Mouse multipotent progenitor 5 cells are located at the interphase between hematopoietic stem and progenitor cells

Short Title: Characterization of MPP5 cells

Pia Sommerkamp[1,2,3,#], Mari Carmen Romero-Mulero[4,#], Andreas Narr[1,2,3,#], Luisa Ladel[1,2,#], Lucie Hustin[5], Katharina Schönberger[4], Simon Renders[1,2], Sandro Altamura[6], Petra Zeisberger[1,2], Karin Jäcklein[4], Daniel Klimmeck[1,2], Alejo Rodriguez-Fraticelli[7,8], Fernando D. Camargo[8,9], Leïla Perié [5], Andreas Trumpp[1,2,10,*] and Nina Cabezas-Wallscheid[1,2,4,*]

[1] Division of Stem Cells and Cancer, German Cancer Research Center (DKFZ) and DKFZ-ZMBH Alliance, 69120 Heidelberg, Germany

[2] Heidelberg Institute for Stem Cell Technology and Experimental Medicine (HI-STEM gGmbH), 69120 Heidelberg, Germany

[3] Faculty of Biosciences, Heidelberg University, 69117 Heidelberg, Germany

[4] Max Planck Institute of Immunobiology and Epigenetics, 79108 Freiburg, Germany

[5] Institut Curie, Université PSL, Sorbonne Université, CNRS UMR168, Laboratoire Physico Chimie Curie, 75005 Paris, France

[6] Department of Pediatric Hematology, Oncology and Immunology, Heidelberg University Medical Center, 69120 Heidelberg, Germany

[7] Department of Pediatrics, Harvard Medical School, Massachusetts, USA

[8] Stem Cell Program, Boston Children's Hospital, Massachusetts, USA

[9] Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, USA

[10] German Cancer Consortium (DKTK), 69120 Heidelberg, Germany

[#] these authors contributed equally

[*] Co-senior authors

Supplemental Methods

Mice

C57BL/6J (CD45.2, CD45.1 or CD45.2/CD45.1) mice were either purchased from Envigo (the Netherlands) or Janvier Labs (France) or bred in-house. For 3'-seq analysis, 6- to 12-week-old female C57BL/6J (CD45.2) mice were used. All mice were bred in-house in the animal facility of the DKFZ or MPI under specific-pathogen-free (SPF) conditions in individually ventilated cages (IVC). According to German guidelines, mice were euthanized by cervical dislocation, and all animal procedures were performed according to protocols approved by the German authorities, Regierungspräsidium Karlsruhe (Nr. A-23/17, Z110/02, DKFZ 299, G-140/13. G-183/17). To reduce animal numbers, remaining bone marrow and cDNA samples generated in this and previous studies were used whenever possible.

Cell Suspensions and Flow Cytometry

Mouse BM cells were isolated, HSC and MPP1-6 progenitors defined by immunophenotype (Lineage- Sca1+ c-Kit+ CD135-/+ CD150-/+ CD48-/+ CD34-/+) were purified by FACS and subsequently subjected to *in vivo* or *in vitro* assays or 3'-Seq analysis. For 10x analysis, LSK cells were sorted. Briefly, BM was isolated from pooled femora, tibiae, ilia and vertebrae by crushing in PBS. If no depletion of lineage-positive cells was performed, lysis of erythrocytes was performed using ACK Lysing Buffer (Thermo Fisher Scientific). To deplete lineage-positive cells, the Dynabeads Untouched Mouse CD4 Cells Kit (Invitrogen) was used. Briefly, total BM was stained for 30 min with a 1:5 dilution of the Lineage Cocktail provided in the Dynabeads Untouched Mouse CD4 Cells Kit (Invitrogen). Labeled cells were then incubated for 20 min with 1.5 mL/mouse of washed polyclonal sheep anti-rat IgG-coated Dynabeads provided in the kit. Cells were depleted using a magnet, enriching for the lineage-negative (Lineage-) cell fraction. To purify HSC and MPP1-6 cells, the Lineage- fraction was stained for 30 min using the following monoclonal antibodies: anti-lineage [anti-CD4 (clone GK1.5), anti-CD8a (53-6.7), anti-CD11b (M1/70), anti-B220 (RA3-6B2), anti-GR1 (RB6-8C5) and anti-TER119 (Ter-119)] all PE-Cy7; anti-CD117/c-Kit (2B8)-APC; anti-Ly6a/Sca-1 (D7)-APC-Cy7; anti-CD34 (RAM34)-FITC; anti-CD150 (TC15-12F12.2)-PE-Cy5; anti-CD48 (HM48-1)-PB; anti-CD135 (A2F10)-PE. To purify LSK cells, the Lineage- fraction was stained for 30 min using the following monoclonal antibodies: anti-lineage [see HSPC staining] all PE-Cy7; anti-CD117/c-Kit (2B8)-BV711; anti-Ly6a/Sca-1 (D7)-APC-Cy7. Monoclonal antibody conjugates were purchased from eBioscience or BioLegend. Cell sorting was performed on a FACS Aria I, II, III or FACS Aria Fusion (Becton Dickinson). Cells were sorted into Complete Stem Cell Medium (StemPro-34 SFM, LifeTechnologies containing 50 ng/mL SCF, 25 ng/mL TPO, 30 ng/mL Flt3-Ligand (all Preprotech), 100 u/mL Penicillin/Streptomycin, 2 mM L-Glutamine and StemPro-34

