

Published in final edited form as:

*Epigenomics*. 2013 June ; 5(3): . doi:10.2217/epi.13.18.

## Characterization of the DNA methylome and its interindividual variation in human peripheral blood monocytes

Hui Shen<sup>1,2</sup>, Chuan Qiu<sup>1,2</sup>, Jian Li<sup>1,2</sup>, Qing Tian<sup>1,2</sup>, and Hong-Wen Deng<sup>\*,1,2,3,4</sup>

<sup>1</sup>Center for Bioinformatics & Genomics, Department of Biostatistics & Bioinformatics, School of Public Health & Tropical Medicine, Tulane University, New Orleans, LA 70112, USA

<sup>2</sup>School of Medicine, University of Missouri-Kansas City, Kansas City, MO 64108, USA

<sup>3</sup>Center of System Biomedical Sciences, University of Shanghai for Science & Technology, Shanghai 200093, China

<sup>4</sup>College of Life Sciences & Bioengineering, Beijing Jiaotong University, Beijing 100044, China

### Abstract

**Aim**—Peripheral blood monocytes (PBMs) play multiple and critical roles in the immune response, and abnormalities in PBMs have been linked to a variety of human disorders. However, the DNA methylation landscape in PBMs is largely unknown. In this study, we characterized epigenome-wide DNA methylation profiles in purified PBMs.

**Materials & methods**—PBMs were isolated from freshly collected peripheral blood from 18 unrelated healthy postmenopausal Caucasian females. Epigenome-wide DNA methylation profiles (the methylome) were characterized by using methylated DNA immunoprecipitation combined with high-throughput sequencing.

**Results**—Distinct patterns were revealed at different genomic features. For instance, promoters were commonly (~58%) found to be unmethylated; whereas protein coding regions were largely (~84%) methylated. Although CpG-rich and -poor promoters showed distinct methylation patterns, interestingly, a negative correlation between promoter methylation levels and gene transcription levels was consistently observed across promoters with high to low CpG densities. Importantly, we observed substantial interindividual variation in DNA methylation across the individual PBM methylomes and the pattern of this interindividual variation varied between different genomic features, with highly variable regions enriched for repetitive DNA elements. Furthermore, we observed a modest but significant excess ( $p < 2.2 \times 10^{-16}$ ) of genes showing a negative correlation between interindividual promoter methylation and transcription levels. These significant genes were enriched in biological processes that are closely related to PBM functions, suggesting that alteration in DNA methylation is likely to be an important mechanism contributing to the interindividual variation in PBM function, and PBM-related phenotypic and disease-susceptibility variation in humans.

© 2013 Future Medicine Ltd

\*Author for correspondence: Tel.: +1 504 988 1310, Fax: +1 504 988 1706, hdeng2@tulane.edu.

### Financial & competing interests disclosure

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

### Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

**Conclusion**—This study represents a comprehensive analysis of the human PBM methylome and its interindividual variation. Our data provide a valuable resource for future epigenomic and multiomic studies, exploring biological and disease-related regulatory mechanisms in PBMs.

## Keywords

DNA methylation; interindividual variation; peripheral blood monocyte

Methylation of cytosine is an important epigenetic mechanism for transcriptional regulation and has profound impacts on human health and disease [1,2]. In human somatic cells, DNA methylation occurs almost exclusively at CpG sites. CpG sites are not distributed evenly along the human genome and their degrees of methylation vary considerably. For instance, many genomic regions termed ‘CpG islands’ (CGIs) contain concentrated CpG sites, which can be found in approximately 70% of promoters of human genes and are largely unmethylated [3]. By contrast, CpG sites in gene bodies and intergenic regions are commonly methylated [4,5]. The etiology of human complex diseases involves interactions between epigenetic, genetic and environmental factors. Epigenetic marks, such as DNA methylation levels, can modify the effects of the genotype, and function as an interface between the genome and the environment [6–8]. Dysregulation of DNA methylation has been implicated in the pathogenesis of various human complex diseases [9–11], such as cancers, cardiovascular disease, obesity and lupus.

Previous DNA methylation studies were commonly performed in cultured cell lines or tissue/blood cells that are composed of a mixture of different cell lineages. However, it has become increasingly evident that different cell types may have distinct DNA methylation profiles [7,12–14]. In addition, previous studies have suggested that there is considerable interindividual variability in DNA methylation, and that the methylation state can be influenced by a number of endogenous and exogenous parameters [15–18], such as gender, age and hormone exposure. Therefore, it is necessary to comprehensively characterize the genome-wide DNA methylation profiles in specific types of primary cells under well-controlled conditions.

Peripheral blood monocytes (PBMs) constitute 3–8% of human leukocytes in the blood and play multiple prominent roles in the immune response. PBMs can migrate from the bloodstream to other tissues (i.e., spleen, liver, lung and bone) and differentiate into various types of tissue-resident cells. Abnormalities in PBMs have been linked to a variety of human disorders, such as osteoporosis [19], alcoholism [20], autoimmune liver disease [21] and hypertension [22]. Using methylated DNA immunoprecipitation coupled with the next-generation sequencing (methylated DNA immunoprecipitation [MeDIP]-seq), we performed the first genome-wide DNA methylation analysis specifically in primary human PBMs in 18 unrelated healthy postmenopausal Caucasian females aged 50–56 years. The results provide novel insights into the PBM DNA methylation landscape and its variable interindividual patterns.

## Materials & methods

### Subjects

The study was approved by the University of Missouri-Kansas City (MO, USA) Institutional Review Board, and each participant signed an informed consent document before entering the study. A total of 18 unrelated healthy postmenopausal Caucasian females, aged 50–56 years, were recruited through the Kansas City Osteoporosis Study, a repertoire of approximately 6000 subjects collected for genomic and epigenomic studies of complex diseases/traits. All subjects were living in Kansas City (MO, USA) and its surrounding

areas, and were self-identified as being of European origin. An extensive set of exclusion criteria (Supplementary Material, see online at [www.futuremedicine.com/doi/suppl/10.2217/EPI.13.18](http://www.futuremedicine.com/doi/suppl/10.2217/EPI.13.18)) was adopted for this study in order to exclude subjects with diseases/conditions that may potentially lead to altered DNA methylation and/or gene expression patterns in PBMs. The basic characteristics of the study subjects are summarized in Supplementary Table 1.

### Sample preparation

Peripheral blood mononuclear cells (PBMCs) were isolated from 60 ml of freshly collected peripheral blood from each subject, using Lymphoprep™ (Axis-Shield, Oslo, Norway). PBMs were then isolated from the PBMCs with the Dynabeads® Untouched™ Human Monocytes kit (Life Technologies, CA, USA) using a previously established and routinely performed protocol [19,23–25]. The kit depleted unwanted cells (i.e., T and B cells) from the PBMCs, leaving PBMs free of the surface-bound antibody and beads with minimum disturbance. The isolated PBMs were visually checked for purity and counted under a microscope. Genomic DNA and total RNA were extracted from the freshly isolated PBMs with the AllPrep® DNA/RNA/Protein Mini Kit (Qiagen, CA, USA).

### MeDIP-seq

MeDIP-seq was performed by Arraystar Inc. (MD, USA) using a protocol based on Down *et al.* with minor modifications [26]. In brief, genomic DNA was sonicated to approximately 100–500 bp with a Bioruptor® sonicator (Diagenode, NJ, USA). A total of 1 µg of sonicated DNA was end repaired, A-tailed and ligated to single-end adapters following the standard Illumina (CA, USA) genomic DNA protocol. Methylated DNAs were enriched by immunoprecipitation with an anti-5-methyl-cytosine monoclonal antibody (MAb-081-100, Diagenode). Precipitated DNA fragments were purified using Qiagen MinElute® columns and PCR amplified using the single-end Illumina PCR primers. The PCR products were separated on 2% agarose gel to select fragments in the approximately 200–300 bp size range. The completed libraries were quantified and checked for quality with Agilent (CA, USA) DNA Bioanalyzer 2100. In addition, an aliquot of each library was analyzed by real-time PCR for a specific methylated site of H19 locus to confirm the enrichment of the methylated region (Supplementary Table 2).

The libraries were denatured with 0.1 M NaOH to generate single-stranded DNA molecules, loaded onto Illumina flow cells, amplified *in situ* using a standard cluster generation kit v2 (Illumina), and subsequently sequenced on an Illumina Genome Analyzer IIx (with single-end, 36-bp reads). Raw sequencing image analysis and base calling were performed using Solexa (CA, USA) pipeline v1.6. The MeDIP-seq data has been uploaded onto the Gene Expression Omnibus under accession number GSE39604 [101].

