

Transcriptional characterization of human megakaryocyte polyploidization and lineage commitment

Fizzah A. Choudry^{1,2,3}  | Frederik O. Bagger^{4,5,6} | Iain C. Macaulay⁵ | Samantha Farrow¹ | Frances Burden¹ | Carly Kempster¹ | Harriet McKinney¹ | Lars R. Olsen^{6,7} | Ni Huang^{4,5} | Kate Downes¹ | Thierry Voet⁵ | Rakesh Uppal^{2,3} | John F. Martin⁸ | Anthony Mathur^{2,3} | Willem H. Ouwehand^{1,5,9} | Elisa Laurenti¹ | Sarah A. Teichmann^{4,5} | Mattia Frontini^{1,9} 

¹Department of Hematology, University of Cambridge, Cambridge, UK

²Barts Heart Centre, Barts Health NHS Trust, London, UK

³Centre for Cardiovascular Medicine and Devices, Queen Mary University of London, London, UK

⁴European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Cambridge, UK

⁵The Wellcome Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK

⁶Centre for Genomic Medicine, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

⁷Department of Health Technology, Technical University of Denmark, Lyngby, Denmark

⁸Division of Medicine, University College London, London, UK

⁹BHF Centre of Excellence, Division of Cardiovascular Medicine, Cambridge University Hospitals, Cambridge, UK

Correspondence

Mattia Frontini, Institute of Biomedical & Clinical Science, College of Medicine and Health, University of Exeter Medical School, RILD Building, Barrack Road, Exeter, EX2 5DW.
Email: mf471@cam.ac.uk

Frederik O. Bagger, European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK.
Email: frederik.otzen.bagger@regionh.dk

Funding information

National Institute for Health Research; NHS Blood and Transplant; Bristol-Myers Squibb; European Commission; British Heart Foundation, Grant/Award Number: RE/13/6/30180; Medical Research Council, Grant/Award Number: MR/K024043/1

Abstract

Background: Megakaryocytes (MKs) originate from cells immuno-phenotypically indistinguishable from hematopoietic stem cells (HSCs), bypassing intermediate progenitors. They mature within the adult bone marrow and release platelets into the circulation. Until now, there have been no transcriptional studies of primary human bone marrow MKs.

Objectives: To characterize MKs and HSCs from human bone marrow using single-cell RNA sequencing, to investigate MK lineage commitment, maturation steps, and thrombopoiesis.

Results: We show that MKs at different levels of polyploidization exhibit distinct transcriptional states. Although high levels of platelet-specific gene expression occur in the lower ploidy classes, as polyploidization increases, gene expression is redirected toward translation and posttranslational processing transcriptional programs, in preparation for thrombopoiesis. Our findings are in keeping with studies of MK ultrastructure and supersede evidence generated using *in vitro* cultured MKs. Additionally, by analyzing

F.A.C. and F.O.B. contributed equally.

Manuscript handled by: Matthew T. Rondina.

Final decision: Matthew T. Rondina and 12-Feb-2021.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Journal of Thrombosis and Haemostasis* published by Wiley Periodicals LLC on behalf of International Society on Thrombosis and Haemostasis.

transcriptional signatures of a single HSC, we identify two MK-biased HSC subpopulations exhibiting unique differentiation kinetics. We show that human bone marrow MKs originate from these HSC subpopulations, supporting the notion that they display priming for MK differentiation. Finally, to investigate transcriptional changes in MKs associated with stress thrombopoiesis, we analyzed bone marrow MKs from individuals with recent myocardial infarction and found a specific gene expression signature. Our data support the modulation of MK differentiation in this thrombotic state.

Conclusions: Here, we use single-cell sequencing for the first time to characterize the human bone marrow MK transcriptome at different levels of polyploidization and investigate their differentiation from the HSC.

KEYWORDS

megakaryocytes, hematopoietic stem cells, platelets, single cell RNA-seq, thrombosis

Essentials

- Full transcript single cell RNA sequencing of primary human megakaryocytes and hematopoietic stem cells.
- Megakaryocyte of different ploidy levels exhibit distinct transcriptional states.
- Two megakaryocyte-biased hematopoietic stem cell populations were identified.
- Megakaryocytes from individuals with myocardial infarction exhibit a specific gene signature.

1 | INTRODUCTION

Megakaryocytes (MKs) comprise <0.01% of all nucleated cells in the human bone marrow. These are large, fragile, up to 150 μm in diameter, and highly polyploid (average 16 N) cells.¹ The current knowledge about their transcriptional landscape is largely based on gene expression array data of *in vitro* differentiated MKs cultured from CD34+ cells obtained from fetal liver, cord blood, and adult blood; all having an average ploidy of 2 N and with severely constrained capacity for platelet production.²⁻⁶ It is therefore likely that some drivers of differentiation, ploidy change, and maturation of *in vivo* MKs are yet to be identified.

MK differentiation from hematopoietic stem cells (HSCs) via the classic route of intermediate progenitors has been challenged and been shown that they can originate directly from the multipotent HSC compartment as the first lineage bifurcation.⁷⁻¹¹ Lineage tracing experiments have established a direct HSC origin, independent from other lineages, for at least one-half of all MKs.¹² MK-primed HSCs have been identified in mouse by their expression of genes for the coagulation protein von Willebrand factor¹³ and glycoprotein (GP) 2b (CD41) of the integrin receptor for fibrinogen (integrin $\alpha 2\text{-}\beta$).¹⁴

The primary physiological function of MKs is thrombopoiesis, where each cell produces up to 6000 platelets that play a pivotal role in hemostasis. During steady-state thrombopoiesis, the circulating platelet mass is maintained constant and an inverse relationship between platelet count and volume exists.¹⁵ Platelet consumption results in the acute release of larger platelets to maintain overall circulating platelet mass and aberrations in platelet count and function result in thrombotic or bleeding disorders.¹⁶ Evidence suggests that stemlike MK-committed progenitors are activated upon acute

inflammatory stress.¹⁷ Moreover, sepsis drives changes in transcription and translation in platelets and transcriptional changes in murine MKs.¹⁸ Stress thrombopoiesis is also thought to occur in myocardial infarction, based on the observation of elevated mean platelet volume and reticulated platelet fraction in the acute setting suggesting an upregulation of the HSC–MK axis.¹⁹

Here, we characterize MKs and HSCs from human bone marrow, down to single-cell resolution, by full-transcript RNA sequencing. Primary MK transcriptome analysis showed that low ploidy states have a high expression of archetypal platelet genes, which are then downregulated as the ploidy increases in favor of genes implicated in translation and protein localization. The origin of MKs was also investigated by performing single-cell RNA sequencing of human bone marrow HSCs from the same individuals. Within the HSC compartment, we identified two subpopulations of cells representing MK-primed HSCs. Finally, by sequencing MKs from individuals undergoing coronary artery bypass grafting following myocardial infarction, we showed changes in primary MK gene expression supporting a role for stress thrombopoiesis in this pathological state.

2 | METHODS

2.1 | MK and HSC isolation

Bone marrow for HSC and MK sorting was obtained from individuals undergoing cardiac surgery at Barts Health NHS Trust, London, after informed consent and ethical approval from London - City & East, REC 13/LO/1760 (BAMI Platelet Sub-study). Inclusion criteria

were: (1) heart valve replacement with no evidence of coronary artery disease on coronary angiography or (2) coronary artery bypass grafting in the context of recent myocardial infarction (within 6 months). Myocardial infarction was defined as acute presentation with chest pain, deviation of ST segments on electrocardiogram (either ST elevation myocardial infarction or non-ST elevation myocardial infarction), and rise in levels of troponin T assay. Exclusion criteria were: presence or history of hematological malignancy and abnormal platelet count and hemoglobin levels <85 g/L. Bone marrow scrapings following median sternotomy were collected into an ethylenediaminetetraacetic acid Vacutainer tube containing Dulbecco's phosphate buffered saline containing 10% human serum albumin.

Red cells were lysed using ammonium chloride lysis. MKs were stained with anti-CD41a-APC, anti-CD42b-PE, and 1 μ g/ml Hoechst 33342. Cells were sorted as single MKs defined as CD41^{Hi}, CD42^{Hi}, and 2 N to 32 N by Hoechst staining using a FACSAria Fusion flow sorter. From each bone marrow sample, the approximate nucleated cell count was 1×10^6 cells. To assess changes in MK transcriptome with increasing ploidy level, MKs were single cell-sorted as single cells and 20 cell pools according to their ploidy level, which was determined by level of Hoechst staining. To assess changes in MK transcriptome between disease (myocardial infarction) and noncoronary artery disease (non-CAD) control groups MKs were single cell-sorted in 50- to 100-cell pools; here, sorting was performed on all MKs without separation by ploidy level, thereby representing with the full distribution of MK ploidy observed. Because of the size of MKs (up to 100 μ M) a 150- μ M nozzle was used for fluorescence-activated cell sorting (FACS). HSCs were stained with anti-Lin-PECy5 (CD2, CD3, CD10, CD11b, CD11c, CD19, CD20, CD56, CD42b, GP6), anti-CD34-AF700, anti-CD38-PerCP-Cy5.5, anti-CD90-PECy7, anti-CD49f-PE, and anti-CD45RA-PB. Cells were sorted as single HSCs defined as Lineage-, CD34+, CD38-, CD45RA-, CD90+, and CD49f+ using a FACSAria Fusion flow sorter. Index sort data were collected for each single cell.

2.2 | RNA sequencing and computational analysis

Single-cell libraries were prepared using the G&T-seq protocol²⁰ and the Illumina Nextera XT DNA preparation kit. ERCC spike-in RNA (Ambion) was added before library preparation. Equal quantities of complementary DNA (cDNA) were used from each single cell or cell pool. Pooled libraries were sequenced by 125-bp paired-end on the Illumina HiSeq 4000 instrument using TruSeq reagents (Illumina). Adapters were trimmed <32 bp using TrimGalore! Reads were mapped to the human reference genome (GRCh37) using STAR.²¹

Quality metrics were assessed for each sample. Expression in all cases was normalized to library size and ERCC spike-ins.