Supplement as recommended) for CFU assays, *in vitro* culture and reconstitution experiments or into lysis buffer (ARCTURUS PicoPure RNA Isolation Kit (Life Technologies, Invitrogen)) for population 3'-Seq and qRT-PCR analysis and stored at -80°C. For 10x cells were sorted in 0.4% BSA in PBS.

For cell cycle analysis, HSPC surface staining was performed on BM cells: anti-lineage [anti-CD4 (clone GK1.5), anti-CD8a (53-6.7), anti-CD11b (M1/70), anti-B220 (RA3-6B2), anti-GR1 (RB6-8C5) and anti-TER119 (Ter-119)] all AF700; anti-CD117/c-Kit (2B8)-BV711; anti-Ly6a/Sca-1 (D7)-APC-Cy7; anti-CD34 (RAM34)-FITC; anti-CD150 (TC15-12F12.2)-PE-Cy5; anti-CD48 (HM48-1)-PE-Cy7; anti-CD135 (A2F10)-PE. Cells were fixed with BD Cytofix/Cytoperm Buffer (Beckton Dickinson). Subsequently, intracellular Ki-67 (BD Biosciences) staining was performed using PermWash solution (Beckton Dickinson). Prior to flow cytometry analysis, cells were stained with DAPI (ThermoFisher).

For comparison of different gating strategies, HSPC surface staining was performed on BM cells: anti-lineage [anti-CD4 (clone GK1.5), anti-CD8a (53-6.7), anti-CD11b (M1/70), anti-B220 (RA3-6B2), anti-GR1 (RB6-8C5) and anti-TER119 (Ter-119)] all AF700; anti-CD117/c-Kit (2B8)-BV711; anti-Ly6a/Sca-1 (D7)-APC-Cy7; anti-CD34 (RAM34)-FITC; anti-CD150 (TC15-12F12.2)-PE-Cy5; anti-CD48 (HM48-1)-PB; anti-CD135 (A2F10)-PE; anti-CD229 (30C7)-APC; anti-CD244 (m2B4 (B6)458.1)-PE-Cy7. Monoclonal antibody conjugates were purchased from BD, eBioscience or BioLegend.

Reconstitution Experiments (long-term)

For reconstitution experiments, 2,000 cells per population (HSC, MPP1-6, HSC^{ST}) were sorted and transplanted into fully irradiated (2x5 Gy) B6 mice (CD45.1) together with 2x10^5 supportive total spleen cells (CD45.1/.2). Contribution of CD45.2 donor cells was monitored in PB approximately at 1 (primary only), 4, 8, 12 and 16 weeks after transplantation in primary and secondary recipients. Outcome was addressed by flow cytometry using the following monoclonal antibodies: anti-CD45.1 (clone A20)-PE-Cy7; anti-CD45.2 (104)–PB; anti-CD4 (GK1.5)-PE, anti-CD8a (53-6.7)-PE, anti-CD11b (M1/70)-AF700, anti-B220 (RA3-6B2)-PE-Cy5, anti-GR1 (RB6-8C5)-APC. For endpoint analysis of chimeras, BM stainings were performed with the following antibodies: anti-lineage [anti-CD4 (clone GK1.5), anti-CD8a (53-6.7), anti-CD11b (M1/70), anti-B220 (RA3-6B2), anti-GR1 (RB6-8C5) and anti-TER119 (Ter-119)] all AF700; anti-CD117/c-Kit (2B8)-APC-Cy7; anti-Ly6a/Sca-1 (D7)-APC; anti-CD34 (RAM34)-FITC; anti-CD150 (TC15-12F12.2)-PE-Cy5; anti-CD48 (HM48-1)-PE; anti-CD45.1 (A20)-PE-Cy7; anti-CD45.2 (104)-PB. In addition, analysis of spleen and bone marrow was performed (anti-CD4 (GK1.5)-PE-Cy5; anti-CD8a (53-6.7)-PE-Cy5; anti-CD11b (M1/70)-APC-