### Sequence alignment & data analyses

Sequencing reads that passed through the Solexa CHASTITY quality filter were aligned to the human genome reference sequence (GRCh37/hg19) using NOVOALIGN (v2.05; Novocraft, Selangor, Malaysia). DNA methylation profiles were inferred from the uniquely aligned reads by using the MEDIPS analysis package [27]. Briefly, each 36 bp read was extended to 250 bp along the plus or minus direction, depending on the strand information of the reads. The number of extended reads in every 50-bp bin throughout the genome was counted as the raw MeDIP-seq signals, which were then transformed into a reads per million format in order to assure that the methylation profiles derived from samples with different amounts of total sequencing reads were comparable. To correct for the influence of MeDIP enrichment by local CpG density [26,28], MeDIPS incorporated the coupling factor for each bin (representing the local CpG densities) into the transformed raw MeDIP-seq signals

(reads per million). The corrected methylation signals were termed relative methylation scores. For any specified genomic region, the mean relative methylation score values of the 50-bp bins that fall into the specified region can be calculated and further corrected for the relative CpG density of the region to produce the absolute methylation score (AMS), which is an absolute methylation estimate allowing for comparisons of methylation profiles between genomic regions with different CpG densities.

To evaluate genome-wide DNA methylation profiles, AMS values were calculated for 1-kb overlapping sliding windows (step size: 0.5 kb) throughout the genome. In addition, the methylation profiles (AMS) were estimated for various classes of genomic features, such as CGIs and promoters. Specifically, promoters were defined as the  $-0.5$  to  $+2$ -kb genomic regions around the transcription start sites (TSSs) of all RefSeq genes [29], and subdivided into three subclasses based on their CpG content [30]: high CpG promoters (HCPs; promoters containing a 500-bp interval with a CpG observed/expected ratio  $\geq 0.6$  and GC content  $\geq 55\%$ ); low CpG promoters (LCPs; promoters containing no 500-bp interval with CpG observed/expected ratio  $\geq 0.4$ ); and intermediate CpG promoters (ICPs; promoters that are not HCPs or LCPs). CGIs were identified as regions of  $\geq 200$  bp in length with a GC content  $\geq 50\%$  and with a CpG observed/expected ratio of  $\geq 0.6$  [31]. The genomic locations of CGIs were obtained from the University of California (CA, USA), Santa Cruz Genome Browser [32]. CGIs were further grouped into three subclasses on the basis of their positions relative to the closest RefSeq genes, including 5'-end CGIs (from  $-1$  to  $+0.3$  kb of a RefSeq gene TSS), intragenic CGIs (from  $+0.3$  kb of TSS to  $-0.3$  kb of the gene transcription end site) and intergenic CGIs, which are CGIs that do not fall into the other two categories. CGI shores were defined as 2-kb regions on either side of the CGIs. Annotations of other genomic features were obtained from the University of California, Santa Cruz Genome Browser.

To determine the common methylation patterns across individual PBM methylomes, the AMS values for a specific class of genomic features (or genome-wide 1-kb windows) were transformed into (0–1000) intervals in each individual sample, with 0 representing the lowest methylation level and 1000 representing the highest methylation level within all examined regions [27]. Based on the bimodal distribution of transformed AMS (trans-AMS) values observed in our samples (data not shown), the methylation status of a specific region can be subsequently defined as follows: trans-AMS  $\geq 600$ : highly methylated;  $460 < \text{trans-AMS} < 600$ : partially methylated; and trans-AMS  $\leq 460$ : lowly methylated.

### MassARRAY® EpiTYPER™ assay

Targeted methylation analysis was performed by using MassARRAY® EpiTYPER™ assays (Sequenom Inc., CA, USA) designed for selected genomic regions. The EpiTYPER assay is a MALDI-TOF mass spectrometry-based method for quantitatively measuring the mean DNA methylation level within each short DNA fragment. The EpiTYPER assays were carried out by the Genomics Shared Resources Facility of Roswell Park Cancer Institute (NY, USA). Briefly, 1  $\mu\text{g}$  of genomic DNA was modified by bisulfite treatment and regions of interest were amplified by PCR with specific primers, by which a T7-promoter tag was incorporated into the 5'-end of the reverse primers. Next, amplified fragments were transcribed *in vitro* into RNAs. The synthesized RNAs were treated with RNase A for base-specific cleavage. The cleavage products differ in molecular weight depending on the methylation state of the original genomic DNA, and the differences can be quantitatively analyzed by the MassARRAY analyzer system. DNA methylation standards (0, 20, 40, 60, 80 and 100% methylated) were used to control for PCR amplification bias.

## Gene ontology analysis

The program GOEAST [33,102] was employed to identify significantly enriched gene ontology terms among list of genes with specific characteristics (e.g., genes with highly methylated promoters across the 18 subjects). The p-values were calculated by hypergeometric tests and adjusted for multiple comparisons by stringent Yekutieli (false discovery rate under dependency) adjustment [34].

## Interindividual variation analysis

To identify genomic regions showing substantial interindividual variability in DNA methylation, we calculated median absolute deviation, a robust estimator of scale, for each 1-kb sliding window across the genome. Specifically, for each 1-kb window (i):

$$\text{Median absolute deviation}_i = \text{median} (| \text{AMS}_{ij} - \text{median} [\text{AMS}_i] |)$$

where  $\text{AMS}_{ij}$  denotes methylation level (AMS value) of individual j at region i, and  $\text{median} (\text{AMS}_i)$  is the median of the AMS values at region i across the subjects.

## Gene expression analysis

Genome-wide gene expression profiles were obtained for seven of the 18 subjects by using 1.0 Human Exon ST arrays (Affymetrix, CA, USA). Microarray data were processed using the robust multiarray analysis algorithm implemented in the R package of 'oligo' [35]. Gene level expression estimates were derived using the 'core' metaprobe list annotation release 21.

## Correlation of the methylation profile with gene expression level

To examine the role of promoter methylation in influencing transcription levels across different genes, we first calculated the average promoter methylation levels and average transcription levels for each gene, and then compared the methylation levels between genes of different transcription levels. In addition to comparison between individual genes, we also divided the assessed genes into quintiles based on the ranking of the mean expression levels of individual genes and compared between the gene groups for the DNA methylation profiles in a  $\pm 1.5$ -kb region surrounding the TSSs. Similar comparisons between gene groups of different expression levels were also carried out separately in genes associated with HCPs, ICPs and LCPs.

To evaluate the correlation between promoter methylation and gene expression levels across individuals, we computed the Pearson's correlation coefficient between promoter methylation levels and gene expression levels for each gene across all individuals. We defined a conservative threshold (Pearson's  $r \leq -0.678$ ) for determining significant negative correlation. This threshold was based on the fifth percentile of the empirical distribution of the correlation coefficient under the null hypothesis of no association between DNA methylation and gene expression levels across individuals, which was obtained from the permutations of the data sets from the seven subjects with both DNA methylation and gene expression data collected.

## Results

### Data generation & quality assessment

Using MeDIP-seq, we generated a total of approximately 553 million 36-bp single-end sequence reads in the 18 PBM samples. Of these raw reads, approximately 283 million reads passed the Solexa CHASTITY quality filter and were uniquely aligned to the human genome (GRCh37/hg19), resulting in an average of approximately 16 million uniquely



aligned high-quality reads per sample (Supplementary Table 1). Based on the high-quality mapped reads, we estimated coverage at genome-wide 50-bp bins using the MEDIPS package [27]. For each individual sample, on average 64% (range: 52–67%) of the 28.2 million CpG sites in the human genome were covered at least once (Supplementary Table 3). In addition, saturation analysis indicated that we had sufficient reads to obtain reproducible genome-wide methylation profiles (Supplementary Figure 1).