For HSCs, low-quality cells were filtered using a previously described support vector machine (SVM) approach²² and HSC cDNA-positive for *GAPDH*. For MKs, filtering of low-quality cells was performed using the same features, but in a five-round training of random forest models.²³ The most highly variable genes were filtered above technical noise using the Scater package²⁴ using previously described methods.²⁵ ERCC spike-ins were used to model the trend in technical variability.²⁶

For both HSC and MK single cells, unsupervised clustering of single cells was performed in a principal component analysis (PCA) with robustness of clusters determined by the Silhouette index.²⁷ Using SC3, cluster marker genes were identified based on their predictive value to separate each cluster from the rest with a *p* value (<0.001) determined using the Wilcoxon signed-rank test.²⁸ Cells were ordered into differentiation trajectories using the Monocle 2 single-cell analysis toolset^{29,30} using five clusters, a maximum of three dimensions, and otherwise default parameters.

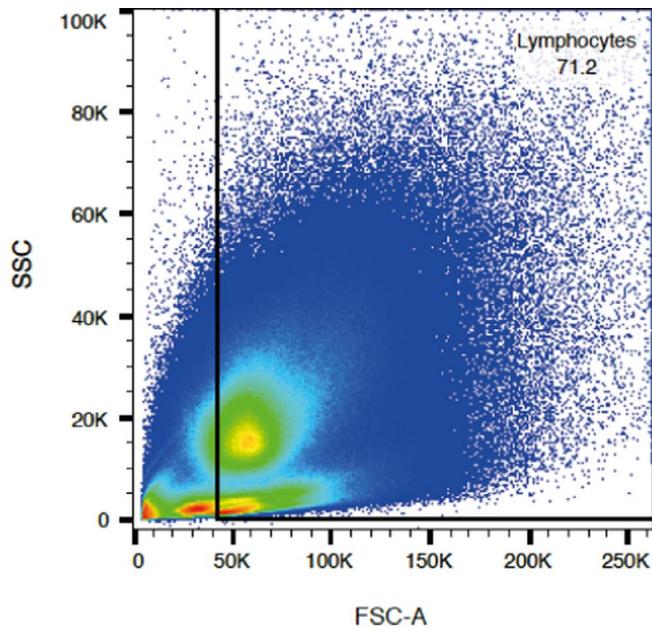
To assess expression of known hematopoietic gene signatures within single-cell clusters, gene signatures for each cell type in the DMAP dataset³¹ were created using the LIMMA package.³² Genes that were considered to be part of the gene signature log₂ fold change of >1 and false discovery rate (FDR) <0.05. DESeq^{23,33} was used to find genes that were differentially expressed between different datasets in single MKs and MK pools with comparisons made between 4 N and 32 N ploidy levels or between disease and non-CAD control. Here, the Wald test was used for hypothesis testing. Differentially expressed genes (significantly upregulated or downregulated) were defined as *p* < .05. Size factors were estimated by scran.³⁴ Gene ontology (GO) analysis was performed using Fidea³⁵ (<http://circe.med.uniroma1.it/fidea>).

FACS data from a previous study by Belluschi et al.³⁶ were used in PCA analysis on scaled data (PCA_B), onto which new samples (BM_HSC) were projected by applying PCA_B scaling target and center. The position was calculated as adot-product of scaled BM_HSC vector and PCA_B rotation. Cells were marked according to differentiated cell type, where cell type perimeter is shown via geom_encircle (<https://CRAN.R-project.org/package=ggalt>) using parameters *s_shape* = 0.6 and *expand* = 0.1.

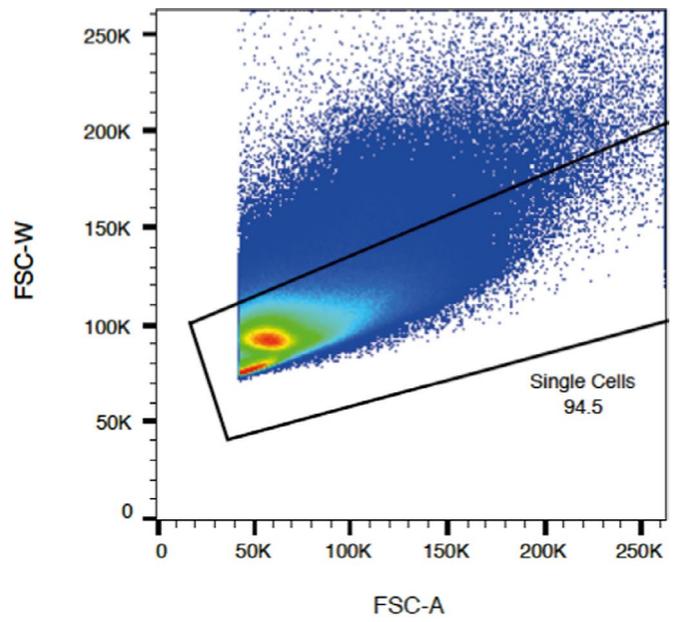
2.3 | Data-sharing statement

All RNA-sequencing (RNA-seq) data have been deposited in the European Genome-phenome Archive under accession number EGAS00001004844.

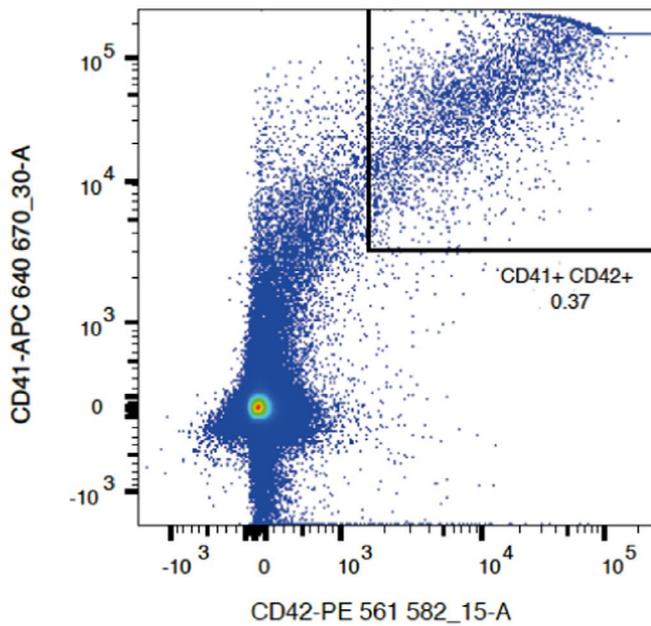
FIGURE 1 Primary human bone marrow MK sorting strategy. Fluorescence-activated cell sorting. Antibody stains used: PECy5 conjugated anti-CD42, APC conjugated anti-CD41, Hoechst 33342. Total bone marrow cells after red cell lysis were sorted as single cells and then identified as CD41+CD42+. Ploidy plot of this population detecting levels of Hoechst 33342 staining showing typical ploidy distribution for human bone marrow MKs. Cells were sorted as 2 N, 4 N, 8 N, 16 N, 32 N, with MKs being defined as 4 N-32 N in case of contamination of other 2 N cells. Number of 20 cell pools for each ploidy class: 2 N: *n* = 3; 4 N: *n* = 6; 8 N: *n* = 7; 16 N: *n* = 8; and 32 N: *n* = 8. MK, megakaryocyte



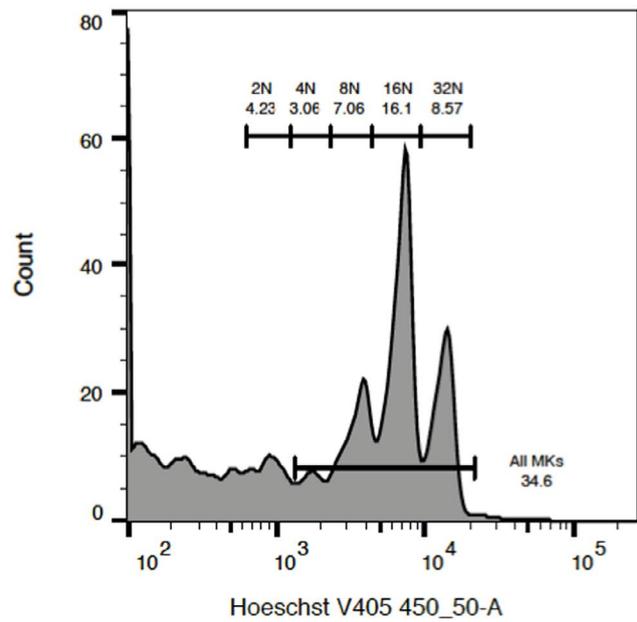
Specimen_001_DFE_002.fcs
 Ungated
 1.87E6
 Lymphocytes : Count : 1.33E6



Specimen_001_DFE_002.fcs
 Lymphocytes
 1.33E6
 Single Cells : Count : 1.26E6



Specimen_001_DFE_002.fcs
 Single Cells
 1.26E6
 CD41+ CD42+ : Count : 4635



Specimen_001_DFE_002.fcs
 CD41+ CD42+
 4635
 2N : Count : 196
 4N : Count : 142
 8N : Count : 327
 16N : Count : 748
 32N : Count : 397

3 | RESULTS

3.1 | Human bone marrow megakaryocyte gene expression varies with ploidy level

To characterize the transcriptome of bone marrow-residing MKs, as well as, the transcriptional changes associated with increased ploidy, we isolated MKs from bone marrow scrapings obtained from individuals. Ploidy distribution analysis by flow cytometry was consistent with previous descriptions of *ex vivo* MKs^{1,16,37} (Figure 1A). We sequenced 32 pools of 20 MKs and 1106 single MKs with ploidy ranging from 2 N to 32 N, isolated from 10 individuals undergoing sternotomy for heart valve replacement (Table S1). Thirty-two pools of 20 MKs were sorted according to ploidy from five individuals and sequenced. The number of pools for each ploidy group were: 2 N: $n = 3$; 4 N: $n = 6$; 8 N: $n = 7$; 16 N: $n = 8$; and 32 N: $n = 8$, with at least one pool from each individual in each ploidy group except 2 N. The 125-bp paired-end sequencing was performed at a depth of approximately 500 million reads/sample. A total of 1106 single MKs were sorted from a further five individuals: 2 N: $n = 220$; 4 N: $n = 158$; 8 N: $n = 200$; 16 N: $n = 272$; and 32 N: $n = 256$. The 125-bp paired-end sequencing was performed at an approximate depth of 50 million reads/sample. Low-quality single cell samples were filtered with a five-round training scheme of random forest models²³ trained on the 20 cell MK pools (Figure S1). A total of 282 single MKs (equally spread between ploidy levels) were taken forward for transcriptome analysis based on quality control measures.

