Megakaryocyte and erythroblast DNA in plasma and platelets

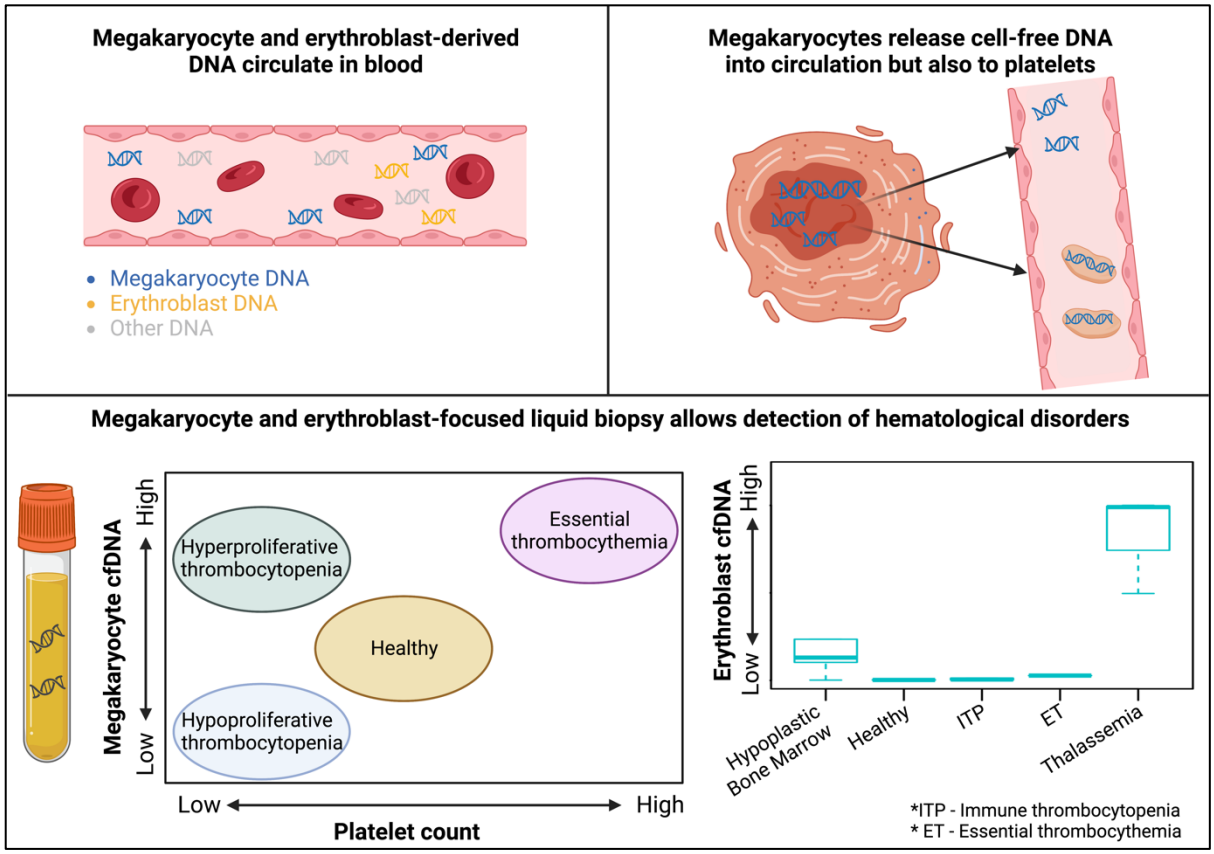
Joshua Moss†[1,2], Roni Ben-Ami†[1], Ela Shai[3], Yosef Kalish[3], Agnes Klochender[1], Gordon Cann[4], Benjamin Glaser[5], Ariela Arad*[3], Ruth Shemer*[1], Yuval Dor*[1]

Affiliations:
[1]Department of Developmental Biology and Cancer Research, Institute for Medical Research Israel-Canada, the Hebrew University-Hadassah Medical School; Jerusalem, Israel
[2]Sharett Institute of Oncology, Hadassah-Hebrew University Medical Center; Jerusalem, Israel
[3]Hematology Department, Hadassah-Hebrew University Medical Center; Jerusalem, Israel
[4]GRAIL, LLC.; Menlo Park, California, United States of America
[5]Endocrinology and Metabolism Service, Hadassah University Medical Center and Faculty of Medicine, the Hebrew University; Jerusalem, Israel
† These authors contributed equally
* Corresponding authors: Yuval Dor, yuvald@ekmd.huji.ac.il; Ruth Shemer, shemer.ru@mail.huji.ac.il; Ariela Arad, arielaar@hadassah.org.il

ABSTRACT

Circulating cell-free DNA (cfDNA) fragments are a biological analyte with extensive utility in diagnostic medicine. Understanding the source of cfDNA and mechanisms of release is crucial for designing and interpreting cfDNA-based liquid biopsy assays. Using cell type-specific methylation markers as well as genome-wide methylation analysis, we determined that megakaryocytes, the precursors of anuclear platelets, are major contributors to cfDNA (~26%), while erythroblasts contribute 1-4% of cfDNA in healthy individuals. Surprisingly, we discovered that platelets contain genomic DNA fragments originating in megakaryocytes, contrary to the general understanding that platelets lack genomic DNA. Megakaryocyte-derived cfDNA is increased in pathologies involving increased platelet production (Essential Thrombocythemia, Idiopathic Thrombocytopenic Purpura) and decreased upon reduced platelet production due to chemotherapy-induced bone marrow suppression. Similarly, erythroblast cfDNA is reflective of erythrocyte production and is elevated in patients with Thalassemia. Megakaryocyte- and erythroblast-specific DNA methylation patterns can thus serve as novel biomarkers for pathologies involving increased or decreased thrombopoiesis and erythropoiesis, which can aid in determining the etiology of aberrant levels of erythrocytes and platelets.

GRAPHICAL ABSTRACT



INTRODUCTION

Circulating cell-free DNA (cfDNA) molecules are thought to be released from dying cells and can be used to monitor tissue turnover rates in health and disease. Fetal DNA present in maternal circulation allows for detection of fetal aneuploidies[1]; donor-derived DNA in the circulation of organ transplant recipients provides a non-invasive marker of graft rejection[2]; and tumor-derived mutant cfDNA allows for detection and monitoring of cancer[3].

Accurate identification of the tissue origins of cfDNA holds great potential for sensitive monitoring of turnover dynamics in specific tissues and cell types, in health and disease. DNA methylation patterns provide means for determining the tissue origins of cfDNA, given the extreme cell type specificity of this epigenetic mark[4]. Indeed, we and others have shown that tissue-specific methylation patterns can serve as cfDNA biomarkers for elevated turnover in specific tissues. For example, cardiomyocyte DNA is present in the plasma of patients following myocardial infarction[5]; hepatocyte cfDNA is present in the plasma of patients with liver damage[6], and exocrine pancreas DNA is found in the plasma of patients with pancreatitis[7]. Additionally, novel assays detect cancer-related methylation aberrations for early diagnosis of cancer[8]. Alternative approaches for determining the tissue origins of cfDNA rely on tissue-specific patterns of nucleosome positioning[9], fragmentation, topology and size of cfDNA[10]. More recently, patterns of histone modification in circulating chromatin fragments were used to infer gene expression and hence identity of the cells that gave rise to cfDNA[11]. Such analyses of the circulating epigenome demonstrated that plasma cfDNA is mostly derived from cells of hematopoietic origin, specifically neutrophils (30%), lymphocytes (12%) and monocytes (11%), as well as vascular endothelial cells (~10%) and hepatocytes (1-3%)[12]. However, there have been contrasting reports regarding the contribution of cfDNA from erythroblasts and megakaryocytes, cells which arise from two related but distinct blood lineages. Lam et al. have identified genomic loci that are unmethylated specifically in erythroid cells and assessed their level in cfDNA. They concluded that ~30% of cfDNA originates from erythroid cells, and that anemia and thalassemia impact the level of erythroid cfDNA[13]. Consistent with this, our own deconvolution of the plasma methylome suggested that erythrocyte progenitors contribute to ~30% of cfDNA in healthy individuals[12]. However, these studies did not take into account the relative contribution of megakaryocytes to cfDNA. Sadeh et al. concluded from immunoprecipitation of circulating chromatin that megakaryocytes, rather than erythroblasts, are major contributors of nucleosomes to healthy plasma[11].

In this study, we identified definitive methylation markers that distinguish megakaryocytes and erythroblasts and used these to characterize the presence of DNA from megakaryocytes and erythroblasts in plasma of healthy individuals and in different pathological scenarios which affect the production of platelets and red blood cells. We found that measuring megakaryocyte and erythroblast cfDNA allows the detection and distinction of pathologies affecting these lineages, even when not reflected in peripheral blood cell counts. Furthermore, after establishing the presence of a significant amount of megakaryocyte-derived DNA in circulation, we went on to question whether megakaryocyte-derived DNA reaches plasma via platelets. This led to an discovery that platelets, not previously believed to contain genomic DNA, do in fact contain genomic DNA, despite the lack of a nucleus.

RESULTS

Methylomes of erythroblasts and megakaryocytes and their representation in healthy plasma

Due to previous reports that non-leukocyte cells of hematopoietic origin are major contributors to cfDNA, we sought to evaluate the relative contribution of erythroid and megakaryocyte genomes. To this end, we isolated DNA from bone-marrow erythroblasts (**Supplementary Figure S1**) as well as multiple types of white blood cells (see methods) and performed whole-genome bisulfite sequencing to obtain a genome-wide methylation profile of each cell type. We also obtained previously published genome-wide methylation profiles of megakaryocytes as well as common megakaryocyte-erythroid progenitors for comparison. This comparison revealed multiple loci that were uniquely methylated or unmethylated in erythroblasts (1884 sites) and in megakaryocytes (97 sites) and could in principle serve as specific biomarkers for DNA derived from these cell types (**Figure 1A-B**). The erythroblast genome was largely unmethylated, in line with previous reports describing gradual genome-wide demethylation of erythroid cells in mice and humans during their terminal differentiation[14,15] (**Supplementary Figure S2**).

