



# Buffy Coat DNA Methylation Profile Is Representative of Methylation Patterns in White Blood Cell Types in Normal Pregnancy

Ranine Ghamrawi<sup>1†</sup>, Igor Velickovic<sup>2†</sup>, Ognjen Milicevic<sup>2</sup>, Wendy M. White<sup>3,4</sup>, Lillian Rosa Thistlethwaite<sup>5</sup>, Julie M. Cunningham<sup>6</sup>, Aleksandar Milosavljevic<sup>5</sup>, Natasa M. Milic<sup>1,2</sup> and Vesna D. Garovic<sup>1,3\*</sup>

## OPEN ACCESS

### Edited by:

Umberto Galderisi,  
University of Campania Luigi Vanvitelli,  
Italy

### Reviewed by:

Ying Cai,  
Albert Einstein College of Medicine,  
United States  
Jing-Woei Li,  
Prince of Wales Hospital, China

### \*Correspondence:

Vesna D. Garovic  
garovic.vesna@mayo.edu

<sup>†</sup>These authors share first authorship

### Specialty section:

This article was submitted to  
Preclinical Cell and Gene Therapy,  
a section of the journal  
Frontiers in Bioengineering and  
Biotechnology

**Received:** 24 September 2021

**Accepted:** 09 December 2021

**Published:** 05 January 2022

### Citation:

Ghamrawi R, Velickovic I, Milicevic O,  
White WM, Thistlethwaite LR,  
Cunningham JM, Milosavljevic A,  
Milic NM and Garovic VD (2022) Buffy  
Coat DNA Methylation Profile Is  
Representative of Methylation Patterns  
in White Blood Cell Types in  
Normal Pregnancy.  
Front. Bioeng. Biotechnol. 9:782843.  
doi: 10.3389/fbioe.2021.782843

<sup>1</sup>Division of Nephrology and Hypertension, Department of Medicine, Mayo Clinic, Rochester, MN, United States, <sup>2</sup>Department of Medical Statistics and Informatics, Medical Faculty University of Belgrade, Belgrade, Serbia, <sup>3</sup>Department of Obstetrics and Gynecology, Mayo Clinic, Rochester, MN, United States, <sup>4</sup>Department of Perinatology, Olmsted Medical Center, Rochester, MN, United States, <sup>5</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, United States, <sup>6</sup>Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, United States

**Background:** We aimed to assess the extent to which the buffy coat DNA methylome is representative of methylation patterns in constitutive white blood cell (WBC) types in normal pregnancy.

**Methods:** A comparison of differential methylation of buffy coat DNA vs DNA isolated from polymorphonuclear (PMN) and lymphocytic fractions was performed for each blood sample obtained within 24 h prior to delivery from 29 normotensive pregnant women. Methylation profiles were obtained using an Illumina Human Methylation 450 BeadChip and CHaMP bioinformatics pipeline. A subset of differentially methylated probes (DMPs) showing discordant methylation were further investigated using statistical modeling and enrichment analysis.

**Results:** The smallest number of DMPs was found between the buffy coat and the PMN fraction (2.96%). Pathway enrichment analysis of the DMPs identified biological pathways involved in the particular leukocyte lineage, consistent with perturbations during isolation. The comparisons between the buffy coat and the isolated fractions as a group using linear modeling yielded a small number of probes (~29,000) with discordant methylation. Demethylation of probes in the buffy coat compared to derived cell lines was more common and was prevalent in shelf and open sea regions.

**Conclusion:** Buffy coat is representative of methylation patterns in WBC types in normal pregnancy. The differential methylations are consistent with perturbations during isolation of constituent cells and likely originate *in vitro* due to the physical stress during cell separation and are of no physiological relevance. These findings help the interpretation of DNA methylation profiling in pregnancy and numerous other conditions.

**Keywords:** buffy coat, DNA methylation, epigenetics, white blood cells, pregnancy

## INTRODUCTION

DNA methylation, the addition of a 5-methyl group to a cytosine nucleotide at a cytosine-phosphate-guanine (CpG) site, is an essential epigenetic modification implicated in switching on and off genes that control early mammalian embryogenesis, including development, differentiation, imprinting, and cellular function (Straussman et al., 2009; Portela and Esteller, 2010; Hodges et al., 2011; Jones, 2012). Many CpG sites are clustered in CpG islands which are flanked by shores and shelves (up to 2, and 2–4 kb from CpG islands, respectively) and are separated by “open sea” regions, which represent the rest of the genome. Tissue-specific DNA methylation tends to be observed within CpG shores rather than islands (Doi et al., 2009). DNA methylation frequently affects a gene promoter leading to gene silencing, while DNA methylation of the gene body indicates active transcription (Hellman and Chess, 2007).

Altered DNA methylation of regulatory regions has been shown to contribute to the control of proliferative, invasive, and immune tolerance mechanisms involved in oncogenesis (Klutstein et al., 2016; Kuss-Duerkop et al., 2018)—a disease process with many parallels to that of normal pregnancy—with the common goal of providing a nutrient supply and immune tolerance to a growing tumor and fetus, respectively. In our previous studies, we described a transient state of hypomethylation in maternal leukocyte DNA occurring in normal early pregnancy (White et al., 2012). We also have demonstrated that in preeclampsia—a pregnancy-specific hypertensive disorder clinically characterized by multisystem involvement and, commonly, proteinuria—maternal leukocyte DNA showed genome-wide differential methylation favoring hypermethylation compared with normotensive pregnant controls (White et al., 2013; White et al., 2016). Similar to other investigators, we have performed our studies on buffy coat, a mixed leukocyte cell population, which is obtained after centrifuging whole anticoagulated blood at low speeds (Houseman et al., 2015). The scientific rigor of such results critically depends on the ability to discern any experimentally introduced methylation changes.

The buffy coat of a healthy, non-pregnant individual contains white blood cells, leukocytes, which are comprised of 70% polymorphonuclear (PMN) leukocytes, also referred to as granulocytes (0.5–1% basophils, ~65% neutrophils and 2–4% eosinophils), and roughly 30% mononuclear cells (3–8% monocytes, and 20–25% lymphocytes). Consequently, buffy coat comparisons may be confounded by shifts in cell type composition, which occur both in physiological conditions and during disease processes. This may be of particular importance for pregnancy, when an increase in the number of PMN leukocytes and monocytes, together with a decrease in the number of lymphocytes, dendritic cells, and natural killer (NK) cells occurs (Naccasha et al., 2001; Luppi et al., 2002a; Luppi et al., 2002b). Preeclampsia has been shown to be associated with an increase in the number of neutrophils (Järemo et al., 2000), along with activation of other PMN leukocytes and monocytes (Sacks et al., 1999). These shifts in white blood cell composition that occur both in normotensive

and preeclamptic pregnancies could affect the DNA methylation results found in the buffy coat (Adalsteinsson et al., 2012). As methylation can vary at specific loci among individual cell types, the shifts in buffy coat cell composition, rather than shifts in cell-intrinsic methylation patterns, may cause methylation differences between buffy coat samples.

To date, limited work has addressed the stability and correlation of DNA methylation patterns in buffy coat compared with the different leukocyte fractions. Given our research focus on pregnancy and its complications, the objective of the current study was to assess to what extent the buffy coat methylome is representative of, or different from, the distinct cell types that it contains, namely PMN leukocytes and lymphocytes. To that end, we compared genome-wide methylation patterns in buffy coat to those of the PMN and lymphocytic fractions in the same pregnant individuals from blood samples collected within 24 h prior to delivery.

## MATERIALS AND METHODS

### Sources for Cases and Blood Samples

Pregnant women were recruited prospectively at Mayo Clinic in Rochester, MN, United States. A convenience sample of 30 ml of blood was drawn into an EDTA tube from normotensive pregnant women admitted for delivery and within the 24 h prior to giving birth. In each of 29 cases, the leukocytes were separated into three groups within 2 hours of collection. Using typical slow centrifugation, a buffy coat was made and immediately frozen. The remaining sample was further subdivided into a PMN and a lymphocytic fraction using a Ficoll gradient and subsequently frozen.