Cy7; anti-Gr1 (RB6-8C5)-APC; anti-Ter119 (Ter-119)-FITC; anti-CD71 (R17217)-PE; anti-B220 (RA3-6B2)-AF700; anti-CD45.1 (A20)-PE-Cy7; anti-CD45.2 (104)-PB). Monoclonal antibody conjugates were purchased from eBioscience or BioLegend.
For secondary transplantations, whole BM was isolated 16 weeks after transplantation and 3x10^6 cells were re-transplanted into fully irradiated (2x5 Gy) B6 mice (CD45.1).

Reconstitution Experiments (short-term)

For short-term transplantation experiments, 4,000 cells per population (HSC, MPP1-5) were sorted and transplanted into sub-lethally irradiated (1x5 Gy) B6 mice (CD45.1). For endpoint analysis after 1 or 2 weeks, whole BM was lineage depleted as described (see Cell Suspensions and Flow Cytometry) and BM stainings were performed with the following antibodies: anti-lineage [anti-CD4 (clone GK1.5), anti-CD8a (53-6.7), anti-CD11b (M1/70), anti-B220 (RA3-6B2), anti-GR1 (RB6-8C5) and anti-TER119 (Ter-119)] all AF700; anti-CD117/c-Kit (2B8)-BV711; anti-Ly6a/Sca-1 (D7)-APC; anti-CD34 (RAM34)-FITC; anti-CD150 (TC15-12F12.2)-PE-Cy5; anti-CD48 (HM48-1)-PE-Cy7; anti-CD45.1 (A20)-APC-Cy7; anti-CD45.2 (104)-PB; anti-CD135 (A2F10)-PE. Monoclonal antibody conjugates were purchased from eBioscience or BioLegend.

Colony-Forming-Unit Assays (CFUs)

1,000 FACS-sorted cells (HSC or MPP1-5) were sorted and cultured in MethoCult M3434 (StemCell Technologies) in technical replicates (500 cells/plate, 1 ml/plate). 7 days after plating of CFUs, colony formation was quantified. Plates were harvested by pipetting and washing using PBS, cells were washed and the cell concentration was determined using a Neubauer chamber. 30,000 cells of each replicate in 100 µl PBS were used for replating and cultured in 1 ml MethoCult M3434 (StemCell Technologies) and quantified 5 days later. For tertiary platings, 30,000 cells were replated and colonies were counted after 5 days.

In vitro ontogeny assay

2,000 cells per analysis timepoint and population per biological replicate were sorted and cultured in Complete Stem Cell Medium (StemPro-34 SFM, LifeTechnologies containing 50 ng/mL SCF, 25 ng/mL TPO, 30 ng/mL Flt3-Ligand (all Preprotech), 100 u/mL Penicillin/Streptomycin, 2 mM L-Glutamine and StemPro-34 Supplement as recommended) in 96-well ultra-low attachment plates. Analysis of cells was performed after 6 h, 16 h, 24 h, and 48 h respectively. Cells were harvested and stained for flow cytometry-based analysis (HSPC staining see Cell Suspensions and Flow Cytometry Sort Panel).

