






Targeted DNA Methylation Analysis Facilitates Leukocyte Counts in Dried Blood Samples

Wouter H.G. Hubens ^{a,b}, Tiago Maié,^c Matthis Schnitker,^{a,b} Ledio Bocova,^{a,b} Deepika Puri,^{a,b} Martina Wessiepe,^d Jan Kramer,^{e,f} Lothar Rink ^g, Steffen Koschmieder ^{h,i}, Ivan G. Costa ^c and Wolfgang Wagner ^{a,b,i,*}

BACKGROUND: Cell-type specific DNA methylation (DNAm) can be employed to determine the numbers of leukocyte subsets in blood. In contrast to conventional methods for leukocyte counts, which are based on cellular morphology or surface marker protein expression, the cellular deconvolution based on DNAm levels is applicable for frozen or dried blood. Here, we further enhanced targeted DNAm assays for leukocyte counts in clinical application.

METHODS: DNAm profiles of 40 different studies were compiled to identify CG dinucleotides (CpGs) with cell-type specific DNAm using a computational framework, CimpleG. DNAm levels at these CpGs were then measured with digital droplet PCR in venous blood from 160 healthy donors and 150 patients with various hematological disorders. Deconvolution was further validated with venous blood ($n = 75$) and capillary blood ($n = 31$) that was dried on Whatman paper or on Mitra microsampling devices.

RESULTS: In venous blood, automated cell counting or flow cytometry correlated well with epigenetic estimates of relative leukocyte counts for granulocytes ($r = 0.95$), lymphocytes ($r = 0.97$), monocytes ($r = 0.82$), CD4 T cells ($r = 0.84$), CD8 T cells ($r = 0.94$), B cells ($r = 0.96$), and NK cells ($r = 0.72$). Similar correlations and precisions were achieved for dried blood samples. Spike-in with a reference plasmid enabled accurate epigenetic estimation of absolute leukocyte counts from dried blood samples, correlating with conventional venous ($r = 0.86$) and capillary ($r = 0.80$) blood measurements.

CONCLUSIONS: The advanced selection of cell-type specific CpGs and utilization of digital droplet PCR analysis provided accurate epigenetic blood counts.

Analysis of dried blood facilitates self-sampling with a finger prick, thereby enabling easier accessibility to testing.

Introduction

Blood differential counts are frequently used to assist in the diagnosis of a variety of disorders or to track treatment. They are typically carried out using either automated cell counters or flow cytometry (1, 2). To this end, fresh blood must be sampled because cell counts are affected within 24 hours, due to loss of cellular integrity, coagulation, disrupted electrical impedance, or poor antibody binding (3–5). By contrast, leukocyte counts from dried blood spots would facilitate self-sampling by a finger prick, without the need for qualified personnel, long-term storage, and shipment of blood specimens, and bears significantly less risk of hemorrhage compared to using venous puncture (6).

Deconvolution of leukocyte subsets is also feasible by epigenetic means. DNA methylation (DNAm) patterns are consistently modulated during differentiation in a cell-type specific manner (7). Epigenetic signatures that integrate DNAm levels at many CG dinucleotides (CpGs) can therefore be used to determine the composition of cell types in tissue (8, 9), or of leukocytes in blood (10–12). Epigenetic signatures for deconvolution of hematopoietic subsets were initially generated for Illumina BeadChip data (10–12). However, since genome-wide DNAm profiling is hardly applicable for routine diagnostics, it is important to define targeted DNAm analysis of individual CpGs for clinical

^aInstitute for Stem Cell Biology, Faculty of Medicine, RWTH Aachen University, Aachen, Germany; ^bHelmholtz-Institute for Biomedical Engineering, RWTH Aachen University, Aachen, Germany; ^cInstitute for Computational Genomics, Faculty of Medicine, RWTH Aachen University, Aachen, Germany; ^dInstitute for Transfusion Medicine, Faculty of Medicine, RWTH Aachen University, Aachen, Germany; ^eDivision of Nephrology and Transplantation Unit, Department of Internal Medicine I, University of Lübeck, Lübeck, Germany; ^fLADR Laboratory Group Dr. Kramer & Colleagues, Geesthacht, Germany; ^gInstitute of Immunology, Faculty of Medicine, RWTH Aachen

University, Aachen, Germany; ^hDepartment of Hematology, Oncology, Hemostaseology and Stem Cell Transplantation, Faculty of Medicine, RWTH Aachen University, Aachen, Germany; ⁱCenter for Integrated Oncology Aachen Bonn Cologne Düsseldorf (CIO ABCD), Aachen, Germany.

*Address correspondence to this author at: Institute for Stem Cell Biology, RWTH Aachen University Medical School, Pauwelsstr. 20, 52074 Aachen, Germany. E-mail wwagner@ukaachen.de. Received May 31, 2023; accepted August 10, 2023 <https://doi.org/10.1093/clinchem/hvad143>

application (13). To this end, many different methods can be used, such as quantitative PCR (qPCR) (14), pyrosequencing (15), barcoded amplicon sequencing (16), or methylation-specific digital droplet PCR (ddPCR) (17). Particularly, ddPCR may facilitate more reliable DNAm measurements as there is no PCR bias between methylated and nonmethylated sequences (18). In addition to selecting a sensitive method, the choice of cell-type specific candidate CpGs is crucial to establish reliable targeted biomarkers. In our previous work, we have selected such genomic sites based on one dataset of sorted leukocyte subsets (15, 19). While these signatures were validated on patient material, they did not show a very strong correlation with conventional blood counts for all leukocyte subsets (17).

In the current study, we have therefore compiled the available Illumina BeadChip DNAm profiles of sorted leukocytes on the NCBI's Gene Expression Omnibus. Based on this, we identified new cell-type specific CpGs using a computational framework named "CimpleG" (20). Epigenetic predictions based on ddPCR measurements at these CpGs clearly improved the accuracy of leukocyte counts in blood samples of healthy donors and patients with hematologic diseases. Furthermore, we demonstrate applicability for dried blood samples.

Materials and Methods

SELECTION OF DIFFERENTIALLY METHYLATED SITES

We searched the Gene Expression Omnibus database for publicly available DNAm profiles (450k or EPIC Illumina BeadChip arrays) of sorted human hematopoietic cells (Supplemental Table 1). These profiles were subsequently analyzed using CimpleG (source code available at github.com/CostaLab/CimpleG) (20). A detailed description of the search strategy and CimpleG analysis is provided in the Supplemental Methods.

BLOOD SAMPLES

All blood samples were collected after informed and written consent, in accordance with the Declaration of Helsinki. Venous blood of 150 patients were collected and stored at the central biobank of the medical faculty of RWTH Aachen University (ethics approval number EK 206/09). Additionally, healthy blood transfusion donors at the University Hospital of RWTH Aachen provided 160 venous blood samples (cryopreserved until used) and 31 capillary blood samples (analyzed fresh) (ethics approval number EK041/15). For a subset of samples, dried blood spots were collected on 2 different microsamplers: Mitra[®] microsampling devices (Neoteryx) and Whatman[®] protein saver cards (Cytiva). We also purchased 36 cryopreserved blood samples from the Referenzinstitut für Bioanalytik

(Bonn, Germany). A detailed description of sample collection and sample processing is provided in the Supplemental Methods and in Supplemental Table 2.