### Clinical Data

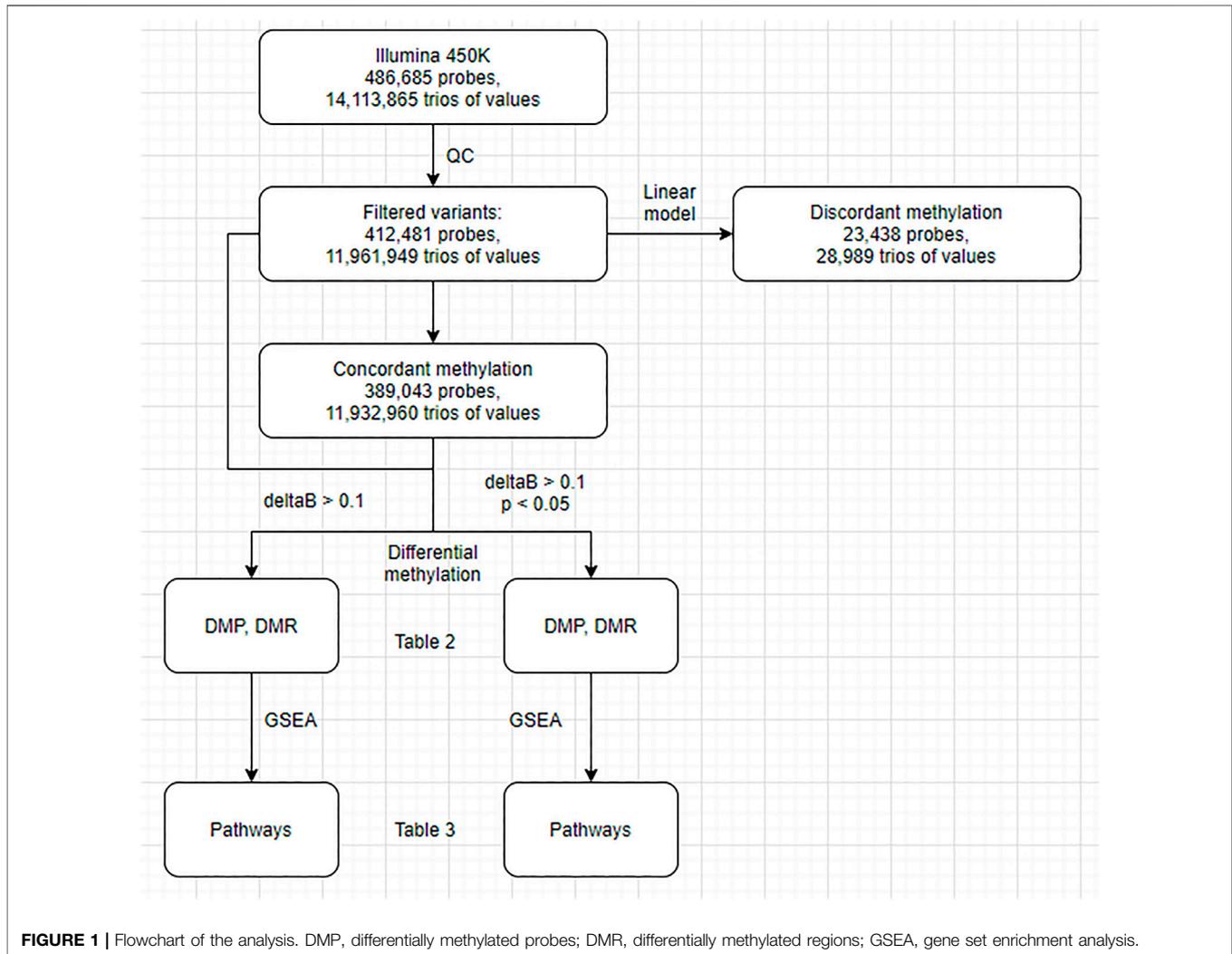
The medical records were abstracted for data including age, ethnicity, gravidity, body mass index (BMI), and gestational age (GA) at the time of delivery.

### DNA Extraction and Processing

In the buffy coat and PMN fraction, DNA was extracted using the AutoGenFlex DNA purification kit. In order to enrich the yield, manual extraction of DNA was performed for the lymphocytic fraction. Genomic DNA was quantified using a NanoDrop spectrophotometer, normalized with standard Pico Green methodology and plated. Bisulfite modification was performed using the EZ DNA Methylation Kit (Zymo Research).

### Methylation Assay

Plate maps were generated to determine the random location for each sample on the plate, as well as the samples that were run in duplicate. All samples were run in a single batch. Bisulfite-treated DNA was hybridized and imaged on an Illumina Infinium Human Methylation 450 K BeadChip that can detect methylation levels at 486,685 CpG dinucleotides across the genome and covers 96% of the CpG islands in the human genome.



## Quality Control

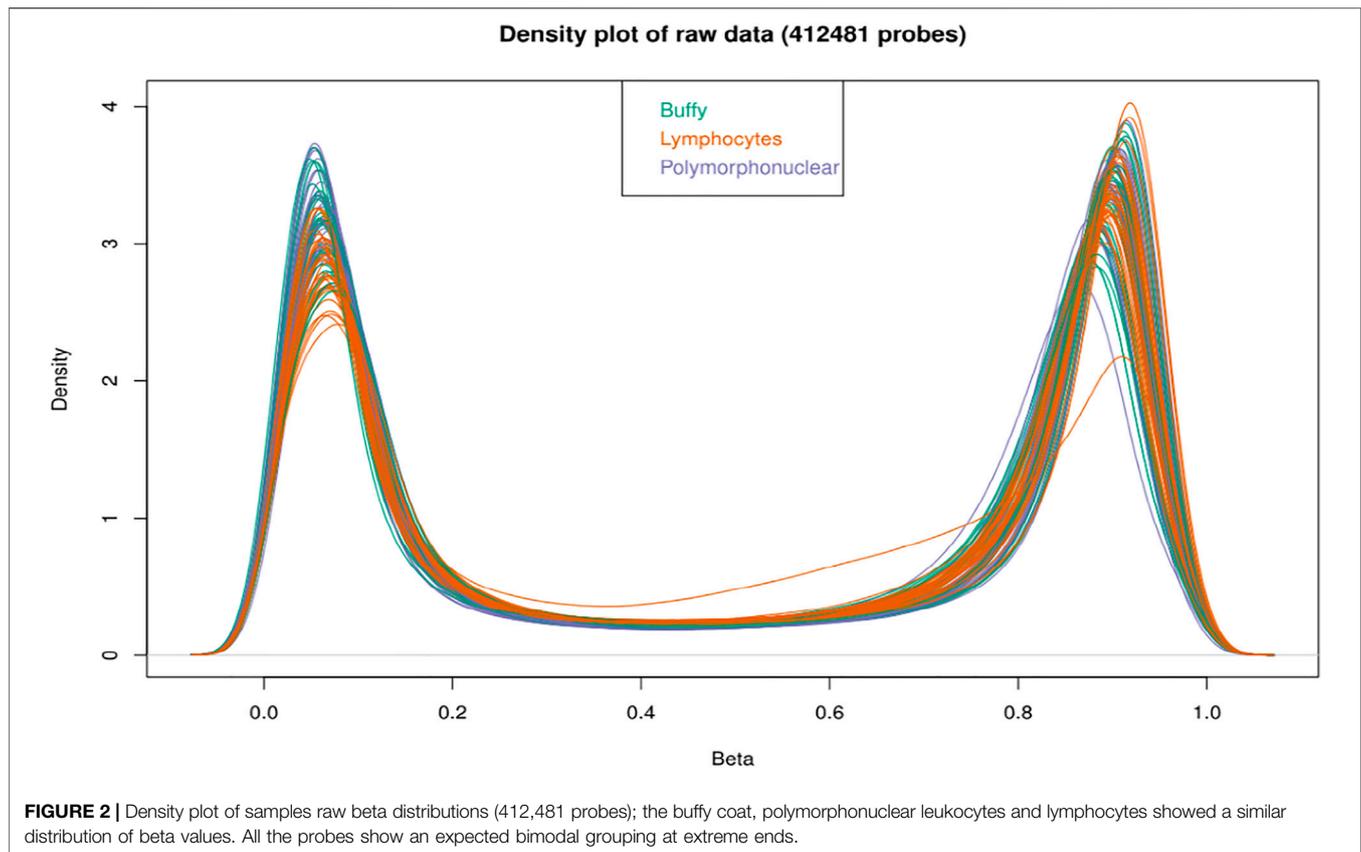
The samples were processed and then scanned using Illumina's iScan instrumentation. The raw data were then analyzed using Illumina's Genome Studio software (version 2011.1), with methylation module (version 1.9.0). Quality assessment of the array was conducted using the "Control Dashboard" in the Genome Studio software package, which includes a graphical inspection of the 10 types of embedded control probes: staining, extension, hybridization, target removal, bisulfite conversion, G/T mismatch, specificity, non-polymorphic controls, negative controls, and restoration controls.

Overall sample performance was determined by the total number of detected CpGs, the average detection  $p$  value across all CpG sites, and the distribution of average beta values for all CpGs. Call rates for each CpG site and sample were determined. Methylation sites and samples were excluded if the unreliable call rate (detection  $p$ -value) was greater than 5%. Technical replicate reproducibility was estimated by the Pearson correlation coefficient.

