Supplemental Materials and Methods

Flow cytometry

The cells were stained with the fluorophore-conjugated monoclonal antibodies c-Kit (Biolegend, San Diego, CA, clone 2B8), CD16/32 (Biolegend, 93), CD34 (BD Biosciences, RAM34), CD49b (Biolegend, DX5), FcεRIα (eBioscience, San Diego, CA, MAR-1), integrin β7 (BD Biosciences, FIB504), Ly6G (Biolegend, 1A8), Sca-1 (Biolegend, D7), SiglecF (BD Biosciences, E50-2440), TER-119 (Biolegend, TER-119), CD48 (eBioscience, HM48-1), CD45 (Biolegend, 30-F11), CD150 (Biolegend, TC15-12F12.2), EPCR (STEMCELL Technologies, RMEPCR1560), CD55 (Biolegend, RIKO-3), CD71 (Biolegend, RI7217), CD105 (Biolegend, MJ7/18), CD49f (Biolegend, GoH3), CD41 (Biolegend and BD Biosciences, MWReg30), and CD11b (Biolegend, M1/70). Fc-block (eBioscience, clone 93) was used where appropriate to prevent non-specific binding. The EasySep Mouse Hematopoietic Progenitor Cell Isolation Kit (STEMCELL Technologies) and fluorophore-conjugated streptavidin (Biolegend) was used to stain lineage-positive cells. Dead cells were excluded using the LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (Thermo Fisher Scientific), DAPI, or 7-aminoactinomycin D (Thermo Fisher Scientific). Doublets were removed from the analysis using the trigger pulse width parameters. The LSRFortessa (BD Biosciences) was used to analyze the cells. The flow cytometry data was analyzed using FlowJo software (Treestar, Ashland, OR), and further processed using RStudio and Prism (GraphPad Software, La Jolla, CA).

HSPC CFU-C assays

100 LK cells were sorted and cultured in a semi-solid medium containing factors to support growth of all myelo-erythroid colony types (M3434, STEMCELL Technologies).

Processing of scRNA-Seq for basophil/mast cell progenitors (BMCPs) and granulocyte/monocyte progenitors (GMPs)

scRNA-Seq was performed using Smart-Seq2 protocol as described previously.[1,2] Two plates of BMCPs and one plate of GMPs, with 47 single cells on each plate, were processed. Sequencing libraries were prepared using the Nextera XT DNA preparation kit (Illumina, San Diego, CA). Pooled libraries were sequenced using the HiSeq 4000 (Illumina). Samples were pooled with additional experiments for sequencing, and a customized genome index was built with GMAP/GSNAP[3] to discard reads caused by index-hopping in the Hoxb8 and Runx1 genes. Reads were aligned using G-SNAP and were aligned to Ensembl genes (release 81)[4] by HTSeq.[5] Analysis of BMCP and GMP scRNA-Seq data was performed using R (https://www.r-project.org). Cells with < 10% of total reads mapping to genes, < 200 000 reads mapping to nuclear genes or > 20% mapped reads mapping to mitochondrial genes were filtered out, leaving 89 BMCPs and 43 GMPs. Single-cell profiles were normalized using scran[6] and 4267 highly variable genes were identified using the ERCC spike-ins as previously described.[7] Counts were transformed to log10(counts+1) for dimensionality reduction and clustering. Diffusion map visualization was performed using the DiffusionMap function from the destiny R package[8] with 'rankcor' distance and k=10. Principal component analysis (PCA) was performed using the prcomp function.

Clustering and differential expression for BMCP and GMP scRNA-Seq profiles

For clustering, highly variable genes were filtered to keep only protein coding genes and genes with mean log-transformed count > 2. Hierarchical clustering was performed on Spearman correlation with average linkage using the hclust function. Investigation of potential batch effects showed that the two BMCP plates contributed equally to the clusters (15 and 14 cells from each plate in cluster 1, and 29 and 31 cells from each plate in cluster 2). All protein coding genes with mean count > 1 expressed in at least 2 cells were tested for differential expression using a Wilcoxon rank sum test with Benjamini-Hochberg correction for multiple testing, with comparisons calculated between cells in cluster 1 versus cluster 2, combined clusters 1 and 3 versus 2 and combined clusters 1 and 2 versus 3. Genes were filtered to those with log2 fold-change > 1 between clusters and ranked by adjusted *P* value to identify the top genes specific to clusters for visualization in the heatmaps.

Alignment and quality control for droplet-based scRNA-Seq

Sample demultiplexing, barcodes processing, and gene counting was performed using the count commands from the Cell Ranger v1.3 pipeline (https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome). After Cell Ranger processing 46 447 unique barcodes were retained for wild type (WT) data, and 14 675 for W^{41}/W^{41} data. Each sample was filtered for potential doublets by simulating synthetic doublets from pairs of scRNA-Seq profiles and assigning scores based on a k-nearest-neighbor classifier on PCA transformed data. The 1% and 4.5% of cells with the highest doublet scores from each Lin^{-} $Sca\text{-}1^{+}$ $c\text{-}Kit^{+}$ (LSK) or Lin^{-} $c\text{-}Kit^{+}$ (LK) sample were removed from further analysis, respectively. This excluded 1452 WT cells and 845 W^{41}/ W^{41} cells. Cells with > 10% of unique molecular identifier (UMI) counts mapping to mitochondrial genes, expressing fewer than 500 genes, or with the total number of UMI counts further than 3 standard deviations from the mean were excluded. After quality control 44 802 WT cells and 13 815 W^{41} cells were retained.

Variable gene selection for droplet-based scRNA-Seq

5032 and 5033 variable genes were identified for the WT and W^{41}/ W^{41} data sets, respectively, following the method of Macosko et al[9] with expression=0.001 and dispersion=0.05 minimum cutoffs. Cells were normalized to the same total count for each cell, and log-transformed (x -> log(x+1)). Each gene was scaled so it was zero-centered. Initial visualizations using diffusion maps[10] revealed a prominent “gap” in the erythroid cells (Figure S1C). Differentially expressed genes between these cells were enriched in cell cycle genes (Figure S1D). To correct this effect we removed any of these differentially expressed genes in the WT data that intersected with a list of cell cycle genes downloaded from Reactome (http://www.reactome.org/) (Figure S1E), along with any other genes that had high correlation (Pearson correlation coefficient > 0.2) with any of the genes in this intersection (Figure S1F). This resulted in 4664 genes carried forward for visualization and clustering analysis on the WT data, and 4745 genes for visualization on the W^{41}/ W^{41} data.

Force-directed graph visualization

For visualization, a k=7 nearest neighbor graph was calculated on the first 50 principal components of PCA on the variable genes. The edge list was exported into Gephi 0.9.1 (https://gephi.org/) and graph coordinates calculated using the ForceAtlas 2 layout.

Visualization of gene expression sets in droplet scRNA-Seq landscape

For a set of genes, a geometric mean based score was calculated on normalized UMI counts, and this score then plotted on the force directed graph embedding. For each cell, this score was given by $\exp\left[\sum_{g\in G}\log\left(x_g+1\right)/m\right]$ for m genes in a set G. To visualize HSC gene expression (Figure S1A) this score was calculated on 29 genes previously identified as being enriched in functional HSCs (MolO genes from table S3, Wilson et al., 2015)[11]. To visualize BMCP cluster specific gene expression in the 44 802 WT cells, values for genes listed in Figure 3D were extracted from the droplet scRNA-Seq, and the above score calculated for each cluster gene set. To calculate G2M marker gene scores, we used the set of 200 Hallmark G2M checkpoint genes downloaded from the Molecular Signatures Database.[12]

Differential expression analysis on droplet-based scRNA-Seq data

Differential expression analysis was performed between WT and W^{41}/W^{41} clusters using edgeR applied to UMI counts.[13] *P* values were corrected for multiple testing (Benjamini & Hochberg correction). Genes were labelled as up/down regulated based on positive/negative fold-change and ranked by adjusted *P* value. The top 200 up- and downregulated genes between each cluster were recorded (Supplemental Table S1).