In order to confirm the validity of our data and the reproducibility of the analyses, we first checked the genome-wide DNA methylation patterns for several previously reported features. In agreement with previous reports [36], we observed that subtelomeric regions were frequently found to be highly methylated and genomic regions associated with lightly stained Giemsa bands often showed higher methylation levels than darkly stained regions (Supplementary Figure 2). We examined methylation signals at specific genomic regions, such as the region harboring the *GAPDH* gene. *GAPDH* is a ubiquitously expressed housekeeping gene, and thus the promoter region of this gene was expected to be unmethylated. Indeed, we consistently observed high levels of *GAPDH* expression (ranked in the top 1% most highly expressed genes among the 20,384 genes assayed on the Affymetrix exon arrays) and a lack of methylation at the *GAPDH* promoter in all samples (Supplementary Figure 3), whereas other regions surrounding the gene showed variable methylation status. Similar results were also observed at several other housekeeping gene regions (data not shown). Additionally, we further validated our MeDIP-seq data by confirming the DNA methylation status at a number of selected genomic regions (five promoters and three CGIs) using MassARRAY EpiTYPER assays, which demonstrated a high concordance rate (Pearson's  $r = 0.74\text{--}0.92$ ) between the two approaches (Supplementary Table 4). Taken together, we have successfully generated comprehensive and reproducible genome-wide DNA methylation profiles in our PBM samples.

### Characteristics of the PBM methylome

Based on the MeDIP-seq data, we characterized DNA methylation patterns at 20 distinct genomic features (Figure 1A). The overall DNA methylation patterns in PBMs were similar to previously reported patterns observed in other cell types [36–38]. For instance, promoters tend to be lowly methylated, whereas gene bodies, particularly protein coding regions, were more frequently found to be highly methylated (Figure 1A). Interestingly, we found that approximately 24% of CGIs were highly methylated. This number may at first seem to be higher than expected since it has been widely accepted that CGIs are largely unmethylated [39]. However, it has become increasingly evident that a significant fraction of CGIs were methylated across a variety of different tissues/cells [36,40–42].

HCPs and LCPs have been shown to exhibit distinct DNA methylation patterns in other human somatic and germline cells [38,43,44]. Accordingly, we performed further DNA methylation analyses for promoters by subdividing the 25,814 promoters into three classes, namely HCP ( $n = 12,416$ ), ICP ( $n = 5510$ ) and LCP ( $n = 7888$ ), based on their CpG content (see 'Materials & methods' section). Consistent with previous findings in other cell types [38,43,44], we observed distinct methylation patterns between HCPs and LCPs (Figure 1B). As such, on average, 84% of HCPs (range: 60–97% in individual samples) were lowly methylated and less than 0.5% of HCPs (range: 0–2.3% in individual samples) were highly methylated. By contrast, on average, 28% of LCPs (range: 9–51%) were highly methylated and an additional 55% of LCPs (range: 39–68%) showed partial methylation. Consequently, highly methylated promoters were almost exclusively (average: 95%; range: 91–98%) detected in LCPs, whereas lowly methylated promoters were predominantly (average: 70%; range: 66–73%) observed in HCPs (Figure 1B).

Similarly, we took a closer examination of the CGIs by subdividing the CGIs into three groups, specifically, 5'-end CGIs (n = 9426), intragenic CGIs (n = 7078) and intergenic CGIs (n = 11,033), on the basis of their positions relative to the nearest RefSeq genes (see 'Materials & methods' section). CGIs located near the 5'-end of RefSeq genes were much less frequently found to be highly methylated than CGIs located in gene bodies and intergenic regions. For instance, on average, only 7% of 5'-end CGIs were found to be highly methylated, in contrast to 32–34% of intra- and inter-genic CGIs. Accordingly, highly methylated CGIs were most often to be found at the intra- and inter-genic regions (average 89%; range 83–96%) than at the 5'-end regulatory regions (average 11%; range 4 – 17% ) (Figure 1C).

Next, we investigated the highly/lowly methylated regions (1-kb sliding windows) that were common to all 18 PBM methylomes (Figure 2A; profiles are saved as bedGraph files, which can be downloaded from the Supplementary Material and visualized on the University of California, Santa Cruz genome browser). Regions that were constitutively highly methylated (n = 128,982) or lowly methylated (n = 246,136) in PBMs were distributed throughout the genome. In general, highly methylated regions have lower CpG density (CpG observed/expected ratios) than lowly methylated regions (Figure 2B;  $p < 2.2 \times 10^{-16}$ ). This is more evident at some long-stretch segments enriched for highly methylated or lowly methylated windows (Figure 2C–D). For instance, a 350-kb segment on chromosome 9q33.1 was found to be significantly enriched for highly methylated windows and a close examination of the sequence in this area revealed that this region has relatively low CpG density and is largely depleted for CGIs (Figure 2C). By contrast, a 500-kb CpG-rich region on chromosome 7q22.1 was largely enriched for constitutively lowly methylated windows (Figure 2D). Out of the 25,814 promoters, we identified 276 promoters that were consistently highly methylated in all 18 subjects, including three ICPs and 273 LCPs (Supplementary Table 5). Gene ontology analysis revealed that genes associated with these constitutively highly methylated promoters were significantly enriched in a number of biological processes (Supplementary Table 6), such as sensory perception ( $p = 5.27 \times 10^{-10}$ ) and keratinization ( $p = 1.21 \times 10^{-7}$ ). As many of these biological processes are not related to PBM function, constitutive silencing of these genes with highly methylated promoters may not be unexpected.

### Interindividual variation in the PBM methylome

Previous studies suggested that there is substantial interindividual variation in DNA methylation and this variability may confer important functional implications for gene regulation and contribute to phenotypic variation in humans. Therefore, in addition to the qualitative characterization of common PBM methylomes, we also assessed the interindividual variability of DNA methylation in normal PBMs in a quantitative manner. Specifically, we measured the interindividual variability with median absolute deviation for each 1-kb sliding window throughout the genome and selected the top 0.5 and 1% most-variable windows for further characterization. These highly variable regions were not evenly distributed along the chromosomes (Figure 3A). Chromosomes 13 and 21 had a significantly higher density of variable regions than predicted from their length, whereas chromosomes 1 and 3 had the lowest density of variable regions across the genome. Furthermore, by annotation of the variable 1-kb windows to their closest genomic features, we found that relative to their sizes in the reference genome, transposable elements, such as satellite and long terminal repeats, were enriched in these highly variable regions; whereas genic components, such as promoters and exons, tend to be under-represented in these highly variable regions (Figure 3B).

## Correlation between DNA methylation & gene expression

The modification of DNA by methylation is an important epigenetic mechanism that affects the spatial and temporal regulation of gene expression. Among the 18 samples used in this study, seven have been previously tested for genome-wide gene expression profile data in PBMs. Subsequently, we investigated the relationship between DNA methylation and gene expression for the 20,684 unique genes that were assessed on the 1.0 Human Exon ST arrays.

First, we compared the promoter DNA methylation and gene expression levels across individual genes (Figure 4A). On the one hand, genes with highly methylated promoters exhibited low expression levels and highly expressed genes showed low promoter methylation, supporting the notion that promoter methylation is correlated with gene silencing. On the other hand, low promoter methylation is not necessarily translated into high gene expression, as many silenced/lowly expressed genes also contained lowly methylated promoters, indicating that various other mechanisms may regulate gene expression independent of the effects of promoter methylation [44,45]. The negative correlation between DNA methylation and gene expression levels was also evident when we split the 20,684 assessed genes into five groups based on the ranking of mean expression levels of individual genes and examined the DNA methylation levels in a 1.5 kb region up- and down-stream of the TSSs (Figure 4B). As such, the top 20% most highly expressed genes exhibited the lowest methylation levels in the TSS-surrounding region and groups of genes with increasingly lower expression levels showed increasingly higher methylation levels. Additionally, we observed a gradual decrease in DNA methylation levels when approaching the TSS from both upstream and downstream, with the bottom of the methylation ‘valley’ at around  $\pm 450$  bp from the TSS (Figure 4B). A similar pattern of DNA methylation around the TSS has been observed in human embryonic stem cells [46], lymphocytes [4] and PBMCs [37]. Since previous studies suggested that the relationship between promoter methylation and gene expression is related to the CpG density of the promoter [44], we further examined the methylation patterns at the TSS-surrounding region, and the correlation between DNA methylation and gene expression in genes associated with HCPs, ICPs and LCPs, respectively (Figure 4C). Similar to what we observed for the combined data, we found that DNA methylation around the TSS region was negatively correlated with gene expression within each gene group associated with a distinct promoter class, even for genes associated with LCPs. However, the DNA methylation ‘valley’ centered at the TSS was only observed for genes associated with HCPs and ICPs, whereas the methylation levels were relatively constant along the TSS-surrounding regions for genes associated with LCPs (Figure 4C). Subsequently, we explored the pattern of CpG density in the TSS-surrounding region in order to determine whether differential CpG density along the region can explain the observed differences in their DNA methylation patterns between genes associated with HCPs/ICPs and LCPs. Indeed, we observed distinct CpG density profiles surrounding TSSs when comparing genes associated with HCPs/ICPs to those associated with LCPs (Figure 4D). In the TSS-surrounding regions of genes associated with HCPs and ICPs, we observed a gradual increase in CpG density when approaching the TSS, with a peak center at the TSS. By contrast, CpG density was relatively constant throughout the TSS-surrounding region for genes associated with LCPs.