DIGITAL DROPLET PCR

Primers were designed using the bisulfite primer seeker tool (Zymo Research) and fluorescent probes were manually designed with melting temperatures several degrees higher than those of the primers. (Supplemental Table 3). Droplets were generated on a QX200 droplet generator (Bio-Rad) for 20 μ L of reaction mix, containing 2 \times Supermix for probes (no dUTPs, Bio-Rad), 1 μ M primers, and 0.25 μ M TaqMan probes (both Metabion) and 10–30 ng of bisulfite-converted DNA. Subsequently, DNA was amplified on a C1000 Touch Thermal Cycler (Bio-Rad) using the following program: 1 \times 95°C for 10 min, 40 \times 94°C for 30 s, 54°C for 30 s, and 1 \times 98°C for 10 min. Droplets were quantified with a QX200 Droplet Digital reader and analyzed using the QuantaSoft analysis Pro software (both Bio-Rad). Additional information on DNA isolation, bisulfite treatment, and primer design is provided in the Supplemental Methods.

EPIGENETIC LEUKOCYTE COUNTS

To estimate the relative leukocyte counts, 50 healthy donors or 50 patient samples were randomly selected to obtain single linear regressions based on the DNAm values vs conventional manual differential cell counts for the respective cell type. The remaining samples were used as independent validation sets. Absolute leukocyte numbers were estimated for a subset of samples in relation to the reference plasmid (17). Based on the detected plasmid copies and the genomic copies, the absolute number of cells was calculated as described in the Supplemental Methods.

DATA ANALYSIS AND STATISTICS

Scatterplots and the principal component analysis plot were created in R. The heatmap was generated with heatmapper (21). Pearson correlation coefficient r and the mean absolute error (MAE) of DNAm were analyzed and plotted in Windows Office 2016 Excel (Microsoft) and GraphPad Prism v.9 (Graph Pad Software Inc.).

Results

IDENTIFICATION OF CELL-TYPE SPECIFIC METHYLATION SITES

For selection of the best candidate CpGs for cell-type specific DNAm, a dataset was compiled with 1303 DNAm profiles of 40 different studies (Fig. 1A; Supplemental Table 1). Using CimpleG, we selected the top 3 ranking CpGs each for granulocytes, monocytes, lymphocytes, CD4 T cells, CD8 T cells, B cells, NK cells, hematopoietic stem and progenitor cells (HSPCs), and nucleated red blood cells (Fig. 1B). The

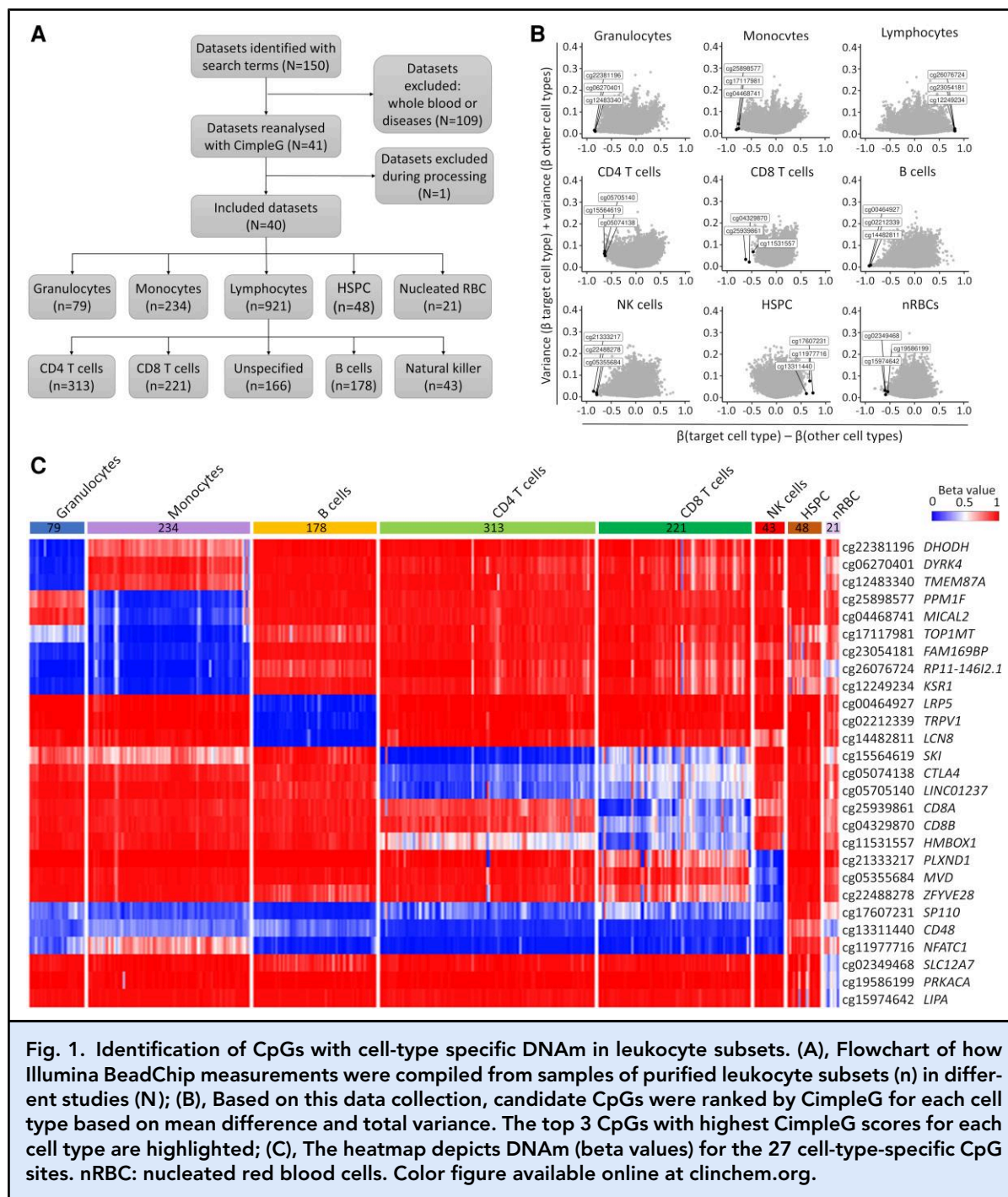


Fig. 1. Identification of CpGs with cell-type specific DNAm in leukocyte subsets. (A), Flowchart of how Illumina BeadChip measurements were compiled from samples of purified leukocyte subsets (n) in different studies (N); **(B),** Based on this data collection, candidate CpGs were ranked by CimpleG for each cell type based on mean difference and total variance. The top 3 CpGs with highest CimpleG scores for each cell type are highlighted; **(C),** The heatmap depicts DNAm (beta values) for the 27 cell-type-specific CpG sites. nRBC: nucleated red blood cells. Color figure available online at clinchem.org.

selected CpGs revealed cell-type specific hypomethylation, except for the pan-lymphocytes and HSPCs, which were hypermethylated. Heatmap analysis also validated the consistency of cell-type specific DNAm differences (Fig 1C). Furthermore, principal component analysis with these CpG sites showed distinct clusters for each cell type (Supplemental Fig. 1). Thus, the selected candidate

CpGs appeared suitable to clearly discern the respective cell types across many different DNAm datasets.

TARGETED DNA METHYLATION ANALYSIS WITH DIGITAL DROPLET PCR

Next, we designed ddPCR assays for targeted DNAm analysis at the selected CpGs. The CpGs for HSPCs

Table 1. Selected candidate CpGs for leukocyte subsets.