In order to evaluate the contribution of these genomes to cfDNA, we obtained genome-wide methylation profiles of purified white blood cells (WBC) and cfDNA from 23 healthy individuals, sequenced at ~85 coverage[16]. Methylation levels of genomic regions uniquely unmethylated in megakaryocytes were ~26% less methylated in cfDNA compared to WBC-derived DNA (**Figure 1A**), suggesting that megakaryocyte genomes are major contributors to cfDNA. In contrast, genomic regions uniquely unmethylated in erythroblasts were only ~4% less methylated in cfDNA compared to WBC-derived DNA (**Figure 1B**), suggesting that erythroblast genomes are minor contributors to cfDNA in healthy individuals. The fact that erythroblast and megakaryocyte markers are methylated in WBC further supports the idea that they originate in cells that are not present in circulation. Targeted bisulfite PCR-sequencing of regions uniquely unmethylated in megakaryocytes or erythroblasts in cfDNA or blood was consistent with the genome-wide analyses, demonstrating elevated levels of megakaryocyte as well as erythroblast DNA in plasma as compared to whole blood (**Figure 1C-D**, **Supplementary Figure S3-S4**).

Platelets contain genomic DNA from MK

We sought to evaluate how megakaryocyte-derived DNA reaches the plasma. Platelets are products of megakaryocytes, that contain megakaryocyte-derived pre-mRNA and mRNA[17] and carry out splicing and protein synthesis, but do not contain a nucleus and are not thought to contain genomic DNA. However it has been reported that platelets contain histone proteins[18], raising the possibility that they may carry some megakaryocyte DNA as well. To test this idea, we isolated platelets (see methods) and performed DNA extraction (**Supplementary Table S2**). We reasoned that DNA found in platelets could be genomic DNA of a megakaryocyte that was trapped within forming platelets, or alternatively could be cfDNA from other cell types that adhered to the external surface of platelets or was internalized, similar to tumor-derived mRNA which has been suggested to be present in platelets of individuals with cancer[19,20]. To distinguish between these possibilities, we performed targeted bisulfite sequencing of regions uniquely unmethylated in megakaryocytes in the platelet DNA concentrates. Strikingly, platelet DNA was found to be mostly unmethylated at these loci (~80%), suggesting megakaryocytes as the main origin of platelet DNA (**Figure 2A**). Platelet concentrates also contained DNA methylation markers of leukocytes and hepatocytes, suggesting the presence of some DNA from these cell types (**Figure 2A**). To distinguish between DNA associated with the external surface of platelets and DNA present within platelets we treated platelet isolates with DNaseI, reasoning that this enzyme would have an effect only on external DNA. Strikingly, DNase treatment did not reduce the concentration of megakaryocyte DNA in platelets but did eliminate leukocyte- and hepatocyte-derived DNA (**Figure 2B and**

Supplementary Figure S5), suggesting that the megakaryocyte DNA is present within the platelet. Interestingly, platelet DNA is present in the form of large molecular weight DNA, in contrast to nucleosome-size fragments of typical cfDNA (**Supplementary Figure S6**). Notably, the total amount of DNA present in platelets (~$2x10^{-6}$ genomes/platelet) suggests that only a small fraction (~0.1%) of a megakaryocyte genome DNA is present in the platelets that derive from this cell (**Supplementary Table S2**).

Genome-wide analysis of platelet DNA

In order to confirm that platelet DNA is megakaryocyte-derived, we performed whole-genome bisulfite sequencing of platelet DNA (n=3 individuals). Platelets contained DNA from all human chromosomes, with an enrichment for mitochondrial DNA (~59% of DNA content). This is consistent with the reported presence of ~5 mitochondria per platelet[21], and suggests that each platelet contains ~56kb of genomic DNA, in accordance with the previous estimate of ~0.1% of the megakaryocyte genome present in platelets on average (given a mean megakaryocyte ploidy of 12N[22] and ~1000 platelets produced per megakaryocyte[23]) (**Figure 3A**). We then focused on regions of the genome uniquely unmethylated in different hematopoietic cell types as well as other cell types which contribute DNA to cfDNA such as endothelial cells and hepatocytes, and evaluated the methylation of these regions in platelet DNA. Importantly, only the megakaryocyte-unmethylated regions were unmethylated in platelets (**Figure 3B-C**). Deconvolution analysis of the DNA methylation profile of platelet DNA further identified megakaryocytes as the major contributors to platelet DNA (**Figure 3D**).

The origins of MK-derived cfDNA

After establishing that megakaryocyte-derived DNA is present in plasma, as well as in platelets, we aimed to determine whether megakaryocyte-derived cfDNA originates directly from megakaryocytes, or alternatively from platelets (either whole platelets that make their way to cfDNA preps, or platelets that release cfDNA). To this end, we obtained plasma samples from females who had received a transfusion of platelets from male donors (n=5). We reasoned that if megakaryocyte cfDNA is derived from platelets, the cfDNA of recipients should contain a large proportion of DNA containing the Y-chromosome. Interestingly, Y-chromosome derived DNA could not be detected in these plasma samples more than 24 hours post-transfusion, even though the half-life of transfused platelets is generally >2 days[24] (**Figure 4 and Supplementary Figure S7**). These findings suggest that megakaryocyte plasma cfDNA is not derived from DNA found in platelets, but rather directly from megakaryocytes.

A survey of megakaryocyte and erythroblast cfDNA in relevant diseases

After establishing that circulating megakaryocyte DNA originates from megakaryocytes and not from platelets, we reasoned that its concentration in plasma may be reflective of megakaryocyte activity and turnover and hypothesized that megakaryocyte cfDNA would increase in cases of hyperproliferation of megakaryocytes and decrease in cases of hypoproliferation. To test this hypothesis, we collected samples from healthy individuals (n=77), from patients with low platelet counts associated with hypoproliferation of megakaryocytes due to chemotherapy treatment (n=5) and from patients with low platelet counts associated with hyperproliferative megakaryocytes due to peripheral destruction of platelets (Immune Thrombocytopenia [ITP]) (n=5). We also collected plasma from patients with high platelet counts due to hyperproliferation of megakaryocytes (Essential Thrombocythemia [ET]) (n=5). Healthy individuals had on average 510 genome equivalents

of megakaryocyte cfDNA/ml plasma (GE/ml) (**Table 1**, results detailed in **Supplementary Tables S3, S4**). ET patients had significantly elevated levels of megakaryocyte cfDNA (average 4242 GE/ml, $p<0.05$). Interestingly, while patients with a hypoproliferative bone marrow and patients with ITP had similar platelets counts, ITP patients had elevated concentrations of circulating megakaryocyte DNA (average 1608 GE/ml, $p<0.05$) while patients with hypoproliferative bone marrows had reduced concentrations of circulating megakaryocyte DNA (average 250 GE/ml, $p<0.05$). These findings are consistent with megakaryocyte cfDNA levels being reflective of megakaryocyte function and turnover (**Figure 5A-B**). Notably, erythroblast cfDNA was not reflective of megakaryocyte-platelet pathologies but was significantly elevated in the plasma of patients with thalassemia major, a disease involving increased, albeit ineffective, production of erythrocytes within the bone marrow (average of 1658 erythroblast GE/ml, $p<0.05$) **(Figure 5-C**). These findings confirm the specificity of our targeted markers and support their potential utility in highly specific identification of altered turnover of the respective cell types of origin.

DISCUSSION

In this work, we demonstrated that megakaryocytes are major contributors to plasma cfDNA in healthy individuals, giving rise to ~26% of total cfDNA. In contrast, erythroblasts give rise to a small but distinct proportion (~0.5-4%) of cfDNA in healthy individuals. These findings contrast with previous reports that identified a major contribution of erythroblasts to cfDNA of healthy individuals. One of these studies, from our own group, relied on a tissue methylome atlas (based on Illumina methylation arrays) which contained the methylome of common megakaryocyte-erythrocyte progenitor cells but not the megakaryocyte methylome[12]. Another study has demonstrated that genomic loci uniquely unmethylated in erythroblasts are also partially unmethylated in plasma, concluding that erythroid cells are a major source of cfDNA in healthy people[13]. A close examination of the methylation status of these marker loci revealed that they are in fact unmethylated not only in erythroblasts but also in megakaryocytes (**Supplementary Figure S8**). Thus, these markers are not sufficient to distinguish between erythroid and megakaryocyte contributions to plasma, and the presence of unmethylated DNA from these loci in healthy plasma is in fact consistent with a significant presence of megakaryocyte DNA in circulation, in line with our current findings and as suggested by analysis of histone modifications in circulating chromatin[11].

Megakaryocytes and erythroblasts have distinct mechanisms of cfDNA release. Megakaryocytes are thought to undergo apoptosis during or immediately subsequent to thrombopoiesis[25]. Our analysis indicates that while a small proportion of the megakaryocyte genome (~0.1%) is trapped within platelets, a fraction is fragmented to nucleosome-size pieces and released to circulation as cfDNA. As for erythroblasts, the process of nuclear extrusion in the final differentiation step of erythrocytes involves efficient removal by local macrophages[26]. Indeed, while 214 billion erythroblast nuclei are extruded every day[27], only a small amount of erythroblast DNA is present in plasma. More experiments are needed to define the exact origin of erythroblast cfDNA.

Additionally, we demonstrated that megakaryocyte- and erythroblast-specific cfDNA concentration is reflective of megakaryocyte and red blood cell precursor activity; thus, it can serve as a biomarker for diseases involving aberrant activity of these blood lineages, not necessarily reflected in peripheral platelet and RBC counts.

Thrombocytopenia in cancer patients can be induced by chemotherapy, bone marrow involvement or peripheral destruction of platelets as a result of immune processes.

Prolonged thrombocytopenia in cancer patients during therapy can be a diagnostic challenge. Bone marrow biopsy is used for differential diagnosis, but the biopsy may be only partially representative. Determining the cause of the thrombocytopenia is crucial, as treatment for such thrombocytopenic patients will differ according to the pathogenesis of the thrombocytopenia. Combined with peripheral blood smear and bone marrow biopsy, cfDNA measurements can potentially distinguish between these clinical scenarios and guide treatment decisions.
ITP is a syndrome characterized by antibody-mediated platelet destruction and variably reduced platelet production[28]. Bone marrow biopsies of ITP patients show normal or increased numbers of megakaryocytes. There is currently no diagnostic tool for ITP. We demonstrated that megakaryocyte cfDNA is elevated in ITP. More experiments are needed in ITP patients and healthy controls to determine whether megakaryocyte methylation biomarkers can serve as a diagnostic tool for ITP.