The MK transcriptome was analyzed in two separate ways. First, we inspected the highly expressed genes. Among those, we found enrichment for mitochondrial and cell metabolism genes such as *MT-RNR1*, *MT-RNR2*, and *CYTB* and the gene encoding cyclooxygenase-1 (*PTGS1*), the protein acetylated by aspirin (Table S2), as previously described in platelet transcriptome studies.³⁸ In the 100 highest expressed genes from MK pools sorted by ploidy level, we observed largely a bimodal distribution in expression, 15 genes were mostly expressed in 4 N and 8 N MKs and were enriched GO terms related to platelet function (Table S3–S8). A second group of genes ($n = 40$) were mostly expressed in 16 N and 32 N cells; these were enriched for GO terms associated with translation and protein processing (Figure 2A). Because of these differences, we performed differential gene expression analysis between 4 N and 32 N MK pools, as well as single cells (Tables S9–S12). This identified 944 differentially expressed genes (DEG), 571 were upregulated in 4 N MK pools compared with 32 N (FDR <0.05, 179 DEG with an FDR <0.0001), whereas 373 were upregulated in 32 N cells compared with 4 N (FDR <0.05; 181 DEG with an FDR <0.0001). GO terms enrichment analysis found GO terms associated with platelet degranulation, coagulation, hemostasis, wound healing, and vesicle-mediated transport in the lower ploidy MK transcriptome, whereas, with increasing polyploidization, we found GO

terms related to translational initiation, elongation and termination, protein localization, and cellular protein complex disassembly (Figure 2B and Tables S13, S14). Comparable results were found in the single-cell differential gene expression/GO analysis (Tables S15, S16).

To investigate signals from the bone marrow niche that might contribute to MK maturation *in vivo*, we annotated genes upregulated with accumulating ploidy in terms of their subcellular localization using the Ensembl database³⁹ (Table 1). Of the 373 upregulated genes, 78 contained transmembrane domains and most localized to the cell membrane. Although these have a wide range of functions, a number of them encode for transmembrane proteins, including: the G protein-coupled receptor *PTGER2*, the tetraspanin *CD63*, the tumor necrosis factor (TNF)-alpha receptor *TNFRSF1B* and lysophosphatidic acid receptor *PPAP2A*. *PTGER2*, *TNFRSF1B*, and *P2Y6* encoding for the pyrimidine nucleotide receptor *P2Y6*, genes encoding for cytokine receptors *IL13RA1* and *IL15RA* as well as the complement receptor *CR1* are highly enriched with increasing MK ploidy (FDR <1E-8). Seventy-two of 78 genes have low levels of expression in previously published datasets of *in vitro*-cultured MKs,⁶ where the MKs are usually of ploidy levels 2 N to 4 N (Figure S1C).

3.2 | MK lineage priming in *ex vivo* human bone marrow HSCs

To investigate the origin of MKs and gain insight into megakaryopoiesis *in vivo*, we characterized the HSC transcriptional landscape and early fate commitment events. We sequenced 884 single phenotypic HSCs (CD34+ CD38- CD45RA- CD90+ CD49f+ [Figure S2], the compartment most enriched in long-term repopulating HSCs⁹) isolated from fresh bone marrow harvested from a further five individuals undergoing sternotomy for heart valve replacement (Table S17). We demonstrated that these phenotypic HSCs do contain cells able to confer long-term hematopoietic reconstitution (Supplementary Methods and Table S19). From the sequenced HSCs, raw reads were filtered to exclude cells that gave low-quality libraries (Figure S3) with a SVM learning method trained on a subset of samples made from cDNA positive for *GAPDH*, as measured by quantitative reverse transcriptase polymerase chain reaction (Table S18). Nineteen single HSCs were taken forward for transcriptome analysis based on quality control measures.

To identify subpopulations of cells within the HSC compartment, we used unsupervised hierarchical clustering of the cells' Pearson correlation coefficients in the PCA space to generate clusters (Figure 3A) where total silhouette score was used to select the number of clusters tested (Figure S4A). The robustness of our clustering strategy was confirmed by inspecting the first four principal components (Figure S4B,C). The cells within the HSC population formed five distinct clusters. Marker genes for each cluster were then identified based on their predictive value to separate that cluster from

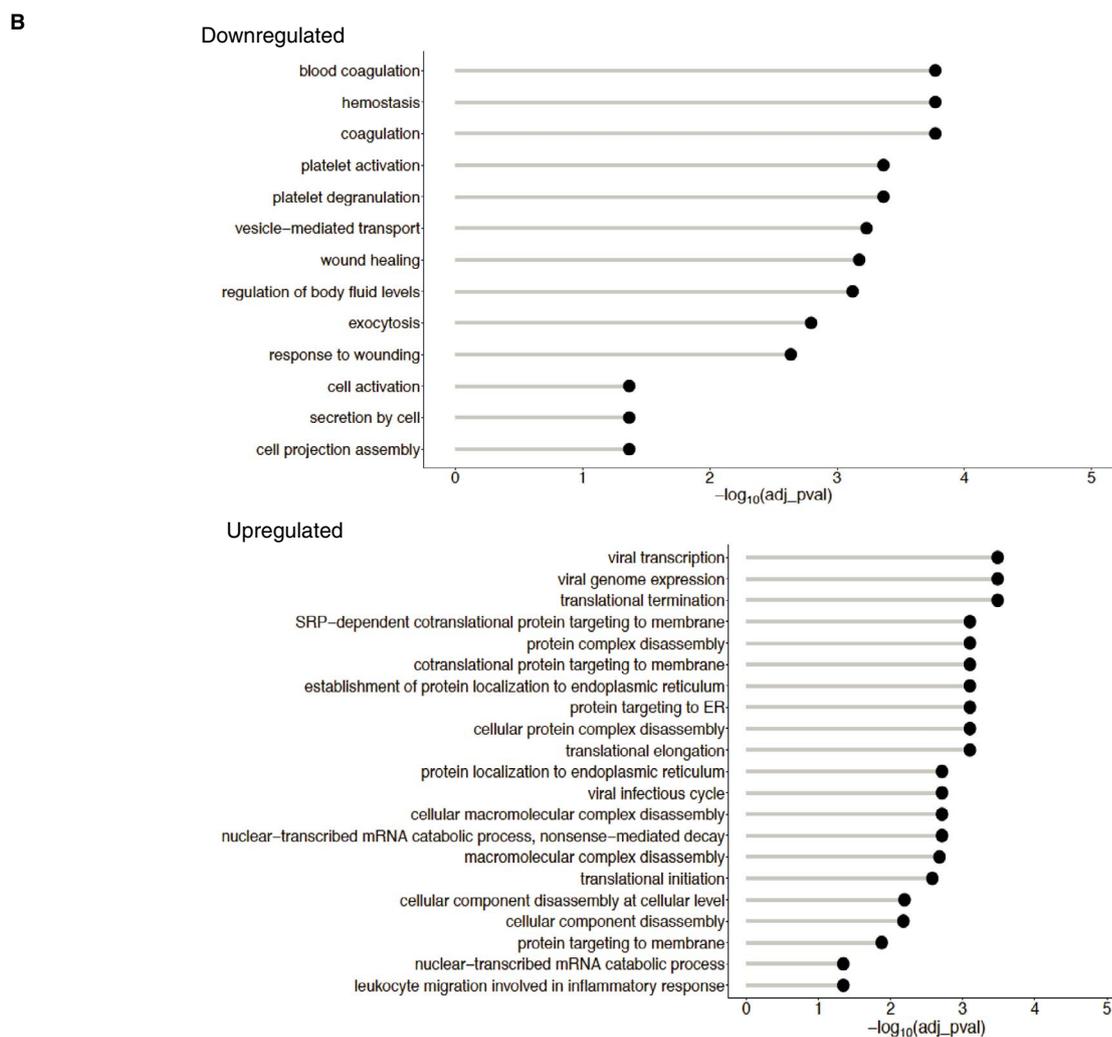
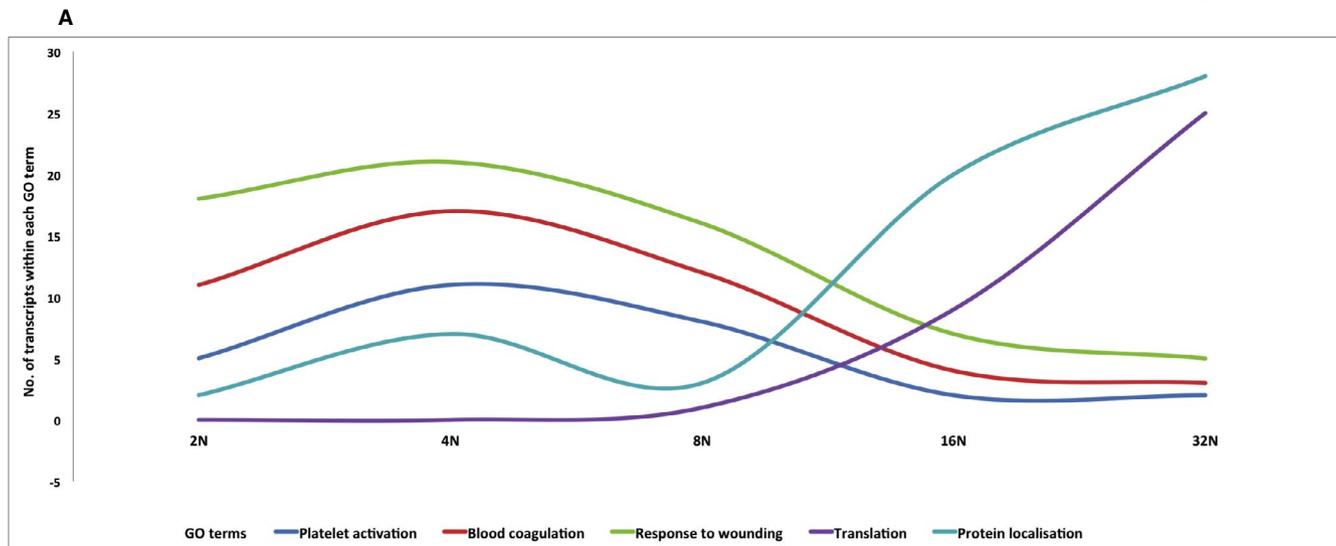


FIGURE 2 Development and polyploidization of bone marrow residing MKs. (A) Number of genes expressed related to specific GO terms at each MK ploidy level; data shown based on 100 most abundantly expressed genes. GO terms: platelet activation, blood coagulation, response to wounding, translation, protein localization. The full GO analysis may be found in Table S10–S14. (B) Two distinct transcriptional states in MK differentiation. Upper panel: Overrepresented GO terms in downregulated and upregulated genes with increasing ploidy, data shown are based on differential expression analysis between 32 N and 4 N MK 20 cell pools. Full GO analysis may be found in Tables S13,S14. GO, gene ontology; MK, megakaryocyte

TABLE 1 MK transcripts upregulated with ploidy: localization within the cell.