Single cell family assay

The assay was performed as described[1]. In short, bone marrow cells were isolated from female C57BL/6J mice of 8-12 weeks and MACS enriched for c-Kit+ cells. The c-Kit enriched cells were then stained in DPBS(14190-094) 10% FCS with CD135-PE (clone A2F10), Sca1 PE-CF594 (clone D7), c-Kit APC (clone 2B8), CD34 AF700 (clone RAM34, Invitrogen 56-0341-82), CD48 APC-Cy7 (clone HM48), CD150 PE-Cy5 (clone TC15-12F12.2) and Lin PE-Cy7, which was constituted of Ter119 PE-Cy7(clone TER-119), CD4 PE-CY7 (clone GK1.5), CD8a PE-CY7 (clone 53-6.7), CD11b PE-CY7 (clone M1/70), Gr1 PE-CY7 (clone RB6-8C5), B220 PE-CY7 (clone RA3-6B2). Monoclonal antibody conjugates were purchased from BD, eBioscience or BioLegend unless otherwise indicated. Subsequently, cells were divided in 4 equal fractions and stained with CTV and/or CFSE before sorting them directly into a 96 well U-bottom plate (Falcon 353077) containing 50 µl Culture Medium (see *In vitro* ontogeny assay) using an Aria III cell sorter (BD Biosciences). For each cell type of interest (HSC, MPP1, MPP2, MPP3, MPP4, MPP5), 4 single cells, 1 cell from each CTV/CFSE fraction, were sorted per well using index sorting. Additionally bulks, constituted of 2,000 LSK and 2,000 Lin- c-Kit+ Sca1- cells, were sorted from each CTV/CFSE fraction to set the gating strategy after cell culture. The sorted cells were then incubated at 37°C 5% CO2 and after 24h, all cells were stained with the same antibody panel as for sorting, and analyzed using a ZE5 Flow cytometer (BioRad). FlowJo was used to perform the analysis. The phenotypic and cell tracer gates were set on the bulks of each CTV/CFSE fraction, then applied to all single cells wells. Since only one cell of each CTV/CFSE fraction was plated per well, all cells collected with the same CTV/CFSE signature from a given well are considered to have the same ancestor and constitute a "family". Impossible family single cells results, such as the presence of 2 cells of the same family CTV/CFSE fraction that have not divided in one well, were removed from the analysis.

qRT-PCR Analysis

For Real-time PCR, total RNA of 5,000-10,000 cells was isolated (ARCTURUS PicoPure RNA isolation kit (Life Technologies, Invitrogen)) and reverse-transcribed using SuperScript VILO cDNA Synthesis Kit (Invitrogen) according to the manufacturer's guidelines. For qRT-PCR analysis, Fast SYBR Green Master Mix was used on a ViiA 7 Real-Time PCR System (Applied Biosystems). RNA expression was normalized to *Oaz1* or *Gapdh* housekeeping gene expression and presented as relative quantification (Ratio = 2^-ΔΔCT). Primers were designed using the Universal ProbeLibrary Assay Design Center (Roche) or ncbi Primer-BLAST (ncbi). For list of primers see supplemental table 4.

RNA-seq

Generation

RNA-seq data of a previous study was used[2]. For MPP5, the RNA-seq dataset was generated as previously described[2]. Briefly, total RNA isolation was performed using the ARCTURUS PicoPure RNA isolation kit (Life Technologies, Invitrogen) according to the manufacturer's instructions. Total RNA was used for quality controls and for normalization of the starting material. Sequencing libraries were generated with approximately 3 ng of total RNA for HSC/MPP using the QuantSeq 3' mRNA-Seq Library Prep Kit REV (Lexogen), 3'-seq method, according to the manufacturer's instructions except for adaptations (version 015UG009V0221: Do not skip step 2; 15 minute incubation time step 4; cycle number was adjusted to 19 using the PCR Add-on Kit for Illumina (Lexogen)). Sequencing was performed with a HiSeq2000 device (Illumina) in paired-end mode reading 125 nucleotides. As the MPP5 libraries were sequenced together with the previously described HSC and MPP1-4 libraries in the same multiplex, no batch correction was necessary.

Downstream Analysis

Libraries were demultiplexed according to barcodes. Quality control of the fastq files was performed with the fastqc package and adapter trimming as well as quality related trimming was performed using the Cutadapt package[3]. For RNA-Seq gene expression analysis, the STAR package[4] was used to map paired-end libraries against the mm10 mouse genome. Reads were annotated and counted using the featureCounts package[5]. For determination of differential expressed genes, the DESeq2 package was used[6]. Genes with FDR < 0.1 were considered as differentially expressed. Hallmark terms were calculated with the package fgsea[7]. Pathways with FDR < 0.05 were considered as differentially regulated. For cluster generation, DEGs were assigned to the cluster with its highest expression in relation to its mean expression in all clusters as previously described[8].

Analysis of differentially used alternative polyadenylation sites and 3'-UTR length were performed as previously described[2].