While all samples were bisulfite modified and run concurrently to avoid batch effects, multiple BeadChips were used, and may have variations in assay integrity leading to the "chip" effect. Thus, data were examined using principal components analysis and subsequent unsupervised hierarchical clustering of obtained components.

## Statistical Analysis

First, we conducted a standard differential methylation analysis to examine DNA methylation differences between the buffy coat and its PMN and lymphocytic fractions. Second, we aimed to differentiate between true differential methylation and discordant methylation, the latter being defined as methylation of specific CPGs in the buffy coat that did not match their methylation status in derived (PMN and lymphocytic) fractions (**Figure 1**). We used linear modeling to determine whether the linear combination of methylation profiles from PMN and lymphocytic fractions predicts the methylation profile of the buffy coat. Assuming that the methylation status of each CPG



site in the buffy coat will agree with the methylation status in the derived fractions, we identified concordant (the presence of prediction/agreement) and discordant (the lack of prediction/agreement) CpGs. Third, we attempted to investigate the factors predicting discordance (e.g., presence of nearby SNPs, GC composition of probes, presence of motifs).

Analysis of the differential methylation at the individual CpG probe level (DMP) was performed using the Limma package (version 3.30.13) (Smyth, 2004). To adjust for multiple comparisons, we used the Benjamini-Hochberg procedure of multiple testing, with  $\alpha = 0.05$ . Analysis of the differentially methylated regions (DMR) as aggregates of individual probes was performed using the DMRcate package (Peters et al., 2015). In DMRcate, a default false discovery rate (FDR) cutoff of 0.05 was used for CpG sites. The data generated by DMR and DMP were used as an input for **Figure 2**. Gene set enrichment analysis (GSEA) was performed using the GOMETH method to avoid probes-per-gene bias (Young et al., 2010; Geeleher et al., 2013; Phipson et al., 2016). Another filtering criterion, in addition to statistical significance, was differences in beta (delta beta) larger than 0.1. For the purposes of enrichment and correlation analyses, we used the GRCh37 reference build. The mappability and uniqueness of probes was calculated using Bismap (Karimzadeh et al., 2018). Normalized beta values and sample annotations are available to the public at [https://osf.io/324ak/?view\\_only=5c1c7cf5b77a40d3bb29b7d9c418f763](https://osf.io/324ak/?view_only=5c1c7cf5b77a40d3bb29b7d9c418f763).

**TABLE 1 |** Study population demographics.

Demographics	All patients (n = 29)
Age, yrs (mean $\pm$ SD)	29.2 $\pm$ 6.0
White, n (%)	24 (82.8)
Gravida, n (%)	
1	13 (44.8)
>1	16 (55.2)
Parity, n (%)	
0	14 (48.3)
1	4 (13.8)
>1	11 (37.9)
BMI, (kg/m <sup>2</sup> ), (mean $\pm$ SD)	29.3 $\pm$ 7.7
GA at delivery, weeks, (mean $\pm$ SD)	39.9 $\pm$ 1.4

BMI, body mass index; GA, gestational age; SD, standard deviation.

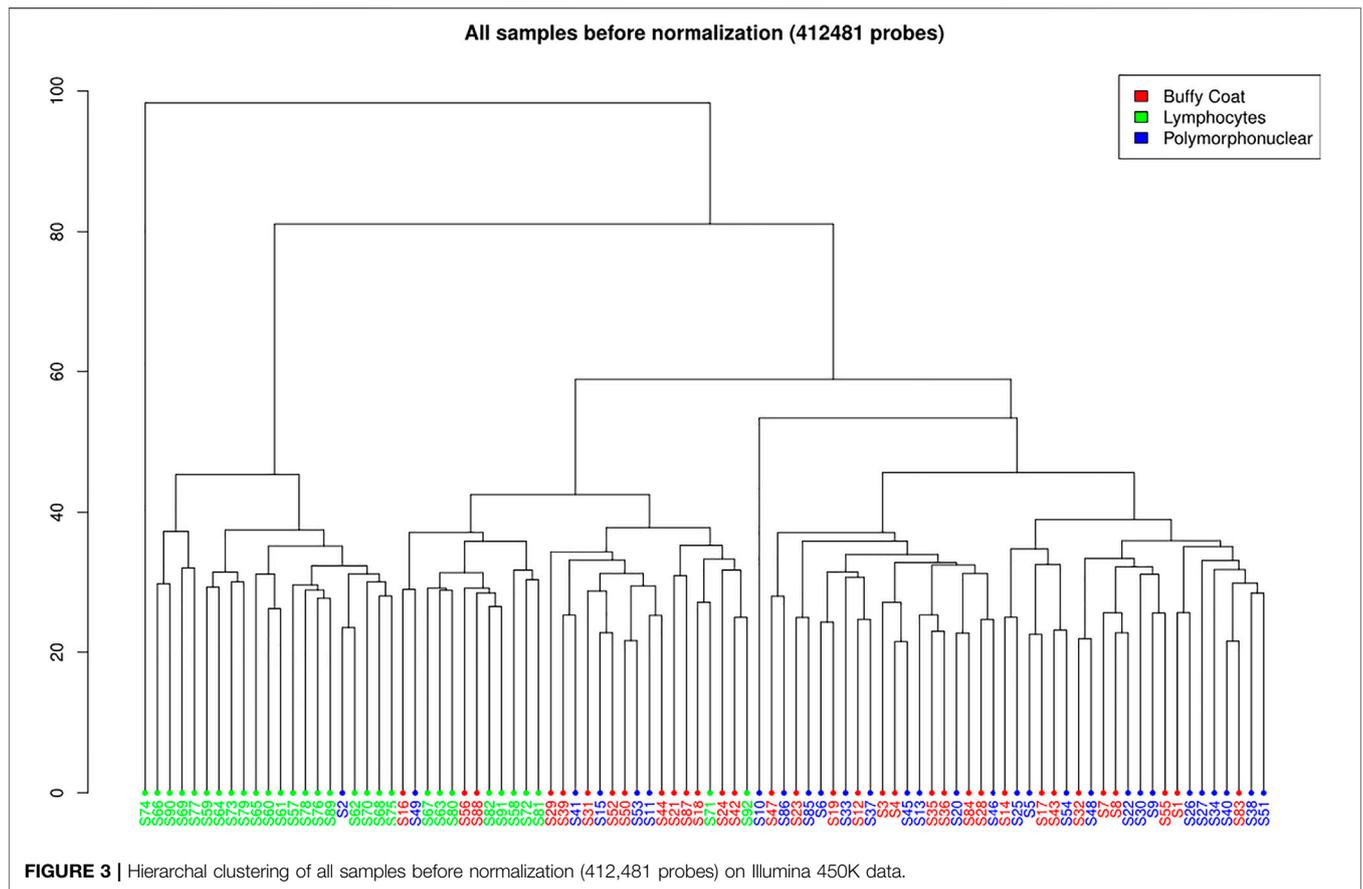
## RESULTS

### Clinical Characteristics

The blood samples of 29 normotensive pregnancies were collected and analyzed. Patients were predominantly white, and the mean gestational age at delivery was 39.9  $\pm$  1.4 weeks. Demographic and clinical characteristics are listed in **Table 1**.

### Differential Methylation

A total of 412,481 probes passed all the filters (**Figure 3**). A pairwise comparison between the buffy coat, PMN leukocytes,



**TABLE 2 |** Number of differentially methylated probes and regions.

Differentially methylated	Significance filters	Original analysis			Without violating probe trios <sup>a</sup>		
		BP	BL	PL	BP	BL	PL
DMP	$p < 0.05$	12,207	125,962	143,097	13,926	124,251	143,048
	$p < 0.05$ and $\Delta\beta > 0.1$	2,450	20,271	26,334	2,963	20,303	26,495
DMR	$p < 0.05$	596	5,270	6,086	627	5,209	6,120
	$p < 0.05$ and $\Delta\beta > 0.1$	0	164	25	0	157	24
GSEA-DMP	$p < 0.05$	15	0	0	14	0	0
	$p < 0.05$ and $\Delta\beta > 0.1$	11	32	17	12	31	23
GSEA-DMR	$p < 0.05$	0	10	19	0	14	23
	$p < 0.05$ and $\Delta\beta > 0.1$	0	1	1	0	0	0