To calculate the significance of overlap between the differentially expressed genes and biologically annotated gene sets we input the top 200 up-/downregulated genes for each cluster into the Molecular Signatures Database online tool (http://software.broadinstitute.org/gsea/msigdb/index.jsp) and computed overlaps between these and the Hallmark gene sets.[12,14]

Supplemental Figure Legends

Figure S1. After cell cycle correction, droplet-based scRNA-Seq of 44 802 hematopoietic stem and progenitor cells reveals entry points to blood lineages. (A) Expression of HSC genes in the droplet-based scRNA-Seq data. A geometric mean score of gene counts was calculated for each cell across genes in the HSC set. The color of cells indicates the value of this sum, with grey being the lowest value and red the highest value. (B) Expression of marker genes plotted on the force-directed graph embedding. Gene expression is plotted on a log(normalized count + 1) scale with grey equal to no counts and dark red the maximum value detected. (C) Diffusion map components (DCs) for droplet scRNA-Seq colored by erythroid gene *Klf1* (left panel) or highlighted by cell in two groups on either side of 'gap' (right panel). Gene expression plotted on log(counts+1) scale. (D) Enriched terms for top 100 differentially expressed genes between groups 1 and 2 from Enrichr (Reactome 2016 category).[15] (E) Intersection between top 100 differentially expressed genes and list of cell cycle genes downloaded from Reactome. (F) Distribution of maximum Pearson correlations between each highly variable gene and 23 differentially expressed genes in intersection in panel E. Genes with correlation > 0.2 (red line) were filtered from analysis for visualization and clustering

Figure S2. Alternative dimensionality reduction techniques can identify rare progenitor populations. (A) t-Distributed Stochastic Neighbor Embedding (t-SNE) of 44 802 droplet-based scRNA-Seq profiles colored by blood stem and progenitor marker genes. (B) Marker genes highlighted on diffusion map embedding of the same cells. Gene expression is plotted on a log(normalized count + 1) scale with grey equal to no counts and dark red the maximum value detected.

Figure S3. Separate force-directed graph embeddings reveal greater heterogeneity in LK cells compared to LSK cells. (A) Expression of marker genes plotted on the force-directed graph embedding calculated on LSK cells alone. (B) Expression of marker genes plotted on the force-directed graph embedding calculated on LK cells alone. Gene expression is plotted on a log(normalized count + 1) scale with grey equal to no counts and dark red the maximum value detected. (C) Principal Component Analysis (PCA) was performed on LSK and LK cells together, and then pairwise distances within the LSK or LK cell populations were calculated. The histograms show the distributions of the median distance to other cells for each cell within either LSK or LK populations.

Figure S4. Lin$^-$ Sca-1$^-$ c-Kit$^+$ integrin β7^{hi} CD16/32^{hi} bone marrow progenitors form mast cells and basophils. (A) Flow cytometry analysis of BMCPs (red) and MPs (black). (B) BMCPs, MPs, and GMPs were cultured for 5 days with IL-3, IL-5, IL-9, GM-CSF, and SCF and analyzed by flow cytometry. The top row shows an example in which BMCPs, MPs, and GMPs were cultured separately, but pooled before analysis, to visualize all cell populations simultaneously. (C) BMCPs and MPs were cultured in myeloid-promoting conditions. Mast cells, basophils, neutrophils, and eosinophils were sorted as described in panel B and stained with May-Grünwald Giemsa. Photo width, 60 µm. (D) May-Grünwald Giemsa-stained monocyte derived from an MP. The mast cells and basophils were sorted from the BMCP population, and the eosinophils and neutrophils were sorted from the MP population. No live/dead staining was used for the cell sort on day 5. Photo width, 60 µm. (E) Single BMCPs were cultured in myeloid-promoting conditions in individual Terasaki plate wells and

the colonies were analyzed with flow cytometry on day 5. Examples of a pure mast cell colony, basophil colony, and a mixed mast cell/basophil colony are shown. (F-H) Single bone marrow progenitors from wild type mice were sorted and cultured under myeloid-promoting conditions. The colony size was determined by visually counting the number of cells in each well at the indicated time point. A colony was defined as at least 5 cells arising from a single progenitor. The whiskers indicate the minimum and maximum values. Each dot represents one colony. The results are pooled from 2 independent sorts for each time point. The two-tailed Mann-Whitney test was used when two groups were compared, and the Kruskal-Wallis test with Dunn's multiple comparisons test was used when three groups were compared. Day 11 GMPs were not analyzed as they mainly constituted dead cells. ***P* value <.01, *****P* value<.0001. Ba, basophil; Eo, eosinophil; MC, mast cell; Mono, monocyte; and N, neutrophil. Images were captured using the Axio Imager.Z2, Axiocam 506, and Zen software (Zeiss, Oberkochen, Germany).

Figure S5. Lin$^-$ Sca-1$^-$ c-Kit$^+$ integrin β7^{hi} CD16/32^{hi} bone marrow progenitors lack erythroid potential. (A) Gating strategy of bone marrow BMCPs and MEPs. Lin$^-$ Sca-1$^-$ c-Kit$^+$ progenitors are visualized in the plots. (B) BMCPs and MEPs were cultured with IL-3, SCF, and EPO, and analyzed by flow cytometry. The gating strategy of basophils, mast cells, and erythroid cells are shown. Ba indicates basophil; Ery, erythroid cell; and MC, mast cell. (C-D) Single bone marrow progenitors from wild type mice were sorted and cultured under erythroid-promoting conditions. The colony size was determined by visually counting the number of cells in each well at the indicated time point. A colony was defined as at least 5 cells arising from a single progenitor. The whiskers indicate the minimum and maximum values. Each dot represents one colony. The results are pooled from 2 independent sorts for each time point. Two-tailed Mann-Whitney test; *****P* value<.0001. Day 5 MEPs were not analyzed as they mainly constituted dead cells. (E) Representative flow cytometry analysis before the first and second sorts.

Figure S6. Basophil/mast cell progenitors (BMCPs) and granulocyte monocyte progenitors (GMPs) show distinct gene expression profiles. (A) Heatmap of the expression of top 100 most upregulated genes in BMCPs versus GMPs. (B) Heatmap of the expression of top 100 most upregulated genes in GMPs versus BMCPs.

Figure S7. Cluster 1 BMCPs display a more differentiated phenotype than cluster 2 BMCPs. Cluster 1 and 2 BMCPs plotted based on the flow cytometry index data. The parameters that are significant (*P* value<.05) based on the two-tailed Mann-Whitney test are shown. The box shows the interquartile range (IQR) and the median is marked with a horizontal line. The whiskers indicate (25% quartile – 1.5 x IQR, 75% quartile + 1.5 x IQR). All cells with index data are shown. SSC indicates side scatter. ****P* value<.0005, ***P* value<.005, **P* value<.05

Figure S8. Cells from c-Kit mutant mice can be mapped to clusters in wild-type compartment. (A) Violin plots of marker gene expression in clusters in wild-type (WT) Lin$^-$ c-Kit$^+$ (LK) cells. (B) Violin plots of marker gene expression in LK cells from W^{41}/W^{41} c-Kit mutant mice mapped to WT clusters. Colors match those in Figure 4 for the same clusters. (C) Bone marrow progenitors from W^{41}/W^{41} mice were sorted and cultured under myeloid-promoting conditions. The colony size was determined by visually counting the number of cells in each well on day 5. A colony

was defined as at least 5 cells arising from a single progenitor. The whiskers indicate the minimum and maximum values. Each dot represents one colony. The results are pooled from 2 independent sorts for each time point. Two-tailed Mann-Whitney test; *****P* value<.0001.