Localization	Significantly Upregulated Genes in 32 N MK Pools Compared with 4 N
Cell membrane	UNC80, ENPEP, PTGER2, TMPPSS7, ASGR2, KCNK5, CACNA2D1, P2RY6, NTRK3, SLC52A3, ADCY5, TNFRSF1B, OPRM1, LRP6, LRP1B, EPCAM, TRPM2, ARHGAP31, CR1, TMCO4, HEPHL1, CACNA1B, SECTM1, IL15RA, PPAP2A, SLC26A3, IGHV1-46, DPEP2, MYO1B, IL13RA1, ADAM29, JAG1, TTYH1, TMEM144, SCRIB, ITGA11, ATRNL1, PDE6G, MS4A1, GRID2IP, RRAS2, IGLV3-25, SLC05A1, RAET1E, IFI6, RAB33A, KCNJ6, MUC16, NFAM1, TREM1, RELT, GRAMD1C, SLC4A7, MGAM, DLGAP2, MSR1, ERBB4, CLECL1, SRGAP2, LRRC37B, SEMA6A, FPR2, SCN3A, PIEZO2, PTPRC, ABCA12, ZAN, CACNB4, PLIN2, ANXA4, RBSN, TMEM176A, CD63, SLC13A5, SLC36A4, LILRB4, ADAM8, FCHO1, OTOA
Cytoplasm	PLA2G16, PDE1C, ARHGEF10L, CLIP4, WNK2, NLRP13, CTH, USP13, KRT23, MYH4, GDPGP1, PRG3, TTLL4, GYS1, GSDMA, NAF1, MYO7A, UNC13C, SPATA21, CAMSAP1, PLEKHS1, ARHGEF26, PRKAA2, OAS3, GAS2, ASZ1, RGS7, KIF20A, APAF1, KLHL22, GALK1, MYO10, SAC3D1, SCLT1, FRY, DNAH8, CHORDC1, SPHK2, RPL38, CNKSR2, TANC2, CDC45, SACS, SEPT14, CCNG2, HS3ST1, PTPN13, DIAPH3, FAM49B, RPLP2, EPB41L4A, GPALPP1, RPS21, PIWIL3, RPLP1, SLFN13, HOOK2, RPL27, PSMB7, UBA6, SWAP70, FRMPD1, PSMA6, DOCK3, PLBD1, TTC6, HINT3, FEZ1, RPS12, PSMB9, CHODL, SLC24A4, RPL27A, RPS17, MTHFD1, RPS26
Internal membrane	GALNTL6, SNX24, CCZ1B, CFTR, YIF1A, CCDC51, POMT2, LDAH, ACAP1, STEAP2, TMEM38B, NCS1, PIPOX, ST6GALNAC1, PIGK, RETSAT, FAM21C, UBIAD1, SMPD4, SNX14, EMC10, RYR1, GCNT1, VPS18, TRAPP8, AGPAT9, SNX13
Mitochondria	GLDC, ACSM2B, MAAA, GRAMD4, OMA1, SFXN4, TFB1 M, MCCC1, GOT2, COX7B, IDH3B, ATP5G2, ABAT, NDUFB11, COX7C, MARC1, SURF1, NDUFC2, CENPBD1P1
Nucleus	RORB, ZNF692, NOC4L, DXO, ESRP1, C2orf83, C16orf86, NUF2, NLE1, ZNF234, KIAA0556, ZNF763, ASF1B, FIGN, SH2D6, ZNF347, ZCCHC4, NUP35, NOVA1, POLR3H, ZNF343, PELO, CCDC155, TATDN1, ZNF780A, PWP2, IQCA1, CHTF18, KIAA0101, AIM1, CDCA5, RHNO1, ZNF174, WDHD1, ZNF543, WDR12, WDFY3, PRIM2, COMMD6, ZMYM5, LRRIQ1, FBXO48, DNASE1L3, NCAPD2, PPARG, ZWILCH, ATAD2, MAP2 K6, ZNF154, LYRM1, BRAT1
Secreted	COCH, APOC1, TIMP2, TLL2, CDNF, AGR3, GDF11, OVCH1, ADAMT51, CES4A, CPAMD8, HSD11B1L, LAMA3, S100A8, OSCAR, VCAN, CFD, S100A9

Note: Using the Ensembl database the genes upregulated with ploidy were annotated in terms of their localization within the cell. Seventy-eight contained transmembrane domains and localized to the cell membrane, 77 were annotated as encoding cytoplasmic proteins, 57 nuclear proteins, 27 internal membrane proteins, 18 mitochondrial proteins, and 18 secreted proteins and for 104 transcripts there was no localization information available. MK, megakaryocyte

the remaining cells (Wilcoxon rank test, $p < 0.001$). The expression of the top 20 genes characterizing each cluster shows distinct expression patterns independently of donor; of note, cluster 3 had no significant marker genes (Figure 3B and Table S20 for the full list of genes). Many of the marker genes for clusters 1 and 4 were found to be highly expressed in previously reported MK datasets.⁶ Cluster 1 cells were marked by the expression of genes known to be highly expressed in all MK precursors (HSC, multipotent progenitors, common myeloid progenitor, MK-erythroid precursor) such as *PRKACB*, *NRIP1*, *PARP1*, *HEMGN*, as well as *ANGPT1* and *IL1b*, which encode proteins directly involved in platelet function. By contrast, cluster 4 cells were characterized only by genes encoding proteins directly involved in platelet function such as: *TUBB1*, *TUBA4A*, *F13A1*, *CCL5*, *PTCRA*, and *GRAP2* (Table S20).

Cluster 1 transcriptome appears to have an ontological bias toward translation, ribosomal function and protein localization, cellular metabolism, coagulation, and hemostasis. Cluster 2 marker genes appear to have an ontological bias toward angiogenesis, endothelial cell function, and wound healing. The overrepresented GO categories for cluster 4 are related to negative regulation of cell death/apoptosis, cell-cycle, protein folding, and exocytosis, whereas the cluster 5 gene signature showed overrepresentation of categories related to immunity and leukocyte function. The full GO analysis is found in Tables S21–S24, overrepresented GO categories identified using a significance cutoff of $p < .05$.

3.3 | Developmental trajectories of early MK lineage priming in HSCs

To investigate if HSCs exhibited lineage priming potential, we performed trajectory analysis, independent of the previous clustering, using Monocle2.³⁰ We obtained potential developmental trajectories and five clusters, very similar to the previous ones, were found. This analysis showed two distinct branching points, with clusters 1 and 4 stemming from the same developmental trajectory (Figure 4A). To determine whether these trajectories could delineate early fate priming in HSC, we derived gene signatures from the highly purified hematopoietic populations in a previously published study³¹ and visualized these signatures onto the trajectory plot. The signature for MK was found enriched in cells localized around branching point 2 and belonging to clusters 1 and 4 (Figure 4B), whereas the gene expression signature for the megakaryocyte erythroid progenitor (MEP) was found enriched only in cells belonging to cluster 1 after the branching point 2 (Figure 4C). These signatures are largely not overlapping and no single gene appears to drive enrichment (Figure S5).

To investigate where the mature MKs transcriptionally fell in terms of differentiation, we added 188 MK pools from 20 individuals undergoing median sternotomy for cardiac surgery to the HSC trajectory analysis discussed previously using Monocle2.³⁰ The addition of mature MK transcriptional signatures to the HSC developmental trajectories showed that these are most similar to and branch