Cell-type specificity	CpG site	Gene
Granulocytes	cg22381196	Dihydroorotate dehydrogenase (<i>DHODH</i>)
Lymphocytes	cg23054181	Family with sequence similarity 169 Member B (<i>FAM169BP</i>)
Monocytes	cg04468741	Microtubule associated monooxygenase, calponin and LIM domain containing 2 (<i>MICAL2</i>)
CD4 T cells	cg05074138	Cytotoxic T-lymphocyte associated protein 4 (<i>CTLA4</i>)
CD8 T cells	cg04329870	CD8B Molecule (<i>CD8B</i>)
B cells	cg02212339	Transient receptor potential cation channel subfamily V member 1 (<i>TRPV1</i>)
NK cells	cg05355684	Mevalonate diphosphate decarboxylase (<i>MVD</i>)

and nucleated red blood cells were not considered for this analysis, as these cell types are rare in venous blood. On the other hand, we additionally designed ddPCR assays for cell-type specific CpGs that we previously selected based on DNAm profiles of a single study (15, 17). As for CD8 T cells, the *CD8A*-associated cg25939861 overlapped in both selections, and we analyzed 27 CpGs. An initial screening was performed in 92 venous blood samples from healthy donors to select the best performing CpGs for granulocytes, monocytes, and pan-lymphocytes. For CD4 T cells, CD8 T cells, B cells, and NK cells, lymphocytes were further stratified by flow cytometry from 21 patients (Supplemental Fig. 2). Overall, correlation of DNAm levels with conventional blood counts was very high ($r > 0.8$ or $r < -0.8$). The best performing candidate CpGs (Table 1) were selected based on these correlations and, in case of similar performance, on their ranking in the CimpleG analysis or on how well the positive ddPCR droplets could be discerned (Supplemental Fig. 3).

The selected CpGs were either localized in the 1500 bp region before the transcription start site (*DHODH*), in the 5' UTR region (*FAM169BP* and *MICAL2*), first exon (*CTLA4*) or the gene body (*CD8B*, *TRPV1*, and *MVD*; Supplemental Fig. 4). Analysis of gene expression in the human protein atlas demonstrated that, with the exception of *CTLA4* and *CD8B*, the corresponding genes do not reveal characteristic upregulation in the respective cell types (Supplemental Fig. 5) (22). Therefore, differential DNAm at these epigenetic biomarkers is not necessarily reflected by gene expression level. To assess whether other epigenetic modifications at these sites contribute to cell specificity, we additionally analyzed ENCODE data (see Supplemental Methods). There were no cell-type specific changes in histone marks related to gene-silencing (H3K27me3 and H3K9me3; Supplemental Fig. 6). Interestingly, we observed that cell-type specific methylation at *MICAL2* and *MVD* coincides with a cell-

type specific enrichment for H3K27Ac and H3K4me1. This indicates that multiple epigenetic regulatory mechanisms at these genomic regions can contribute to lineage-specific differentiation.

EPIGENETIC ESTIMATION OF GRANULOCYTES, MONOCYTES, AND LYMPHOCYTES

For epigenetic estimation of the fractions of granulocytes, monocytes, and pan-lymphocytes, we randomly selected 50 samples of healthy donors to train the single linear regression models for each cell type (Fig. 2A, Supplemental Table 4). These models could accurately predict granulocyte ($r = 0.81$; MAE = 3.7) and pan-lymphocyte counts ($r = 0.82$; MAE = 3.4), and moderately predict monocyte counts ($r = 0.40$; MAE = 2.1) in 110 independent donor samples that were measured on the same Cell Dyn Emerald cell counter (Fig. 2B). However, when we tested these predictors on 150 patient samples analyzed on a Sysmex XS800i in another department, there was a systematic underestimation of granulocytes (MAE = 9.5) and a systematic overestimation of lymphocytes (MAE = 8.0) and monocytes (MAE = 3.7) (Fig. 2C). Yet, the very high correlations with conventional counts ($r = 0.95$, $r = 0.96$, and $r = 0.86$, respectively) indicated that this offset can be attributed to the reference measurements on different cell counters. Notably, the CpG for monocytes provided a better correlation with automated cell counts in the patient samples, indicating that these measurements might be more precise. Therefore, the models were retrained on a randomly selected set of 50 patient samples that were measured on a Sysmex XS800i (Supplemental Table 4).

To further benchmark these assays, we participated in an interlaboratory ring trial with >400 other institutes using various conventional automated cell counters. The ddPCR measurements clearly correlated with the overall mean cell count (Fig. 2D). The ring trial