This study also provides the first evidence that platelets contain megakaryocyte-derived nuclear DNA. An important implication of this situation is that it provides a non-invasive window into the genome and epigenome of the megakaryocyte, a cell type particularly difficult to interrogate due to its relative scarceness within the bone marrow. The presence of somatic mutations in megakaryocytes, as well as epigenomic aberrations, can therefore be interrogated by analysis of platelet samples. Additionally, gene expression of the megakaryocyte, as reflected by transcription related epigenetic modifications, may be indirectly evaluated via the platelet epigenome.
We have not addressed whether platelet DNA is concentrated in a subset of platelets or if it is distributed randomly throughout the platelet population. Additionally, the fate and function of the megakaryocyte genomic fragments trapped within platelets is not clear. It is well established that platelets contain RNA which undergoes translation to protein, yet this RNA is generally understood to have been transcribed in the megakaryocyte prior to platelet formation or from mitochondrial DNA within the platelet [29]. Given our findings that platelets contain nuclear DNA, the question as to whether transcription of nuclear DNA occurs within the platelet should be revisited.

MATERIALS AND METHODS

Platelet isolation

Platelets were isolated from donors via single donor plateletpheresis for the purpose of platelet transfusion at the Hadassah Medical Center Blood Bank. Platelets which had not been transfused within one week were obtained for research purposes. In order to subsequently purify platelets from surrounding plasma, the platelets were diluted in PBS noCa^{2+} & Mg^{2+} or tyrode's buffer (1/3 SDP+2/3 PBS). Platelet activation was inhibited with citric acid (5ul of 1M citric acid per 1ml or 10% citrate from collection tubes). Subsequently, centrifugation was performed at 750g (3000rpm) for 8 minutes at room temperature. A platelet pellet was obtained which was diluted in PBS with citric acid (5 ul citric acid/ml PBS). Optionally, Dnase I was added at a concentration of 100µg/ml and samples were incubated for 20 minutes in a heatbath at 37 degrees Celsius. Samples were then centrifuged again at 750g for 8 min and then pellets were suspended in 1.4 ml PBS (no citrate).

Subject enrollment

This study was conducted according to protocols approved by the Institutional Review Board at the study site, with procedures performed in accordance with the Declaration of Helsinki. Blood samples were obtained from donors who have provided written informed consent.

Subject characteristics are presented in Supplementary Table S3 and in Supplementary Table S4. Patients were recruited with acute ITP, with less than 35000 platelets/ul (n=5), with Essential Thrombocytopenia (n=5, >500000 platelets/ul); 5 patients with Acute Myeloid Leukemia (AML), 14 days after the start of induction chemotherapy with cytarabine and daunorubicin (7+3 regimen) were recruited as patients with hypoplastic bone marrow. Additionaly, 3 patients were recruited with thalassemia major. 77 healthy controls were recruited. Healthy controls were excluded if platelet count was less than 150000 platelets/ul or greater than 500000 platelets/ul.

Blood sample collection and processing

Blood samples were collected by routine venipuncture in 10 ml EDTA Vacutainer tubes or Streck blood collection tubes and stored at room temperature for up to 4 hr or 5 days, respectively. Tubes were centrifuged at 1500× *g* for 10 min at 4°C (EDTA tubes) or at room temperature (Streck tubes). The supernatant was transferred to a fresh 15 ml conical tube without disturbing the cellular layer and centrifuged again for 10 min at 3000× *g*. The supernatant was collected and stored at –80°C. cfDNA was extracted from 2 to 4 ml of plasma using the QIAsymphony liquid handling robot (Qiagen). cfDNA concentration was determined using Qubit double-strand molecular probes kit (Invitrogen) according to the manufacturer's instructions.

DNA derived from all samples was treated with bisulfite using EZ DNA Methylation-Gold (Zymo Research), according to the manufacturer's instructions, and eluted in 24 µl elution buffer.

Whole-genome bisulfite sequencing of reference cell types

Previously published single-cell genome-wide methylation profiles of megakaryocytes and genome-wide profiles of megakaryocyte-erythrocyte progenitors[30] of three individuals were obtained from Blueprint Epigenome. Single-cell methylation profiles were grouped by individual for further analysis.

Erythroblasts were isolated as follows: Bone marrow was diluted with 3 equivalent volumes of PBS, filtered through a 100um cell strainer and laid over Lymphoprep density gradient medium (Stemcell Technologies). After centrifugation, interphase bone marrow mononuclear cells were transferred to PBS, washed and red blood cells (RBC) were lysed using standard RBC lysis buffer.

CD45-negative, CD235a and CD71-positive erythrocyte precursors were FACS-sorted on a BD FACSAriaTM III flow cytometer.

Granulocytes, monocytes, B cells, NK cells, CD3+ T cells, CD4+ T cells, CD8+ T cells, hepatocytes and endothelial cells were isolated as previously described[31].

Up to 75 ng of sheared gDNA was subjected to bisulfite conversion using the EZ-96 DNA Methylation Kit (Zymo Research; Irvine, CA), with liquid handling on a Hamilton MicroLab STAR (Hamilton; Reno, NV). Dual indexed sequencing libraries were prepared using Accel-NGS Methyl-Seq DNA library preparation kits (Swift BioSciences; Ann Arbor, MI) and custom liquid handling scripts executed on the Hamilton MicroLab STAR. Libraries were quantified using KAPA Library Quantification Kits for Illumina Platforms (Kapa Biosystems; Wilmington, MA). Four uniquely dual indexed libraries, along with 10% PhiX v3 library (Illumina; San Diego, CA), were pooled and clustered on a Illumina NovaSeq 6000 S2 flow cell followed by 150-bp paired-end sequencing.

Selection of megakaryocyte and erythroblast methylation markers

CpGs were identified as being cell type specific if they were unmethylated (<20% methylation) in all samples of the cell type of interest and methylated in all other samples

(>80% methylation) or if they were methylated in all samples of the cell type of interest and unmethylated in other samples.
For selection of markers for analysis by targeted bisulfite sequencing, CpGs with neighboring CpGs which were also cell type specific and located in a CpG dense genomic region (>=5 CpGs within 120 bp) were selected for further analysis. DNA methylation of candidate regions were then compared to DNA methylation at these regions in genome-wide methylation profiles of multiple tissues published as part of Roadmap Epigenomics, using the same criteria as described above.
Genome coordinates of cell-type specific methylation markers and primer sequences are included in **Supplementary Table S1**.

PCR

To efficiently amplify and sequence multiple targets from bisulfite-treated cfDNA, we used a two-step multiplexed PCR protocol, as described[32]. In the first step, up to 10 primer pairs were used in one PCR reaction to amplify regions of interest from bisulfite-treated DNA, independent of methylation status. Primers were 18–30 base pairs (bp) with primer melting temperature ranging from 58°C to 62°C. To maximize amplification efficiency and minimize primer interference, the primers were designed with additional 25 bp adaptors comprising Illumina TruSeq Universal Adaptors without index tags. All primers were mixed in the same reaction tube. For each sample, the PCR was prepared using the QIAGEN Multiplex PCR Kit according to manufacturer's instructions with 7 µl of bisulfite-treated cfDNA. Reaction conditions for the first round of PCR were: 95°C for 15 min, followed by 30 cycles of 95°C for 30 s, 57°C for 3 min and 72°C for 1.5 min, followed by 10 min at 68°C.
In the second PCR step, the products of the first PCR were treated with Exonuclease I (ThermoScientific) for primer removal according to the manufacturer's instructions. Cleaned PCR products were amplified using one unique TruSeq Universal Adaptor primer pair per sample to add a unique index barcode to enable sample pooling for multiplex Illumina sequencing. The PCR was prepared using 2× PCRBIO HS Taq Mix Red Kit (PCR Biosystems) according to manufacturer's instructions. Reaction conditions for the second round of PCR were: 95°C for 2 min, followed by 15 cycles of 95°C for 30 s, 59°C for 1.5 min, 72°C for 30 s, followed by 10 min at 72°C. The PCR products were then pooled, run on 3% agarose gels with ethidium bromide staining, and extracted by Zymo GEL Recovery kit.

NGS and analysis of PCR products

Pooled PCR products were subjected to multiplex NGS using the *NextSeq* 500/550 v2 Reagent Kit (Illumina). Sequenced reads were separated by barcode, and aligned to the target sequence with Bismark, using a computational pipeline available (https://github.com/Joshmoss11/btseq). CpGs were considered methylated if 'CG' was read and unmethylated if 'TG' was read. Proper bisulfite conversion was assessed by analyzing methylation of non-CpG cytosines. We then determined the fraction of molecules in which all CpG sites were unmethylated. The fraction obtained was multiplied by the concentration of cfDNA measured in each sample, to obtain the concentration of tissue-specific cfDNA from each donor. Given that the mass of a haploid human genome is 3.3 pg, the concentration of cfDNA could be converted from units of ng/ml to haploid GE/ml by multiplying by a factor of 303.
Sequenced molecules were considered to be unmethylated (for specifically unmethylated markers) if all CpGs on the sequenced DNA fragment were unmethylated. The fraction of DNA derived from the cell type of interest for each marker (for specifically unmethylated markers) was calculated as the fraction of completely unmethylated molecules among all sequenced molecules from the marker region.

Whole-genome bisulfite sequencing of platelet DNA

DNA was extracted from three platelet samples using the QIAsymphony liquid handling robot (Qiagen). Subsequently, whole-genome bisulfite sequencing libraries were generated with Swift AccelNGS Methyl-Seq
DNA Library preparation protocol (Swift Biosciences, Ann Arbor, MI). Paired-end sequencing was performed on the Illumina Nextseq 550 System of 300 bp per read at an average depth of 2.54x.