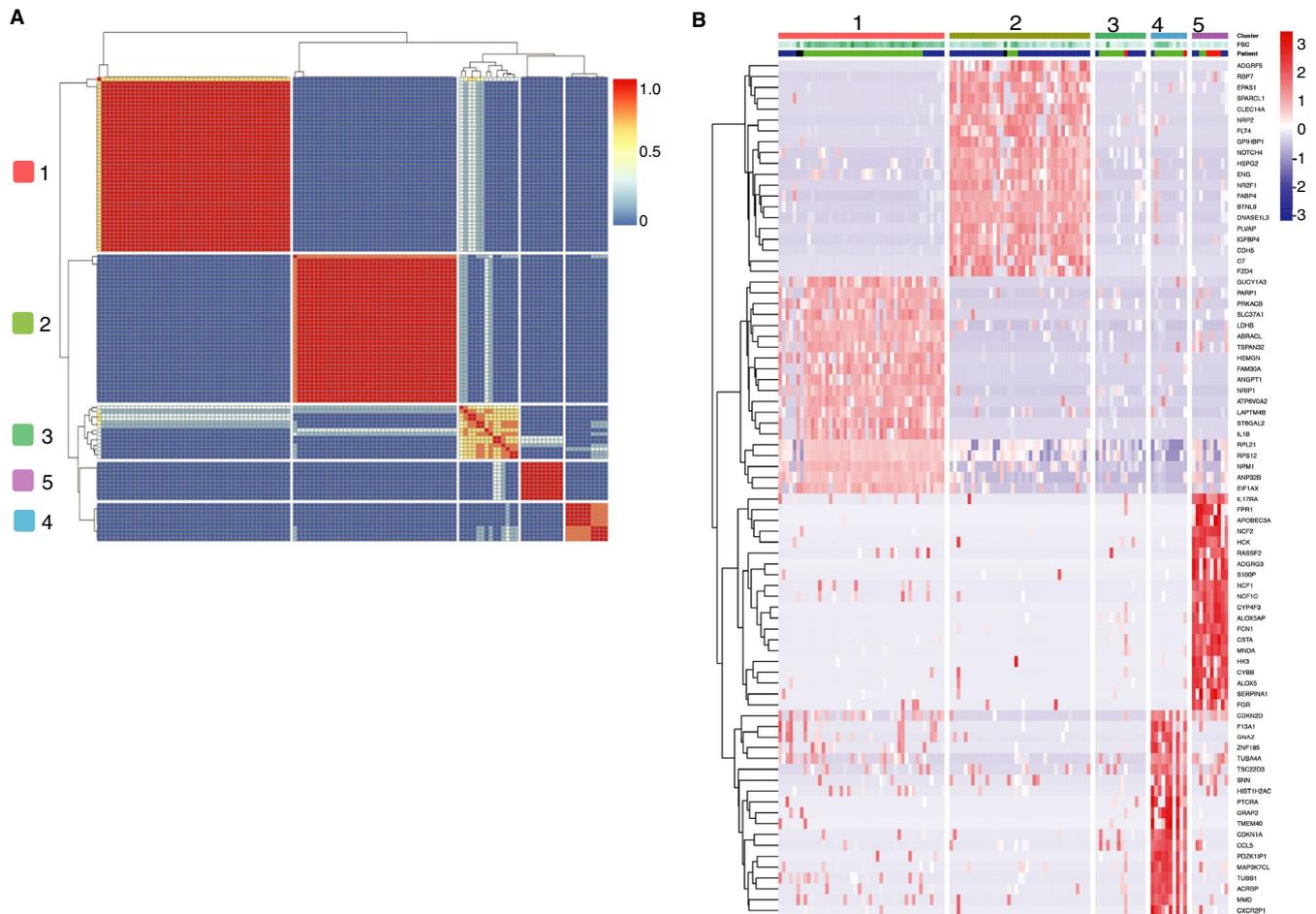


FIGURE 3 Unsupervised clustering of early lineage priming in single HSCs. (A) Pearson's correlation map of HSCs in unsupervised clusters by distance within the PCA space; dendrograms are formed by hierarchical clustering on the Euclidean distances between cells (B) Heatmap of mean-centered normalized and corrected log-expression values for the top 20 marker genes for each cluster. Dendrograms are formed by hierarchical clustering on the Euclidean distances between genes (row). Column colors represent the cluster to which each cell is assigned. HSC, hematopoietic stem cell; PCA, principal component analysis

from HSC clusters 1 and 4 with the MKs forming their own trajectory in the dimensionality reduced space (Figure 4D).

3.4 | Functional potential of MK primed clusters

We then retrospectively analyzed FACS index data and found that the cells that belong to different clusters based on transcriptome data also show differences based on cell surface markers fluorescence intensity (Figure 5A). Clusters 1 and 4 could be distinguished on the basis of FSC-A and CD34 with cluster 1 having CD34^{Lo} and FSC-A^{Hi}, whereas cluster 4 has CD34^{Hi} and FSC-A^{Hi}, with FSC-A^{Hi} enriched for clusters 1 and 4 (Figure 5B). Analysis of top PCA vector loadings showed that the cell surface expression of CD34, FSC-A, and CD49f drive most of the variance observed between HSCs on PC2 and PC3 (Figure 6A).

To characterize the lineage differentiation potential of clusters 1 and 4, we compared our HSC index FACS profiles with those of Belluschi et al,³⁶ who previously determined the lineage potential

of index sorted single HSCs from human cord blood using an optimized single-cell colony-forming cell assay, which supported myeloid/erythroid and MK lineage differentiation. We therefore transposed index sorting surface marker data from Belluschi et al onto Figure 6A (Figure 6B). Although only a few cells from Belluschi et al differentiated to form MK colonies (violet triangle, Figure 6B), these few cells shared cell surface marker expression with HSC clusters 1 and 4 from our work (pink triangle and blue triangle, respectively).

3.5 | MK signature in myocardial infarction

Myocardial infarction has been suggested as a model for stress or accelerated thrombopoiesis accompanied by elevated mean platelet volume and reticulated platelet count.¹⁹ We compared the transcriptomes of MKs obtained from seven individuals with severe coronary disease and recent myocardial infarction in the last 6 months (undergoing sternotomy for coronary artery bypass grafting) with

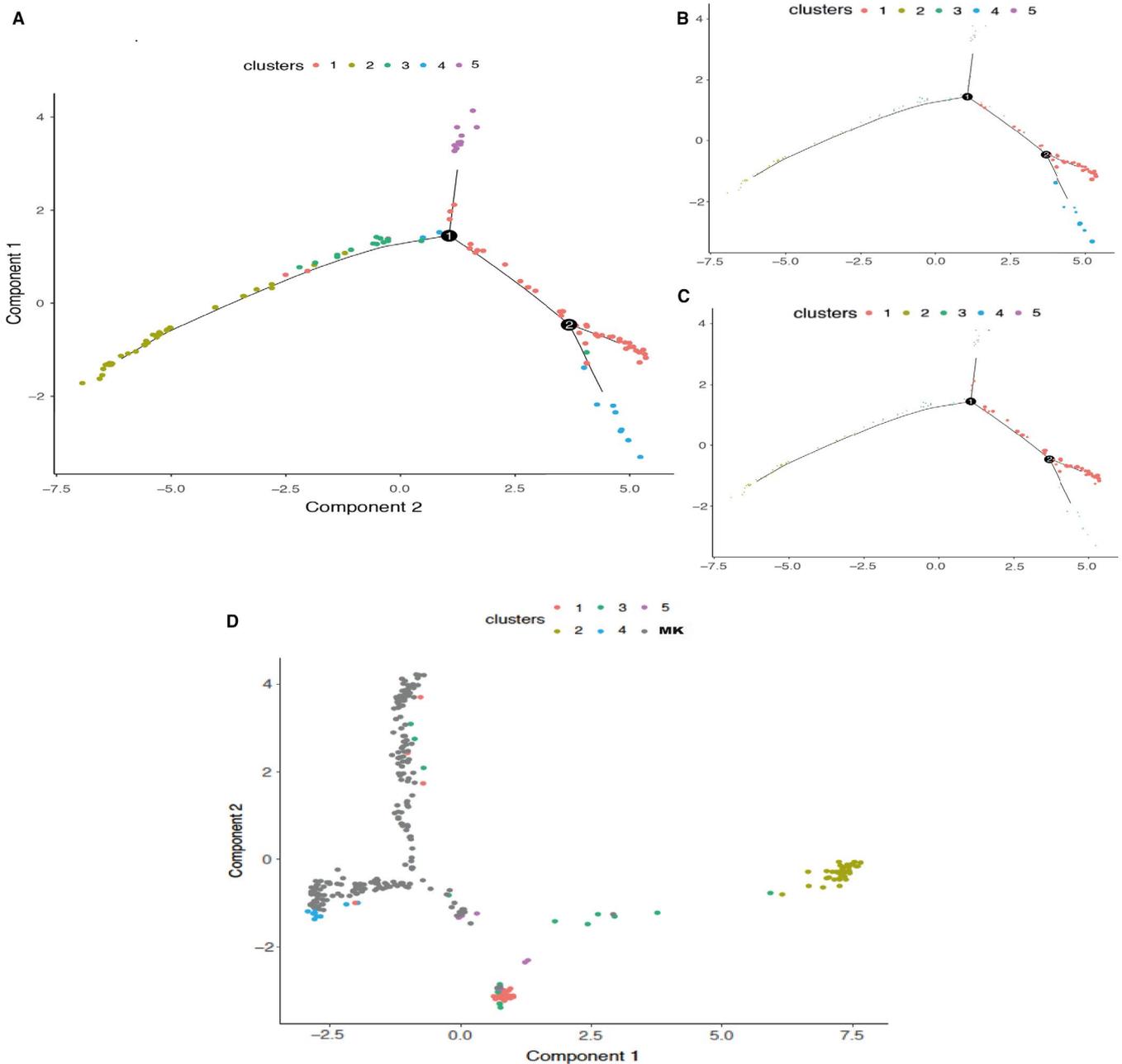


FIGURE 4 Developmental trajectories of early lineage priming in single HSCs. (A) Ordering single-cell differentiation between cell clusters using Monocle 2. Individual cells are connected by a minimum spanning tree with branch points (thin lines), representing the differentiation trajectory and clusters are differentiated by color. (B) Single-cell differentiation trajectory showing expression of MK gene signature from DMAP dataset31. Expression of gene signature is indicated by the size of each single cell. (C) Single-cell differentiation trajectory showing expression of MEP gene signature from DMAP dataset31. Expression of gene signature is indicated by the size of each single cell. (D) Ordering differentiation trajectories between HSC cell clusters using Monocle 2 with the addition of the MK 20- to 100-cell pools. HSC clusters and MKs are differentiated by color. HSC, hematopoietic stem cell; MEP, megakaryocyte erythroid progenitor; MK, megakaryocyte

those eight individuals from a non-CAD control group (undergoing sternotomy for heart valve replacement). Baseline characteristics including demographics, platelet parameters, platelet activation, MK numbers, and ploidy distribution are shown in Table S25. Essentially, the disease and non-CAD control groups are similar in demographics, with an increased incidence of statin use and antiplatelet use in the myocardial infarction group. There is no significant difference in

mean platelet volume, reticulated platelet count, or MK ploidy between the two groups. When a larger clinical cohort was assessed for platelet parameters and MK FACs characteristics only without transcriptome (baseline characteristics in Table S26 [myocardial infarction $n = 19$, non-CAD controls $n = 27$]), the same pattern was revealed; however, there was an increase in platelet mass in the infarct group with increased baseline expression of P-selectin; there