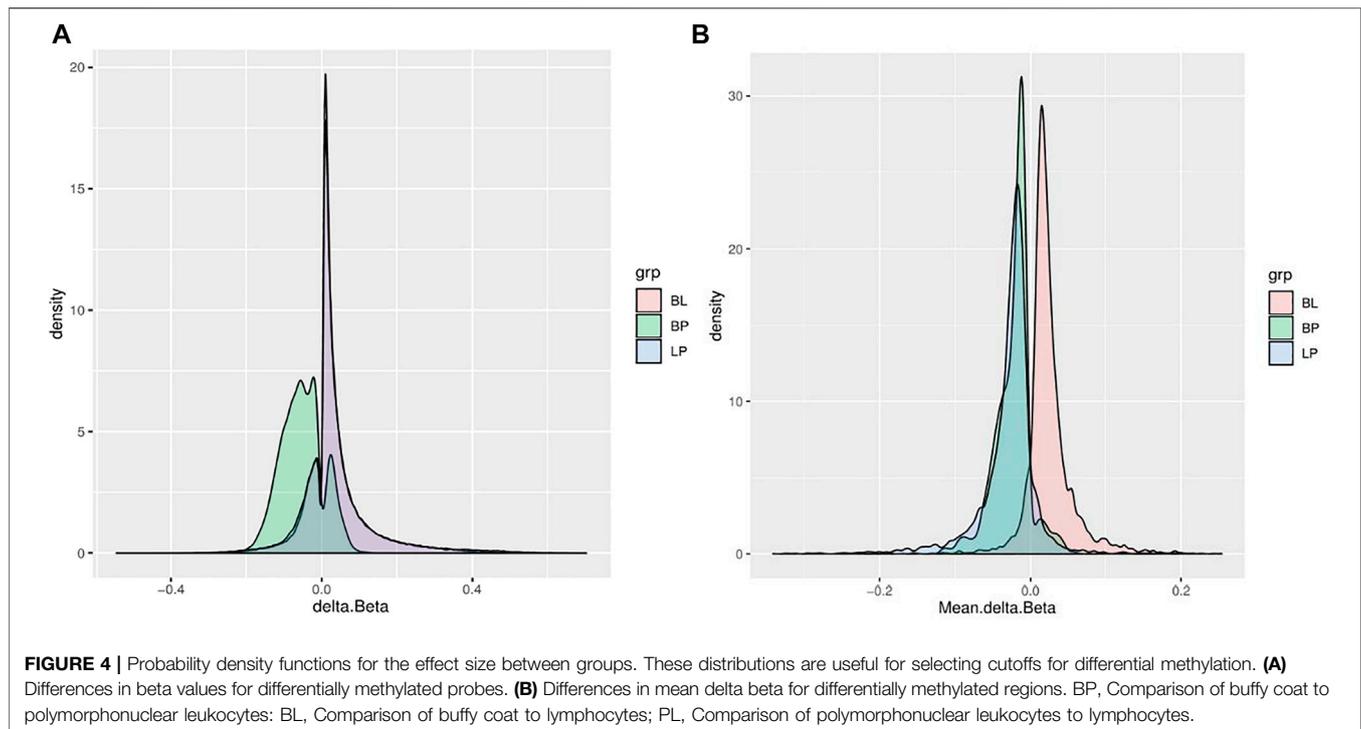
<sup>a</sup>Analysis of differentially methylated probes after elimination of outliers.

DMP, differentially methylated probes; DMR, differentially methylated regions; GSEA, gene set enrichment analysis.

BP, Comparison of buffly coat to polymorphonuclear leukocytes; BL, comparison of buffly coat to lymphocytes; PL, comparison of polymorphonuclear leukocytes to lymphocytes.

and lymphocytes was performed to identify differentially methylated probes (DMP). When considering the  $\alpha = 0.05$  significance level with Benjamini-Hochberg multiple testing corrections, the greatest number of differentially methylated sites was found between PMN and lymphocytes (143,097; 34.69%), and the smallest between the buffly coat and PMN (12,207; 2.96%), as represented in **Table 2**. This finding is not unexpected, considering that the buffly coat contains predominantly PMN leukocytes, so in this case only a small

proportion of other cells (mainly lymphocytes) was present and contributed to differential signals. The opposite rationale can be drawn for the comparison of buffly coat and lymphocyte fractions. Clustering in **Figure 3** corroborates this with a clearly distinct lymphocyte fraction and an overlap of the other two fractions. The analysis was subsequently repeated for differentially methylated regions (DMR), but these results have moderate interpretability due to the variable sizes of a region, and they primarily served as an input for gene set enrichment analysis



(GSEA). The magnitudes of differences for both the DMP and DMR analyses are shown in **Figure 4A** (delta beta) and 4 B (mean beta fold change). In addition, the number of differentially methylated probes/regions between the groups are shown when the applied threshold for delta beta was 0.1 (Original Analysis in **Table 2**).

### Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) was performed using either only  $p < 0.05$ , or by combining  $p < 0.05$  with delta beta  $> 0.1$ . When using only  $p$ -values, DMPs give enriched gene sets only for the buffy coat to polymorphonuclear leukocytes (BP) pairing, where the least differences in methylation are expected. Matching DMR analyses give opposite (and expected) results with no gene sets for the BP and a larger number of sets in the polymorphonuclear leukocytes to lymphocyte (PL) pairing. The pathways identified tended to be general in nature, e.g., *gene silencing by miRNA*, *mRNA binding involved in post-transcriptional gene silencing*, and *micro-ribonucleoprotein complex* (**Table 2**). In contrast, when we used delta beta values as an additional filter, GSEA generated several pathways for DMP and virtually no pathways for DMR, and the resulting pathways were indicative of the leukocyte lineages, e.g., *innate immune response*, *inflammatory response*, and *neutrophil degranulation* (**Table 3**).

### Consistency of Results Between Leukocyte Fractions and the Buffy Coat

Under the assumption that the buffy coat is mostly the weighted sum of PMN and lymphocytes, buffy coat methylations are

essentially estimated twice. We took advantage of the redundant methylation information to test whether we could extract the ratio of PMN and lymphocytes in the peripheral blood. Without additional constraints, we fit the model *buffy coat*  $\sim$  *lymphocytes* + *PMN leukocytes* + *error* for each subject and obtained coefficients for both lymphocytes and PMN summing to 0.98–0.99. This shows that the model is valid and reflects the physiological setup. The coefficients were interpreted as the differential blood count, with an expectancy of 20–40% lymphocytes and 40–80% PMN leukocytes.

We next performed linear modeling to determine whether the weighted combination of methylation profiles from isolated cellular fractions predicts the methylation profile of the buffy coat and identified the discordant CpG probes where the weighted combination deviates from methylation measured on the buffy coat. We applied the fitted model to each probe and each trio of samples and calculated the residuals (**Figure 5**). Although most probes centered around zero in a unimodal distribution, there was a prominent peak around the negative extreme (**Figure 6**), which corresponds to CpGs that are less methylated in the buffy coat than predicted from the relatively high CpG methylation in constituent fractions. Positive residuals correspond to methylated loci in buffy coat for which either of the fractions is hypomethylated and we can see several clusters between 0.3 and 0.8. Notably, the negative peak at the far left is dominant and has no matching positive counterpart. These discordant values with absolute residuals larger than expected could be a consequence of random probe errors or may be a result of physical/chemical stress occurring during cell separation, thus leading to methylation in isolated cellular fractions of

**TABLE 3** | Gene sets obtained.

Method	Sig. Filters	Original analysis			Without violating probe trios*		
		BP	BL	PL	BP	BL	PL
GSEA-DMP	$p < 0.05$	1.neutrophil degranulation 2.specific granule membrane 3.specific granule lumen 4.signal transduction 5.tertiary granule membrane 6.external side of plasma membrane 7.inflammatory response 8.cellular response to lipopolysaccharide 9.humoral immune response 10.innate immune response 11.positive regulation of interleukin-10 production 12.actin binding 13.plasma membrane 14.antimicrobial humoral response 15.tertiary granule lumen			1.neutrophil degranulation 2.specific granule membrane 3.tertiary granule membrane 4.external side of plasma membrane 5.actin binding 6.immune response 7.signal transduction 8.specific granule lumen 9.innate immune response 10.positive regulation of interleukin-10 production 11.humoral immune response 12.tertiary granule lumen 13.positive regulation of I-kappaB kinase/ NF-kappaB signaling 14.cellular response to lipopolysaccharide		
	$p < 0.05$ & $\Delta\beta > 0.1$	1.neutrophil degranulation 2.specific granule lumen 3.specific granule membrane 4.antimicrobial humoral response 5.innate immune response 6.tertiary granule lumen 7.cytosol 8.tertiary granule membrane 9.defense response to bacterium 10.azurophil granule lumen 11.protein binding	1.neutrophil degranulation 2.inflammatory response 3.specific granule membrane 4.innate immune response 5.cytokine-mediated signaling pathway 6.immune response 7.signal transduction 8.specific granule lumen 9.tertiary granule membrane 10.lipopolysaccharide-mediated signaling pathway 11.positive regulation of GTPase activity 12.B cell receptor signaling pathway	1.neutrophil degranulation 2.inflammatory response 3.specific granule membrane 4.innate immune response 5.signal transduction 6.cytokine-mediated signaling pathway 7.immune response 8.lipopolysaccharide-mediated signaling pathway 9.tertiary granule membrane 10.external side of plasma membrane 11.chemotaxis 12.signaling receptor activity	1.neutrophil degranulation 2.specific granule lumen 3.specific granule membrane 4.cytosol 5.antimicrobial humoral response 6.tertiary granule lumen 7.innate immune response 8.defense response to bacterium 9.protein binding 10.azurophil granule lumen 11.tertiary granule membrane 12.endoribonuclease activity	1.neutrophil degranulation 2.inflammatory response 3.specific granule membrane 4.innate immune response 5.cytokine-mediated signaling pathway 6.signal transduction 7.immune response 8.specific granule lumen 9.lipopolysaccharide-mediated signaling pathway 10.tertiary granule membrane 11.positive regulation of GTPase activity 12.positive regulation of I-kappaB kinase/ NF-kappaB signaling	1.neutrophil degranulation 2.inflammatory response 3.innate immune response 4.specific granule membrane 5.signal transduction 6.cytokine-mediated signaling pathway 7.immune response 8.lipopolysaccharide-mediated signaling pathway 9.tertiary granule membrane 10.cellular response to lipopolysaccharide 11.external side of plasma membrane 12.positive regulation of I-kappaB kinase/ NF-kappaB signaling