<u>scRNAseq sequencing</u>

Generation

SiC RNA-sequencing was performed on the 10X Genomics platform using the Single Cell 3' Reagent Kit V3 (10x Genomics) following the manufacturer's instructions. Briefly, 20,000 LSK (CD4/CD8a/CD11b/GR1/B220/TER119- lineage negative, Sca-1+, c-Kit+) cells were pooled and sorted from 2 3-months-old female C57Bl/6J mice into 0.4% BSA in PBS. Cells were loaded according to the manufacturer's instructions aiming for a targeted cell recovery of 8,000 cells. The quality of the obtained cDNA library upon adapter ligation and index PCR (13 cycles),

was assessed by Bioanalyzer fragment analysis (HS DNA Kit, Agilent). Sequencing was performed on a HiSeq3000 device (Illumina).

Downstream Analysis

Raw UMI-based data files were mapped against the mm10 reference genome using the scRNAseq tool from the bioinformatics pipeline snakePipes, with the 10xV3 mode[9]. In this tool, STARsolo was used to i) map, ii) UMI-deduplicate and iii) count reads, to create the BAM files and a Seurat object with the gene counts[4]. The quality of this data was checked by running Deeptools QC[10]. After the preprocessing of data, R package Seurat was used to perform the scRNAseq analysis[11]. The Seurat object was imported and cell filtering was performed, selecting the cells which i) contained more than 35,000 counts and ii) expressed between 1,000 and 6,000 genes to avoid analysis of doublets or empty droplets. Moreover, low-quality and dying cells with a percentage of mitochondrial mRNA higher than 10% were filtered out. A log-normalization and scaling of the data was applied prior to the linear dimension reduction with the Principal Component Analysis (PCA). After applying the JackStraw and Elbow plot procedures from Seurat, a clustering analysis was performed selecting the first 20 PCs, resulting in 11 communities that were visualized with the Uniform Manifold Approximation and Projection (UMAP) technique[12]. From those, the last four defined clusters were filtered out based on the doublets scoring calculated with the doubletCells function from scran package[13], the enrichment of LSK and LS-K signatures described previously[14] applying the AddModuleScore function and the presence of differentiated cells markers, detected with the FindAllMarkers function (min.pct = 0, logfc.threshold = 0.15). A total of 5,520 LSK cells distributed in 7 clusters were further analyzed. The enrichment of the MolO and NoMO signatures[15] and those from the HSC/MPP1/2/3/4/5 signatures from the bulk RNA-seq analysis previously performed[2] and expanded in this paper were represented with the FeaturePlot and DotPlot functions. Other genes of interest were represented in the UMAP graph. Cell cycle phases assignment was performed by calculating the G1 and G2M scores defined by the cyclone function from scran package, using default parameters[16]. Lineage trajectory analysis was analyzed using Slingshot package[17] with the PCA embedding and represented in the 2-dimensional UMAP graph. The diffusion pseudotime (DPT) analysis was performed by applying the DiffusionMap function from the destiny package[18] to the log normalized counts, and represented in a diffusion map.

To validate that our bulk RNA-seq gene signatures provide a meaningful annotation in a LSK single cell landscape, a published dataset was analyzed[19]. In the quality control step, i) cells with less than 10 % mitochondrial mRNA, ii) those that contained more than 30,000 counts or iii) expressed between 1,000 and 5,500 genes were selected. After the preprocessing step,

five clusters were not considered for further analysis based on the doublets scoring and the high enrichment of LS-K signatures, leading to analysis of 21,218 LSK cells belonging to 8 different clusters. The annotation of HSC/MPP1/2/3/4/5 populations based on the enrichment of our bulk RNA-seq signatures was shown in a UMAP figure. A high consistency was shown in the results, which further confirmed the high quality of our analysis methods and sequencing data.

Reanalysis of LARRY dataset

Processed data for LARRY barcoding[20] was obtained from GSE140802 and reanalyzed using Scanpy v.1.6. Briefly, normalized count matrices, metadata and clonal information were imported. Genes were filtered (min_counts=100), and HSPC signatures were calculated with the sc.tl.score_genes() function using the gene signatures obtained from bulk analysis. The signatures were partially filtered to exclude genes that were poorly expressed in HSPCs. We then subselected the day 2 cells, identified variable genes sc.pp.highly_variable_genes(), calculated principal components sc.pp.pca(), computed neighborhood graphs sc.pp.neighbors with n_neighbors=10, and created UMAP representations with default parameters. Leiden clustering was performed using a resolution=1. We then calculated the average signature score for each cluster, and calculated the z-score across clusters. For each cluster, the HSPC identity was assigned by choosing the HSPC signature with the top z-score. At the end of this procedure, the MPP1 identity was consistently not assigned to any clusters. To estimate a cluster distribution for MPP1s, we pooled the top 3 clusters with the highest z-score, which shared identity with HSC and MPP2. For each cluster that shared identities, clones were randomly shuffled between the two HSPC identities (i.e. MPP1 and MPP2).