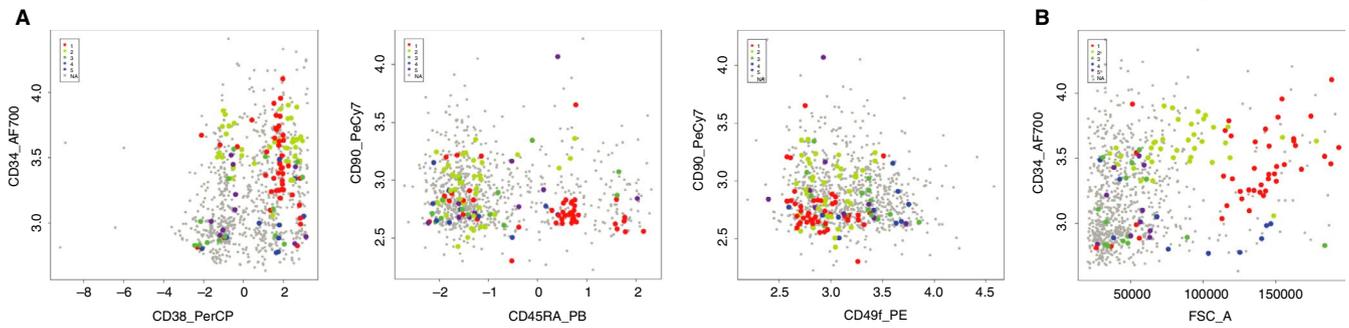


FIGURE 5 Index sorting of single HSCs. (A) Retrospective analysis of FACS sorting strategy for each individual HSC sorted. Left panel: CD34 vs CD38; middle panel: CD90 vs CD45RA; right panel: CD90 vs CD49 f. Individual surface marker expression was normalized. Cluster 1: red; cluster 2: yellow; cluster 3: green; cluster 4: blue; cluster 5: magenta; cells filtered because of poor quality: gray. (B) Retrospective analysis of FACS index data: FSC-A vs CD34 for each HSC sorted. Cluster 1 (red): FSC-A^{Hi} CD34^{Hi}; cluster 4 (blue): FSC-A^{Hi} CD34^{Lo}; cluster 2 (yellow): FSC-A^{Lo} CD34^{Hi}. Individual surface marker expression was normalized. Cluster 1: red; cluster 2: yellow; cluster 3: green; cluster 4: blue; cluster 5: magenta; cells filtered because of poor quality: gray. FACS, fluorescence-activated cell sorting; HSC, hematopoietic stem cell

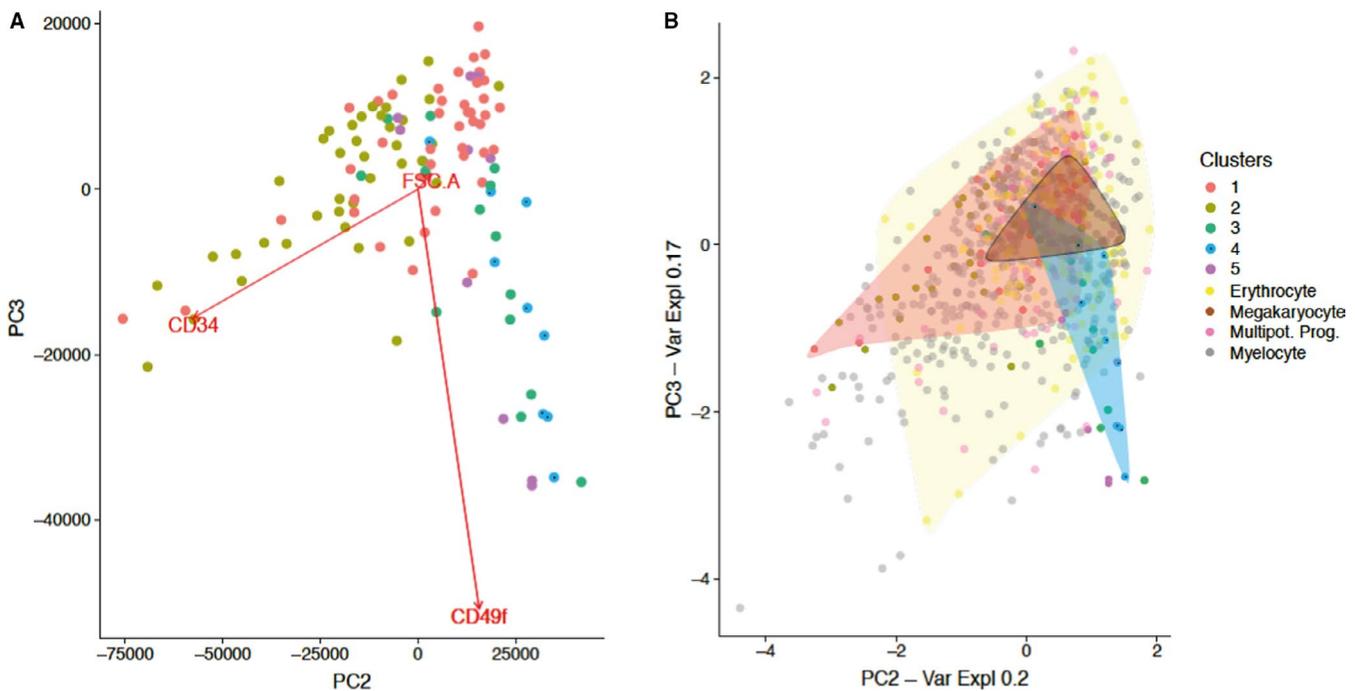


FIGURE 6 Prospective identification of HSC clusters. (A) Principal component analysis (PC2/PC3) of FACS surface marker expression for each individual HSC sorted, including: FSC-A, SSC-A, Lin, CD34, CD38, CD45RA, CD90, CD49 f. Vector loading for FSC-A, CD34, and CD49f are shown. Cluster 1: red; cluster 2: olive; cluster 3: green; cluster 4: blue; cluster 5: magenta. (B) Projection of FACS surface marker expression of HSCs characterized by differentiation assay and their lineage output from Belluschi et al³⁶ onto Figure 6A. As in Figure 6A: cluster 1: red; cluster 2: olive; cluster 3: green; cluster 4: blue; cluster 5: magenta. Differentiation assay outputs: erythrocyte: yellow; MK: violet; multipotent progenitor: pink; myelocyte: gray. Shaded areas: red: cluster 1; blue: cluster 4; yellow: erythrocyte output; violet: MK. The violet triangle in the middle of the figure shows the HSCs from Belluschi et al that formed MKs. The blue shaded area encompasses the position of cells in cluster 4; the pink shaded area encompasses the position of the cells in cluster 1. FACS, fluorescence-activated cell sorting; HSC, hematopoietic stem cell; MK, megakaryocyte

were again no differences in the MK numbers and ploidy distribution between groups (Figure S6).

We sequenced a total of 156 pools of 50 to 100 MKs (irrespective of ploidy); 101 pools from seven individuals with severe coronary disease and recent myocardial infarction and 55 pools from eight non-CAD controls. Differential gene expression analysis revealed 139 upregulated (FDR <0.05; 21 DEG with an FDR <0.00010)

and 679 downregulated (FDR <0.05; 62 DEG with an FDR <0.0001) genes in MKs from patients with severe coronary disease and recent myocardial infarction (Tables S27, S28). A number of upregulated genes were directly related to platelet activation and proteins secreted by the alpha granule including the neutrophil chemoattractant CXCL7 (PPBP), THBS1, and RAP1B as well as the cell surface glutamate receptor GRIA1 (Table 2).

Localization	Upregulated Genes in MK Pools from Myocardial Infarction Compared with Non-CAD Control
Cell membrane	BCAM, TMEFF2, GNG11, DUOX2, MARVELD2, NCMAP, NLGN4Y, RAP1B, COLCA1, TSPAN10, GRIA1, MOG, FAT3, SGCD, HSD17B2, RND1, NFASC, KCNIP4
Cytoplasm	CFAP74, SAMD4A, RPE65, TRIM2, EIF4ENIF1, NME2, MTMR7, PMP2, DPF1, TUB, PRKCI, TEX14, FUT6, CABP1, SPIRE2, MMACHC
Internal membrane	FITM2, LAPT4B, TMEM130, ARMC10, THBS1, ELOVL7
Mitochondria	ECHDC3, KIAA0391, COQ3, TOMM70A
Nucleus	TIAF1, ZNF732, CASZ1, CHEK1, DZIP1, RAD51B, SUV420H2, HIST1H2BJ
Secreted	MTRNR2L12, HMCN2, PPBP, PRELP, OLFML2A, C7, GREM1, SPX, IGHG4

TABLE 2 MK transcripts upregulated in myocardial infarction: localization within the cell

Note: Using the Ensembl database the genes upregulated with ploidy were annotated in terms of their localization within the cell. Eighteen contained transmembrane domains and localized to the cell membrane, 16 were annotated as encoding cytoplasmic proteins, 8 nuclear proteins, 6 internal membrane proteins, 4 mitochondrial proteins, and 9 secreted proteins and for 78 transcripts there was no localization information available. MK, megakaryocyte

4 | DISCUSSION

Here, we used full-transcript single-cell and low-input RNA sequencing to characterize the stages of human bone marrow MK polyploidization and chart their transcriptional journey from HSC early lineage priming, where we identified two separate MK primed clusters.

Our work presents the first interrogation of the transcriptional landscape of human bone marrow MKs. The knowledge of the MK transcriptome until now being based largely on *in vitro*-derived MKs from CD34+ cells.^{2-6,40-43} There has, however, been some recent work on primary bone marrow MK in mouse and primary MK progenitors but not MKs in human.^{44,45} We charted transcriptional changes in human bone marrow MKs through different stages of polyploidization. Microarray data of increasing ploidy in *in vitro*-derived MKs previously showed an overall pattern of upregulation of genes involved with platelet, coagulation, and hemostatic pathways and a downregulation of cell cycle-associated genes.⁵ Our data instead provide a model of MK development and polyploidization whereby two distinct transcriptional states exist for low- and high-ploidy MKs and the cells transition from one to the other with successive endomitotic replication. With increasing ploidy, we observed a downregulation of genetic programs related to platelet functionality, although these genes are still transcribed and remain expressed within the higher ploidy MKs. Additionally, with increasing ploidy level, we observed a marked upregulation of genes encoding ribosomal subunits, those associated with protein translation, and protein localization. These findings suggest that at lower ploidy levels, many of the proteins that are transcribed relate to platelet function and platelet and MK cell-surface expression. However, at higher ploidy levels, energy is redirected to messenger RNA translation and appropriate localization of the newly generated proteins into alpha granules, dense granules, and other vesicles, readying the cell for thrombopoiesis. Bone marrow MKs' high-energy requirement, particularly those of higher ploidy, is the likely explanation for the high

levels of genes involved in cellular metabolism since the production of ribosomes consumes large amounts of energy.⁴⁶ Our data are in keeping with previously observed changes in MK ultrastructure during *in vivo* maturation. Although stage 2 MKs (low ploidy) have free ribosomes, smooth endoplasmic reticulum, and few granules, suggesting lower levels of translation, stage 3 MKs (high ploidy) are characterized by rough endoplasmic reticulum, formed alpha and dense granules, increases in heterochromatin and smaller nucleoli, indicating a reduction of transcription.⁴⁷⁻⁴⁹ The differences observed with previous studies are likely ascribed to the fact that these were characterizing the final steps of MK differentiation rather than MK maturation.