To explore how differential promoter methylation was correlated with variation in gene expression across individuals, we calculated the Pearson’s correlation coefficient between methylation and gene expression levels for each gene across all individuals. We observed a modest but significant excess ( $p < 2.2 \times 10^{-16}$ ) of genes showing negative correlation between promoter methylation and transcription levels across individuals, particularly for genes associated with HCPs (Figure 5). We selected genes with correlation coefficients of  $r \leq -0.678$  (equivalent to the fifth percentile of the empirical distribution of the correlation



coefficient) as significant genes. Gene ontology analysis for these significant genes revealed an enrichment of biological processes that are closely related to PBM function (Table 1 & Supplementary Table 7), suggesting that the interindividual variation in expression levels for genes involved in these functions may be under strong influence from individual differences in DNA methylation levels.

## Discussion

In this study, we reported a genome-wide DNA methylation analysis in primary human PBMs freshly isolated from a sample cohort of unrelated healthy individuals that were closely matched for ethnicity, age, gender and menopause status. Owing to the diverse functions of PBMs and the broad relevance to a number of human complex disorders [19–22], the PBM DNA methylation profiles characterized in this study will provide a useful resource for future epidemiological and biological functional studies.

In agreement with earlier reports in other cells/tissues [36–38,44], we found that promoters in PBMs were largely unmethylated. When dividing the promoters into subclasses by their CpG density, it was clear that HCPs were rarely found to be highly methylated and the methylated promoters predominantly consisted of LCPs. Interestingly, the protein coding regions were significantly more frequently found to be highly methylated than all other genic regions. This is consistent with previous reports showing relatively high methylation levels over gene bodies in mammals [4,37,46]. A high level of gene body methylation may improve the transcription efficiency of actively transcribed genes by repressing spurious initiation of transcription within active genes [3,5]. We also found that a considerable fraction of CGIs, predominantly CGIs located in intragenic and intergenic regions, were highly methylated. Our results confirmed the emerging evidence that methylation of CGIs is more common than previously appreciated [36,40–42]. For instance, approximately 26% of CGIs were found to be heavily methylated in human B cells, including 36% of intragenic CGIs [37]. Similarly, Rakyan *et al.* examined the DNA methylation profiles in 13 normal somatic tissues and found that unmethylated CGIs represented only a small proportion of nonpromoter CGIs (20–48%) [42].

It has long been implicated that DNA methylation in the promoter region plays an important role in the regulation of gene expression. By comparing genes with different expression levels, we observed a clear negative correlation between promoter methylation and gene transcription levels and this trend was also found at all three promoter subclasses. Interestingly, we observed different DNA methylation patterns at TSS-surrounding regions when comparing genes associated with HCPs/ICPs with those associated with LCPs. As such, a ‘valley’ of methylation level centered at the TSS was only seen at genes associated with HCPs and ICPs, but not at those associated with LCPs. This observed difference likely reflected the higher CpG density distribution along the TSS regions in HCPs/ICPs compared with that in LCPs. Taken together, these findings suggested that DNA methylation at promoters, including LCPs, contributes to the intraindividual variability in gene expression, even though promoters with different CpG density exhibited distinct DNA methylation patterns.

Several previous studies have implied that there is significant interindividual variation in DNA methylation in the general population [47–51] and that the DNA methylation variation, along with other epigenetic variations, may contribute to the interindividual differences in disease susceptibility [52–54]. In this study, we observed considerable variation in DNA methylation levels, even within individuals that were matched for ethnicity, age, sex and menopausal status, suggesting that many other intrinsic and extrinsic factors may give rise to variations in DNA methylation status. This is consistent with

previous reports showing that genetic variations [48,55,56] and many environmental factors besides sex, age, and ethnicity may contribute to DNA methylation variability, such as nutrition [57], cigarette smoking [58], perceived stress and early-life socioeconomic status [18]. In addition, we should bear in mind that the interindividual variations in cell composition may produce spurious results in the association between epigenomic factors (e.g., DNA methylation) and the diseases/traits of interest, when the epigenomic factors were assessed in a heterogeneous mixture of cells (e.g., most tissue samples and leukocytes) [13,18]. Therefore, the potential cell composition differences between individuals must be carefully considered for epigenomic studies, even for studies using immunomagnetic-enriched cells (e.g., the PBMs used in this study), since these cell-enrichment approaches normally cannot yield complete purification of a single type of cell.

The extent of this interindividual variation was not the same across the genome and varied between different genomic features, with highly variable regions enriched for repetitive DNA elements, such as satellite DNA repeats and long terminal repeats, compared with the genic regions. Interestingly, a large degree of interindividual variability in the DNA methylation of satellite repeats and other repetitive elements has also been reported in other types of human cells/tissues [49,59]. The higher variability in repetitive elements may reflect the fact that these regions were under less stringent evolutionary constraint [60,61], as seen in the case of extensive genetic variations in these regions [62,63]. However, these results should be interpreted with caution, as alignment with repetitive sequence regions is known to be difficult, especially for short sequencing reads generated by most of the current next-generation sequencing platforms [64] and, thus, the diminished alignment accuracy in repetitive regions may, to some extent, contribute to the observed variability in DNA methylation levels.

We also evaluated whether interindividual variation in promoter methylation levels was correlated with interindividual variation in gene expression levels. The results indicated a modest but significant excess of genes whose differential expression levels across individuals were negatively correlated with their promoter methylation levels in different individuals. We also found that genes showing significant negative correlation between interindividual promoter methylation and gene expression levels mainly consisted of genes with HCPs, suggesting that DNA methylation at HCPs is more likely to play a significant role in regulating interindividual gene expression levels and affecting individual susceptibility in health and disease conditions. In addition, these significant genes were enriched in biological processes that are closely related to PBM function, suggesting that alteration in DNA methylation is likely to be an important mechanism contributing to interindividual variation in PBM function, and PBM-related phenotype and disease-susceptibility variation in humans. The results call for a focus on DNA methylation-mediated epigenetic mechanisms for the future study of complex diseases that are closely linked to PBM function.

## Conclusion

In summary, we reported a genome-wide DNA methylation analysis for primary PBMs in a group of healthy postmenopausal Caucasian females. Our results delineated the methylome landscape of PBMs and provided novel insights into the intra- and inter-individual correlation between DNA methylation and gene expression levels.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The authors wish to thank M Gui at Arraystar Inc. (MD, USA) for his assistance in setting up and performing the methylated DNA immunoprecipitation coupled with the next-generation sequencing, and L Chavez at the Max-Planck-Institute for Molecular Genetics (Berlin, Germany) for providing consultation for data analysis.

The investigators of this work were partially supported by grants from the NIH (P50AR055081, R01AG026564, R01AR050496, R01AR057049 and R03TW008221), the Franklin D Dickson/Missouri Endowment from University of Missouri-Kansas City, and the Edward G Schlieder Endowment from Tulane University. The work also benefited from the Shanghai Leading Academic Discipline Project (S30501).