For *in vitro* label transfer quantification, clones that appeared in multiple identities at the initial timepoint (day 2) and clones that did not have a pair at the second time point (day 4) were excluded. For each HSPC identity, and for each clone at day 2, the distribution of identities at day 4 was calculated to estimate the fate transitions. These were then summed for each HSPC identity at day 2, and the fraction of transitions corresponding to each day 4 fate is represented as a heat map.

For *in vivo* fate analysis, clones that appeared in multiple identities at the initial timepoint (day 2) and clones that did not have a pair at the second time point (day 9-16) were excluded. Data from day 9-16 were pooled. For each HSPC identity, and for each clone at day 2, the distribution of fates *in vivo* was calculated as follows. First, for each fate, the fraction of cells corresponding to each clone *i* was calculated:

$$f_{i,fateA} = \frac{\pi_{i,fateA}}{\sum \pi_{fateA}}$$

Then, for each clone, fate bias was calculated for each fate:

$$P_{i,fateA} = \frac{f_{i,fateA}}{f_i}$$

Then, for each HSPC identity, the fate bias was averaged between all clones corresponding to that identity.

Quantification and Statistical Analysis

Statistical analysis was performed by unpaired Student's t test or two-way ANOVA without correction for multiple comparison (Fisher LSD test). All data are presented as mean +SD. Please see figure legends for detailed information. Significance levels were set at p* < 0.05, p** < 0.01 and p*** < 0.001. For statistical analysis, GraphPad Prism was used.

References

1. Tak T, Prevedello G, Simon G, Paillon N, Duffy KR, Perié L. Simultaneous tracking of division and differentiation from individual hematopoietic stem and progenitor cells reveals within-family homogeneity despite population heterogeneity. *bioRxiv*. 2019:586354.
2. Sommerkamp P, Altamura S, Renders S, et al. Differential Alternative Polyadenylation Landscapes Mediate Hematopoietic Stem Cell Activation and Regulate Glutamine Metabolism. *Cell Stem Cell*. 2020.
3. Martin M. CUTADAPT removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal*. 2011;17.
4. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
5. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930.
6. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
7. Korotkevich G, Sukhov V, Sergushichev A. Fast gene set enrichment analysis. *bioRxiv*. 2019:060012.
8. Cabezas-Wallscheid N, Klimmeck D, Hansson J, et al. Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell*. 2014;15(4):507-522.
9. Bhardwaj V, Heyne S, Sikora K, et al. snakePipes: facilitating flexible, scalable and integrative epigenomic analysis. *Bioinformatics*. 2019;35(22):4757-4759.
10. Ramirez F, Ryan DP, Gruning B, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016;44(W1):W160-165.
11. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol*. 2015;33(5):495-502.
12. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*. 2018;3(29):861.
13. Lun AT, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res*. 2016;5:2122.
14. Klimmeck D, Cabezas-Wallscheid N, Reyes A, et al. Transcriptome-wide profiling and posttranscriptional analysis of hematopoietic stem/progenitor cell differentiation toward myeloid commitment. *Stem Cell Reports*. 2014;3(5):858-875.
15. 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.
16. Scialdone A, Natarajan KN, Saraiva LR, et al. Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods*. 2015;85:54-61.