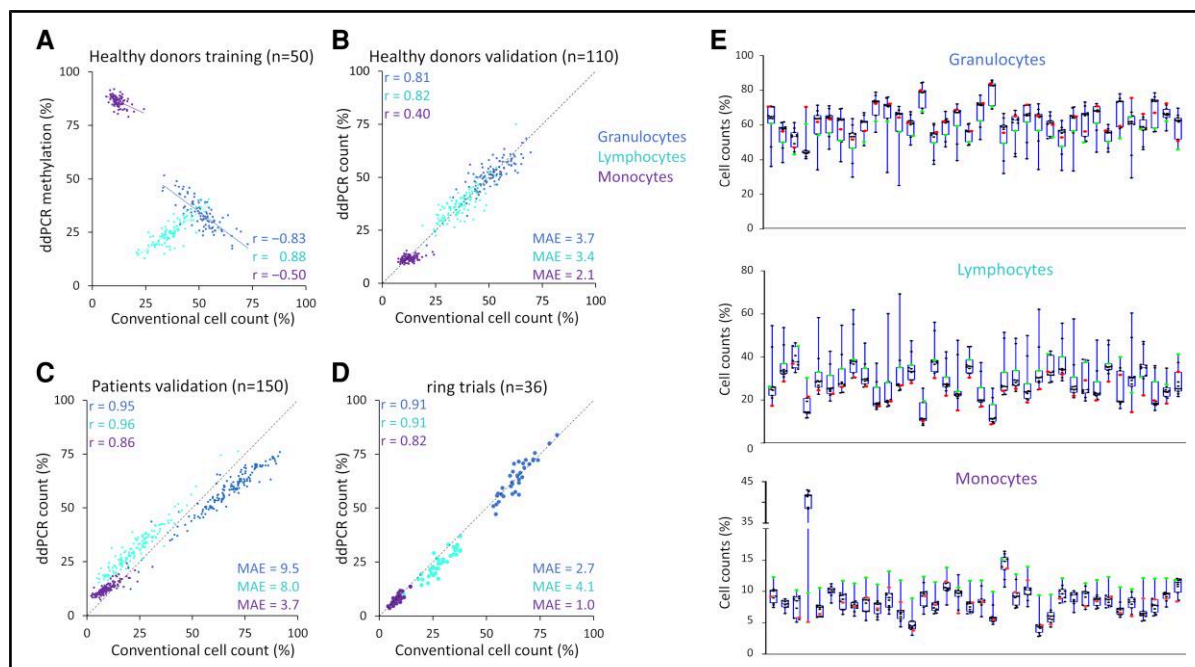


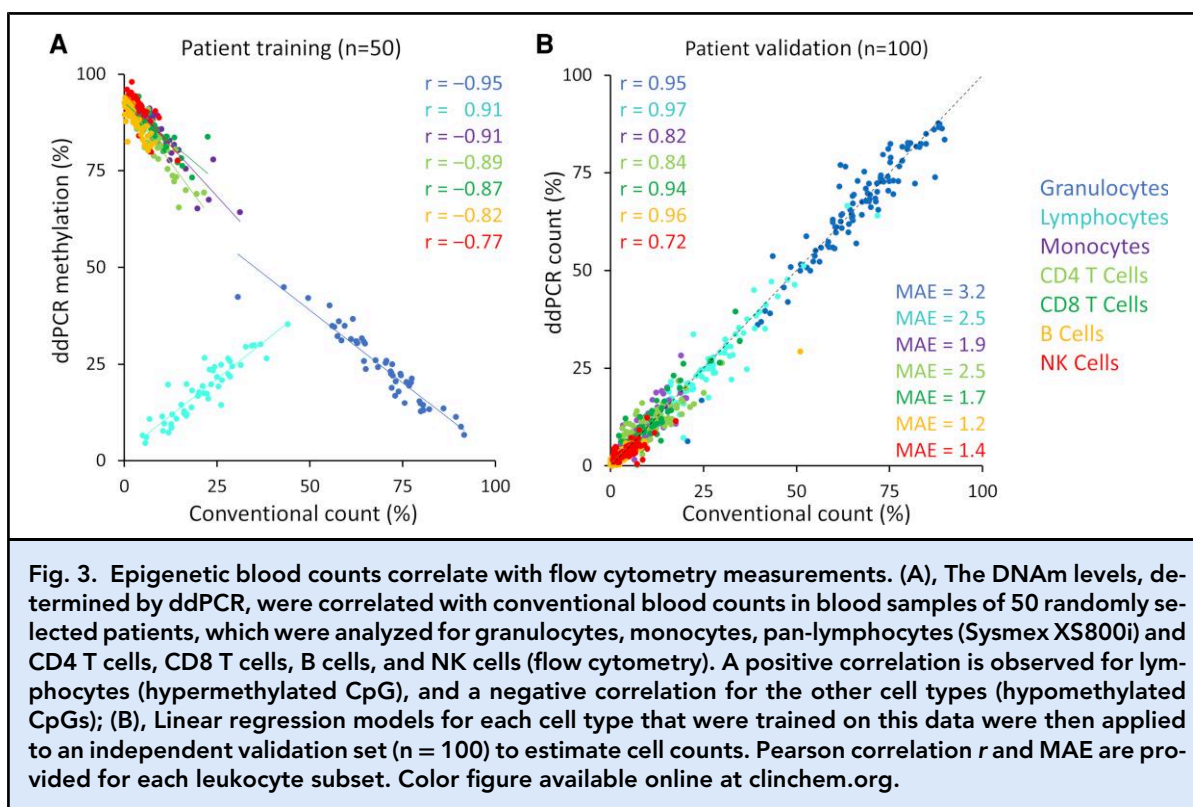
Fig. 2. Epigenetic estimates of granulocytes, lymphocytes, and monocytes. (A), Single linear regression models for monocytes, lymphocytes, and granulocytes were trained for 50 healthy donors based on DNAm measurements (ddPCR) vs conventional cell counts (measured on Cell Dyn Emerald). An inverse correlation is shown for monocytes and granulocytes (hypomethylated CpGs) and a positive correlation for lymphocytes (hypermethylated CpG); (B, C), These models were tested on 110 independent blood samples of healthy donors (B, also measured on Cell Dyn Emerald) and 150 patients (C, measured on Sysmex XS800i). Owing to the systemic offset observed for epigenetic estimations of samples measured on Sysmex XS800i, we retrained single linear regression models for this device on 50 randomly selected samples; (D, E), ddPCR measurements were benchmarked on 36 ring trial samples that were measured by about 400 laboratories, using various conventional cell counters from different manufacturers. The epigenetic predictions trained on Sysmex XS800i was compared against the average cell counts obtained from the ring trial (D). Furthermore, we compared alternative models for epigenetic predictions that were either trained on Cell Dyn Emerald (green) or on Sysmex XS800i (red) for each individual sample (E). Pearson correlation r and MAE are given for each cell type. Color figure available online at clinchem.org.

results revealed a large variation between the different devices (Fig. 2E). This exemplifies the importance of precise reference measurements to fine-tune the linear regression models for ddPCR. While ddPCR measurements are robust, our predictions trained on Cell Dyn measurements (green) overestimated lymphocytes and monocytes, whereas the predictions trained on Sysmex XS800i measurements (red) rather overestimated granulocytes (Fig. 2E).

EPIGENETIC BLOOD COUNTS CORRELATE WITH FLOW CYTOMETRY IN PATIENT SAMPLES

To benchmark the performance of CpGs specific for CD4 T cells, CD8 T cells, B cells, and NK cells, we analyzed blood samples of 150 patients with various hematological

disorders. These blood samples were measured with a Sysmex XS800i for granulocytes, monocytes, and pan-lymphocytes, and the lymphocytes were further stratified by flow cytometry. As indicated above, a subset of 50 patients was randomly selected for single linear regression models for each of these cell types (Fig. 3A; Supplemental Table 4). Using these models on the independent set of 100 patient samples, we could accurately predict the fraction of granulocytes ($r = 0.95$, MAE = 3.2), pan-lymphocytes ($r = 0.97$; MAE = 2.5), monocytes ($r = 0.82$; MAE = 1.9), CD4 T cells ($r = 0.84$; MAE = 2.5), CD8 T cells ($r = 0.94$; MAE = 1.7), B cells ($r = 0.96$; MAE = 1.2), and NK cells ($r = 0.72$; MAE = 1.4) (Fig. 3B). Accuracy of the predictions was not significantly affected by age or gender (Supplemental Fig. 7). Furthermore, no clear age-associated DNAm changes



where observed at any of our CpG sites when we explored this in data from the Generation Scotland cohort ($n = 4450$) (23).

To further assess the consistency of our estimations independent of the conventional cell counters, we determined whether predictions of the individual cell types added up to 100% (Supplemental Fig. 8). The estimated fractions of granulocytes, monocytes, and pan-lymphocytes added up to $99.3 \pm 5.2\%$ in patients, which was similar to the total in healthy donors ($100.1 \pm 2.6\%$), and for both groups the total was not significantly different from the expected 100% ($P = 0.12$ and $P = 0.78$, respectively). The total was also not significantly different from 100% ($99.2 \pm 6.8\%$; $P = 0.18$) when we further stratified pan-lymphocytes into B cells, CD4 T cells, CD8 T cells, and NK cells. Thus, epigenetic estimations are robust and can also provide consistent results for patient samples.