Figure S9. The frequency of late erythroid progenitors is increased in W^{41}/W^{41} mice. (A) Gating strategy for the identification of populations P1-P5 in bone marrow. The analysis was performed without red blood cell lysis. (B) Log_2-fold change of the frequency of a cell population in the W^{41}/W^{41} mice compared with wild type mice. Left-hand and right-hand bars indicate fold-changes for samples from two separate W^{41}/W^{41} mice. The mean of samples from two separate wild type mice was used for the comparison. (Ci) CFU assays of Lin^- $c\text{-}Kit^+$ (LK) cells. W^{41}/W^{41} and WT LK cells were sorted and cultured for 7 (blue), 10 (black), or 12 (red) days and the colony types were determined. W^{41}/W^{41} CFUs are reported as percentage of the mean number of wild type colonies of the same type within an experiment. The results are derived from 6 W^{41}/W^{41} and 4 WT mice from 3 independent sorts. Whiskers display minimum and maximum values, and all 6 W^{41}/W^{41} mice are shown. Two-tailed one-sample t-test different from 100 %; **P* value<.05. (Cii) Single E-SLAMs were cultured for 14 days with stem cell factor and IL-11 and analyzed by flow cytometry. Data concatenated from 32 WT and 10 W^{41}/W^{41} colonies is shown. Mo, monocyte; Neu, neutrophil; MK, megakaryocyte. (D) Expression of G2/M marker genes (Hallmark set) in the wild-type (WT) and W^{41}/W^{41} droplet-based scRNA-Seq data of LK cells. A geometric mean score of gene counts was calculated for each cell across genes in this set. The color of cells indicates the value of this sum, with blue being the lowest value and red the highest value. (E) Violin plots of G2M score from C in WT and W^{41}/W^{41} clusters in LK cells. Colors match those in Figure 4. (F) Correlation between pseudotime ordering of WT cells from clusters 1, 6, 3, 2 & 5 (stem cells to erythroid) calculated on variable genes (x-axis) and variable genes with Hallmark G2M genes removed (y-axis). Pseudotime ordering was calculated with diffusion pseudotime (DPT).[16] (G) Violin plots of erythroid genes marking different stages of differentiation in WT and W^{41}/W^{41} clusters 1, 6, 3, 2 & 5 to confirm differentiation order of these clusters. (H,I) Violin plots showing the distribution of selected genes in WT and corresponding W^{41}/W^{41} clusters, as measured by scRNA-Seq. Colors correspond to those in Figure 4.

References

1. Wilson NK, Kent DG, Buettner F, et al. Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell*. 2015;16(6):712–724.
2. Picelli S, Faridani OR, Björklund ÅK, et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 2014;9:171.
3. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*. 2010;26(7):873–881.
4. Yates A, Akanni W, Amode MR, et al. Ensembl 2016. *Nucleic Acids Res.* 2016;44(D1):D710–D716.
5. Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166–169.
6. Lun ATL, Bach K, Marioni JC. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* 2016;17(1):75.
7. Brennecke P, Anders S, Kim JK, et al. Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods*. 2013;10(11):1093–1095.
8. Angerer P, Haghverdi L, Büttner M, et al. destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics*. 2016;32(8):1241–1243.
9. Macosko EZ, Basu A, Satija R, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*. 2015;161(5):1202–1214.
10. Haghverdi L, Buettner F, Theis FJ. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics*. 2015;1–10.
11. Wilson NK, Kent DG, Buettner F, et al. Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell*. 2015;712–724.
12. Liberzon A, Birger C, Thorvaldsdóttir H, et al. The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* 2015;1(6):417–425.
13. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–140.
14. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* 2005;102(43):15545–15550.
15. Kuleshov M V., Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016;44(W1):W90–W97.
16. Haghverdi L, Büttner M, Wolf FA, Buettner F, Theis FJ. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods*. 2016;13(10):845–848.

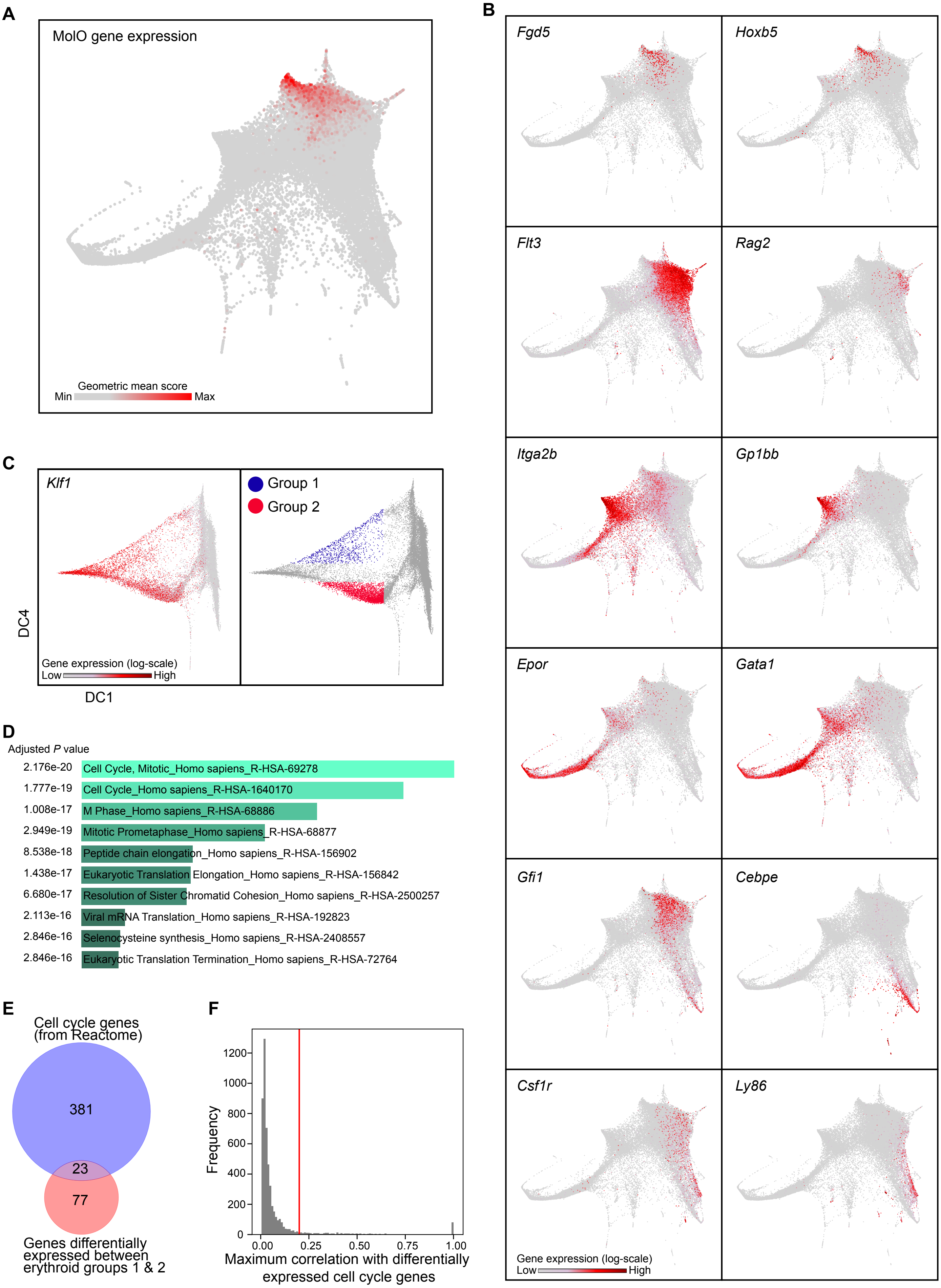
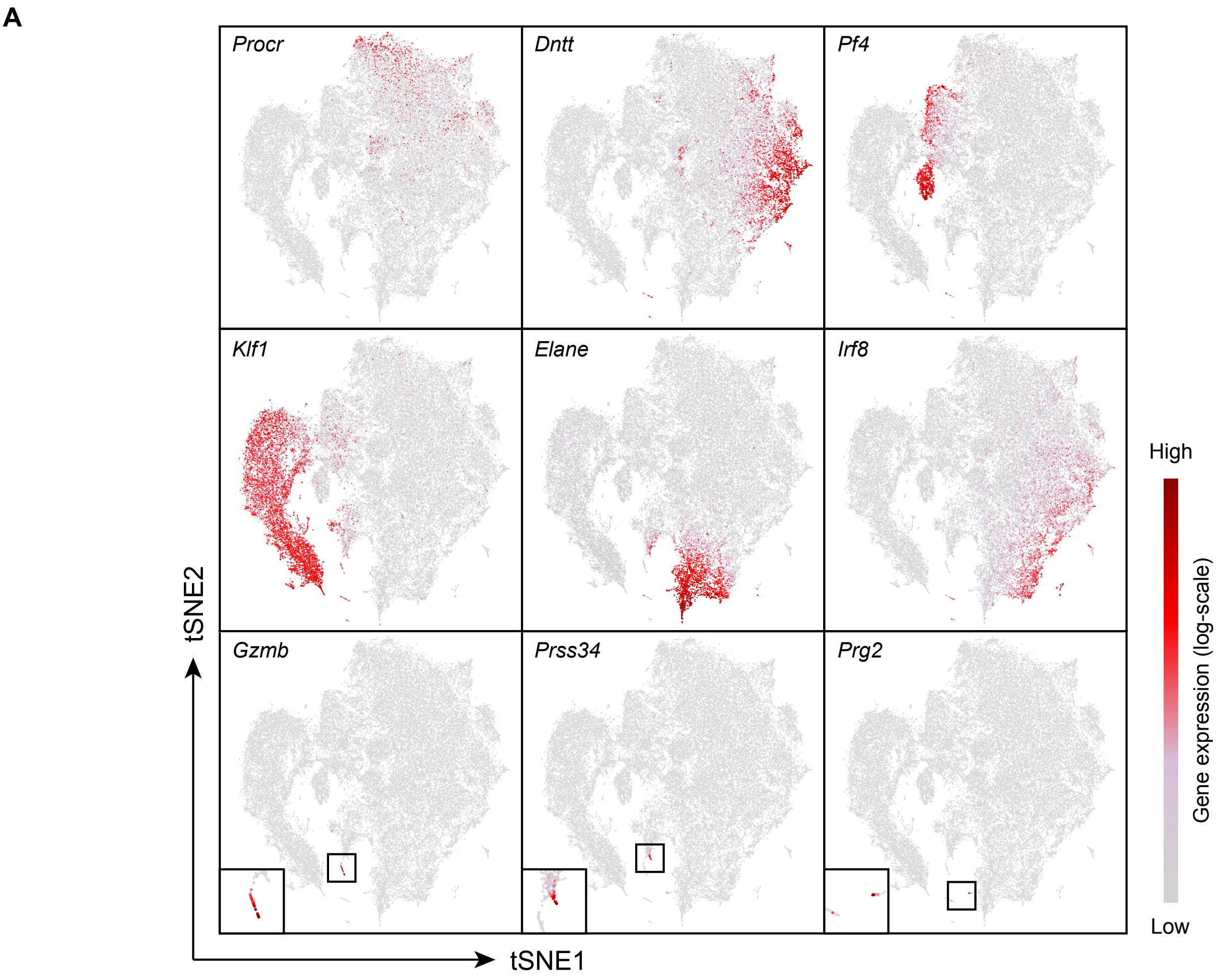