(Continued on following page)

**TABLE 3 |** (Continued) Gene sets obtained.

Method	Sig. Filters	Original analysis			Without violating probe trios*		
		BP	BL	PL	BP	BL	PL
GSEA- DMR	$p < 0.05$		13.cellular response to lipopolysaccharide	13.cellular response to lipopolysaccharide		13.cellular response to lipopolysaccharide	13.B cell receptor signaling pathway
			14.positive regulation of I-kappaB kinase/NF-kappaB signaling	14.positive regulation of GTPase activity		14.B cell receptor signaling pathway	14.positive regulation of interleukin-2 biosynthetic process
			15.secretory granule membrane	15.positive regulation of I-kappaB kinase/NF-kappaB signaling		15.secretory granule membrane	15.cell surface
			16.cell surface	16.secretory granule membrane		16.T cell activation	16.signaling receptor activity
			17.T cell activation	17.T cell activation		17.chemotaxis	17.chemotaxis
			18.perinuclear region of cytoplasm			18.cell surface	18.secretory granule membrane
			19.chemotaxis			19.negative regulation of tumor necrosis factor production	19.cell-cell adhesion
			20.Golgi apparatus			20.tertiary granule lumen	20.T cell activation
			21.negative regulation of tumor necrosis factor production			21.cellular defense response	21.positive regulation of GTPase activity
			22.tertiary granule lumen			22.T cell costimulation	22.focal adhesion
			23.cellular defense response			23.adaptive immune response	23.negative regulation of tumor necrosis factor production
			24.T cell costimulation			24.cytosol	
			25.adaptive immune response			25.platelet activation	
			26.platelet activation			26.GTPase activator activity	
			27.regulation of immune response			27.Golgi apparatus	
			28.GTPase activator activity			28.regulation of immune response	
			29.cytosol			29.perinuclear region of cytoplasm	
			30.external side of plasma membrane			30.extracellular exosome	
			31.Golgi membrane			31.Golgi membrane	
			32.extracellular exosome				
			1.mRNA binding involved in posttranscriptional gene silencing	1.gene silencing by miRNA		1.mRNA binding involved in posttranscriptional gene silencing	1.gene silencing by miRNA
			2.micro-ribonucleoprotein complex	2.mRNA binding involved in posttranscriptional gene silencing		2.micro-ribonucleoprotein complex	2.mRNA binding involved in posttranscriptional gene silencing
			3.gene silencing by miRNA	3.micro-ribonucleoprotein complex		3.gene silencing by miRNA	3.micro-ribonucleoprotein complex
			4.extracellular space	4.extracellular space		4.extracellular space	4.extracellular space
			5.DNA-binding transcription activator activity, RNA polymerase II-specific	5.DNA-binding transcription activator activity, RNA polymerase II-specific		5.DNA-binding transcription activator activity, RNA polymerase II-specific	5.extracellular region
			6.uterus development	6.extracellular region		6.extracellular region	6.DNA-binding transcription activator activity, RNA polymerase II-specific

(Continued on following page)

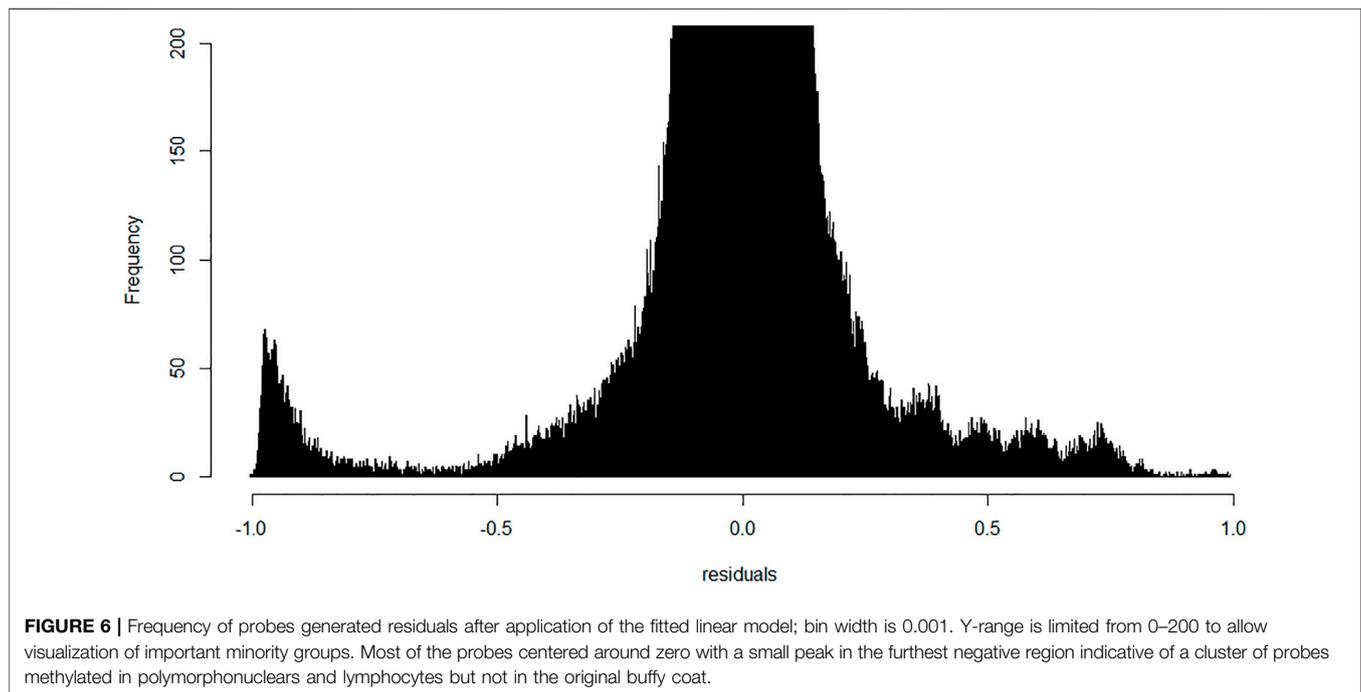
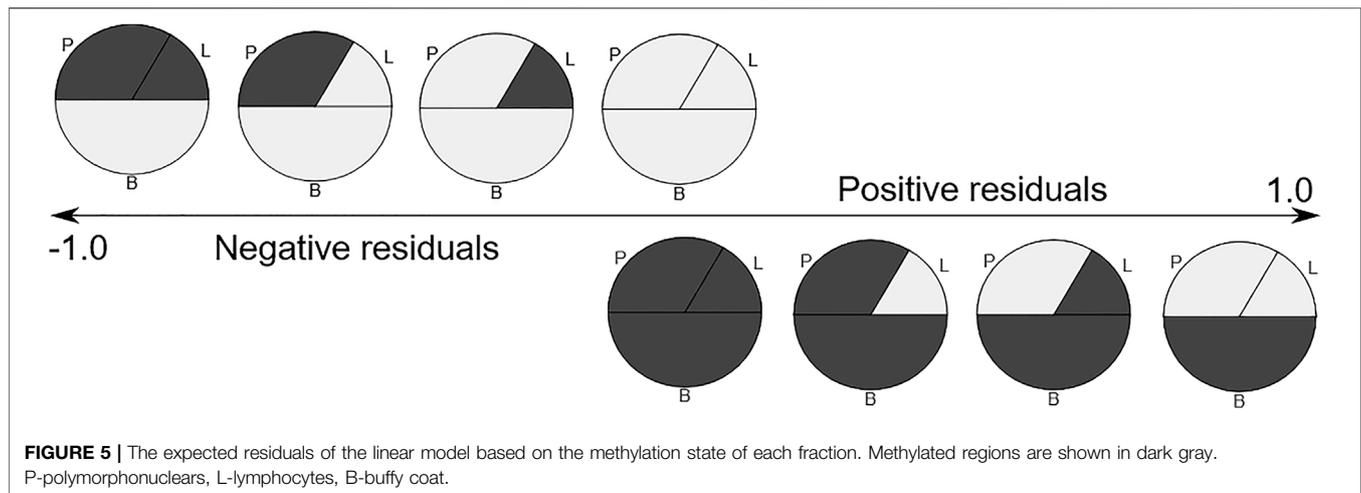
**TABLE 3** | (Continued) Gene sets obtained.

