Dev 1995;5:359–64.

107. Dzierzak E, Medvinsky A. Mouse embryonic hematopoiesis. Nat Rev Genet 2005;6:173–82.

108. Shendure J, Zhong S. The 4D Nucleome Project. Nature 2017; 549:219–26.

109. Wang H, Lou D, Wang Z. Crosstalk of genetic variants. Front Genet 2018;9:969.

110. Perera F, Herbstman J. Prenatal environmental exposures, epigenetics, and disease. Reprod Toxicol 2011;31:363–73.

111. Liyanage V, Jarmasz J, Murugeshan N, Del Bigio M, Rastegar M, Davie J. DNA modifications: function and applications in normal and disease States. Biology (Basel) 2014;3:670–723.

112. Breton CV, Marsit CJ, Faustman E, Nadeau K, Goodrich JM, Dolinoy DC, Herbstman J, Holland N, LaSaille JM, Schmidt R, Yousefi P, Perera F, Joubert BR, Wiemels J, Taylor M, Yang IV, Chen R, Hew KM, Freeland DMH, Miller R, Murphy SK. Small-magnitude effect sizes in epigenetic end points are important in children’s environmental health studies: the children’s environmental health and disease prevention research center’s epigenetics working group. Environ Health Perspect 2017;125:511–26.

113. Meissner A. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res 2005;33:5868–77.

114. Pedersen JS, Valen E, Velazquez AMV, Parker BJ, Rasmussen M, Lindgreen S, Lilje B, Tobin DJ, Kelly TK, Vang S, Andersson R, Jones PA, Hoover CA, Tikhonov A, Prokhortchouk E, Rubin EM, Sandelin A, Gilbert MTP, Krogh A, Willerslev E, Orlando L. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. Genome Res 2014;24:454–66.

115. Andersson R, Enroth S, Rada-Iglesias A, Wadelius C, Komorowski J. Nucleosomes are well positioned in exons and carry characteristic histone modifications. Genome Res 2009;19:1732–41.

116. Orkin SH. Development of the hematopoietic system. Curr Opin Genet Dev 1995;5:359–64.

117. Dzierzak E, Medvinsky A. Mouse embryonic hematopoiesis. Nat Rev Genet 2005;6:173–82.

118. Shendure J, Zhong S. The 4D Nucleome Project. Nature 2017; 549:219–26.

119. Wang H, Lou D, Wang Z. Crosstalk of genetic variants. Front Genet 2018;9:969.

120. Perera F, Herbstman J. Prenatal environmental exposures, epigenetics, and disease. Reprod Toxicol 2011;31:363–73.
121. Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted variation in microarray data. Biostatistics 2012; 13:539–52.

122. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 2012;13:86.

123. Zhu Y, Li Y, Lou D, Gao Y, Yu J, Kong D, Zhang Q, Jia Y, Zhang H. Sodium arsenate exposure inhibits histone acetyltransferase p300 for attenuating H3K27ac at enhancers in mouse embryonic fibroblast cells. Toxicol Appl Pharmacol 2018;357:70–9.

124. Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Zhu Y, Li Y, Lou D, Gao Y, Yu J, Kong D, Zhang Q, Jia Y, Zhang H. Sodium arsenate exposure inhibits histone acetyltransferase p300 for attenuating H3K27ac at enhancers in mouse embryonic fibroblast cells. Toxicol Appl Pharmacol 2018;357:70–9.

125. Felix JF, Joubert BR, Baccarelli AA, Sharp GC, Almqvist C, Svensson V, Vento-Tormo R, Teichmann SA. Exponential scaling of single-cell RNA-seq in the past decade. Nat Protoc 2018;13:599–604.

126. Karlsson R, Karmaus W, Grote V, Kajantie E, Magnus MC, Örtqvist AK, Eng C, Liu AH, Kull I, Jaddoe VV, Sunyer J, Kere J, Hoyo C, Annesi-Maesano I, Arshad SH, Richards AJ, Bergström A, Sharp GC, Felix JF, Burstamante M, Gruzieva O, Maguire RL, Gilliland F, Raizada K, Magnus P, Maguire RL, Marsit CJ, McArdle W, Mele´n E, Karlsson R, Karmaus W, Kuchars LK, Rifas-Shiman SL, Liu C, Burrows K, Zhao S, Magnus MC, Ha˚berg SE, Nystad W, Nohr EA, Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 2012;13:86.

127. Sharp GC, Arathimos R, Reese SE, Page CM, Felix J, Küpers LK, Rivas-Shiman SL, Liu C, Burrows K, Zhao S, Magnus MC, Duijts L, Corpeleijn E, DeMeo DL, Litonjua AA, Baccarelli A, Oken E, Holloway JW, Karmas W, Arshad SH, Dabelea D, Sørensen TIA, Laivuori H, Raikkonen K, Felix JF, London SJ, Hivert M-F, Gaunt TR, Lawlor DA, Retel CL. Hypertensive disorders of pregnancy and DNA methylation in newborns. Hypertension 2019;74:375–83.

128. Kazmi N, Sharp GC, Reese SE, Vehmeijer FO, Lahti J, Page CM, Zhang W, Rivas-Shiman SL, Rezwan FI, Simpkin AJ, Burrows K, Richardson TG, Santos Ferreira DL, Fraser A, Harmon QE, Zhao S, Jaddoe VVW, Vento-Tormo R, Teichmann SA. Exponential scaling of single-cell RNA-seq in the past decade. Nat Protoc 2018;13:599–604.

129. Kupers LK et al. Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birth weight. Nat Commun 2019;10:1893.

130. Reese SE, Xu C-J, den Dekker HT, Lee MK, Sikdar S, Ruiz-Arenas C, Merid SR, Rezwan FI, Simpkin AJ, Page CM, Perron P, Karagas MR, Arshad SH, Baccarelli A, Hoyo C, Annesi-Maesano I, Arshad SH, Koppelman GH, London SJ, Relton CL, Zuccolo L. Maternal alcohol consumption and offspring DNA methylation: findings from six general population studies in neonates reveals widespread differential DNA methylation associated with birth weight. Nat Commun 2019;10:1893.

131. Sharp GC, Salas LA, Monnereau C, Allard C, Yousef P, Eveson TV, Bohlin J, Xu Z, Huang R-C, Reese SE, Xu C-J, Baiz N, Hoyo C, Agha G, Roy R, Holloway JW, Ghantous A, Merid SR, Bakermans-Kranenburg MJ, Binder EB, Bouchard L, Lawlor DA, Maguire RL, Barcellos LF, Davey Smith G, Eskenazi B, Karmaus W, Marsit CJ, Mele´n E, Duijts L, Koppelman GH, London SJ, Epigenome-wide meta-analysis of DNA methylation and childhood asthma. J Allergy Clin Immunol 2019;143:2062–74.

132. Group AB-BS et al. Understanding biological mechanisms underlying adverse birth outcomes in developing countries: protocol for a prospective cohort (AMANHI bio-banking) study. J Glob Health 2017;7:021202.

133. Bhatnagar S, Majumder PP, Salunke DM. A pregnancy cohort protocol for a prospective cohort (AMANHI bio-banking) study. J Glob Health 2017;7:021202.

134. Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted variation in microarray data. Biostatistics 2012; 13:539–52.