A
MolO gene expression
Geometric mean score
Min
Max
B
Fgd5
Hoxb5
Flt3
Rag2
Itga2b
Gp1bb
Epor
Gata1
Gfi1
Cebpe
Csf1r
Ly86
Gene expression (log-scale)
Low
High
C
Klf1
Group 1
Group 2
DC4
DC1
Gene expression (log-scale)
Low
High
D
Adjusted P value
2.176e-20 Cell Cycle, Mitotic_Homo sapiens_R-HSA-69278
1.777e-19 Cell Cycle_Homo sapiens_R-HSA-1640170
1.008e-17 M Phase_Homo sapiens_R-HSA-68886
2.949e-19 Mitotic Prometaphase_Homo sapiens_R-HSA-68877
8.538e-18 Peptide chain elongation_Homo sapiens_R-HSA-156902
1.438e-17 Eukaryotic Translation Elongation_Homo sapiens_R-HSA-156842
6.680e-17 Resolution of Sister Chromatid Cohesion_Homo sapiens_R-HSA-2500257
2.113e-16 Viral mRNA Translation_Homo sapiens_R-HSA-192823
2.846e-16 Selenocysteine synthesis_Homo sapiens_R-HSA-2408557
2.846e-16 Eukaryotic Translation Termination_Homo sapiens_R-HSA-72764
E
Cell cycle genes (from Reactome)
381
23
77
Genes differentially expressed between erythroid groups 1 & 2
F
Frequency
1200
1000
800
600
400
200
0
0.00
0.25
0.50
0.75
1.00
Maximum correlation with differentially expressed cell cycle genes

Figure S1


A
Procr
Dntt
Pf4
Klf1
Elane
Irf8
Gzmb
Prss34
Prg2
tSNE2
tSNE1
High
Gene expression (log-scale)
Low

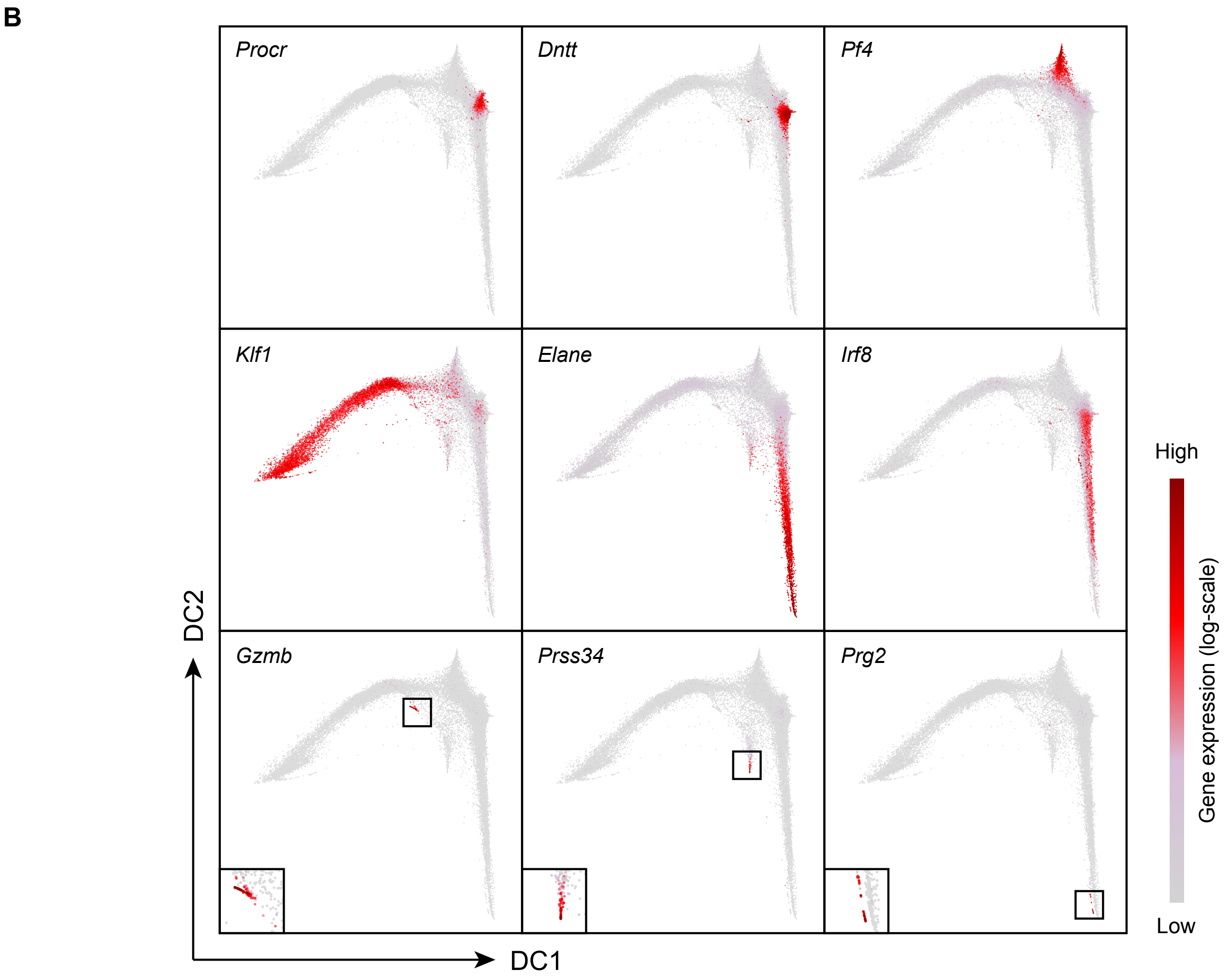

B
Procr
Dntt
Pf4
Klf1
Elane
Irf8
Gzmb
Prss34
Prg2
DC2
DC1
High
Gene expression (log-scale)
Low