Whole-genome bisulfite sequencing computational processing and analysis

Paired-end FASTQ files were mapped to the human (hg19) genome using bwa-meth (V 0.2.0), with default parameters[33], then converted to BAM files using SAMtools (V 1.9)[34].
Duplicated reads were marked by Sambamba (V 0.6.5), with parameters "-l 1 -t 16 --sort-buffer-size 16000 --overflow-list-size 10000000"[35].
Reads with low mapping quality, duplicated, or not mapped in a proper pair were excluded using SAMtools view with parameters -F 1796 -q 10.
Reads were stripped from non-CpG nucleotides and converted to BETA files using wgbstools (V 0.1.0) (https://github.com/nloyfer/wgbs_tools)[31].
For comparison of platelet DNA methylation to DNA methylation of other cell types, bone marrow-residing immune cell progenitor cells were obtained from Blueprint Epigenome[36]. genome-wide methylation profiles of methylation profiles of hepatocytes, endothelial cells and immune cells, which have been found to release DNA to plasma under normal conditions[12], were obtained as described[31]. Regions with methylation unique to each cell type were identified by segmenting the genome into multi-sample homogenous blocks as previously described[31], and identifying regions unmethylated (<20%) or methylated (>80%) in a specific cell type. Deconvolution of the platelet DNA methylation profiles at the resultant regions was performed by NNLS as previously described[12].

Statistics

To determine the significance of differences between groups we used a non-parametric two-tailed Mann-Whitney test. P-values were considered significant at < 0.05. Samples that were detected as outliers were excluded. All statistical analyses were performed with R (version 4.1)[37].

List of Supplementary Materials

Fig S1 to S8
Tables S1 to S4

References

1. Lo, Y. M. D. *et al.* Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl. Med.* (2010). doi:10.1126/scitranslmed.3001720

2. De Vlaminck, I. *et al.* Circulating cell-free DNA enables noninvasive diagnosis of heart transplant rejection. *Sci. Transl. Med.* (2014). doi:10.1126/scitranslmed.3007803

3. Wan, J. C. M. *et al.* Liquid biopsies come of age: Towards implementation of circulating tumour DNA. *Nature Reviews Cancer* **17**, 223–238 (2017).

4. Dor, Y. & Cedar, H. Principles of DNA methylation and their implications for biology and medicine. *Lancet* (2018). doi:10.1016/S0140-6736(18)31268-6

5. Zemmour, H. *et al.* Non-invasive detection of human cardiomyocyte death using

methylation patterns of circulating DNA. *Nat. Commun.* (2018). doi:10.1038/s41467-018-03961-y

6. Lehmann-Werman, R. *et al.* Monitoring liver damage using hepatocyte-specific methylation markers in cell-free circulating DNA. *JCI Insight* (2018). doi:10.1172/jci.insight.120687

7. Lehmann-Werman, R. *et al.* Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E1826–E1834 (2016).

8. Liu, M. C. *et al.* Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann. Oncol.* (2020). doi:10.1016/j.annonc.2020.02.011

9. Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M. & Shendure, J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **164**, 57–68 (2016).

10. Lo, Y. M. D., Han, D. S. C., Jiang, P. & Chiu, R. W. K. Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. *Science (80-.).* **372**, (2021).

11. Sadeh, R. *et al.* ChIP-seq of plasma cell-free nucleosomes identifies cell-of-origin gene expression programs. *bioRxiv* (2019). doi:10.1101/638643

12. Moss, J. *et al.* Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat. Commun.* **9**, 5068 (2018).

13. Lam, W. K. J. *et al.* DNA of Erythroid Origin Is Present in Human Plasma and Informs the Types of Anemia. *Clin. Chem.* **63**, 1614–1623 (2017).

14. Shearstone, J. R. *et al.* Global DNA Demethylation During Erythropoiesis in vivo. *Science* **334**, 799 (2011).

15. Schulz, V. P. *et al.* A Unique Epigenomic Landscape Defines Human Erythropoiesis. *Cell Rep.* **28**, 2996-3009.e7 (2019).

16. Fox-Fisher, I. *et al.* Remote immune processes revealed by immune-derived circulating cell-free dna. *Elife* **10**, (2021).

17. Bray, P. F. *et al.* The complex transcriptional landscape of the anucleate human platelet. *BMC Genomics* **14**, 1–15 (2013).

18. Frydman, G. H. *et al.* Megakaryocytes contain extranuclear histones and may be a source of platelet-associated histones during sepsis. *Sci. Rep.* (2020). doi:10.1038/s41598-020-61309-3

19. Nilsson, R. J. A. *et al.* Blood platelets contain tumor-derived RNA biomarkers. *Blood* **118**, 3680–3683 (2011).

20. In 't Veld, S. G. J. G. *et al.* Detection and localization of early- and late-stage cancers using platelet RNA. *Cancer Cell* **40**, 999-1009.e6 (2022).

21. Shuster, R. C., Rubenstein, A. J. & Wallace, D. C. Mitochondrial DNA in anucleate human blood cells. *Biochem. Biophys. Res. Commun.* **155**, 1360–1365 (1988).

22. Bessman, J. D. The relation of megakaryocyte ploidy to platelet volume. *Am. J. Hematol.* **16**, 161–170 (1984).

23. Zadehkoochak, M. *et al.* Clinical Physics and Physiological Measurement Analysis of the sensitivity method of reconstruction using spectral expansion Platelet thermophysiology: a new field of investigation dependent upon an improved sub-

ambient platelet aggregometer The origin Of platelet count and volume. *Clin. Phys. Physiol. Meas* **5**, 145 (1984).

24. Cesar, J. M., Vecino, A. M., Jesu´, J. & Cesar, J. M. Survival and function of transfused platelets. Studies in two patients with congenital deficiencies of platelet membrane glycoproteins. *https://doi.org/10.1080/09537100902751925* **20**, 158–162 (2009).

25. McArthur, K., Chappaz, S. & Kile, B. T. Apoptosis in megakaryocytes and platelets: the life and death of a lineage. *Blood* **131**, 605–610 (2018).

26. Chasis, J. A. & Mohandas, N. Erythroblastic islands: niches for erythropoiesis. *Blood* **112**, 470–478 (2008).

27. Sender, R. & Milo, R. Brief CommuniCation The distribution of cellular turnover in the human body. *Nat. Med.* **27**, 45–48 (2021).

28. Cines, D. B., Cuker, A. & Semple, J. W. Pathogenesis of immune thrombocytopenia. *Presse Med.* **43**, e49–e59 (2014).

29. Schubert, S., Weyrich, A. S. & Rowley, J. W. A tour through the transcriptional landscape of platelets. *Blood* **124**, 493–502 (2014).

30. Farlik, M. *et al.* DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation. *Cell Stem Cell* (2016). doi:10.1016/j.stem.2016.10.019

31. Loyfer, N. *et al.* A human DNA methylation atlas reveals principles of cell type-specific methylation and identifies thousands of cell type-specific regulatory elements. *bioRxiv* 2022.01.24.477547 (2022). doi:10.1101/2022.01.24.477547

32. Neiman, D. *et al.* Multiplexing DNA methylation markers to detect circulating cell-free DNA derived from human pancreatic β cells. *JCI Insight* **5**, (2020).

33. Pedersen, B. S., Eyring, K., De, S., Yang, I. V & Schwartz, D. A. Fast and accurate alignment of long bisulfite-seq reads. (2014).

34. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

35. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).

36. Adams, D. *et al.* BLUEPRINT to decode the epigenetic signature written in blood. *Nat. Biotechnol. 2012 303* **30**, 224–226 (2012).

37. R Core Team. R: A Language and Environment for Statistical Computing. (2018).

Acknowledgments

We acknowledge the donors of the biological samples presented in this study for their benevolent contribution to science. We thank Dr. Abed Nasereddin and Dr. Idit Shiff of the Core Research Facility of the Faculty of Medicine of the Hebrew University of Jerusalem for performing the targeted bisulfite sequencing experiments and WGBS of platelets presented in the article, and the blood bank of the Hadassah Medical Center for aid in obtaining platelet samples. We thank Prof. Neta Goldschmidt for recruiting patients to the study.

Funding

This study was supported by grants from The Ernest and Bonnie Beutler Research Program of Excellence in Genomic Medicine (FREEDOME), The Helmsley Charitable Trust,

Alzheimer's Drug Discover Foundation, a Merck Grant for Multiple Sclerosis Innovation (GMSI), The Israel Science Foundation, the Waldholtz/Pakula family, and the Robert M. and Marilyn Sternberg Family Charitable Foundation (to YD). YD holds the Walter and Greta Stiel Chair and Research grant in Heart studies.

Author contributions

Conceptualization: JM, YD, RS, AA, RB
Methodology: JM, RB, YD, GC, BG, RS, AK, ES, YK
Investigation: JM, RB, AK, AA, ES
Visualization: JM, AK
Funding acquisition: YD
Project administration: YD, RS, AA, JM
Supervision: YD, RS, AA
Writing – original draft: JM
Writing – review & editing: JM, YD, BG, AA, YK, RB, GC, RS

Competing interests

JM, RS, BG, and YD are inventors of a patent filing describing analysis of platelet DNA as well as megakaryocyte methylation markers and their use for cfDNA analysis. All remaining authors have declared no conflicts of interest.

Data and materials availability

Results of targeted bisulfite sequencing for erythroblast and megakaryocyte methylation markers are included in Supplementary Table S3.
Whole-genome bisulfite sequencing data of platelet DNA has been deposited at GEO, accession number GSE206818.