EVALUATION OF LEUKOCYTE FRACTIONS FROM DRIED BLOOD

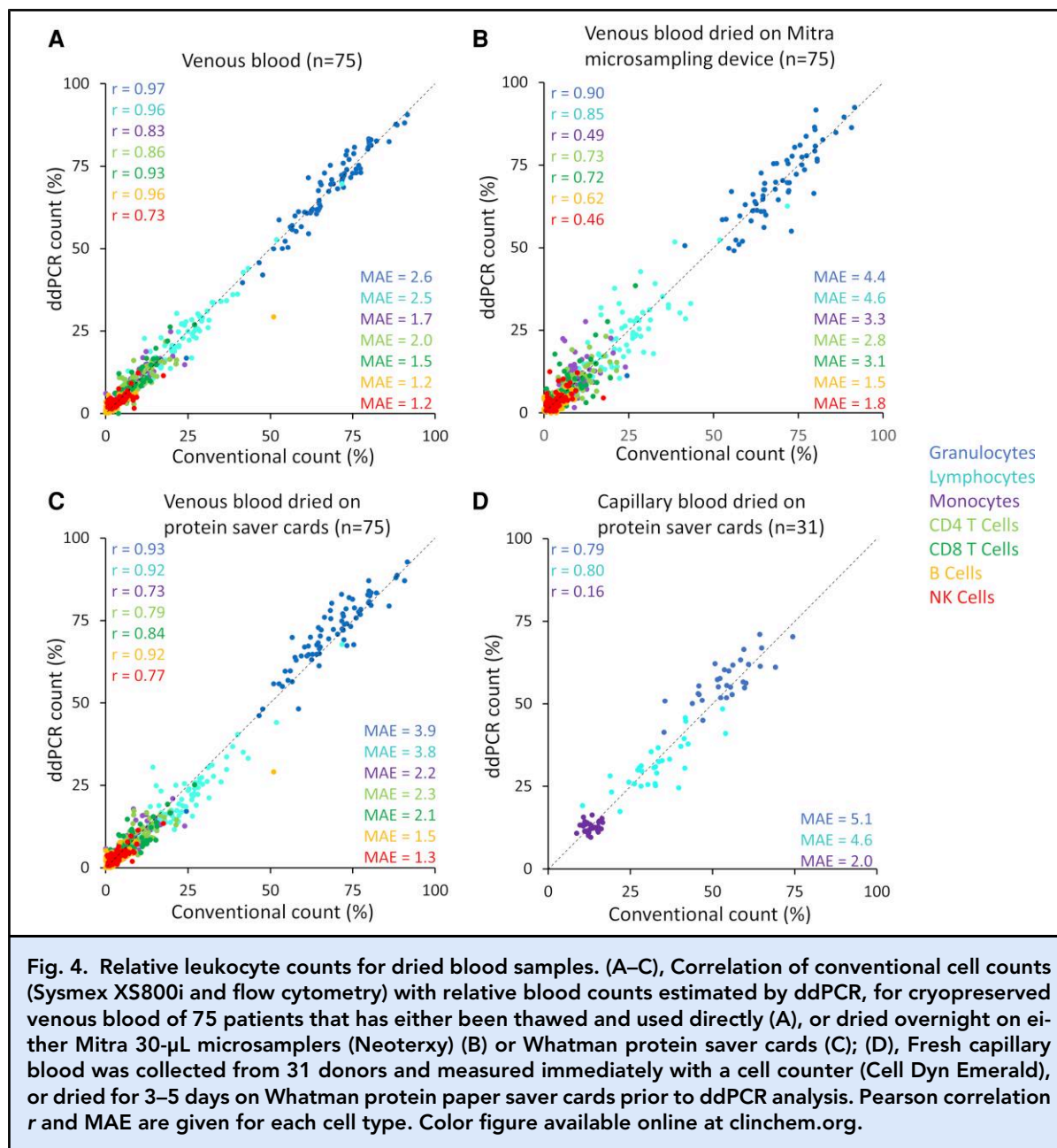
Next, we assessed the feasibility of using epigenetic blood counts for dried blood samples. To this end, 30 μL of venous blood from 75 patients was pipetted and dried overnight on 2 alternative sampling devices: Whatman protein saver cards and Mitra 30- μL micro-sampling devices. Notably, the predictions of individual

cell fractions revealed a similar precision as observed for 150 μL of venous blood, particularly when blood was dried on Whatman protein saver cards that slightly outperformed the Mitra microsampling devices (Fig. 4A–C).

Next, we tested the applicability for capillary blood, which was harvested by finger pricks from 31 healthy donors. A fresh sample was measured with a Cell Dyn Emerald device, and in parallel, blood spots were dried on Whatman protein saver cards for up to 5 days before epigenetic blood counts. Good correlations were observed in particular for granulocytes ($r = 0.79$, MAE = 5.1) and pan-lymphocytes ($r = 0.80$, MAE = 4.6) (Fig. 4D). A lower correlation was observed for monocytes ($r = 0.16$, MAE = 2.0), which might be due to the inconsistency of reference counts on the Cell Dyn Emerald device, as also indicated previously. Overall, the results demonstrate that our method is applicable to dried capillary blood harvested with a finger prick.

ABSOLUTE QUANTIFICATION OF LEUKOCYTE COUNTS

Since every leukocyte has 2 copies of DNA, the absolute cell number is expected to correlate with the DNA concentration. This can be estimated with the number of detected copies in ddPCR of unique genomic regions in nonbisulfite-converted DNA. In fact, for DNA



isolated from 150 μ L of venous blood, the number of positive droplets in ddPCR for genomic region “R5” correlated with leukocyte counts ($r=0.95$; Supplemental Fig. 9). Yet, for dried blood, several outliers were observed, in particular for blood dried on Mitra microsampling devices ($r=0.58$). Direct comparison of the positive ddPCR events in 30 μ L of fresh blood compared to 30 μ L of the same sample dried overnight revealed that for both Whatman protein saver cards ($67.7\% \pm 21.4\%$) and Mitra microsampling devices ($33.5\% \pm 27.6\%$) significantly fewer copies were

detected, and thus the DNA isolation efficiency was lower and more variable for dried blood spots.

To improve absolute quantification from dried blood samples, we aimed to correct for variability in DNA isolation efficiency by spiking the samples with a reference plasmid of known quantity before DNA isolation, as described before (Fig. 5A) (17). Cell numbers can then be estimated based on the ratio of genomic vs plasmid DNA copies. This approach provides a higher correlation with conventional white blood counts of venous blood dried on either Whatman protein saver