Figure S2

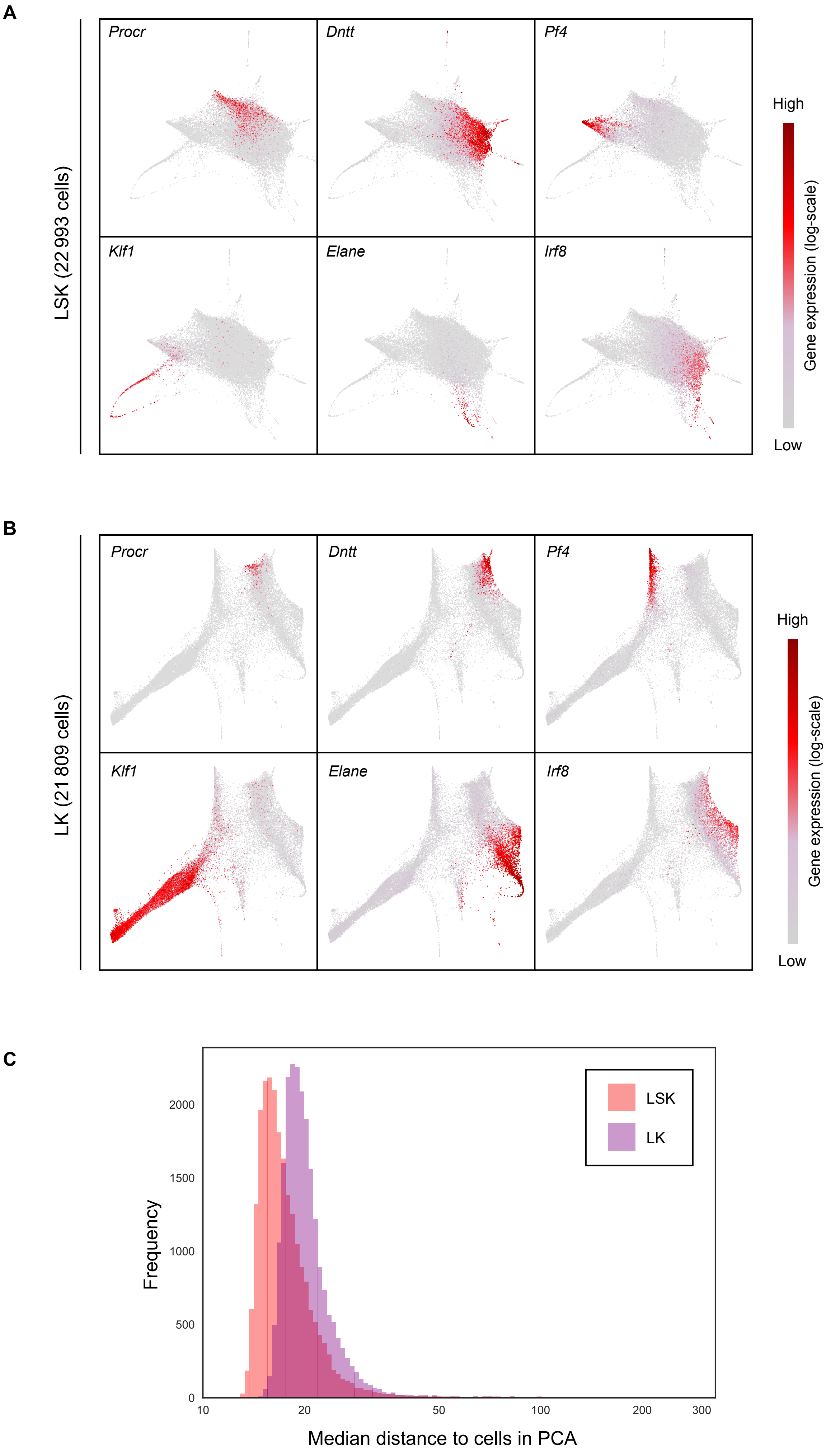

A
LSK (22 993 cells)
Procr
Dntt
Pf4
Klf1
Elane
Irf8
High
Gene expression (log-scale)
Low
B
LK (21 809 cells)
Procr
Dntt
Pf4
Klf1
Elane
Irf8
High
Gene expression (log-scale)
Low
C
Frequency
0
500
1000
1500
2000
10
20
50
100
200
300
Median distance to cells in PCA
LSK
LK

Figure S3

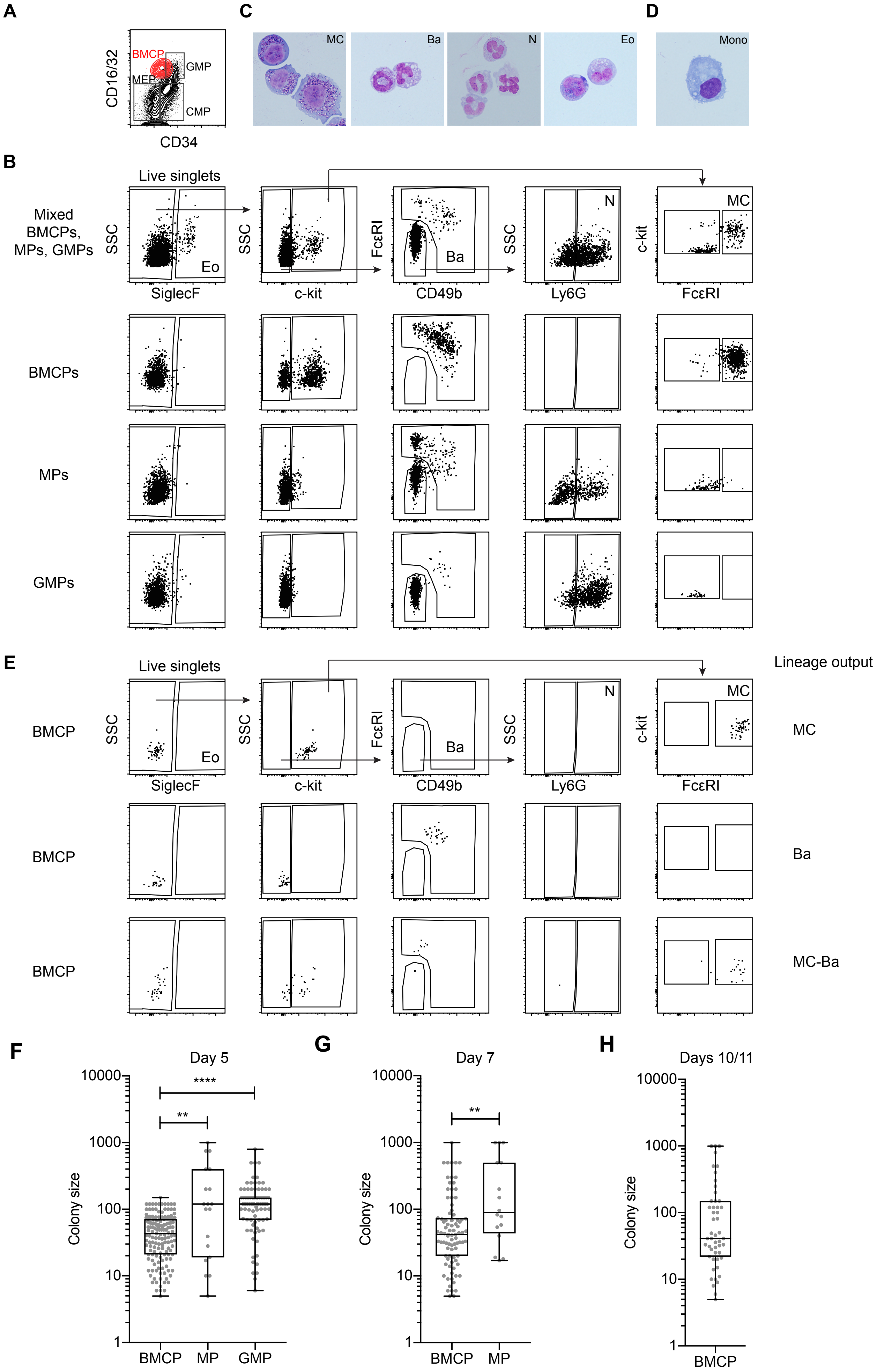

A
CD16/32
BMCP
GMP
MEP
CMP
CD34
C
MC
Ba
N
Eo
D
Mono
B
Live singlets
Mixed BMCPs, MPs, GMPs
SSC
Eo
SiglecF
SSC
c-kit
FcεRI
Ba
CD49b
SSC
N
Ly6G
c-kit
MC
FcεRI
BMCPs
MPs
GMPs
E
Live singlets
Lineage output
BMCP
SSC
Eo
SiglecF
SSC
c-kit
FcεRI
Ba
CD49b
SSC
N
Ly6G
c-kit
MC
FcεRI
MC
BMCP
Ba
BMCP
MC-Ba
F
Day 5
10000
1000
100
10
1
Colony size
****
**
BMCP
MP
GMP
G
Day 7
10000
1000
100
10
1
Colony size
**
BMCP
MP
H
Days 10/11
10000
1000
100
10
1
Colony size
BMCP