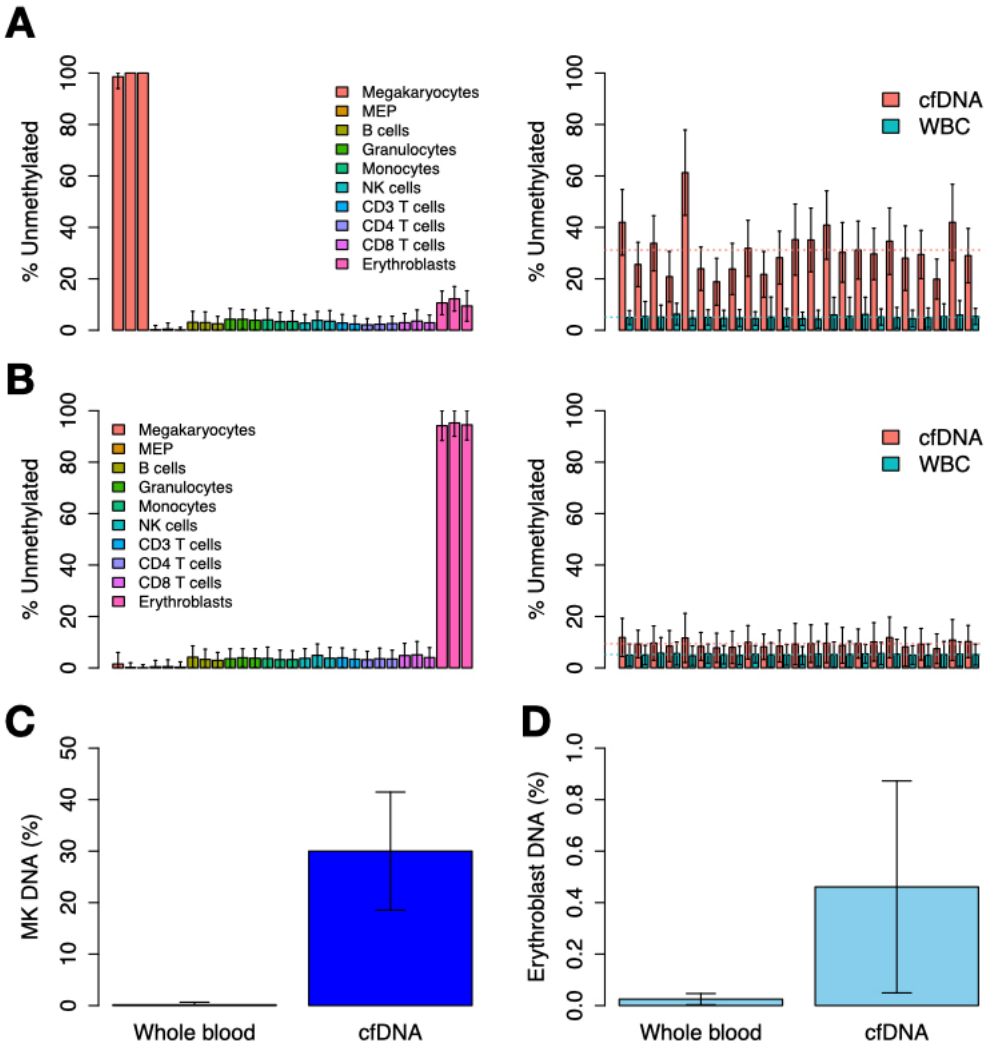
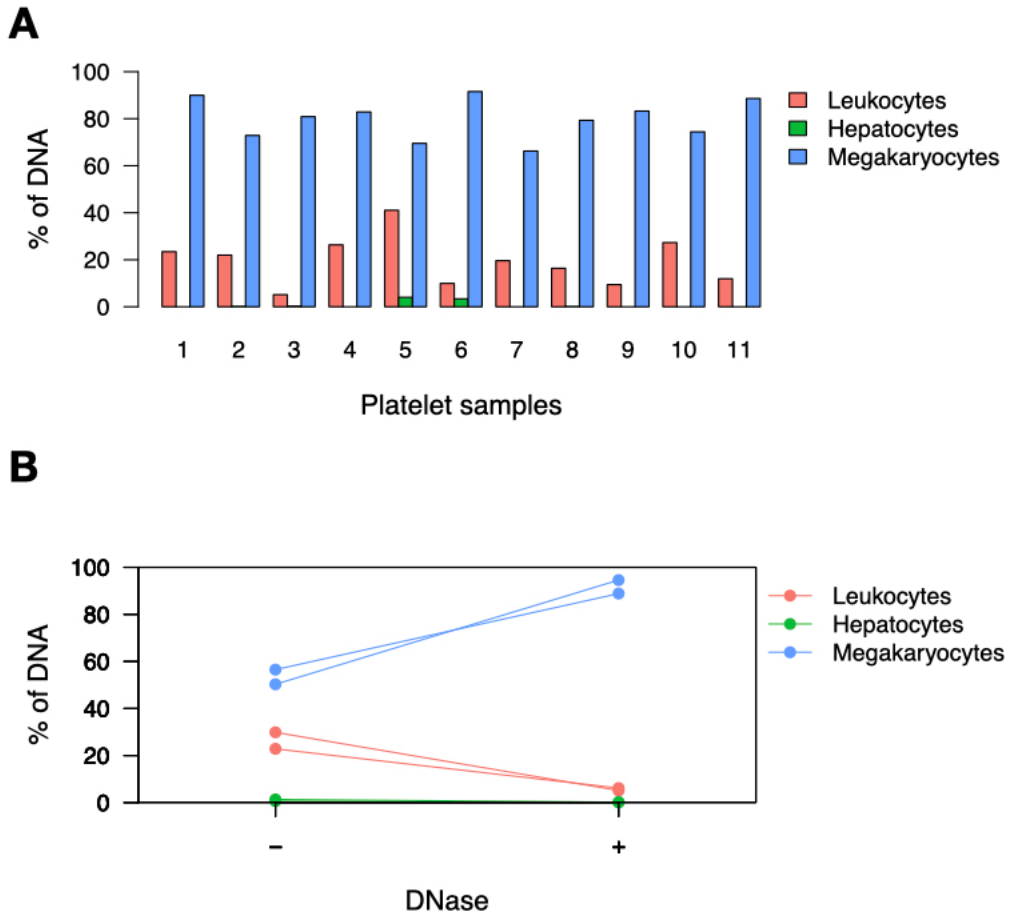


Fig. 1. **Comparative analysis of erythroblast and megakaryocyte methylomes, and their representation in plasma.** (**A**) Genomic loci unmethylated specifically in megakaryocytes (n=97 loci) were identified by comparison to other cell types of hematopoietic origin (left) and subsequently evaluated by WGBS of white blood cells (WBC) and cell-free DNA (cfDNA) of 23 individuals (right). These sites were, on average, 5% unmethylated in WBC and 31% unmethylated in cfDNA. For each individual, the sites were significantly hypomethylated in cfDNA ($p<0.05$). Average percent of unmethylated sites is marked by a dotted line for cfDNA (red) and WBC (blue). (**B**) Genomic loci unmethylated specifically in erythroblasts (n=1884 loci) were identified by comparison to other cell types of hematopoietic origin (left) and subsequently evaluated by WGBS of white blood cells (WBC) and cell-free DNA (cfDNA) of 23 individuals (right). These sites were, on average, 5% unmethylated in WBC and 9% unmethylated in cfDNA. For each individual, the sites were significantly hypomethylated in cfDNA ($p<0.05$). Average percent of unmethylated sites is marked by a dotted line for cfDNA (red) and WBC (blue). (**C**) Targeted bisulfite-sequencing of megakaryocyte-specific unmethylated regions in whole blood (n=26) and cfDNA (n=62) of healthy individuals, demonstrating a higher percentage of megakaryocyte-derived DNA in cfDNA compared to WBC ($p<0.05$). (**D**) Targeted bisulfite-sequencing of erythroblast-specific unmethylated regions in whole blood (n=15) and cfDNA (n=71) of healthy

602 individuals, demonstrating a higher percentage of erythroblast-derived DNA in cfDNA
603 compared to WBC ($p<0.05$).



604

605 **Fig. 2. Platelets contain genomic DNA derived from megakaryocytes.** (**A**) DNA extracted from
606 platelet concentrates contains DNA derived from megakaryocytes, leukocytes and
607 hepatocytes, as measured by targeted bisulfite sequencing of cell-type specific unmethylated
608 genomic regions. (**B**) Uncentrifuged platelet concentrates (platelets in plasma) were analyzed
609 with or without DNase treatment, demonstrating that DNase treatment reduces leukocyte and
610 liver, but not MK markers in platelets.