The identification of genes upregulated with increasing ploidy also has implications for improving our understanding of platelet release and potential therapeutic implication for *in vitro* platelet production. Currently, the number of platelets produced for each *in vitro*-generated MK is ~3 log orders lower than what observed *in vivo*, which severely hinder platelet production for clinical transfusion using induced pluripotent stem cell-derived MKs.⁵⁰ This is likely a direct result of the differences between MKs produced in culture from pluripotent stem cells and MKs in the bone marrow niche in terms of ploidy level, cytoplasmic maturation, or extracellular signals. A number of genes upregulated with increasing ploidy were identified to encode transmembrane receptors (PTGER2, TNFRSF1B, IL13RA1, and PPAP2A), which upon binding to their ligands, might modulate functional effects such as MK chemotactic migration from the osteoclastic niche to the vascular sinusoidal space or indeed initiate platelet production. These may represent important novel drivers of MK maturation and platelet release. PTGER2, the G-protein coupled receptor for prostaglandin E2 known to be present on the platelet membrane, is an important regulator of HSC expansion and function in the bone marrow niche⁵¹ and has been shown to specifically promote MK lineage recovery after radiation injury in mouse.⁵² Binding of TNFRSF1B

by TNF-A, also present on the platelet membrane, suppresses apoptosis and has been shown to induce megakaryopoiesis in hematopoietic progenitors, which could, in part, account for increased platelet mass in inflammation.¹⁷ IL13RA1 mediates the effects of IL13, which significantly increase MK colony formation *in vitro*⁵³ and in murine models.⁵⁴ PPAP2A is an integral membrane glycoprotein that degrades lysophosphatidic acid by dephosphorylation. Lysophosphatidic acid gradient has been proposed in regulating localization within the osteoclastic niche and has been shown to inhibit megakaryopoiesis *in vitro*.⁵⁵ Future work on these transmembrane receptors and their ligands, to investigate their role in MK maturation, would be imperative in increasing the efficiency of *in vitro* platelet production from induced pluripotent stem cell-derived MKs.

To investigate the origin of the MK, we identified five transcriptionally distinct HSC subsets in human bone marrow supported by two independent clustering methods. Two of these subsets (clusters 1 and 4) displayed transcriptional priming toward the MK lineage; when overlaid onto data from single-cell assays,³⁶ they showed a functional MK lineage bias. Cells in these two clusters expressed genes that have been shown to be expressed in the MK-platelet lineage.^{3,6,31,38} The high proportion of HSCs displaying priming to the MK lineage is in keeping with existing evidence indicating that this is the first lineage bifurcation in the multipotent HSC compartment⁷⁻¹¹ and the role of MKs in HSC regulation within the stem cell niche.⁵⁶⁻⁵⁸ Plotting these cells in pseudotime differentiation trajectories showed that clusters 1 and 4 in fact branch from the same trajectory. Comparison with other datasets^{6,31} demonstrated that, although cells in cluster 1 were enriched for genes involved in HSC function, megakaryopoiesis and genes commonly expressed between MK and MEP populations, cells in cluster 4, in contrast, were enriched for genes specific to MKs with a number of marker genes for this cluster directly related to platelet function. We recreated a differentiation trajectory combining the HSCs and MKs transcriptomic data together and showed that human bone marrow MKs branch directly from phenotypic HSCs, specifically cells in clusters 1 and 4 supporting the notion of transcriptional priming of these clusters toward the MK lineage. Because HSCs in cluster 1 display gene expression common to intermediate progenitors as well as MKs, we hypothesize that these may be primed for MK differentiation via a classical differentiation pathway, whereas HSCs in cluster 4 might be those that exhibit direct MK differentiation bypassing the classic MEP, which has been shown in several studies.⁷⁻¹¹ It is also possible that these two HSC clusters may give rise to two phenotypically different types of mature MKs.

We also compared MK transcriptional signatures in myocardial infarction as a model for stress thrombopoiesis with non-CAD controls. Despite finding no significant differences in platelet parameters and MK ploidy in myocardial infarction, our data showed that a number of genes upregulated in MKs from individuals with severe coronary disease and myocardial infarction compared with non-CAD controls were related to thrombus formation, such as *PPBP*, the gene

encoding chemokine CXCL7, which plays a critical role in leukocyte migration through thrombi by binding its receptor CXCR1/2 on neutrophils and other leukocytes.^{59,60} Similar higher levels of thrombospondin 1 (*THBS1*) have been found to be associated with peripheral artery disease and *RAP1B* that encodes a small GTP-binding protein involved in outside-in signalling via the collagen receptor integrin alpha2-beta1 (CD49b/29) on platelets; it also plays an important role in the cross-talk between alpha2-beta1 and integrin alphaIIb/beta3 (CD41/61), the receptor for fibrinogen on MKs and platelets.⁶¹ Hence, *RAP1B* has been suggested as a potential therapeutic target for a novel class of platelet inhibitors in myocardial infarction.⁶² There is now compelling evidence for enhanced megakaryopoiesis as a pathogenic driver for atherosclerosis and myocardial infarction.^{63,64} The clinical trial CANTOS⁶⁵ demonstrated improved cardiovascular outcomes in patients with myocardial infarction with Canakinumab, a monoclonal antibody targeting IL1B, a known driver of megakaryopoiesis both *in vitro* and *in vivo*.⁶⁶ We found that the gene encoding the AMPA glutamate receptor, *GRIA1*, was upregulated in MKs in individuals with myocardial infarction. As glutamate serum levels are increased in thrombosis⁶⁷ and interruption of glutamate binding to the NMDA glutamate receptor in megakaryocyte cell lines resulted in impaired megakaryopoiesis⁶⁸; this observation raises the possibility of a positive feedback mechanism of glutamate signaling leading to increased platelet production perpetuating a prothrombotic state. Therefore, our data also support a pathological role of stress thrombopoiesis in acute coronary thrombosis.

Limitations of this transcriptional study should be acknowledged, including the baseline characteristics of the patient populations. Advanced age leading to clonal hematopoiesis, male predominance particularly in the myocardial infarction group, and statin and anti-platelet use can all lead to changes in gene expression.

Taken together, our work charts the transcriptional journey from MK lineage commitment through to the mature polyploid MK, describes 2 HSC subpopulations exhibiting MK lineage priming and direct branching of MK differentiation from these HSC subpopulations bypassing intermediate progenitors such as the MEP. Furthermore, the MK transcriptional changes in the setting of myocardial infarction are then explored. Ultimately, better understanding of MK lineage commitment, differentiation from the HSC, polyploidization, and platelet release are key to deciphering pathological modulation leading to thrombosis and bleeding disorders.

ACKNOWLEDGMENTS

The authors thank the Barts Heart Centre research and surgical teams for their support and help in this project; the Cambridge NIHR BRC Cell Phenotyping Hub, particularly Anna Petrunkina-Harrison and Esther Perez for their flow cytometry advice; Elisabeth Cramer-Borde, Université Paris Descartes for her advice on with isolating bone marrow MKs; and The Lundbeck Foundation [R193-2015-1611, R182-2014-3881 to F.O.B.]

CONFLICT OF INTEREST

There are no conflicts of interest to declare.

AUTHOR CONTRIBUTIONS

Fizzah A. Choudry designed and performed experiments and wrote the manuscript; Frederik O. Bagger performed bioinformatic analysis and wrote the manuscript. Iain C. Macaulay supervised experiments and parts of the study. Samantha Farrow, Frances Burden, Carly Kempster, Harriet McKinney, and Kate Downes carried out experiments. Lars R. Olsen and Ni Huang provided bioinformatic analysis and edited the manuscript. Anthony Mathur, John F. Martin, and Rakesh Uppal supervised the clinical aspects of the study. Thierry Voet, Willem H. Ouwehand, Elisa Laurenti, and Sarah Teichmann supervised parts of the study and edited the manuscript. Mattia Frontini designed and supervised experiments and wrote the manuscript.