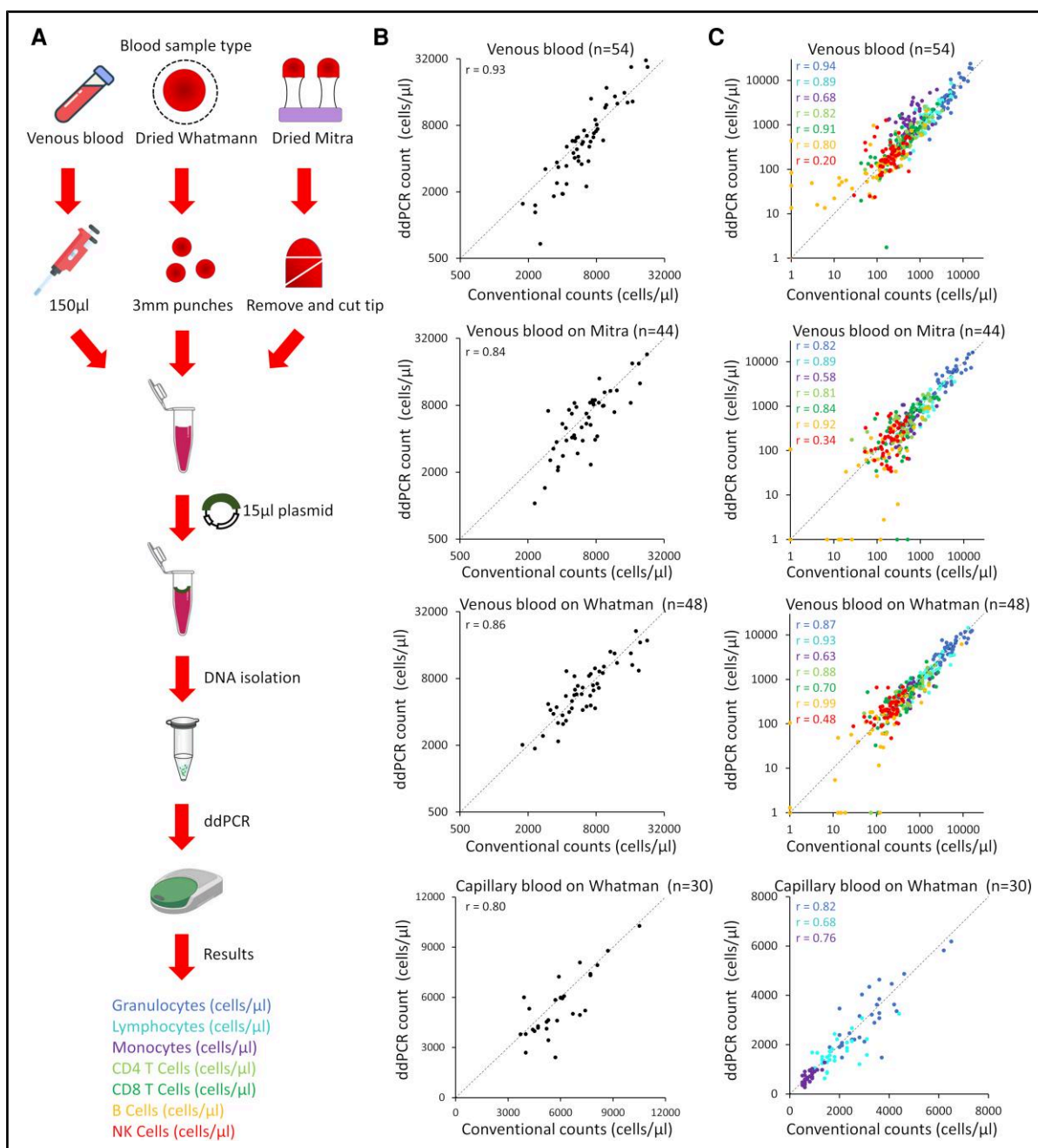


Fig. 5. Total leukocyte counts by ddPCR with a reference plasmid. (A), Scheme of the workflow for absolute quantification of leukocytes. Venous blood, punches of Whatman protein saver cards with dried blood, or tips of the Mitra microsampling devices with dried blood were mixed with a reference plasmid before DNA isolation and ddPCR; (B), Cell numbers were estimated based on the ratio of genomic vs plasmid DNA; (C), Combination of estimates for relative leukocyte counts with absolute cell number. Pearson correlation between epigenetic estimations and conventional measurements are provided. Color figure available online at clinchem.org.

cards ($r = 0.86$) or Mitra microsamplers ($r = 0.84$) (Fig. 5B). Furthermore, this was also applicable to capillary blood that was dried on Whatman paper ($r = 0.80$).

Combining estimates of leukocyte numbers with relative counts based on cell-type specific DNAm, we were able to calculate the absolute numbers for each of the

leukocyte subsets. For each cell type, we observed a clear correlation between conventional absolute cell counts and epigenetic estimations (Fig. 5C).

Discussion

In this study, we further improved epigenetic leukocyte counts through a comprehensive selection of cell-type specific candidate CpGs and optimized ddPCR assays. The results outperformed the precision of our previous models based on different CpGs and measurements with pyrosequencing (17). Furthermore, we demonstrate that dried blood spots, harvested by a finger prick, can be used for epigenetic leukocyte counts.

For selection of candidate CpGs, we compiled a very large dataset of DNAm profiles (1303 cell-type specific DNAm profiles, of 726 donors, derived from 40 different studies). By contrast, our previous selection of cell-type specific CpGs was based on DNAm profiles of only 6 donors (19). In another study, Baron and coworkers initially focused on functionally relevant genes to select candidate CpGs, specifically for CD4 T cells (*CD4* gene) and CD8 T cells (*CD8B* gene) (14). Notably, despite the very different selection approaches, our candidate CpG for CD8 T cells (cg04329870) was also found within this *CD8B* region. Furthermore, they analyzed sorted cells of 2 donors to identify cell-type specific CpGs for B cells, NK cells, and granulocytes (14). Remarkably, our candidate CpG for NK cells (cg05355684) was also selected in their work. More recently, we have tested cell-type specific CpGs for HSPCs that were identified based on DNAm profiles of CD34⁺ cells of 5 donors (24) in comparison to DNAm profiles of other sorted leukocyte subsets of 6 donors (19). Based on this, we identified the top candidate CpGs for HSPCs (25). While we have not validated the HPSCs in our current study, since CD34⁺ cells are extremely rare in normal peripheral blood, there was a striking overlap of 2 of the 3 top CpGs in the different selections: *SPI40* (cg17607231) and *CD48* (cg13311440) were also identified in that independent selection and correlated with CD34⁺ cells in mobilized peripheral blood and blast counts in leukemia (25). The finding that different studies select the same genomic region or even the exact same CpGs, indicates that genomic regions with well-suited candidate CpGs for cell-type specific DNAm may be limited.

For a more systemic selection of the top candidate CpGs we used our new CimpleG pipeline (20). We did not consider DNA of nonleukocyte cell types, because our application focuses specifically on analysis of blood. While it is theoretically possible that skin cells, endothelial cells, or even cell-free DNA impact our predictions (8, 9, 26), most DNA in blood clearly stems

from leukocytes. Notably, some of the selected CpGs did not reveal a clear correlation with cell counts in the ddPCR measurements (*PPM1F* and *TOP1MT* for monocytes; *LCN8* for B cells; and *PLXND1* for NK cells). Thus, either the probe sets of the Illumina BeadChip array or the corresponding ddPCR assays did not provide specific or accurate DNAm measurements. Such discrepancy between arrays and targeted assays has been reported previously (27, 28) and substantiates the need to validate such biomarkers.

Multiple methods are available for targeted DNAm measurements (18). In our previous work, epigenetic blood counts were primarily based on pyrosequencing (17, 19), whereas Baron et al. established their assays based on methylation-specific qPCR (14). It has been shown that the DNAm levels can vary considerably in measurements with pyrosequencing vs qPCR (29). We anticipate that the improved precision of epigenetic blood counts in our current study can partly be attributed to the use of ddPCR. In contrast to qPCR, there is less PCR bias for methylated or nonmethylated sequences in ddPCR, since individual droplets are only classified as positive or negative (30). In fact, it has been shown for epigenetic estimations of CD3 T lymphocytes that ddPCR provided higher precision and greater accuracy, especially in samples with low copy numbers of the target genes (31, 32).

A big advantage of epigenetic blood counts is that the blood samples can be taken with a simple finger prick and shipped as dried blood spots. This facilitates self-testing without the need for trained medical personnel—e.g., for elderly patients who have difficulties in visiting clinicians. Furthermore, the method is applicable for newborn screening for severe combined immunodeficiency (14, 33). Other clinical, relevant examples are enumeration of CD4 T cells during HIV infection (14), or monitoring neutropenia after chemotherapy. Here, we have tested 2 alternative sampling devices for dried blood. Filter papers, such as Whatman protein saver cards, are commonly used and more cost effective, whereas microsampling devices, such as Mitra microsamplers, may sample a more defined blood volume. Surprisingly, in our experiments, epigenetic cell counts were more precise with Whatman protein saver cards, which might be due to less efficient DNA isolation from Mitra tips.