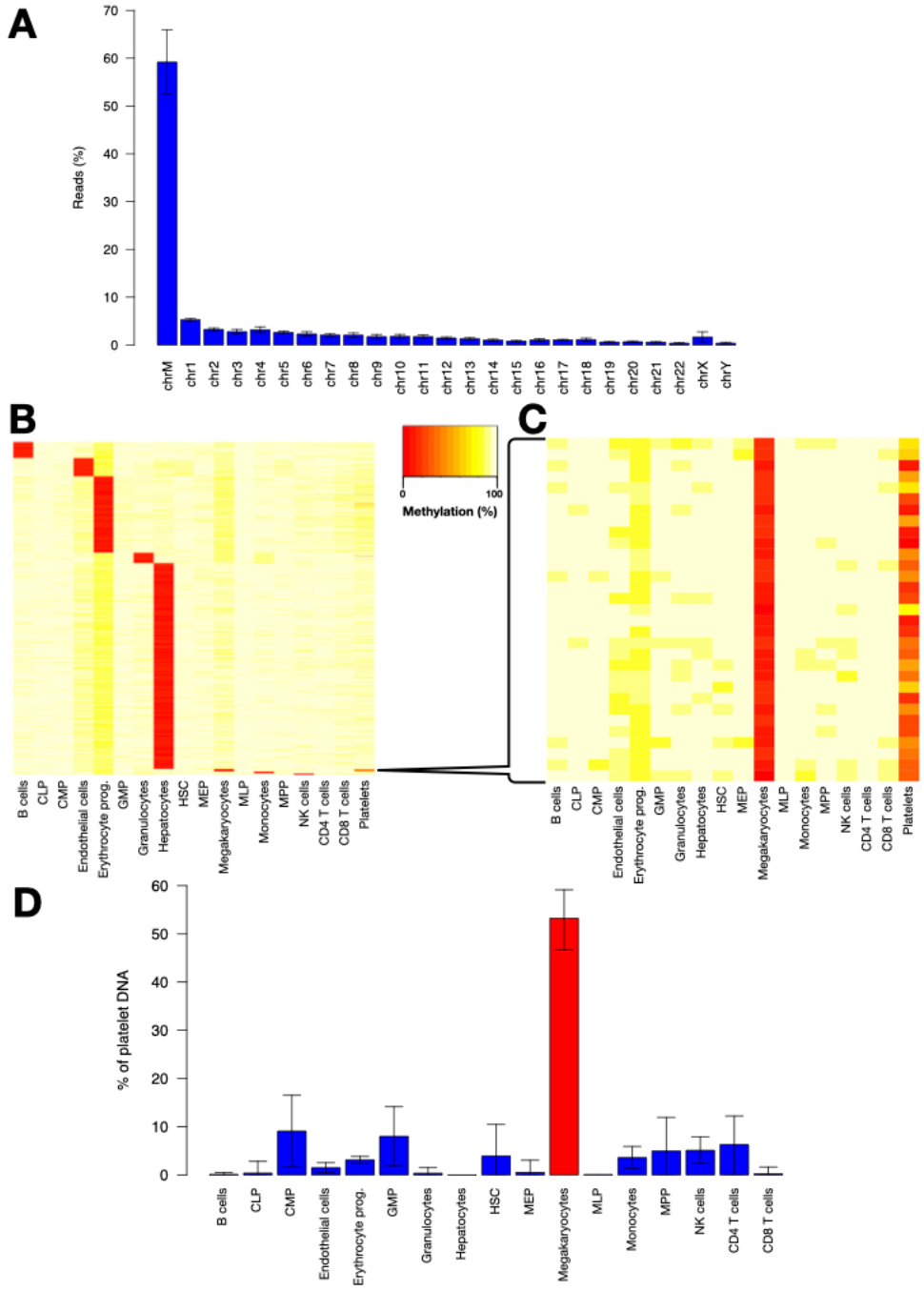


Fig. 3. Genome-wide analysis of platelet DNA supports megakaryocyte origin. (**A**) Platelet DNA is derived from all human chromosomes; however, the majority is mitochondrial DNA. The mean of three platelet samples is plotted with error bars representing standard deviations. (**B-C**) Regions uniquely unmethylated in cell types of hematopoietic origin, hepatocytes and endothelial cells were identified as described (Methods). Megakaryocyte-unique unmethylated regions are unmethylated in platelets as well. HSC – Hematopoietic stem cell, MPP – Multipotent progenitor, CMP – Common myeloid progenitor, CLP – Common lymphoid progenitor, MLP – Multi-lymphoid progenitor, GMP – Granulocyte macrophage progenitor, MEP – Megakaryocyte erythroid progenitor. (**D**) Deconvolution of platelet DNA methylation demonstrated megakaryocyte DNA as the main component of platelet DNA. Error bars represent 90% confidence intervals calculated by bootstrapping over 10000 iterations.

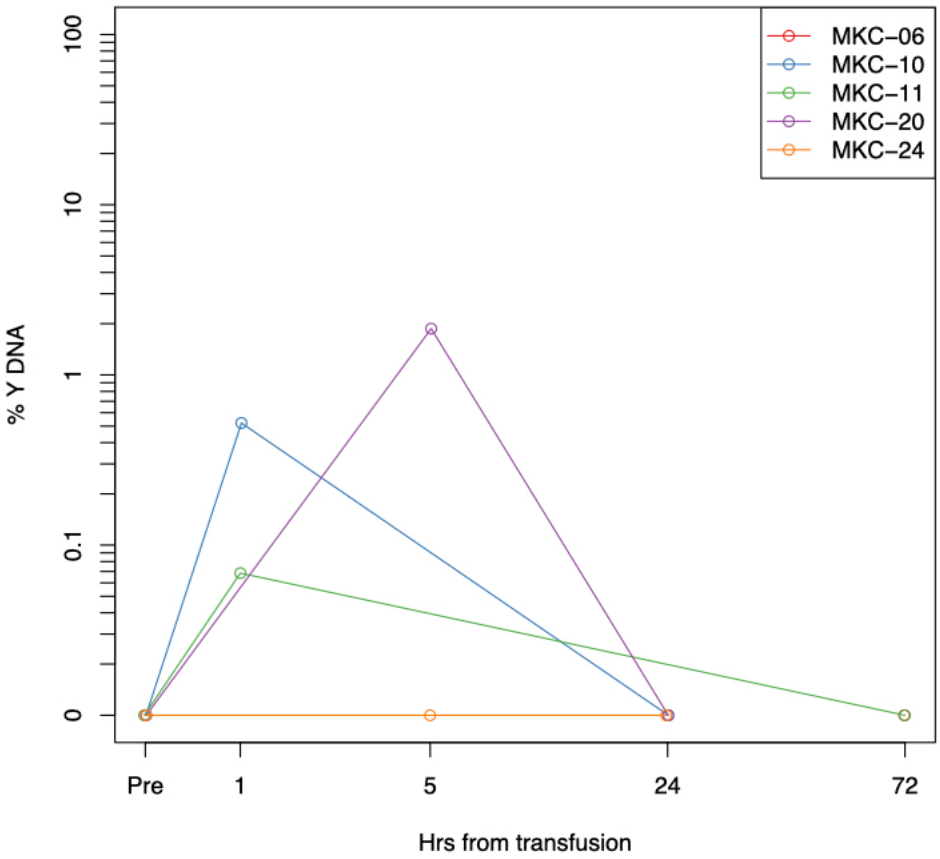
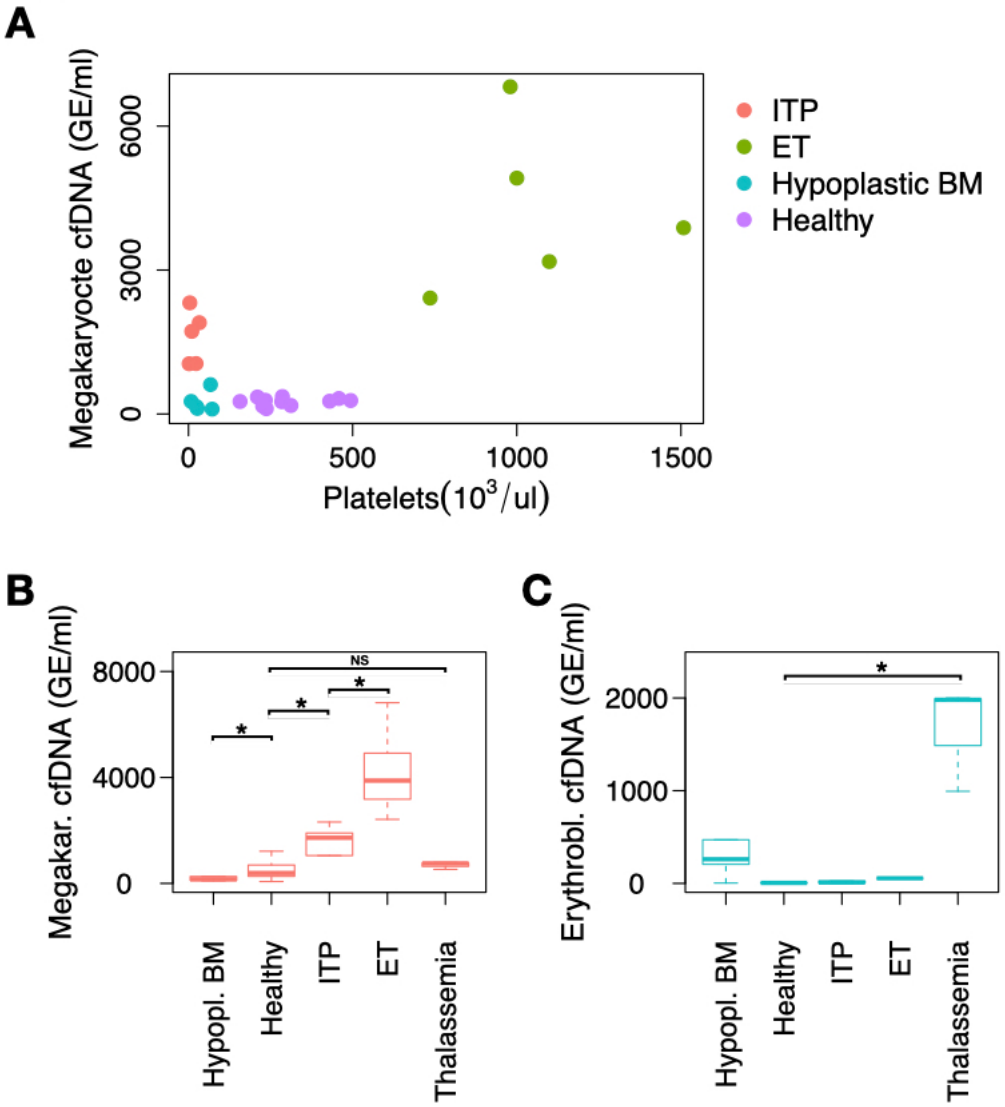


Fig. 4. Gender-mismatched platelet transfusions suggest that platelets are not the source of megakaryocyte DNA. The ratio of concentration of DNA from the SRY gene, located on the Y-chromosome, to DNA from the Beta-actin gene, located on chromosome 7, was evaluated in the plasma of female recipients of platelets from male donors, by massive parallel sequencing. At 24-72 hours after transfusion there was no remaining male DNA detectable.



630

631 **Fig. 5. Targeted analysis of megakaryocyte (MK) and erythroblast methylation markers in**
632 **plasma samples.** (**A**) Genome equivalents (GE) per milliliter plasma of megakaryocyte DNA
633 in samples from healthy donors and patients with ITP, ET and hypoplastic bone marrow (after
634 chemotherapy). (**B**) MK DNA is present in significantly different concentrations between
635 healthy, ITP, ET and hypoplastic bone marrow ($p<0.05$). MK DNA is not significantly
636 different in Thalassemia as compared to healthy individuals. (**C**) Erythroblast cfDNA is
637 significantly elevated in Thalassemia compared to healthy individuals, as well as compared to
638 ET ($p<0.05$). * $p<0.05$

639

640

641

642

Table 1. Description and results of clinical samples analyzed for megakaryocyte and erythroblast cfDNA concentrations.