## References

Papers of special note have been highlighted as:

■ of interest

■ ■ of considerable interest

1. van der Maarel SM. Epigenetic mechanisms in health and disease. *Ann. Rheum. Dis.* 2008; 67(Suppl. 3):iii97–iii100. [PubMed: 19022824]
2. Feinberg AP. Epigenetics at the epicenter of modern medicine. *JAMA.* 2008; 299(11):1345–1350. [PubMed: 18349095]
3. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* 2008; 9(6):465–476. [PubMed: 18463664]
4. Ball MP, Li JB, Gao Y, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* 2009; 27(4):361–368. [PubMed: 19329998]
5. Maunakea AK, Nagarajan RP, Bilenky M, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature.* 2010; 466(7303):253–257. [PubMed: 20613842]
6. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature.* 2007; 447(7143):433–440. [PubMed: 17522677]
7. Ozanne SE, Constancia M. Mechanisms of disease: the developmental origins of disease and the role of the epigenotype. *Nat. Clin. Pract. Endocrinol. Metab.* 2007; 3(7):539–546. [PubMed: 17581623]
8. Jiang YH, Bressler J, Beaudet AL. Epigenetics and human disease. *Ann. Rev. Genomics Hum. Genet.* 2004; 5:479–510. [PubMed: 15485357]
9. Bjornsson HT, Cui H, Gius D, Fallin MD, Feinberg AP. The new field of epigenomics: implications for cancer and other common disease research. *Cold Spring Harb. Symp. Quant. Biol.* 2004; 69:447–456. [PubMed: 16117680]
10. Migliore L, Coppede F. Genetics, environmental factors the emerging role of epigenetics in neurodegenerative diseases. *Mutat. Res.* 2008; 667(1–2):82–97. [PubMed: 19026668]
11. Esteller M. Epigenetics in cancer. *N. Engl. J. Med.* 2008; 358(11):1148–1159. [PubMed: 18337604]
12. Johannes F, Colot V, Jansen RC. Epigenome dynamics: a quantitative genetics perspective. *Nat. Rev. Genet.* 2008; 9(11):883–890. [PubMed: 18927581]
13. Reinius LE, Acevedo N, Joerink M, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS ONE.* 2012; 7(7):e41361. [PubMed: 22848472] ■ ■ Reports considerable variation in the DNA methylation profiles of the main blood cell types and suggests that the differences resulting from varying proportions of white blood cell types should be carefully considered for epigenomic studies using whole blood as the sample source.
14. Koestler DC, Marsit CJ, Christensen BC, et al. Peripheral blood immune cell methylation profiles are associated with nonhematopoietic cancers. *Cancer Epidemiol. Biomarkers Prev.* 2012; 21(8):1293–1302. [PubMed: 22714737]
15. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* 2003; 33(Suppl.):245–254. [PubMed: 12610534]

16. Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. *Nat. Rev. Genet.* 2007; 8(4):253–262. [PubMed: 17363974]
17. Fraser HB, Lam LL, Neumann SM, Kobor MS. Population-specificity of human DNA methylation. *Genome Biol.* 2012; 13(2):R8. [PubMed: 22322129]
18. Lam LL, Emberly E, Fraser HB, et al. Factors underlying variable DNA methylation in a human community cohort. *Proc. Natl Acad. Sci. USA.* 2012; 109(Suppl. 2):17253–17260. [PubMed: 23045638] ■ Broadly assesses and characterizes potential factors underlying the variation in DNA methylation patterns in a human community cohort.
19. Liu YZ, Dvornyk V, Lu Y, et al. A novel pathophysiological mechanism for osteoporosis suggested by an *in vivo* gene expression study of circulating monocytes. *J. Biol. Chem.* 2005; 280(32):29011–29016. [PubMed: 15965235]
20. Laso FJ, Vaquero JM, Almeida J, Marcos M, Orfao A. Production of inflammatory cytokines by peripheral blood monocytes in chronic alcoholism: relationship with ethanol intake and liver disease. *Cytometry B Clin. Cytom.* 2007; 72(5):408–415. [PubMed: 17266151]
21. Longhi MS, Mitry RR, Samyn M, et al. Vigorous activation of monocytes in juvenile autoimmune liver disease escapes the control of regulatory t-cells. *Hepatology.* 2009; 50(1):130–142. [PubMed: 19437492]
22. Dorffel Y, Latsch C, Stuhlmuller B, et al. Preactivated peripheral blood monocytes in patients with essential hypertension. *Hypertension.* 1999; 34(1):113–117. [PubMed: 10406833]
23. Deng FY, Lei SF, Zhang Y, et al. Peripheral blood monocyte-expressed *ANXA2* gene is involved in pathogenesis of osteoporosis in humans. *Mol. Cell. Proteomics.* 2011; 10(11) M111.011700.
24. Deng FY, Liu YZ, Li LM, et al. Proteomic analysis of circulating monocytes in chinese premenopausal females with extremely discordant bone mineral density. *Proteomics.* 2008; 8(20): 4259–4272. [PubMed: 18924182]
25. Lei SF, Wu S, Li LM, et al. An *in vivo* genome wide gene expression study of circulating monocytes suggested *GBP1*, *STAT1* and *CXCL10* as novel risk genes for the differentiation of peak bone mass. *Bone.* 2009; 44(5):1010–1014. [PubMed: 19223260]
26. Down TA, Rakyan VK, Turner DJ, et al. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat. Biotechnol.* 2008; 26(7):779–785. [PubMed: 18612301]
27. Chavez L, Jozefczuk J, Grimm C, et al. Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage. *Genome Res.* 2010; 20(10):1441–1450. [PubMed: 20802089] ■ References [26,27] describe two analytical tools for quality assessment and quantitative analysis of methylated DNA immunoprecipitation coupled with chromatin immunoprecipitation and/or next-generation sequencing data, both controlling for the influence of local CpG density.
28. Weber M, Davies JJ, Wittig D, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* 2005; 37(8):853–862. [PubMed: 16007088]
29. Pruitt KD, Tatusova T, Brown GR, et al. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* 2012; 40:D130–D135. [PubMed: 22121212]
30. Mikkelsen TS, Ku M, Jaffe DB, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature.* 2007; 448(7153):553–560. [PubMed: 17603471]
31. Illingworth RS, Bird AP. CpG islands – ‘a rough guide’. *FEBS Lett.* 2009; 583(11):1713–1720. [PubMed: 19376112]
32. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res.* 2002; 12(6):996–1006. [PubMed: 12045153]
33. Zheng Q, Wang XJ. GOEAST: a web-based software toolkit for gene ontology enrichment analysis. *Nucleic Acids Res.* 2008; 36:W358–W363. (Web Server Issue). [PubMed: 18487275]
34. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple hypothesis testing under dependency. *Ann. Statistics.* 2001; 29:1165–1188.
35. Carvalho B, Irizarry RA, Speed TP, et al. Exploration, normalization, and genotype calls of high density oligonucleotide SNP array data. *Biostatistics.* 2007; 8(2):485–499. [PubMed: 17189563]

36. Rauch TA, Wu X, Zhong X, Riggs AD, Pfeifer GP. A human B cell methylome at 100-base pair resolution. *Proc. Natl Acad Sci USA*. 2009; 106(3):671–678. [PubMed: 19139413] ■ Interesting study that comprehensively characterizes the DNA methylation profile of human B cells.
37. Li Y, Zhu J, Tian G, et al. The DNA methylome of human peripheral blood mononuclear cells. *PLoS Biol*. 2010; 8(11):e1000533. [PubMed: 21085693] ■ Interesting study that comprehensively characterizes the DNA methylation profile of human peripheral blood mononuclear cells.
38. Meissner A, Mikkelsen TS, Gu H, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*. 2008; 454(7205):766–770. [PubMed: 18600261]
39. Bird AP. CpG-rich islands and the function of DNA methylation. *Nature*. 1986; 321(6067):209–213. [PubMed: 2423876]
40. Shen L, Kondo Y, Guo Y, et al. Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet*. 2007; 3(10):e181.
41. Illingworth R, Kerr A, Desousa D, et al. A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol*. 2008; 6(1):e22. [PubMed: 18232738]
42. Rakyan VK, Down TA, Thorne NP, et al. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res*. 2008; 18(9):1518–1529. [PubMed: 18577705]
43. Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl Acad. Sci. USA*. 2006; 103(5):1412–1417. [PubMed: 16432200]
44. Weber M, Hellmann I, Stadler MB, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet*. 2007; 39(4):457–466. [PubMed: 17334365]
45. Zhang Y, Rohde C, Tierling S, et al. DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. *PLoS Genet*. 2009; 5(3):e1000438. [PubMed: 19325872]
46. Laurent L, Wong E, Li G, et al. Dynamic changes in the human methylome during differentiation. *Genome Res*. 2010; 20(3):320–331. [PubMed: 20133333]
47. Rakyan VK, Hildmann T, Novik KL, et al. DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol*. 2004; 2(12):e405. [PubMed: 15550986]
48. Bell JT, Pai AA, Pickrell JK, et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol*. 2011; 12(1):R10. [PubMed: 21251332]
49. Flanagan JM, Popendikyte V, Pozdniakovaite N, et al. Intra- and interindividual epigenetic variation in human germ cells. *Am. J. Hum. Genet*. 2006; 79(1):67–84. [PubMed: 16773567]
50. Bock C, Walter J, Paulsen M, Lengauer T. Interindividual variation of DNA methylation and its implications for large-scale epigenome mapping. *Nucleic Acids Res*. 2008; 36(10):e55. [PubMed: 18413340]
51. Byun HM, Siegmund KD, Pan F, et al. Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue- and individual-specific DNA methylation patterns. *Hum. Mol. Genet*. 2009; 18(24):4808–4817. [PubMed: 19776032]
52. Feinberg AP, Irizarry RA, Fradin D, et al. Personalized epigenomic signatures that are stable over time and covary with body mass index. *Sci. Transl. Med*. 2010; 2(49):49ra67.
53. Gervin K, Vigeland MD, Mattingsdal M, et al. DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis: identification of epigenetically dysregulated genes. *PLoS Genet*. 2012; 8(1):e1002454. [PubMed: 22291603]
54. Movassagh M, Choy MK, Goddard M, Bennett MR, Down TA, Foo RS. Differential DNA methylation correlates with differential expression of angiogenic factors in human heart failure. *PLoS ONE*. 2010; 5(1):e8564. [PubMed: 20084101] ■■ Interesting study in which separate analyses of DNA methylation and gene expression data revealed no significant results, but a combined analysis identified epigenetic alterations that potentially contribute to the development of psoriasis.