Figure S4

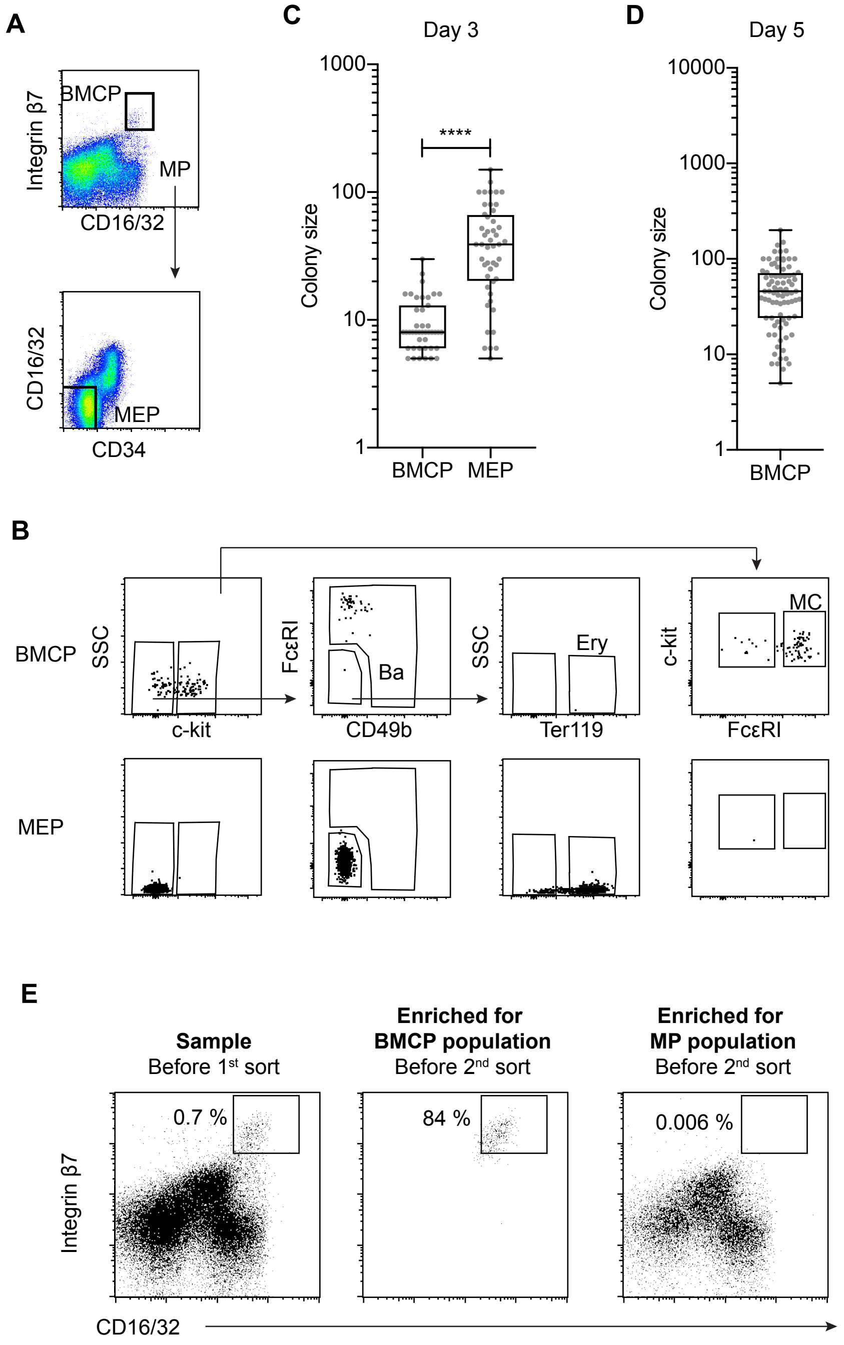

A
Integrin β7
BMCP
MP
CD16/32
CD16/32
MEP
CD34
C
Day 3
1000
100
10
1
Colony size
****
BMCP
MEP
D
Day 5
10000
1000
100
10
1
Colony size
BMCP
B
BMCP
SSC
c-kit
FcεRI
Ba
CD49b
SSC
Ery
Ter119
c-kit
MC
FcεRI
MEP
E
Sample
Before 1st sort
Enriched for
BMCP population
Before 2nd sort
Enriched for
MP population
Before 2nd sort
0.7 %
84 %
0.006 %
Integrin β7
CD16/32