Group	Count	Age (avg.)	Age (SD)	Female (n)	Platelet count (n/ul)	Megakaryocyte cfDNA (avg. GE/ml)	Erythroblast cfDNA (avg. GE/ml)
ITP	5	42	25	3	14400	1608	21
ET	5	67	15	3	1064600	4242	112
Hypoplastic BM	5	60	21	2	39600	250	2126
Healthy	77	36	15	40	302545	510	7
Thalassemia	3	48	7	2	619667	690	1658

Supplementary Figures

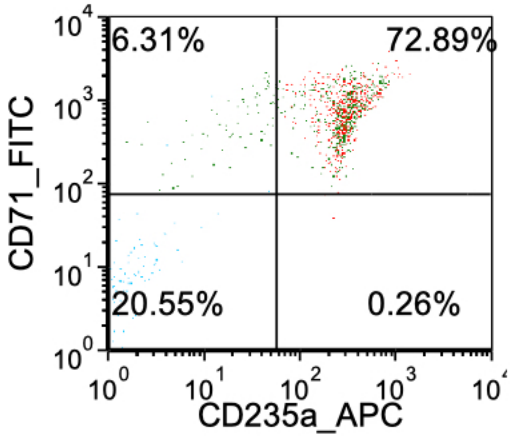


Fig S1. FACS plot of erythroid cells sorted from bone marrow. CD45-negative, CD235a and CD71-positive erythrocyte precursors, derived from bone marrow were FACS-sorted on a BD FACSAriaTM III flow cytometer.

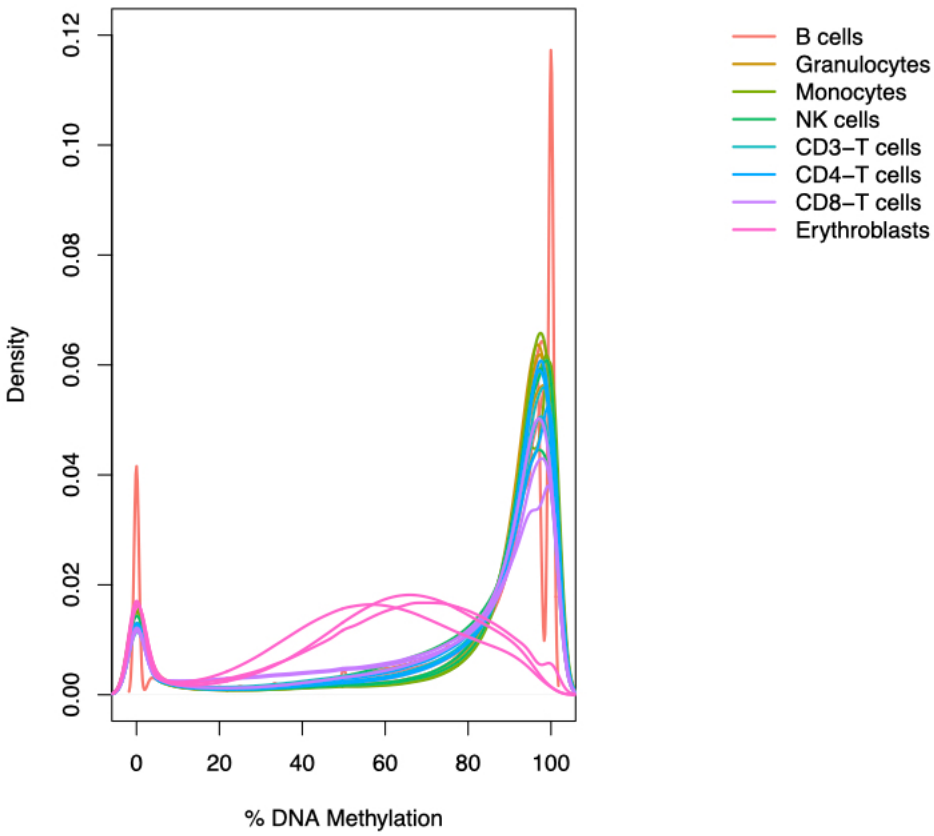


Fig S2. Density plot of methylomes of hematopoietic lineage demonstrating global demethylation in erythroblasts as compared to other cell types.

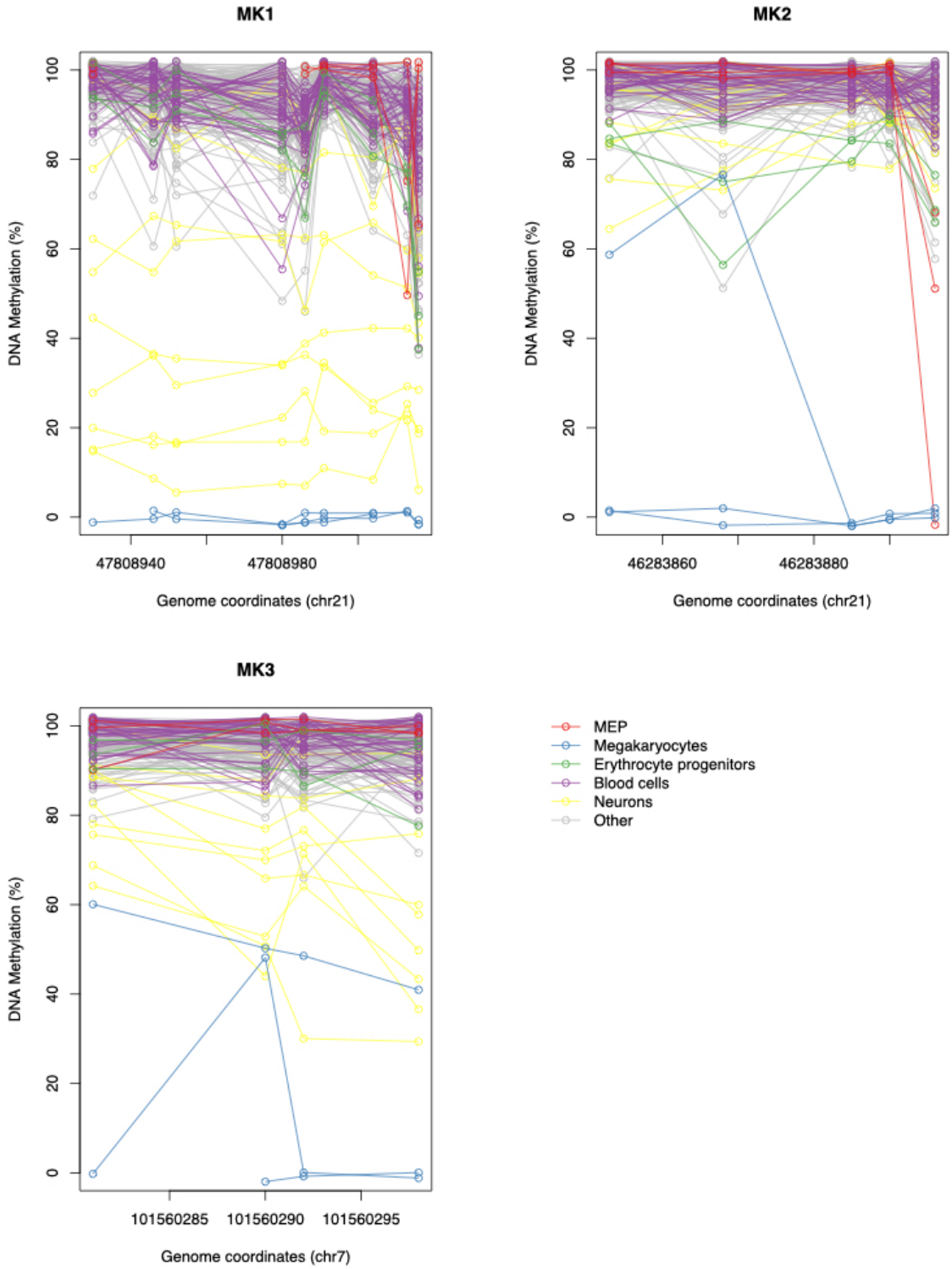
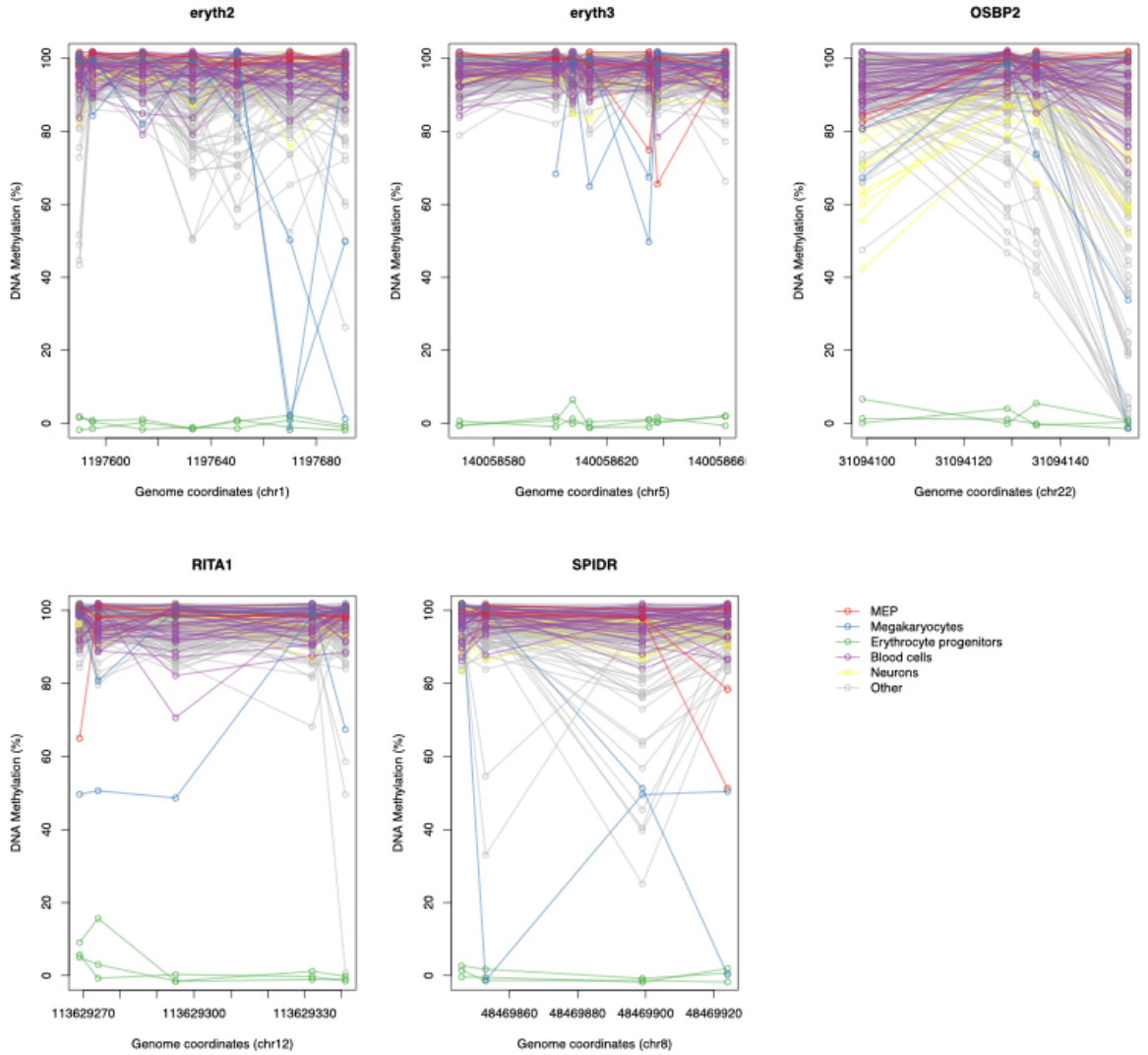
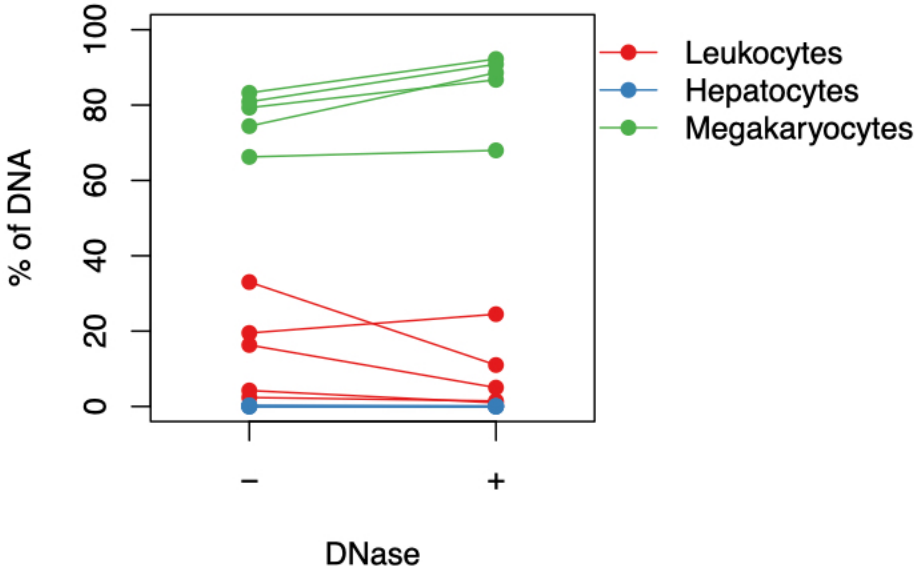