Method	Sig. Filters	Original analysis			Without violating probe trios*		
		BP	BL	PL	BP	BL	PL
			7.DNA-binding transcription factor activity, RNA polymerase II-specific	7.integral component of plasma membrane		7.uterus development	7.integral component of plasma membrane
			8.extracellular region	8.collagen-containing extracellular matrix		8.embryonic forelimb morphogenesis	8.collagen-containing extracellular matrix
			9.odontogenesis	9.DNA-binding transcription factor activity, RNA polymerase II-specific		9.odontogenesis	9.sequence-specific DNA binding
			10.embryonic forelimb morphogenesis	10.embryonic forelimb morphogenesis		10.positive regulation of neuron differentiation	10.DNA-binding transcription factor activity, RNA polymerase II-specific
				11.dopaminergic neuron differentiation		11.collagen-containing extracellular matrix	11.embryonic forelimb morphogenesis
				12.homophilic cell adhesion via plasma membrane adhesion molecules		12.dopaminergic neuron differentiation	12.homophilic cell adhesion via plasma membrane adhesion molecules
				13.sequence-specific DNA binding		13.integral component of plasma membrane	13.dopaminergic neuron differentiation
				14.uterus development		14.DNA-binding transcription factor activity, RNA polymerase II-specific	14.uterus development
				15.negative regulation of sprouting angiogenesis			15.negative regulation of sprouting angiogenesis
				16.calcium ion binding			16.chemical synaptic transmission
				17.extracellular matrix organization			17.cell junction
				18.neuropeptide signaling pathway			18.neuropeptide signaling pathway
				19.cell junction			19.extracellular matrix organization
							20.pancreas development
							21.calcium ion binding
							22.DNA-binding transcription factor activity
							23.growth factor activity
	$p < 0.05$ & $\Delta\beta > 0.1$		1.neutrophil degranulation	1.neutrophil degranulation			

particular site(s), which were not methylated in the initial, pre-separation analysis of the buffy coat.

For subsequent analysis, we calculated the  $p < 0.05$  cutoff using Bonferroni correction (z-score 5.29) and labeled as discordant all probes that fell outside of this range as outliers when standardized. The total number of discordant probe trios was 28,989 (0.24%) out of  $n^*412,481$  ( $n = 29$ ) included in the model. These values were filtered out of the

data from all three types of samples. All analyses were then repeated with the new “filtered” dataset (called without discordant probes) to determine whether the differentially methylated regions or probes were actually false positives. We found that, regardless of the filtering criteria used, the changes were minor in terms of DMP and DMR (**Table 2** left, without discordant probes) and did not seem to affect the top scoring GSEA sites (**Table 2** right).

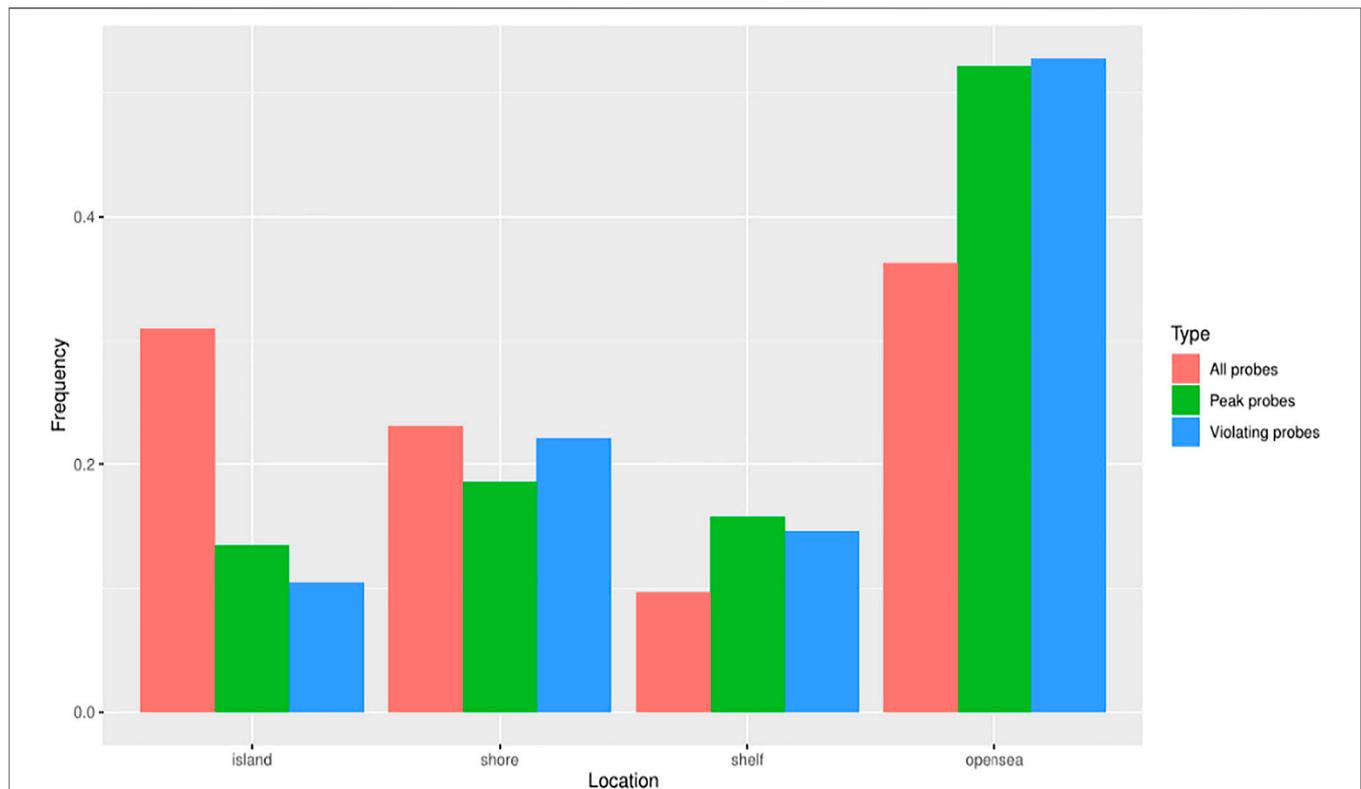


## Analysis of Discordant Probes

The number of discordant probes for which prediction of the methylation in the buffy coat was discordant from the prediction based on constituent fractions (PMN leukocytes and lymphocytes) was 23,438 (4.83% of probes). The remaining 389,043 probes that passed filters were labeled as “concordant.” We hypothesized that there would be something in the design or location of those probes that would predict discordance. We downloaded the Illumina 450K manifest and analyzed the properties that differed between the two datasets.

After intersecting the locations of discordant probes with the unique Bismarq ranges, only 11 of 23,438 discordant probes

proved to map to non-unique segments in the bisulfite-converted reference, which are expected to yield more reliable results. On the other hand, 212 of 389,043 concordant probes proved to be multi-mapped in the concordant probes. Both types of probes had identical uniqueness (99.95%) rounded to four significant digits, suggesting that systematic mis-mapping does not explain discordance. Interestingly, 11,730 of 73,031 (16.06%) probes that were filtered out by ChAMP default QC mapped to non-unique 50-mers according to this methodology. This high uniqueness of virtually all the probes that passed filters indicates that the probes are hybridizing with the expected sequences and that the discordance is an *in-vitro*, pre-measurement phenomenon.



**FIGURE 7 |** Genomic loci classification of the probes. Violating and peak probes were less frequently located in CpG islands compared to all probes. Violating probes - probes with at least one discordant beta value. Peak probes - probes appearing in the peak of residuals below  $-0.6$ .

**TABLE 4 |**  $p$ -values obtained by chi-squared proportion tests on different types of discordant probes with the pool of all probes, stratified by the CpG region. All results are significant at  $\alpha = 0.05$

Comparison with all probes	Islands	Shore	Shelf	Open sea
Violating probes	< 2.2e-16	9.479e-08	< 2.2e-16	< 2.2e-16
Peak probes	< 2.2e-16	1.726e-15	< 2.2e-16	< 2.2e-16

We also noted a prominent difference in guanine-cytosine (GC) content, and the discordant probes in the linear model had approximately 3% lower GC. One of the probes was discordant in 14 of 29 trio samples but had only a moderate GC content of 55%. Similarly, many of the probes with a higher error rate had lower GC values.