While our estimations from dried blood are promising, there are still limitations and challenges that need to be addressed before the method can be used in clinical application: (a) One limitation is that erythrocytes and thrombocytes cannot be addressed since they do not comprise genomic DNA. (b) A particular challenge for epigenetic blood counts is the detection of rare cell types (<5%) or closely related cell types, as their differential DNA methylation will only have a small impact on

the mean methylation values in bulk DNA. (c) The method is relatively time consuming. With our current protocol, turnaround time is around 2 days, due to an overnight (12–16 h) bisulfite conversion. However, with alternative bisulfite conversion procedures directly applicable on blood, the entire workflow can be reduced to about 8 hours. (d) For absolute quantification, in particular for dried blood, DNA extraction efficiency has to be improved. Here, we addressed the variability in DNA isolation efficiency with a reference plasmid, but its stability at defined concentration needs to be further validated in independent cohorts. (e) While we did not observe that aging had a significant impact on the epigenetic blood counts, there may be offsets between samples of children and adults. In addition, leukemia is diverse and although the estimations in our patients were barely affected by disease, we cannot exclude that disease-associated changes in the epigenetic makeup may influence the results (16). Furthermore, chromosomal abnormalities need to be taken into consideration in the future. For example, deletion of the short arm of chromosome 17 occurs in many hematopoietic malignancies (34) and a loss of one copy of *TRPV1* at this region may influence estimation of B cells. Thus, the method will need to be specifically validated for specific clinical applications. (f) Last but not least, there are regulatory hurdles to overcome. The biomarkers need to be validated and accredited according to local regulatory demands; for example, in Europe according to the new directive for in vitro diagnostic devices (35). In general, it should be feasible to fulfill these demanding requirements as similar procedures have been clinically accredited, the instrumentation for DNA isolation and ddPCR has already been CE certified, and the required analysis software, which in our case is based on relatively simple linear regression models, could also be certified accordingly.

Despite these limitations and challenges, our study showed that epigenetic leukocyte counts by ddPCR is an accurate alternative for cryopreserved blood or dried blood obtained from a finger prick. Future studies will be required to validate and approve the procedure for specific clinical applications.

Supplemental Material

Supplemental material is available at *Clinical Chemistry* online.

Nonstandard Abbreviations: DNAm, DNA methylation; CpGs, CG dinucleotides; qPCR, quantitative PCR; ddPCR, digital droplet PCR; MAE, mean absolute error; HSPCs, hematopoietic stem and progenitor cells.

Human Genes: *DHODH*, dihydroorotate dehydrogenase; *TMEM87A*, transmembrane protein 87A; *DYRK4*, dual specificity tyrosine phosphorylation regulated kinase 4; *WDR20*, WD repeat domain 20; *FAM169BP*, pseudogene of family with sequence similarity 169 member B, pseudogene; *FYN*, FYN proto-oncogene, Src family tyrosine kinase; *KSRI*, kinase suppressor of ras 1; *MICAL2*, microtubule associated monoxygenase calponin and LIM domain containing 2; *CENPA*, centromere protein A; *PPM1F*, protein phosphatase Mg²⁺/Mn²⁺ dependent 1F; *TOP1MT*, DNA topoisomerase I mitochondrial; *CTLA4*, cytotoxic T-lymphocyte associated protein 4; *LINC02137*, long intergenic non-protein coding RNA 2137; *SKI*, SKI proto-oncogene; *CD4*, CD4 molecule; *CD8B*, CD8 subunit beta; *CD8A*, CD8 subunit alpha; *HMBOX1*, homeobox telomere-binding protein 1; *TRPV1*, transient receptor potential cation channel subfamily V member 1; *LRP5*, LDL receptor related protein 5; *WIPI2*, WD repeat domain, phosphoinositide interacting 2; *LCN8*, lipocalin 8; *MVD*, mevalonate diphosphate decarboxylase; *ZFYVE28*, zinc finger FYVE-type containing 28; *SLC15A4*, solute carrier family 15 member 4; *PLXND1*, plexin D1.

Author Contributions: *The corresponding author takes full responsibility that all authors on this publication have met the following required criteria of eligibility for authorship: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved. Nobody who qualifies for authorship has been omitted from the list.*

Wouter Hubens (conceptualization—equal, data curation—lead, formal analysis—lead, methodology—lead, writing—original draft—lead, writing—review and editing—lead), Tiago Maie (data curation—supporting, formal analysis—equal, methodology—supporting, software—lead, visualization—supporting, writing—review and editing—supporting), Matthis Schnitker (conceptualization—supporting, data curation—equal, formal analysis—equal, investigation—equal, methodology—supporting, writing—review and editing—supporting), Ledio Bocova (conceptualization—supporting, data curation—equal, investigation—supporting, methodology—supporting, writing—review and editing—supporting), Deepika Puri (data curation—supporting, formal analysis—supporting, writing—review and editing—supporting), Martina Wessiepe (project administration—supporting, resources—lead, writing—review and editing—supporting), Jan Kramer (project administration—supporting, resources—lead, writing—review and editing—supporting), Lothar Rink (project administration—supporting, resources—equal, writing—review and editing—supporting), Steffen Koschmieder (project administration—supporting, resources—lead, visualization—supporting, writing—review and editing—supporting), Ivan Costa (formal analysis—supporting, funding acquisition—supporting, methodology—supporting, software—lead, supervision—supporting, writing—review and editing—supporting), and Wolfgang Wagner (Conceptualization—lead, data curation—supporting, formal analysis—supporting, funding acquisition—lead, investigation—supporting, project administration—lead, resources—lead, supervision—lead, visualization—lead, writing—original draft—lead, writing—review & editing—lead).

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form.*

Disclosures: W. Wagner is cofounder of Cygenia GmbH that can provide service for various epigenetic signatures and has received travel support from AACCC. W. Wagner, W.H.G. Hubens, and L. Bocova are nameholders on patent applications by RWTH Aachen University Medical School for patents for epigenetic blood counts and absolute quantification

with reference plasmids (EP17163798.6, Method for determining blood counts based on DNA methylation; DE102017004108A, Verfahren zur Bestimmung der Zellzahl von eukaryotischen Zellen; DE10201912 4828A1, Verfahren zur Bestimmung der Zellzahl mittels einer Referenz-DNA). J. Kramer is CEO of LADR Laboratory Group Dr. Kramer & Colleagues, Geesthacht, Germany. S. Koschmieder, received research funding, advisory board honoraria, honoraria, and other financial support (e.g., travel support) unrelated to the work described in this manuscript from Janssen and Geron; and holds patent DE102019129527A1, Inhibitoren für Bromodmänen.

Research Funding: This research was supported by the ForTra GmbH für Forschungstransfer der Else Kröner-Fresenius-Stiftung (EKFS:

2020_EKTP12), by the Federal Ministry of Education and Research (BMBF: VIP + Epi-Blood-Count; Fibromap), and by German Research Foundation (WA 1706/12-1; WA 1706/14-1; GE 2811/4-1).

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, preparation of manuscript, or final approval of manuscript.

Acknowledgments: The authors would like to thank all patients and clinicians for providing the blood samples for this study. Furthermore, we acknowledge the central biobank of the medical faculty of RWTH Aachen (RWTH cBMB).