55. Gertz J, Varley KE, Reddy TE, et al. Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. *PLoS Genet.* 2011; 7(8):e1002228. [PubMed: 21852959]
56. Quon G, Lippert C, Heckerman D, Listgarten J. Patterns of methylation heritability in a genome-wide analysis of four brain regions. *Nucleic Acids Res.* 2013; 41(4):2095–2104. [PubMed: 23303775] ■ Assesses the inheritance of genome-wide DNA methylation patterns in humans and demonstrates that the influence of genotype on patterns of DNA methylation is widespread in the genome and greatly exceeds the influence of imprinting on genome-wide methylation patterns.
57. Gomes MV, Toffoli LV, Arruda DW, et al. Age-related changes in the global DNA methylation profile of leukocytes are linked to nutrition but are not associated with the *MTHFR* C677T genotype or to functional capacities. *PLoS ONE.* 2012; 7(12):e52570. [PubMed: 23285094]
58. Talikka M, Sierro N, Ivanov NV, et al. Genomic impact of cigarette smoke, with application to three smoking-related diseases. *Crit. Rev. Toxicol.* 2012; 42(10):877–889. [PubMed: 22989067]
59. Sandovici I, Kassovska-Bratinova S, Loredó-Ostí JC, et al. Interindividual variability and parent of origin DNA methylation differences at specific human Alu elements. *Hum. Mol. Genet.* 2005; 14(15):2135–2143. [PubMed: 15972727]
60. Gogvadze E, Buzdin A. Retroelements and their impact on genome evolution and functioning. *Cell. Mol. Life Sci.* 2009; 66(23):3727–3742. [PubMed: 19649766]
61. Kazazian HH Jr. Mobile elements: drivers of genome evolution. *Science.* 2004; 303(5664):1626–1632. [PubMed: 15016989]
62. Xing J, Zhang Y, Han K, et al. Mobile elements create structural variation: analysis of a complete human genome. *Genome Res.* 2009; 19(9):1516–1526. [PubMed: 19439515]
63. Huang CR, Schneider AM, Lu Y, et al. Mobile interspersed repeats are major structural variants in the human genome. *Cell.* 2010; 141(7):1171–1182. [PubMed: 20602999]
64. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* 2012; 13(1):36–46. [PubMed: 22124482]

## Websites

101. NCBI. Gene expression omnibus. [www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo)
102. Gene Ontology Enrichment Analysis Software Toolkit. <http://omicslab.genetics.ac.cn/GOEAST/index.php>

### Future perspective

With the continuous advancement in experimental and analytic approaches over the coming years, we will witness a surge of epigenome-wide DNA methylation studies in various cell types for a broad range of diseases and complex traits. These studies will unveil novel insights into the epigenetic basis of human complex disorders. Additionally, by integrating the DNA methylation data with the genome sequencing, transcriptomic and proteomic profiling, as well as other epigenomic (e.g., histone modification) profiling data, we will be able to build an increasingly unified biological view of dynamic gene regulation, and reveal the complex molecular mechanisms underlying cellular processes and disease pathogenesis.

### Executive summary

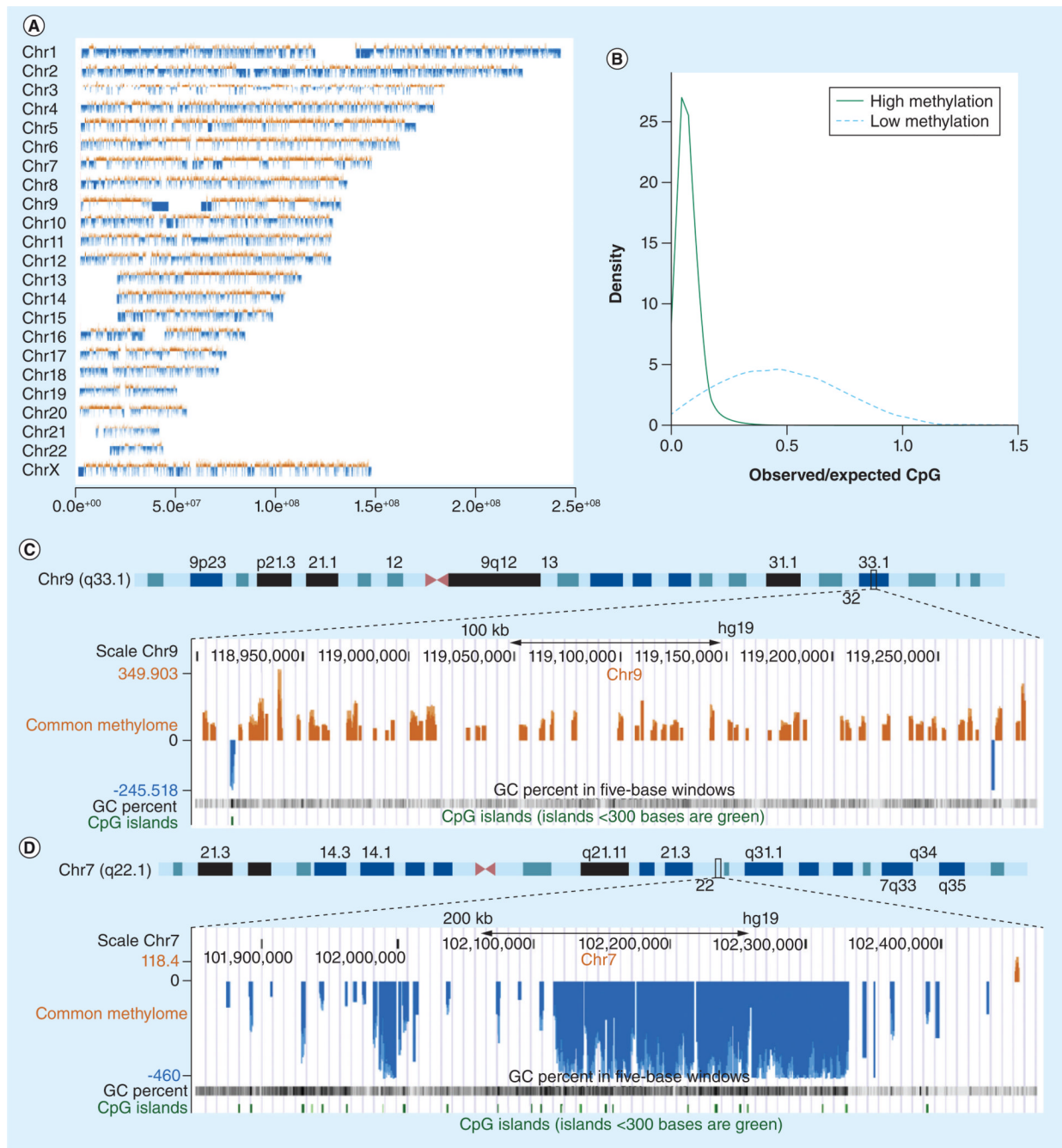
- Peripheral blood monocytes (PBMs) play critical roles in multiple biological processes and abnormalities in PBMs have been linked to a variety of human disorders.
- Distinct patterns were revealed at different genomic features.
- We observed a negative correlation between promoter methylation levels and gene transcription levels consistently across promoters with high to low CpG densities.
- Importantly, we observed substantial interindividual variation in DNA methylation across individual PBMs and revealed a modest but significant excess ( $p < 2.2 \times 10^{-16}$ ) of genes showing negative correlation between interindividual promoter methylation and transcription levels.
- Our results contribute to comprehensive characterization of the human PBM methylation landscape and its interindividual variation, and provide an invaluable resource for future epigenomic studies of PBM-related biological process and diseases.