As mappability and GC content strongly differ along the genome, and presumably correlate with the presence of CpG islands, we investigated the location of discordant probes and tried to determine any preference for genomic loci relative to CpG islands. To avoid effect of overrepresentation of individual probes without the understanding of the underlying process, we counted each probe only once regardless of the number of occurrences in the investigated groups. Resorting to the standard four groups—CpG islands, shores, shelves, and open sea—we plotted and tested the proportion of groups for all probes

(412,481), discordant probes (23,438), and the subset of probes giving the peak of residuals below  $-0.6$  (peak probes, 4,535). We noted that the genomic location of both discordant groups (discordant and peak probes) was significantly different from the group of all probes (Figure 7; Table 4). These differences indicated that the discordant trios were most likely to be located in open sea regions, and least likely to be present in CpG islands.

## DISCUSSION

Our present study reports several novel findings obtained through a comparative analysis of the methylation profile in buffy coat versus PMN and lymphocyte cell lines. First, the buffy coat methylation profile was representative of methylation patterns of derived cell lines. We characterized the genome-wide methylation profile of the buffy coat, PMN and the lymphocytic fractions drawn from the same individual in 29 normotensive pregnancies across  $>450,000$  CpG sites in genes across the entire genome. We performed pairwise comparisons that yielded a number of probes that are differentially methylated, but with relatively small differences in beta values. We found a very small percentage of differentially methylated CpG sites when the buffy coat was compared to the PMN fraction (2.96%) and a greater percentage of differentially methylated CpG sites between PMN and lymphocytes (34.69%), consistent with the fact that the

PMN fraction is the main constituent of the buffy coat. Second, differential methylation occurred in biological pathways that are specific for the derived cell lines, such as neutrophil degranulation and cytokine-mediated signaling pathways, consistent with the sensitivity of these cell-type specific pathways to perturbations during the cell separation process. Third, using a linear model we were able to identify approximately 29,000 probes for which the prediction of CpG methylation within the buffy coat is discordant from the prediction from constituent fractions. These discordant probes had a lower GC content, and the CpG sites involved also preferentially affected the open sea and shelf regions; discordant methylation seems to be an *in vitro* phenomenon and likely due to the separation process of the buffy coat.

To date, few studies have provided comparative analyses of different epigenetic profiles across different blood fractions in pregnancy. One notable example is a study of the cord blood, which showed that methylation in whole blood is reasonably comparable to buffy coat in a small number ( $n = 8$ ) of paired samples (Dou et al., 2018). However, the concordance between special cell populations (such as lymphocytes or monocytes) with either whole blood or buffy coat was not studied. To address cell type composition of complex tissues, we have previously developed and characterized *in silico* epigenomic deconvolution methods (Onuchic et al., 2016; Decamps et al., 2020) that infer DNA methylation of constituent cell types by minimizing residuals.

We now extend residual analysis to address the complexity of the blood samples with their multiple cell types by providing a novel approach to analyzing and interpreting methylation profiles from the tissue of origin (whole blood, buffy coat) and derived cell lines (PMN and lymphocyte cell lines). Negative residuals in the overall linear model indicate that at least one of the fractions changed its state from unmethylated to methylated in derived cell lines (i.e., PMN and lymphocytes), while the positive residuals representing the opposite process, were few. Most residuals in the interval between  $-0.7$  and  $0.7$  seem balanced, in accordance with the expected normal distribution as a result of unobserved variation. On the other hand, close to 50% of discordant probes are grouped in a clear peak below  $-0.8$  and have no substantial positive counterpart, i.e., the same probe being methylated in the buffy coat, but unmethylated in derived cell line(s), indicating strong preference towards methylation in PMN and lymphocytes. This targeted DNA inactivation, preferentially targeting isolated CpGs in the open-sea regions, could be the result of the separation process, and the mechanical/chemical stress exerted on the cells removed from their normal medium.

While our study was performed on a relatively small sample size, a pairwise comparison of samples drawn from the same individual should limit the effects of potential confounders. We studied only normotensive pregnancies within 24 h prior to delivery, thus limiting the extrapolation of our results to other conditions with vastly different white blood cell composition or with selective methylation of a certain white blood cell fraction.

Our results have the potential to improve rigor and reproducibility of studies involving epigenomic profiling of buffy coat samples. For example, in the context of one of our

previous studies, we identified a state of transient hypomethylation in normal early pregnancy compared with non-pregnancy (White et al., 2012). This hypomethylation would spontaneously revert after delivery. Our sample included patients close to delivery, suggesting a potential tendency towards global hypomethylation. DNA methylation can be affected by many factors including age (Bell et al., 2012), race, BMI, smoking, gestational diabetes (Wu et al., 2018), and preeclampsia (White et al., 2013). The patients in our study had healthy normotensive pregnancies, with a similar age range, a relatively homogeneous ethnicity, and free of major medical comorbidities that could affect DNA methylation. Some fetal DNA contamination was possible, but it likely represented a very small fraction (up to 6% of total measured DNA) (Bischoff, 2002) and, therefore, it was unlikely to have significantly changed our results. However, in our original study, we could not firmly establish the relation between the DNA methylation profiles in buffy coat samples and that of constituent cell types. Our current results help establish this relation, thus improving both the rigor and biological interpretability of our results. Moreover, our results will help integrate results obtained from buffy coat samples and those obtained from profiling of isolated constituent cell types. Finally, our results are not limited to studies of pregnancy and have implications for numerous other studies involving DNA methylation profiling of buffy coat samples and constituent cell types.

A potential limitation of this study is that we did not account for the effect of DNA extraction techniques. DNA was isolated in samples using the AutoGenFlex DNA purification kit for the buffy coat and the neutrophil fraction, and manual extraction of DNA was performed for the lymphocytic fraction. The use of different methods for DNA isolation may affect methylation results, although the methylation mark is considered quite stable, but this is currently minimally discussed in the literature (Hjorthaug et al., 2018). Blood sample processing techniques, such as using Ficoll density centrifugation, can also potentially confound methylation. Finally, we used the Illumina 450K array, which covers only  $\sim 2\%$  of total CpGs within the human genome. The Illumina Infinium Human Methylation 450, however, controls for the confounding and takes SNPs into account when analyzing output data. Future directions include using a different sample cohort and methodology to confirm our results. We hope to reproduce our findings in a larger, independent sample, and in a different study population that would confirm the stability of the methylome in the buffy coat and its sub-fractions.

Despite these limitations, our results demonstrate that the buffy coat methylation profile is representative of the methylation patterns in white blood cell types in normal pregnancy obtained using Illumina Human Methylation 450 BeadChip. Small differences in the buffy coat composition may confound the methylation analysis at a very small number of CpG sites, but this is not likely to affect most results. Current methods to adjust for cellular heterogeneity, either by excluding these differentially methylated genes, or, better yet, adjusting methylation data to account for these differences in buffy coat composition, improve the robustness of methylome analysis in buffy coat. Overall, our

results support DNA methylation profiling of buffy coats as an acceptable approach for epigenomic profiling in pregnancy research and suggest that separation is likely only needed when studying lineage-specific diseases.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in an online repository. The name of the repository and accession link can be found below: OSF, [https://osf.io/324ak/?view\\_only=5c1c7cf5b77a40d3bb29b7d9c418f763](https://osf.io/324ak/?view_only=5c1c7cf5b77a40d3bb29b7d9c418f763).