References

- Pitoiset F, Cassard L, El Soufi K, Boselli L, Grivel J, Roux A, et al. Deep phenotyping of immune cell populations by optimized and standardized flow cytometry analyses. *Cytometry A* 2018;93:793–802.
- Bruegel M, Nagel D, Funk M, Fuhrmann P, Zander J, Teupser D. Comparison of five automated hematology analyzers in a university hospital setting: Abbott cell-dyn sapphire, beckman coulter DxH 800, siemens advia 2120i, sysmex XE-5000, and sysmex XN-2000. *Clin Chem Lab Med* 2015;53:1057–71.
- Navas A, Giraldo-Parra L, Prieto MD, Cabrera J, Gómez MA. Phenotypic and functional stability of leukocytes from human peripheral blood samples: considerations for the design of immunological studies. *BMC Immunol* 2019;20:5.
- McGann LE, Yang HY, Walterson M. Manifestations of cell damage after freezing and thawing. *Cryobiology* 1988;25:178–85.
- Herzenberg LA, Tung J, Moore WA, Herzenberg LA, Parks DR. Interpreting flow cytometry data: a guide for the perplexed. *Nat Immunol* 2006;7:681–5.
- Jimenez Vera E, Chew YV, Nicholson L, Burns H, Anderson P, Chen HT, et al. Standardisation of flow cytometry for whole blood immunophenotyping of islet transplant and transplant clinical trial recipients. *PLoS One* 2019;14:e0217163.
- Mattei AL, Bailly N, Meissner A. DNA Methylation: a historical perspective. *Trends Genet* 2022;38:676–707.
- Moss J, Magenheim J, Neiman D, Zemmour H, Loyfer N, Korach A, et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat Commun* 2018;9:5068.
- Schmidt M, Maié T, Dahl E, Costa IG, Wagner W. Deconvolution of cellular subsets in human tissue based on targeted DNA methylation analysis at individual CpG sites. *BMC Biol* 2020;18:178.
- Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012;13:86.
- Accomando WP, Wiencke JK, Houseman EA, Nelson HH, Kelsey KT. Quantitative reconstruction of leukocyte subsets using DNA methylation. *Genome Biol* 2014;15:R50.
- Salas LA, Koestler DC, Butler RA, Hansen HM, Wiencke JK, Kelsey KT, Christensen BC. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the illumina HumanMethylationEPIC BeadArray. *Genome Biol* 2018;19:64.
- Wagner W. How to translate DNA methylation biomarkers into clinical practice. *Front Cell Dev Biol* 2022;10:854797.
- Baron U, Werner J, Schildknecht K, Schulze JJ, Mulu A, Liebert UG, et al. Epigenetic immune cell counting in human blood samples for immunodiagnosics. *Sci Transl Med* 2018;10:eaan3508.
- Frobel J, Božić T, Lenz M, Uciechowski P, Han Y, Herwartz R, et al. Leukocyte counts based on DNA methylation at individual cytosines. *Clin Chem* 2018;64:566–75.
- Božić T, Kuo CC, Hapala J, Franzen J, Eipel M, Platzbecker U, et al. Investigation of measurable residual disease in acute myeloid leukemia by DNA methylation patterns. *Leukemia* 2022;36:80–9.
- Sontag S, Bocova L, Hubens WHG, Nüchtern S, Schnitker M, Look T, et al. Toward clinical application of leukocyte counts based on targeted DNA methylation analysis. *Clin Chem* 2022;68:646–56.
- Han Y, Franzen J, Stiehl T, Gobs M, Kuo CC, Nikolić M, et al. New targeted approaches for epigenetic age predictions. *BMC Biol* 2020;18:71.
- Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlén SE, Greco D, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One* 2012;7:e41361.
- Maié T, Schmidt M, Erz M, Wagner W, G Costa I. Cimplog: finding simple CpG methylation signatures. *Genome Biol* 2023;24:161.
- Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A, Wishart DS. Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res* 2016;44:W147–53.
- Karlsson M, Zhang C, Méar L, Zhong W, Digre A, Katona B, et al. A single-cell type transcriptomics map of human tissues. *Sci Adv* 2021;7:eabh2169.
- Bernabeu E, McCartney DL, Gadd DA, Hillary RF, Lu AT, Murphy L, et al. Refining epigenetic prediction of chronological and biological age. *Genome Med* 2023;15:12.
- Aranyi T, Stockholm D, Yao R, Poinignon C, Wiart T, Corre G, et al. Systemic epigenetic response to recombinant lentiviral vectors independent of proviral integration. *Epigenetics Chromatin* 2016;9:29.
- Bocova L, Hubens W, Engel C, Koschmieder S, Jost E, Wagner W. Quantification of hematopoietic stem and progenitor cells by targeted DNA methylation analysis. *Clin Epigenetics* 2023;15:105.
- Neuberger EW, Sontag S, Brahmer A, Philipp KFA, Radsak MP, Wagner W, Simon P. Physical activity specifically evokes release of cell-free DNA from granulocytes thereby affecting liquid biopsy. *Clin Epigenetics* 2022;14:29.
- Cheung K, Burgers MJ, Young DA, Cockell S, Reynard LN. Correlation of Infinium HumanMethylation450K and MethylationEPIC BeadChip arrays in cartilage. *Epigenetics* 2020;15:594–603.
- Roessler J, Ammerpohl O, Gutwein J, Hasemeier B, Anwar SL, Kreipe H, Lehmann U. Quantitative cross-validation and content analysis of the 450K DNA methylation array from illumina, inc. *BMC Res Notes* 2012;5:210.
- De Chiara L, Leiro-Fernandez V, Rodríguez-Girondo M, Valverde D, Botana-Rial MI, Fernandez-Villar A. Comparison of bisulfite pyrosequencing and methylation-specific qPCR for methylation assessment. *Int J Mol Sci* 2020;21:9242.
- Warnecke PM, Stirzaker C, Melki JR, Millar DS, Paul CL, Clark SJ. Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res* 1997;25:4422–6.
- Wiencke JK, Bracci PM, Hsuang G, Zheng S, Hansen H, Wrensch MR, et al. A comparison of DNA methylation specific droplet digital PCR (ddPCR) and real time

qPCR with flow cytometry in characterizing human T cells in peripheral blood. *Epigenetics* 2014;9:1360–5.

- 32.** Malic L, Daoud J, Geissler M, Boutin A, Lukic L, Janta M, et al. Epigenetic subtyping of white blood cells using a thermoplastic elastomer-based microfluidic emulsification device for multiplexed, methylation-specific digital droplet PCR. *Analyst* 2019;144:6541–53.
- 33.** Blom M, Pico-Knijnenburg I, Imholz S, Vissers L, Schulze J, Werner J, et al. Second tier testing to reduce the number of non-actionable secondary findings and false-positive referrals in newborn screening for severe combined immunodeficiency. *J Clin Immunol* 2021; 41:1762–73.
- 34.** Chen M, Yang Y, Liu Y, Chen C. The role of chromosome deletions in human cancers. In: Zhang Y, editor. *Chromosome translocation*. Singapore: Springer Singapore; 2018. p. 135–48.
- 35.** Lubbers BR, Schilhabel A, Cobbaert CM, Gonzalez D, Dombink I, Bruggemann M, et al. The new EU regulation on in vitro diagnostic medical devices: implications and preparatory actions for diagnostic laboratories. *Hemasphere* 2021;5: e568.