**Figure 1. Distinct DNA methylation profiles in different genomic features**

Relative fraction of highly, partially and lowly methylated regions in (A) different genomic features, (B) promoters with low, intermediate and high CpG content, and (C) intragenic, intergenic and 5'-end CGIs.

CDS: Coding DNA sequence; CGI: CpG island; HCP: High CpG promoter containing a 500-bp interval with a CpG observed/expected ratio 0.6 and GC content 55%; ICP: Intermediate CpG promoter that is not a HCP or LCP; L1: Long interspersed nuclear element 1; L2: Long interspersed nuclear element 2; L3: Long interspersed nuclear element 3; L4: Long interspersed nuclear element 4; LCP: Low CpG promoter containing no 500-bp interval with a CpG observed/expected ratio 0.4; LTR: Long terminal repeat; MIR: Mammalian wide interspersed repeat.

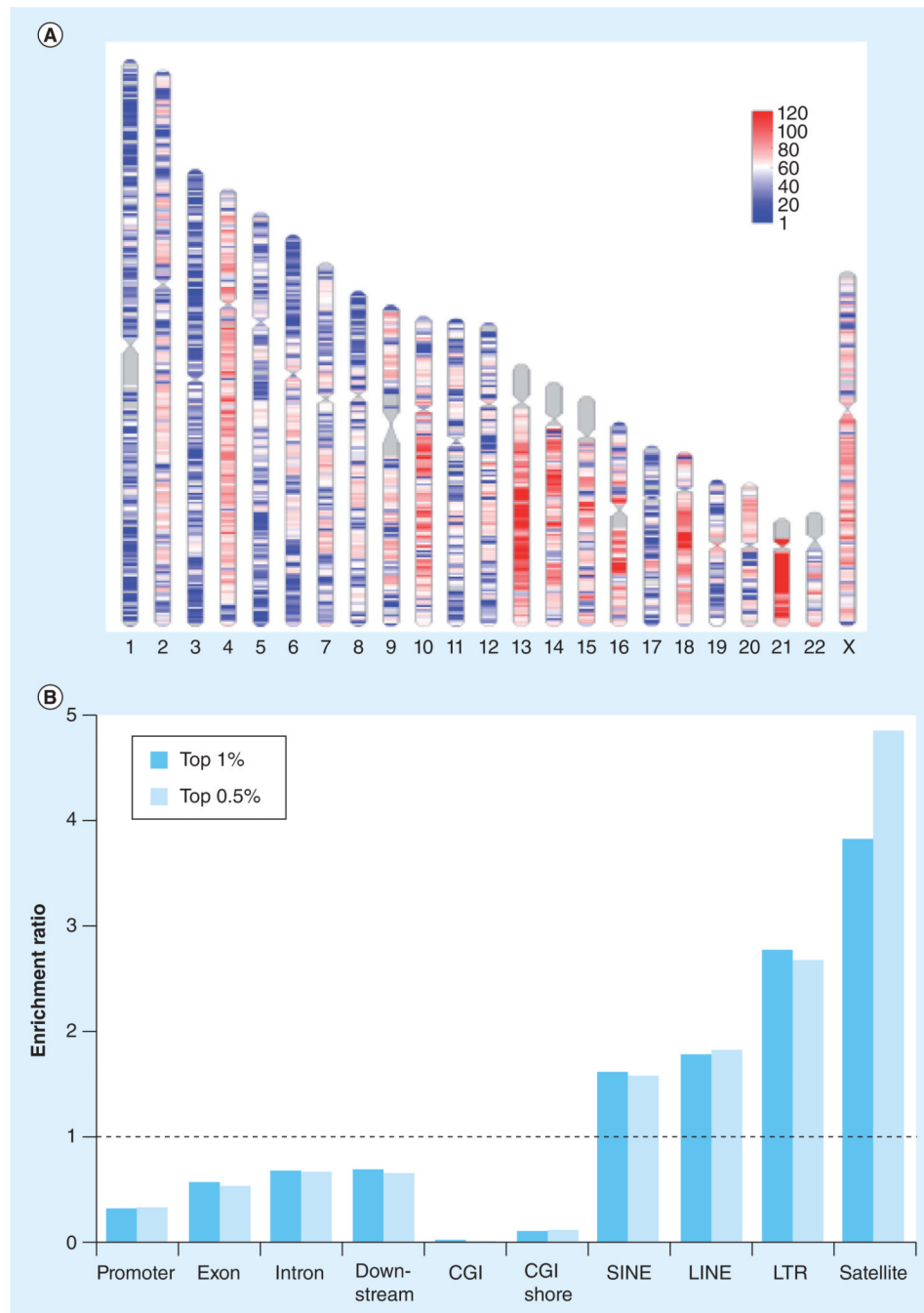


**Figure 2. Characterization of regions that were constitutively highly methylated or lowly methylated in peripheral blood monocytes**

**(A)** Genomic distribution of constitutively highly methylated and lowly methylated regions. The orange lines denote the genomic positions of 1-kb windows that were found to be highly methylated in all 18 individuals and the blue lines denote regions that were lowly methylated in all subjects. The height of the line is relative with an average transformed absolute methylation score of 600 for the highly methylated regions and an average transformed absolute methylation score of 460 for the lowly methylated regions. **(B)** Distributions of CpG observed/expected ratios for highly and lowly methylated regions. **(C)**



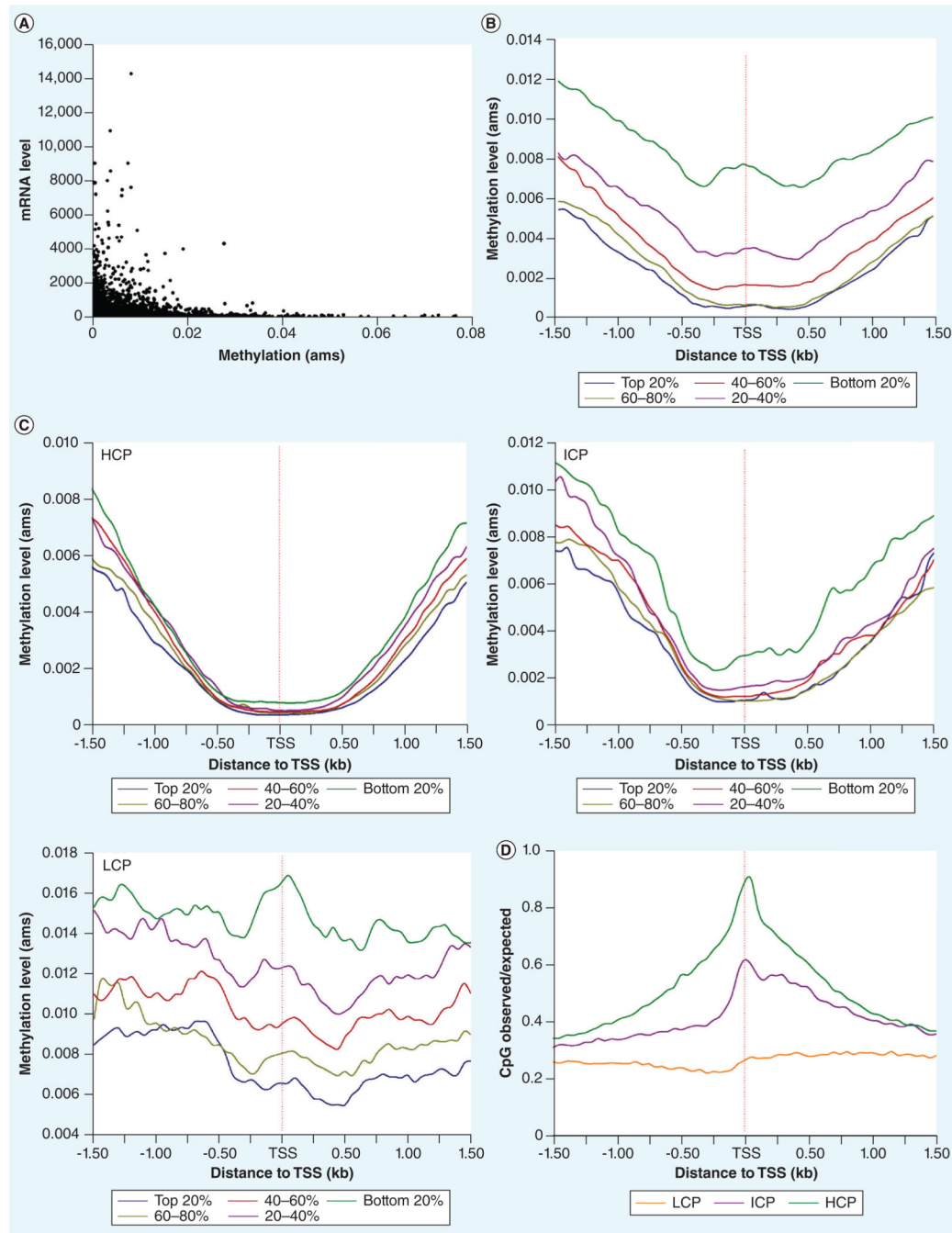
**& D)** Examples of long-stretch segments enriched for constitutively highly methylated or lowly methylated windows. Chr: Chromosome.