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Mayo Clinic Institutional Review Board. The

## REFERENCES

- Adalsteinsson, B. T., Gudnason, H., Aspelund, T., Harris, T. B., Launer, L. J., Eiriksdottir, G., et al. (2012). Heterogeneity in white Blood Cells Has Potential to Confound DNA Methylation Measurements. *PLoS One* 7 (10), e46705. doi:10.1371/journal.pone.0046705
- Bell, J. T., Tsai, P.-C., Yang, T.-P., Pidsley, R., Nisbet, J., Glass, D., et al. (2012). Epigenome-wide Scans Identify Differentially Methylated Regions for Age and Age-Related Phenotypes in a Healthy Ageing Population. *Plos Genet.* 8 (4), e1002629. doi:10.1371/journal.pgen.1002629
- Bischoff, F. Z. (2002). Cell-free Fetal DNA and Intact Fetal Cells in Maternal Blood Circulation: Implications for First and Second Trimester Non-invasive Prenatal Diagnosis. *Hum. Reprod. Update* 8 (6), 493–500. doi:10.1093/humupd/8.6.493
- Decamps, C., Privé, F., Privé, F., Bacher, R., Jost, D., Waguet, A., et al. (2020). Guidelines for Cell-type Heterogeneity Quantification Based on a Comparative Analysis of Reference-free DNA Methylation Deconvolution Software. *BMC Bioinformatics* 21 (1), 16. doi:10.1186/s12859-019-3307-2
- Doi, A., Park, I.-H., Wen, B., Murakami, P., Aryee, M. J., Irizarry, R., et al. (2009). Differential Methylation of Tissue- and Cancer-specific CpG Island Shores Distinguishes Human Induced Pluripotent Stem Cells, Embryonic Stem Cells and Fibroblasts. *Nat. Genet.* 41 (12), 1350–1353. doi:10.1038/ng471
- Dou, J., Schmidt, R. J., Benke, K. S., Newschaffer, C., Hertz-Picciotto, I., Croen, L. A., et al. (2018). Cord Blood Buffy Coat DNA Methylation Is Comparable to Whole Cord Blood Methylation. *Epigenetics* 13 (1), 108–116. doi:10.1080/15592294.2017.1417710
- Geeleher, P., Hartnett, L., Egan, L. J., Golden, A., Raja Ali, R. A., and Seoighe, C. (2013). Gene set Analysis Is Severely Biased when Applied to Genome-wide Methylation Data. *Bioinformatics* 29 (15), 1851–1857. doi:10.1093/bioinformatics/btt311
- Hellman, A., and Chess, A. (2007). Gene Body-specific Methylation on the Active X Chromosome. *Science* 315 (5815), 1141–1143. doi:10.1126/science.1136352
- Hjorthaug, H. S., Gervin, K., Mowinckel, P., and Munthe-Kaas, M. C. (2018). Exploring the Influence from Whole Blood DNA Extraction Methods on Infinium 450K DNA Methylation. *PLoS One* 13 (12), e0208699. doi:10.1371/journal.pone.0208699
- Hodges, E., Molaro, A., Dos Santos, C. O., Thekkat, P., Song, Q., Uren, P. J., et al. (2011). Directional DNA Methylation Changes and Complex Intermediate States Accompany Lineage Specificity in the Adult Hematopoietic Compartment. *Mol. Cell* 44 (1), 17–28. doi:10.1016/j.molcel.2011.08.026
- Houseman, E. A., Kim, S., Kelsey, K. T., and Wiencke, J. K. (2015). DNA Methylation in Whole Blood: Uses and Challenges. *Curr. Envir Health Rpt* 2 (2), 145–154. doi:10.1007/s40572-015-0050-3
- patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

Conception and design of the work: RG, IV, OM, WW, LT, JC, AM, NM, VG; Acquisition, analysis or interpretation of data for the work: RG, IV, OM, WW, LT, JC, AM, NM, VG; Drafting the article: RG, IV, OM, WW, LT, CJ, AM, NM, VG; Revising article critically for important intellectual content: RG, IV, OM, WW, LT, JC, AM, NM, VG.

## FUNDING

This work was supported by the National Institutes of Health R01-HL136348.

- Järemo, P., Lindahl, T. L., Lennmarken, C., and Forsgren, H. (2000). The Use of Platelet Density and Volume Measurements to Estimate the Severity of Pre-eclampsia. *Eur. J. Clin. Invest.* 30 (12), 1113–1118. doi:10.1046/j.1365-2362.2000.00753.x
- Jones, P. A. (2012). Functions of DNA Methylation: Islands, Start Sites, Gene Bodies and beyond. *Nat. Rev. Genet.* 13 (7), 484–492. doi:10.1038/nrg3230
- Karimzadeh, M., Ernst, C., Kundaje, A., and Hoffman, M. M. (2018). Umap and Bismap: Quantifying Genome and Methyloome Mappability. *Nucleic Acids Res.* 46 (20), e120. doi:10.1093/nar/gky677
- Klutstein, M., Nejman, D., Greenfield, R., and Cedar, H. (2016). DNA Methylation in Cancer and Aging. *Cancer Res.* 76 (12), 3446–3450. doi:10.1158/0008-5472.can-15-3278
- Kuss-Duerkop, S. K., Westrich, J. A., and Poon, D. (2018). DNA Tumor Virus Regulation of Host DNA Methylation and its Implications for Immune Evasion and Oncogenesis. *Viruses* 10 (2), 82. doi:10.3390/v10020082
- Luppi, P., Haluszczak, C., Betters, D., Richard, C. A., Trucco, M., and DeLoia, J. A. (2002). Monocytes Are Progressively Activated in the Circulation of Pregnant Women. *J. Leukoc. Biol.* 72 (5), 874–884. doi:10.1189/jlb.72.5.874
- Luppi, P., Haluszczak, C., Trucco, M., and DeLoia, J. A. (2002). Normal Pregnancy Is Associated with Peripheral Leukocyte Activation. *Am. J. Reprod. Immunol.* 47 (2), 72–81. doi:10.1034/j.1600-0897.2002.10041.x
- Naccasha, N., Gervasi, M.-T., Chaiworapongsa, T., Berman, S., Yoon, B. H., Maymon, E., et al. (2001). Phenotypic and Metabolic Characteristics of Monocytes and Granulocytes in normal Pregnancy and Maternal Infection. *Am. J. Obstet. Gynecol.* 185 (5), 1118–1123. doi:10.1067/mob.2001.117682
- Onuchic, V., Hartmaier, R. J., Boone, D. N., Samuels, M. L., Patel, R. Y., White, W. M., et al. (2016). Epigenomic Deconvolution of Breast Tumors Reveals Metabolic Coupling between Constituent Cell Types. *Cell Rep.* 17 (8), 2075–2086. doi:10.1016/j.celrep.2016.10.057
- Peters, T. J., Buckley, M. J., Statham, A. L., Pidsley, R., Samarasinghe, K., V Lord, R., et al. (2015). De Novo identification of Differentially Methylated Regions in the Human Genome. *Epigenetics & Chromatin* 8, 6. doi:10.1186/1756-8935-8-6
- Phipson, B., Maksimovic, J., and Oshlack, A. (2016). missMethyl: an R Package for Analyzing Data from Illumina's HumanMethylation450 Platform. *Bioinformatics* 32 (2), 286–288. doi:10.1093/bioinformatics/btv560
- Portela, A., and Esteller, M. (2010). Epigenetic Modifications and Human Disease. *Nat. Biotechnol.* 28 (10), 1057–1068. doi:10.1038/nbt.1685
- Sacks, G., Sargent, I., and Redman, C. (1999). An Innate View of Human Pregnancy. *Immunol. Today* 20 (3), 114–118. doi:10.1016/s0167-5699(98)01393-0
- Smyth, G. K. (2004). Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Stat. Appl. Genet. Mol. Biol.* 3, 1–25. doi:10.2202/1544-6115.1027

- Straussman, R., Nejman, D., Roberts, D., Steinfeld, I., Blum, B., Benvenisty, N., et al. (2009). Developmental Programming of CpG Island Methylation Profiles in the Human Genome. *Nat. Struct. Mol. Biol.* 16 (5), 564–571. doi:10.1038/nsm.1594
- White, W. M., Brost, B. C., Sun, Z., Rose, C., Craici, I., Wagner, S. J., et al. (2012). Normal Early Pregnancy: a Transient State of Epigenetic Change Favoring Hypomethylation. *Epigenetics* 7 (7), 729–734. doi:10.4161/epi.20388
- White, W. M., Brost, B., Sun, Z., Rose, C., Craici, I., Wagner, S. J., et al. (2013). Genome-wide Methylation Profiling Demonstrates Hypermethylation in Maternal Leukocyte DNA in Preeclamptic Compared to Normotensive Pregnancies. *Hypertens. Pregnancy* 32 (3), 257–269. doi:10.3109/10641955.2013.796970
- White, W. M., Sun, Z., Borowski, K. S., Brost, B. C., Davies, N. P., Rose, C. H., et al. (2016). Preeclampsia/Eclampsia Candidate Genes Show Altered Methylation in Maternal Leukocytes of Preeclamptic Women at the Time of Delivery. *Hypertens. Pregnancy* 35 (3), 394–404. doi:10.3109/10641955.2016.1162315
- Wu, P., Farrell, W. E., Haworth, K. E., Emes, R. D., Kitchen, M. O., Glossop, J. R., et al. (2018). Maternal Genome-wide DNA Methylation Profiling in Gestational Diabetes Shows Distinctive Disease-Associated Changes Relative to Matched Healthy Pregnancies. *Epigenetics* 13 (2), 122–128. doi:10.1080/15592294.2016.1166321
- Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene Ontology Analysis for RNA-Seq: Accounting for Selection Bias. *Genome Biol.* 11 (2), R14. doi:10.1186/gb-2010-11-2-r14
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Ghamrawi, Velickovic, Milicevic, White, Thistlethwaite, Cunningham, Milosavljevic, Milic and Garovic. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.