Figure S5

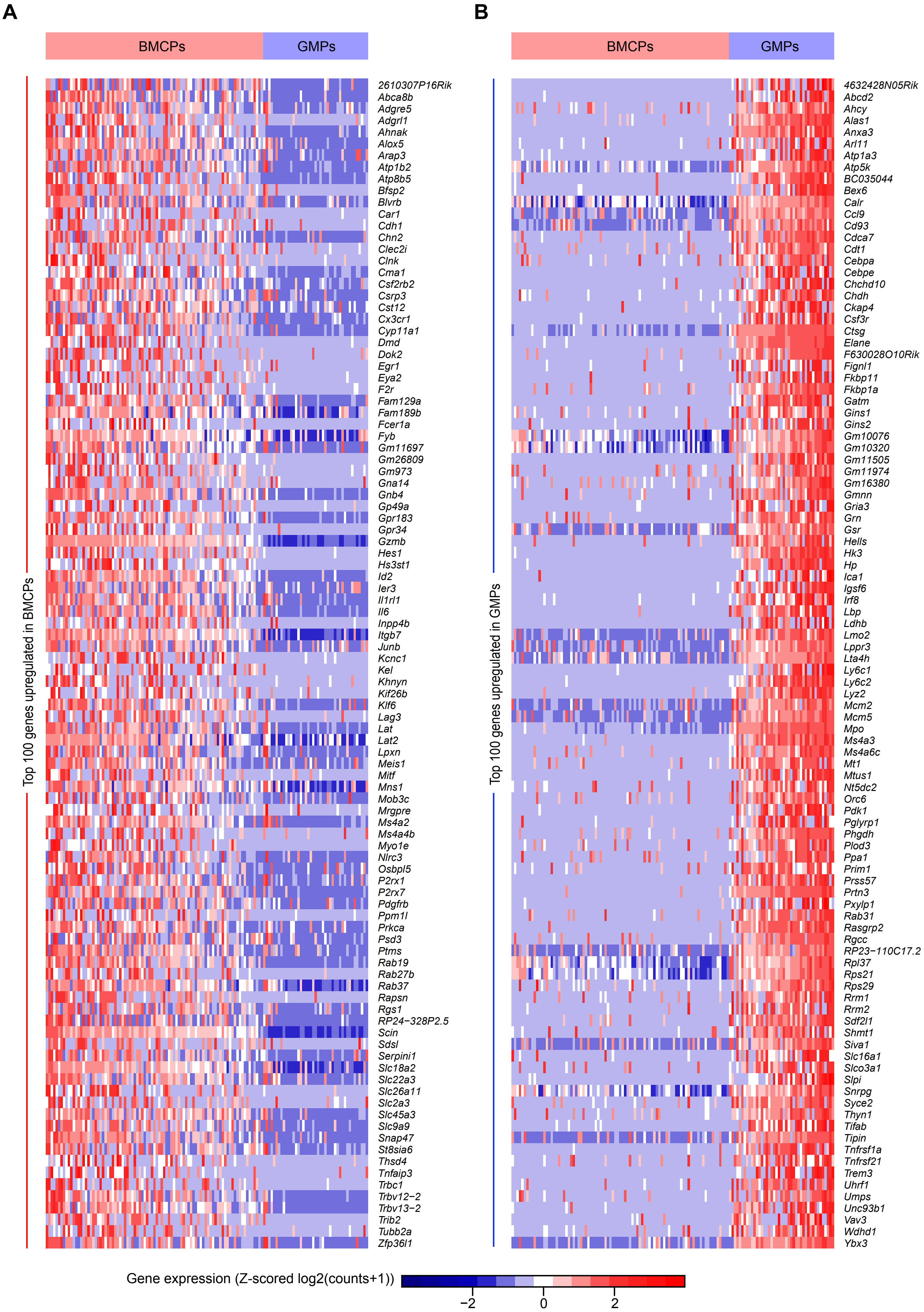

A
BMCPs
GMPs
Top 100 genes upregulated in BMCPs
2610307P16Rik
Abca8b
Adgre5
Adgrl1
Ahnak
Alox5
Arap3
Atp1b2
Atp8b5
Bfsp2
Blvrb
Car1
Cdh1
Chn2
Clec2i
Clnk
Cma1
Csf2rb2
Csrp3
Cst12
Cx3cr1
Cyp11a1
Dmd
Dok2
Egr1
Eya2
F2r
Fam129a
Fam189b
Fcer1a
Fyb
Gm11697
Gm26809
Gm973
Gna14
Gnb4
Gp49a
Gpr183
Gpr34
Gzmb
Hes1
Hs3st1
Id2
Ier3
Il1rl1
Il6
Inpp4b
Itgb7
Junb
Kcnc1
Kel
Khnyn
Kif26b
Klf6
Lag3
Lat
Lat2
Lpxn
Meis1
Mitf
Mns1
Mob3c
Mrgpre
Ms4a2
Ms4a4b
Myo1e
Nlrc3
Osbpl5
P2rx1
P2rx7
Pdgfrb
Ppm1l
Prkca
Psd3
Ptms
Rab19
Rab27b
Rab37
Rapsn
Rgs1
RP24−328P2.5
Scin
Sdsl
Serpini1
Slc18a2
Slc22a3
Slc26a11
Slc2a3
Slc45a3
Slc9a9
Snap47
St8sia6
Thsd4
Tnfaip3
Trbc1
Trbv12−2
Trbv13−2
Trib2
Tubb2a
Zfp36l1
B
BMCPs
GMPs
Top 100 genes upregulated in GMPs
4632428N05Rik
Abcd2
Ahcy
Alas1
Anxa3
Arl11
Atp1a3
Atp5k
BC035044
Bex6
Calr
Ccl9
Cd93
Cdca7
Cdt1
Cebpa
Cebpe
Chchd10
Chdh
Ckap4
Csf3r
Ctsg
Elane
F630028O10Rik
Fignl1
Fkbp11
Fkbp1a
Gatm
Gins1
Gins2
Gm10076
Gm10320
Gm11505
Gm11974
Gm16380
Gmnn
Gria3
Grn
Gsr
Hells
Hk3
Hp
Ica1
Igsf6
Irf8
Lbp
Ldhb
Lmo2
Lppr3
Lta4h
Ly6c1
Ly6c2
Lyz2
Mcm2
Mcm5
Mpo
Ms4a3
Ms4a6c
Mt1
Mtus1
Nt5dc2
Orc6
Pdk1
Pglyrp1
Phgdh
Plod3
Ppa1
Prim1
Prss57
Prtn3
Pxylp1
Rab31
Rasgrp2
Rgcc
RP23−110C17.2
Rpl37
Rps21
Rps29
Rrm1
Rrm2
Sdf2l1
Shmt1
Siva1
Slc16a1
Slco3a1
Slpi
Snrpg
Syce2
Thyn1
Tifab
Tipin
Tnfrsf1a
Tnfrsf21
Trem3
Uhrf1
Umps
Unc93b1
Vav3
Wdhd1
Ybx3
Gene expression (Z-scored log2(counts+1))
−2
0
2

Figure S6

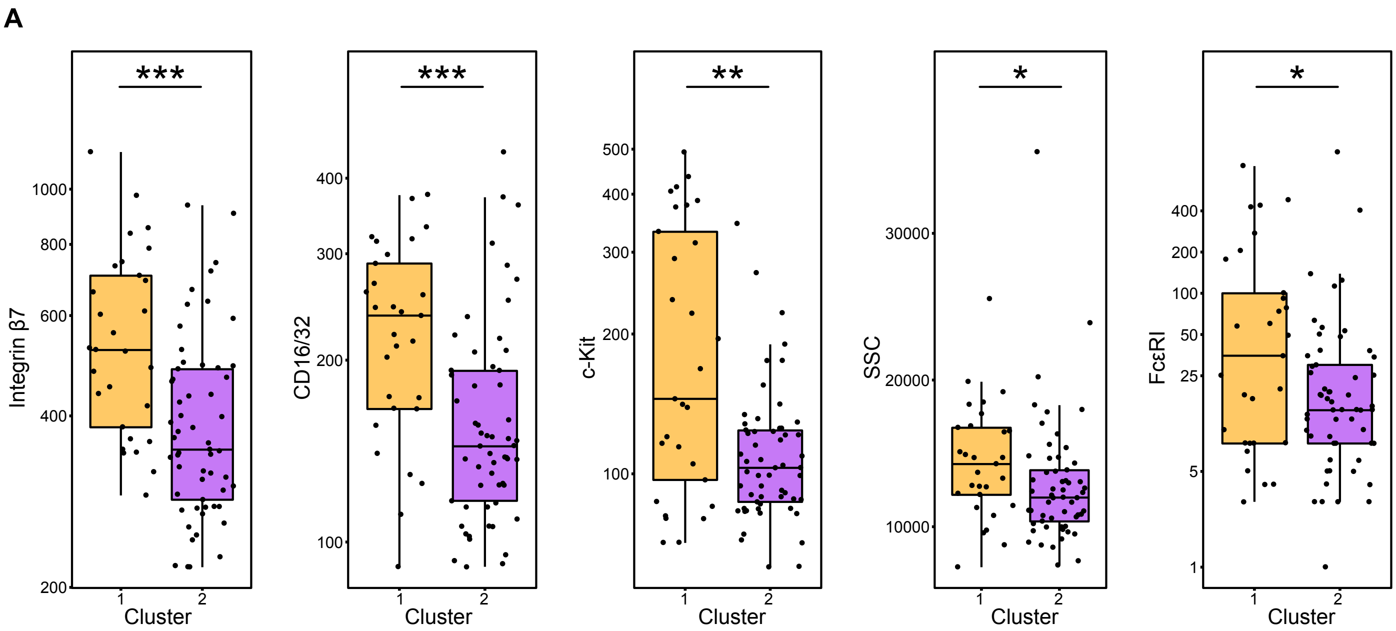

A
***
Integrin β7
200
400
600
800
1000
1
2
Cluster
***
CD16/32
100
200
300
400
1
2
Cluster
**
c-Kit
100
200
300
400
500
1
2
Cluster
*
SSC
10000
20000
30000
1
2
Cluster
*
FcεRI
1
5
25
50
100
200
400
1
2
Cluster