**Figure 3. Characterization of regions showing large interindividual variability in DNA methylation levels**

**(A)** Distributions of the top 1% of most-variable regions among different chromosomes. The number of variable regions were calculated in nonoverlap 1-Mb windows across the human genome and plotted in ideograms. The diversities are illustrated by colors, with red indicating higher numbers and blue indicating lower numbers. Genomic regions in which no reference sequences could be determined are shown in gray. **(B)** The relative enrichment of most-variable regions in distinct genomic features. The enrichment ratio was calculated for each genomic feature as a percentage of the genomic feature in top variable regions/

percentage of the size of the genomic feature relative to the size of the reference genome.  
The dotted line marks an enrichment ratio of 1.  
CGI: CpG island; LTR: Long terminal repeat.

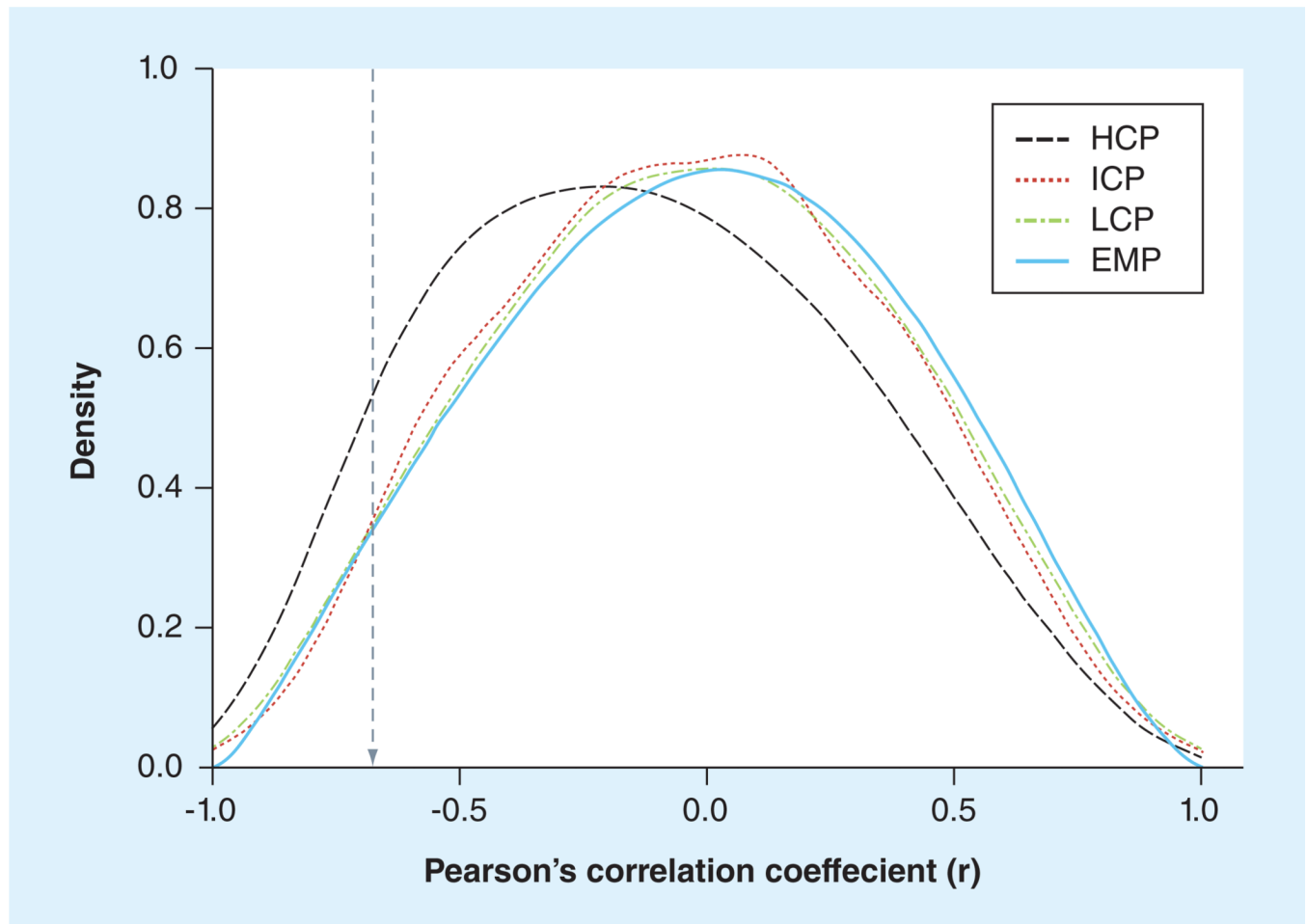


**Figure 4. Correlation between DNA methylation and transcription levels across genes (facing page)**

(A) Promoter DNA methylation levels were plotted against the gene expression levels across individual genes. DNA methylation level for each gene was calculated as the mean absolute methylation score value across tested individuals. Mean gene expression level was computed based on the normalized expression values of the exon array data. (B) Genes were divided into five groups based on the ranking of average expression levels. The average methylation levels in 50-bp bins throughout the  $\pm 1.5$ -kb TSS-surrounding region were computed for each gene group and the data for adjacent 50-bp bins were connected to form a smooth curve. (C) Genes were first separated into three categories according to the CpG

density of their promoters. Within each gene category, genes were further divided into quintiles based on their relative gene expression levels, and the methylation patterns were examined at the  $\pm 1.5$ -kb TSS-surrounding regions. **(D)** CpG density at the  $\pm 1.5$ -kb TSS-surrounding regions of genes associated with promoters containing high, intermediate and low CpG content. ams: Absolute methylation signals; HCP: High CpG promoter containing a 500-bp interval with a CpG observed/expected ratio  $\geq 0.6$  and GC content  $\geq 55\%$ ; ICP: Intermediate CpG promoter that is not a HCP or LCP; LCP: Low CpG promoter containing no 500-bp interval with a CpG observed/expected ratio  $\geq 0.4$ ; TSS: Transcription start site.





**Figure 5. Correlation between promoter methylation and transcription levels across individuals**

The distribution of Pearson's correlation coefficient between interindividual methylation and gene expression levels was plotted for genes associated with promoters of high, intermediate and low CpG content, respectively. The solid line represents the EMP of no association between DNA methylation and gene expression levels across individuals. The vertical dashed line with an arrow marks the threshold for declaring significant correlation (i.e., the fifth percentile of the empirical distribution).

EMP: Empirical distribution of the correlation coefficient under the null hypothesis; HCP: High CpG promoter containing a 500-bp interval with a CpG observed/expected ratio 0.6 and GC content 55%; ICP: Intermediate CpG promoter that is not a HCP or LCP; LCP: Low CpG promoter containing no 500-bp interval with a CpG observed/expected ratio 0.4.

**Table 1**

The top five most significant gene ontology terms enriched for genes showing significant interindividual correlation between promoter methylation and transcription levels.

GO_ID	Term	p-value
GO:2000332	Regulation of blood microparticle formation	$<1.18e^{-38}$
GO:0055095	Lipoprotein particle-mediated signaling	$<1.18e^{-38}$
GO:0071447	Cellular response to hydroperoxide	$<1.18e^{-38}$
GO:0060099	Regulation of phagocytosis, engulfment	$<1.18e^{-38}$
GO:0070339	Response to bacterial lipopeptide	$<1.18e^{-38}$

Gene ontology enrichment analysis was carried out by using the GOEAST program [31], which gave 'p-value > 0' when the obtained p-value is less than the minimum float value ( $1.17549435082229e^{-38}$ ).