Fig S3. Megakaryocyte DNA methylation markers in megakaryocytes and other cell types.



662
663 **Fig S4. Erythroblast DNA methylation markers in megakaryocytes and other cell types.**



664
665 **Fig S5. MK DNA in centrifuged platelets treated with DNase.**

The percentage of MK DNA isolated from centrifuged platelets is significantly increased when treating with DNase, supporting the presence of MK DNA within platelets as opposed to leukocyte and liver DNA ($p<0.05$).

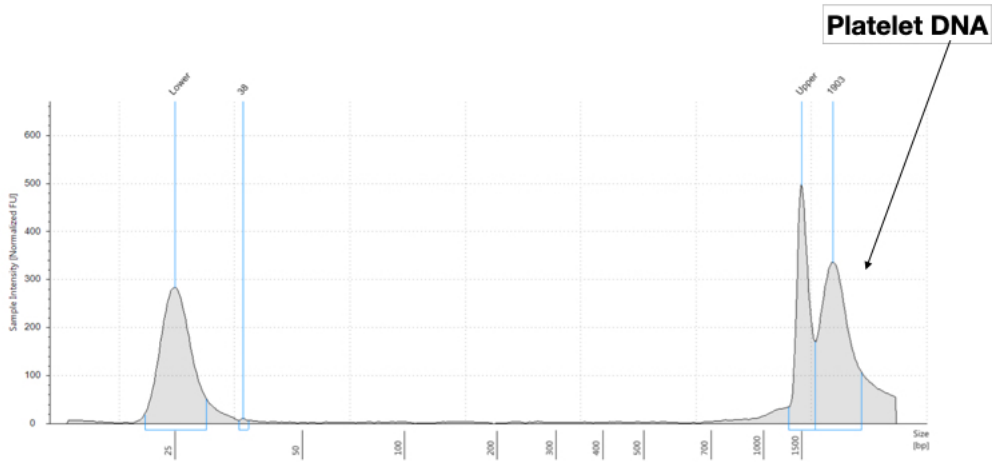


Fig S6. Platelet DNA is composed of high molecular weight DNA, as measured by Agilent High Sensitivity D1000 system.

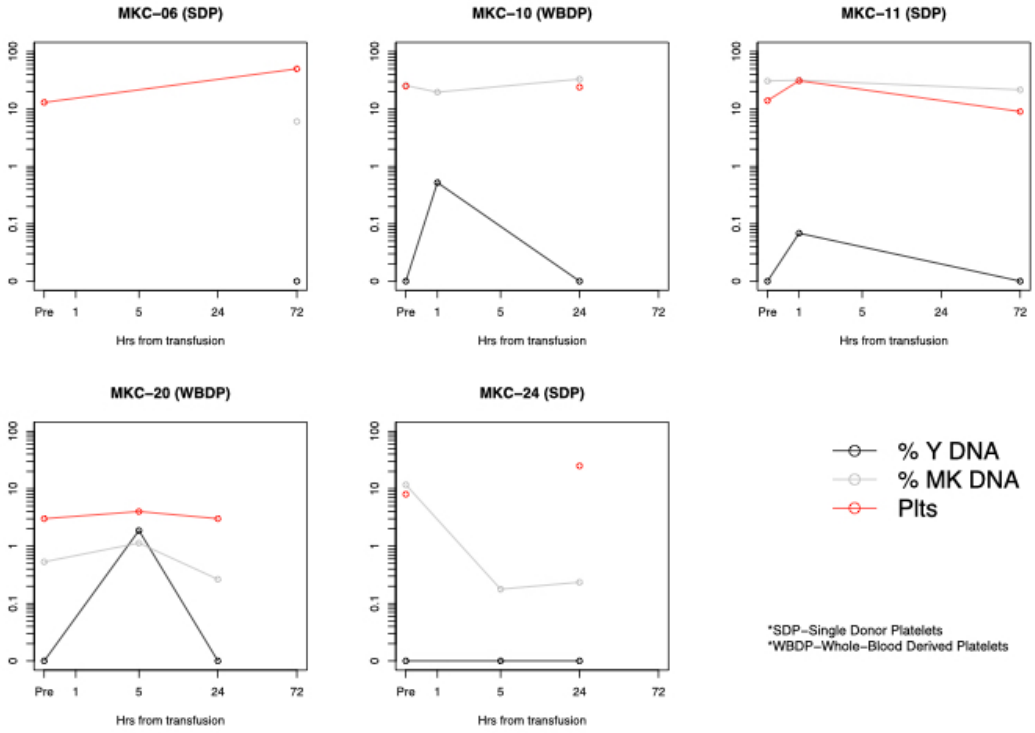


Fig S7. %Y chromosome DNA (SRY/b-Actin), platelet counts and MK DNA for females receiving male donor platelets.

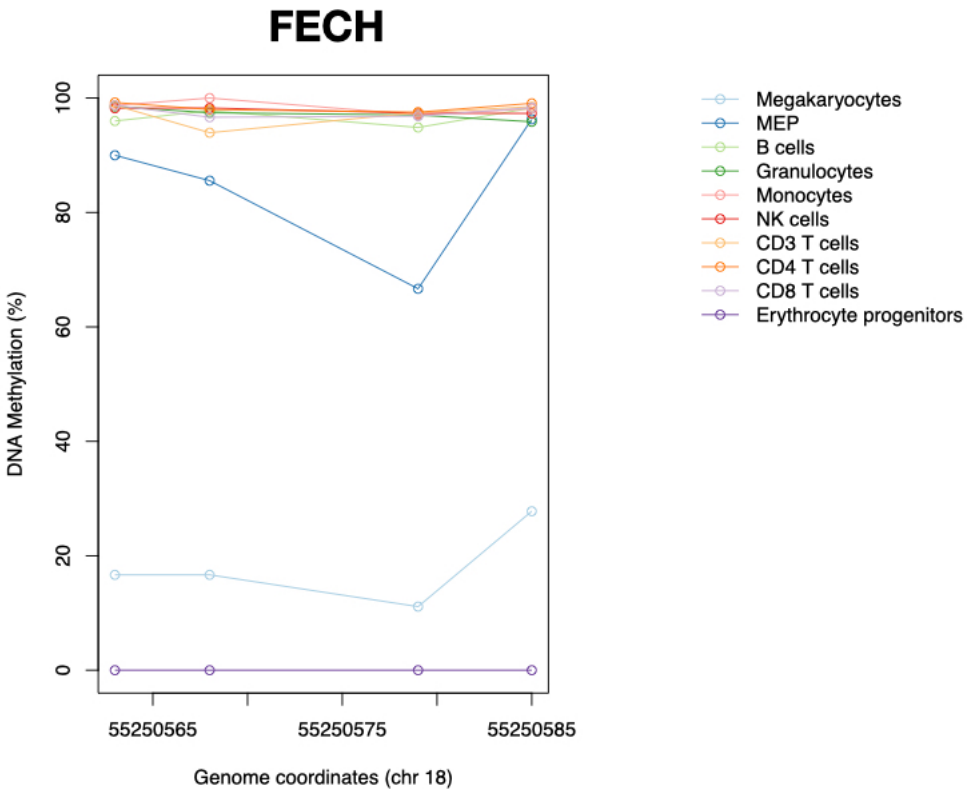


Fig S8. A previously reported erythroid-specific unmethylated marker is unmethylated in both erythroblasts and megakaryocytes.