Figure S7

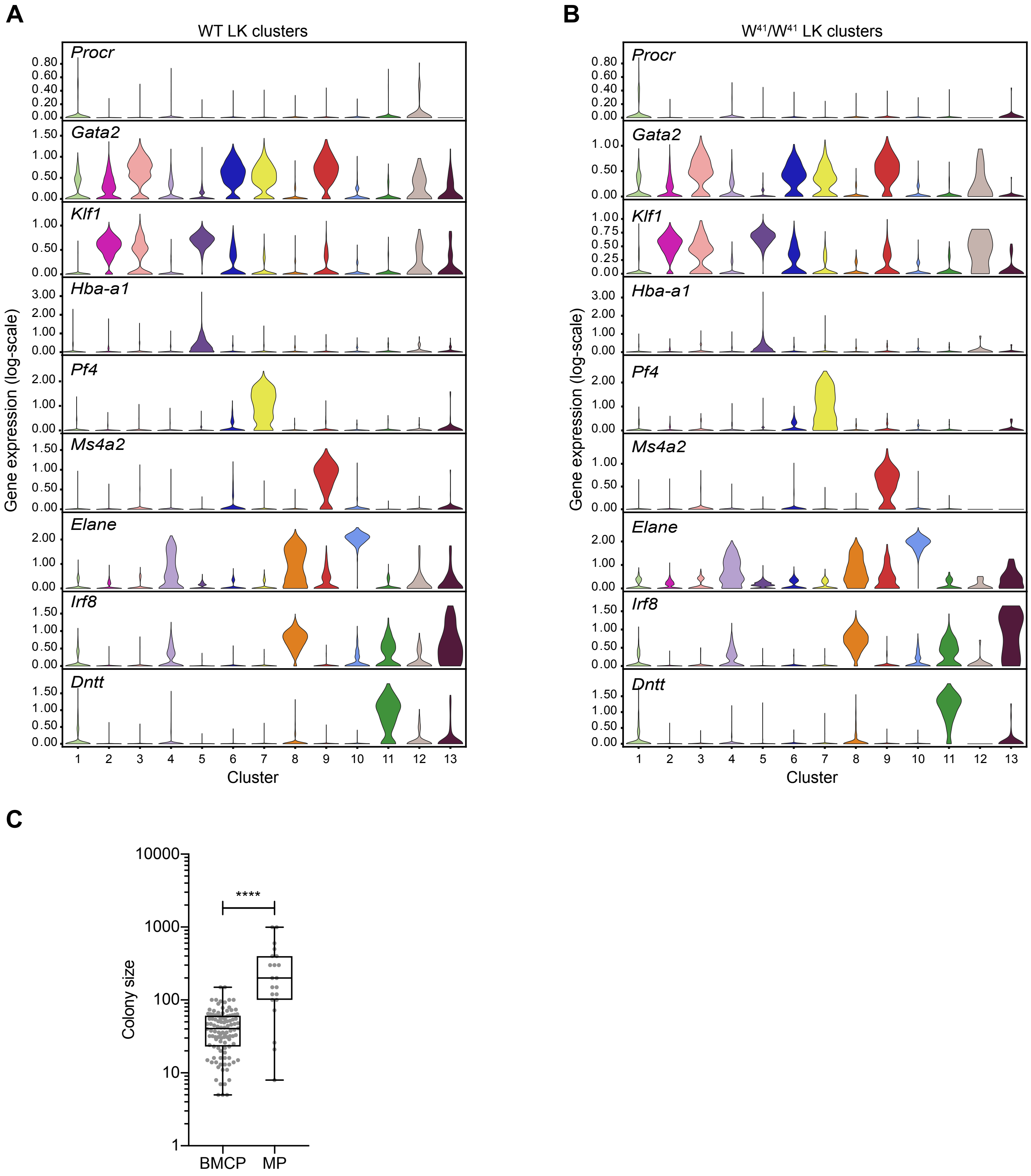

A
WT LK clusters
B
W41/W41 LK clusters
Procr
Gata2
Klf1
Hba-a1
Pf4
Ms4a2
Elane
Irf8
Dntt
0.00
0.20
0.40
0.60
0.80
0.25
0.50
0.75
1.00
1.50
2.00
3.00
Gene expression (log-scale)
1
2
3
4
5
6
7
8
9
10
11
12
13
Cluster
C
Colony size
1
10
100
1000
10000
****
BMCP
MP

Figure S8

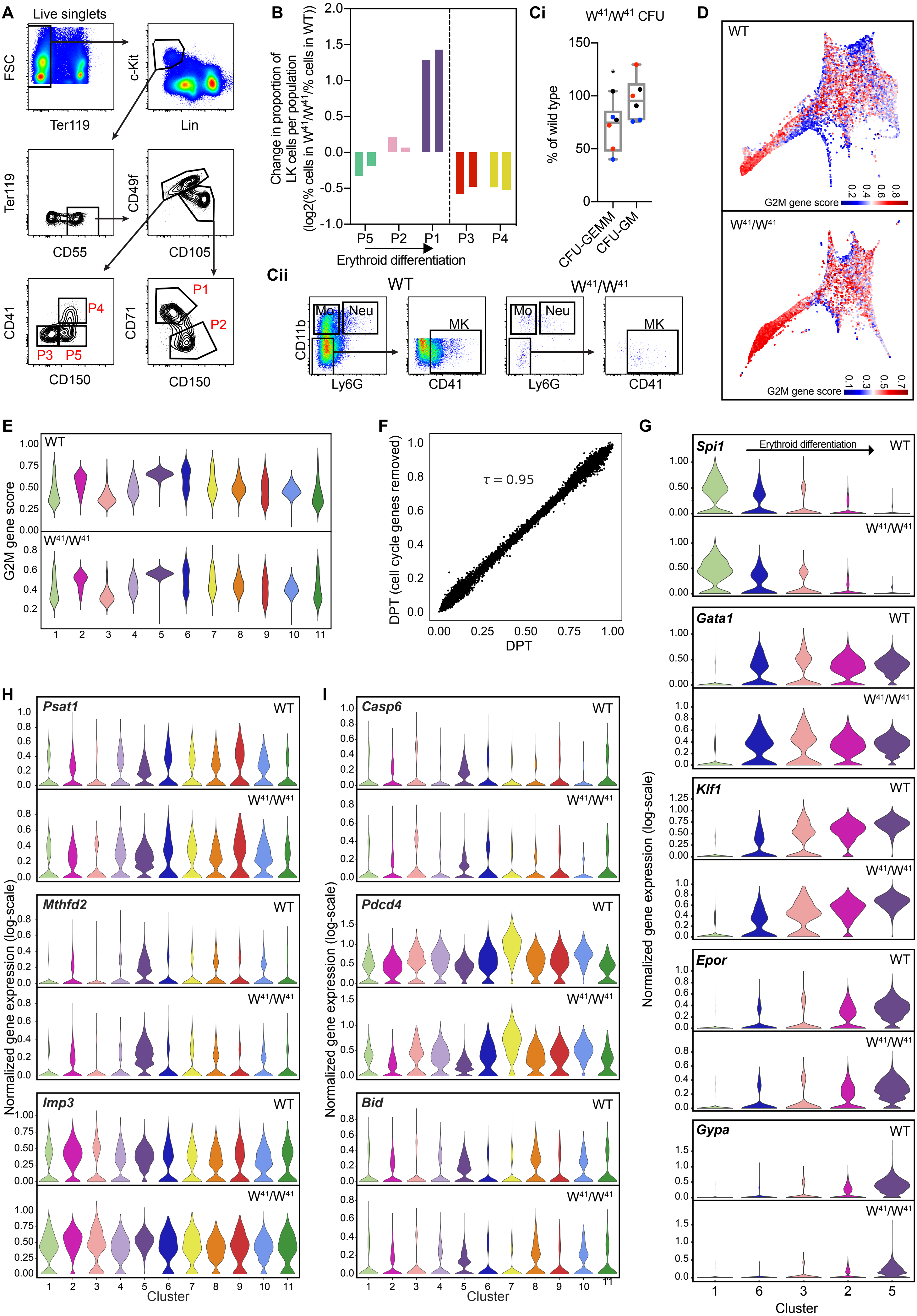

A
Live singlets
FSC
Ter119
c-Kit
Lin
Ter119
CD55
CD49f
CD105
CD41
CD150
P4
P3
P5
CD71
CD150
P1
P2
B
Change in proportion of LK cells per population (log2(% cells in W41/W41/% cells in WT))
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
P5
P2
P1
P3
P4
Erythroid differentiation
Ci
W41/W41 CFU
% of wild type
150
100
50
0
*
CFU-GEMM
CFU-GM
Cii
WT
W41/W41
CD11b
Mo
Neu
Ly6G
MK
CD41
D
WT
W41/W41
G2M gene score
0.2
0.4
0.6
0.8
0.1
0.3
0.5
0.7
E
G2M gene score
WT
W41/W41
1.00
0.75
0.50
0.25
0.00
0.8
0.6
0.4
0.2
1
2
3
4
5
6
7
8
9
10
11
F
DPT (cell cycle genes removed)
τ = 0.95
DPT
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.25
0.50
0.75
1.00
G
Spi1
Erythroid differentiation
Gata1
Klf1
Epor
Gypa
WT
W41/W41
Normalized gene expression (log-scale)
1
6
3
2
5
Cluster
H
Psat1
Mthfd2
Imp3
Normalized gene expression (log-scale)
Cluster
I
Casp6
Pdcd4
Bid
Normalized gene expression (log-scale)
Cluster

Figure S9