ORCID

Fizzah A. Choudry  <https://orcid.org/0000-0002-9475-5257>

Mattia Frontini  <https://orcid.org/0000-0001-8074-6299>

REFERENCES

- Tomer A, Harker LA, Burstein SA. Flow cytometric analysis of normal human megakaryocytes. *Blood*. 1988;71:1244-1252.
- Bluteau O, Langlois T, Rivera-Munoz P, et al. Developmental changes in human megakaryopoiesis. *J. Thromb. Haemost.* 2013;11(9):1730-1741.
- Shim MH, Hoover A, Blake N, et al. Gene expression profile of primary human CD34+CD38lo cells differentiating along the megakaryocyte lineage. *Exp Hematol*. 2004;32:638-648.
- Macaulay IC, Tijssen MR, Thijssen-Timmer DC, et al. Comparative gene expression profiling of in vitro differentiated megakaryocytes and erythroblasts identifies novel activatory and inhibitory platelet membrane proteins. *Blood*. 2007;109(8):3260-3269.
- Raslova H, Kauffmann A, Sekkaï D, et al. Interrelation between polyploidization and megakaryocyte differentiation: a gene profiling approach. *Blood*. 2007;109(8):3225-3234.
- Chen L, Kostadima M, Martens JHA, et al. Transcriptional diversity during lineage commitment of human blood progenitors. *Science*. 2014;345(6204):e1251033.
- Notta F, Zandi S, Takayama N, et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science*. 2016;351(6269):2116.
- Velten L, Haas SF, Raffel S, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol*. 2017;19(4):271-281.
- Yang J, Tanaka Y, Seay M, et al. Single cell transcriptomics reveals unanticipated features of early hematopoietic precursors. *Nucleic Acids Res*. 2017;45(3):1281-1296.
- Popescu DM, Botting RA, Stephenson E, et al. Decoding human fetal liver haematopoiesis. *Nature*. 2019;574:365-371.
- Pellin D, Loperfido M, Baricordi C, et al. A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nat Commun*. 2019;10:2395.
- Rodriguez-Fraticelli AE, Wolock SL, Weinreb CS, et al. Clonal analysis of lineage fate in native haematopoiesis. *Nature*. 2018;553(7687):212-216.
- Sanjuan-Pla A, Macaulay IC, Jensen CT, et al. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature*. 2013;502(7470):232-236.
- Gekas C, Graf T. CD41 expression marks myeloid biased adult hematopoietic stem cells and increases with age. *Blood*. 2013;121(22):4463-4472.
- Astle W, Elding H, Jiang T, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell*. 2016;167(5):1415-1429.
- Tomer A, Friese P, Conklin R, et al. Flow cytometric analysis of megakaryocytes from patients with abnormal platelet counts. *Blood*. 1989;74(2):594-601.
- Haas SF, Hansson J, Klimmeck D, et al. Inflammation-induced emergency megakaryopoiesis driven by hematopoietic stem cell-like megakaryocyte progenitors. *Cell Stem Cell*. 2015;17(4):422-434.
- Middleton EA, Rowley JW, Campbell RA, et al. Sepsis alters the transcriptional and translational landscape of human and murine platelets. *Blood*. 2019;134(12):911-923.
- Martin JF, Kristensen SD, Mathur A, et al. The causal role of megakaryocyte-platelet hyperactivity in acute coronary syndromes. *Nat Rev Cardiol*. 2012;9(11):658-670.
- Macaulay IC, Haerty W, Kumar P, et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat Methods*. 2015;12(6):519-522.
- Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
- Ilicic T, Kim JK, Kolodziejczyk AA, et al. Classification of low quality cells from single-cell RNA-seq data. *Genome Biol*. 2016;17:29.
- Breiman L. Random forests. *Mach Learn*. 2001;45:5-32.
- McCarthy DJ, Campbell KR, Lun ATL, et al. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics*. 2017;33(8):1179-1186.
- Lun ATL, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data. *F1000Research*. 2016;5:2122.
- Lun ATL, Calero-Nieto FJ, Haim-Vilmovsky L, et al. Assessing the reliability of spike-in normalization for analyses of single-cell RNA sequencing data. *Genome Res*. 2017;27:1795-1806.
- Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math*. 1987;20:53-65.
- Kiselev VY, Kirschner K, Schaub MT, et al. SC3: consensus clustering of single-cell RNA-Seq data. *Nat Methods*. 2017;14(5):483-486.
- Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*. 2014;32(4):381-386.
- Qiu X, Mao Q, Tang Y, et al. Reversed graph embedding resolves complex single-cell developmental trajectories. *Nat Methods*. 2017;10:979-982.
- Novershtern N, Subramanian A, Lawton LN, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*. 2011;144(2):296-309.
- Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
- Lun ATL, Bach K, Marioni JC. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol*. 2016;17:75.
- D'Andrea D, Grassi L, Mazzapioda M, et al. FIDEA: a server for the functional interpretation of differential expression analysis. *Nucleic Acids Res*. 2013;41:W84-88.
- Belluschi S, Calderbank EF, Ciaurro V, et al. Myelo-lymphoid lineage restriction occurs in the human haematopoietic stem cell compartment before lymphoid-primed multipotent progenitors. *Nat Commun*. 2018;9:4100.
- Lichtman MA, Shafer MS, Felgar RE, Wang N. Lichtman's atlas of hematology 2016, www.accessmedicine.com
- Rowley JW, Oler AJ, Tolley ND, et al. Genome-wide RNA-seq analysis of human and mouse platelet transcriptomes. *Blood*. 2011;118(14):e101-111.

39. Zerbino DR, Achuthan P, Akanni W, et al. Ensembl 2018. *Nucleic Acids Res.* 2017;46(D1):D754-D761.
40. Gieger C, Radhakrishnan A, Cvejic A, et al. New gene functions in megakaryopoiesis and platelet formation. *Nature.* 2011;480(7376):202-208.
41. Petersen R, Lambourne JJ, Javierre BM, et al. Platelet function is modified by common sequence variation in megakaryocyte super enhancers. *Nat Commun.* 2017;8:16058.
42. Cheng E, Luo Q, Brussels EM, et al. Role for MKL1 in megakaryocytic maturation. *Blood.* 2009;113(12):2826-2834.
43. Cecchetti L, Tolley ND, Michetti N, et al. Megakaryocytes differentially sort mRNAs for matrix metalloproteinases and their inhibitors into platelets: a mechanism for regulating synthetic events. *Blood.* 2011;118(7):1903-1911.
44. Davizon-Castillo P, McMahon B, Aguila S, et al. TNF- α -driven inflammation and mitochondrial dysfunction define the platelet hyperreactivity of aging. *Blood.* 2019;134:727-740.
45. Lu YC, Sanada C, Xavier-Ferrucio J, et al. The molecular signature of megakaryocyte-erythroid progenitors reveals a role for the cell cycle in fate specification. *Cell Rep.* 2018;25:3229.
46. Lane N, Martin W. The energetics of genome complexity. *Nature.* 2010;467:929-934.
47. Kosaki G, Kambayashi J. Thrombocytogenesis by megakaryocyte; Interpretation by protoplatelet hypothesis. *Proc Japan Acad Ser. B.* 2011;87:254-273.
48. Ru YX, Zhao SX, Dong SX, et al. On the maturation of megakaryocytes: a review with original observations on human in vivo cells emphasizing morphology and ultrastructure. *Ultrastruct Pathol.* 2015;39:79-87.
49. Tanum G, Engeset A. Low ploidy megakaryocytes in steady-state rat bone marrow. *Blood.* 1983;62:87-91.
50. Moreau T, Evans A, Vasquez L, et al. Large-scale production of megakaryocytes from human pluripotent stem cells by chemically defined forward programming. *Nat Commun.* 2016;7:11208.
51. North TE, Goessling W, Walkley CR, et al. Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis. *Nature.* 2007;447(7147):1007-1011.
52. Porter RL, Georger MA, Bromberg O, et al. Prostaglandin E2 increases hematopoietic stem cell survival and accelerates hematopoietic recovery after radiation injury. *Stem Cells.* 2013;31(2):372-383.
53. Xi X, Schlegel N, Caen JP, et al. Differential effects of recombinant human interleukin-13 on the in vitro growth of human hematopoietic progenitor cells. *Br J Haematol.* 1995;90(4):921-927.
54. Lai YH, Heslan JM, Poppema S, et al. Continuous administration of IL-13 to mice induces extramedullary hemopoiesis and monocytosis. *J Immunol.* 1996;156:3166-3173.
55. Lin KH, Yao CL, Lee H. Lysophosphatidic acid inhibits megakaryocyte differentiation in CD34+ hematopoietic stem cells. *FASEB J.* 2013;27:216-224.
56. Bruns I, Lucas D, Pinho S, et al. Megakaryocytes regulate hematopoietic stem cell quiescence through CXCL4 secretion. *Nat Med.* 2014;20(11):1315-1320.
57. Heazlewood SY, Neaves RJ, Williams B, et al. Megakaryocytes co-localise with hemopoietic stem cells and release cytokines that up-regulate stem cell proliferation. *Stem Cell Res.* 2013;11(2):782-792.
58. Zhao M, Perry JM, Marshall H, et al. Megakaryocytes maintain homeostatic quiescence and promote post-injury regeneration of hematopoietic stem cells. *Nat Med.* 2014;20(11):1321-1326.
59. Ghasemzadeh M, Kaplan ZS, Alwis I, et al. The CXCR1/2 ligand NAP-2 promotes directed intravascular leukocyte migration through platelet thrombi. *Blood.* 2013;121(22):4555-4566.
60. Grassi L, Pourfarzad F, Ullrich S, et al. Dynamics of transcription regulation in human bone marrow myeloid differentiation to mature blood neutrophils. *Cell Rep.* 2018;24(10):2784-2794.
61. Bernardi B, Guidetti GF, Campus F, et al. The small GTPase Rap1b regulates the cross talk between platelet integrin α 2 β 1 and integrin α IIb β 3. *Blood.* 2006;107(7):2728-2735.
62. Smyth S, Woulfe DS, Weitz JI, et al. G-protein-coupled receptors as signaling targets for antiplatelet therapy. *Arterioscler Thromb Vasc Biol.* 2009;29(4):449-457.
63. Murphy AJ, Bijl N, Yvan-Charvet L, et al. Cholesterol efflux in megakaryocyte progenitors suppresses platelet production and thrombocytosis. *Nat Med.* 2013;19(5):586-594.
64. Murphy AJ, Sarrazy V, Wang N, et al. Deficiency of ATP-binding cassette transporter b6 in megakaryocyte progenitors accelerates atherosclerosis in mice. *Arterioscler Thromb Vasc Biol.* 2014;34(4):751-758.
65. Ridker PM, Everett BM, Thuren T, et al. Anti-inflammatory therapy with canakinumab for atherosclerotic disease. *N Engl J Med.* 2017;377(12):1119-1131.
66. Beaulieu LM, Lin E, Mick E, et al. Interleukin 1 receptor 1 and interleukin 1 β regulate megakaryocyte maturation, platelet activation, and transcript profile during inflammation in mice and humans. *Arterioscler Thromb Vasc Biol.* 2014;34(3):552-564.
67. Mallolas JO, Hurtado O, Castellanos M, et al. A polymorphism in the EAAT2 promoter is associated with higher glutamate concentrations and higher frequency of progressing stroke. *J Exp Med.* 2006;203(3):711-717.
68. Genever PG, Wilkinson DJ, Patton AJ, et al. Expression of a functional n-methyl-d-aspartate-type glutamate receptor by bone marrow megakaryocytes. *Blood.* 1999;93(9):2876-2883.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Choudry FA, Bagger FO, Macaulay IC, et al. Transcriptional characterization of human megakaryocyte polyploidization and lineage commitment. *J Thromb Haemost.* 2021;19:1236-1249. <https://doi.org/10.1111/jth.15271>