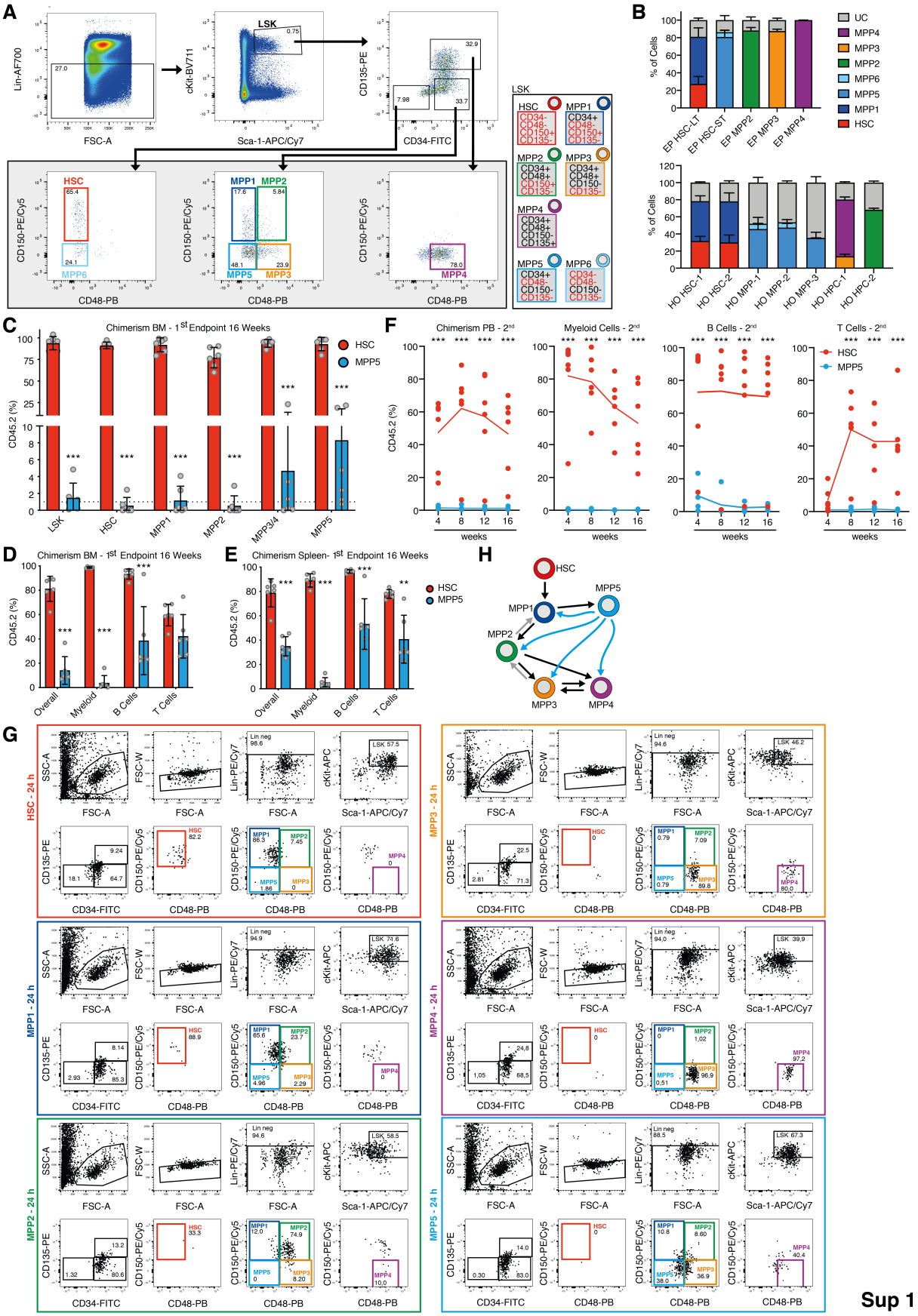
17. Street K, Risso D, Fletcher RB, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*. 2018;19(1):477.
18. Angerer P, Haghverdi L, Buttner M, Theis FJ, Marr C, Buettner F. destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics*. 2016;32(8):1241-1243.
19. Dong F, Hao S, Zhang S, et al. Differentiation of transplanted haematopoietic stem cells tracked by single-cell transcriptomic analysis. *Nat Cell Biol*. 2020.
20. Weinreb C, Rodriguez-Fraticelli A, Camargo FD, Klein AM. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science*. 2020;367(6479).
21. Pietras EM, Reynaud D, Kang YA, et al. Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. *Cell Stem Cell*. 2015;17(1):35-46.
22. Oguro H, Ding L, Morrison SJ. SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. *Cell Stem Cell*. 2013;13(1):102-116.
23. Dahlin JS, Hamey FK, Pijuan-Sala B, et al. A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit mutant mice. *Blood*. 2018;131(21):e1-e11.

Supplemental Tables

Supplemental Table 4: List of Primers

target	forward	reverse
CD28	gttttgggcactggtcgt	tttgaaggagtctgttccttctact
CD69	cccttgggctgtgttaatagtg	aacttctcgtacaagcctggg
CD74	aggcaggaactgggtcaag	gggaacataagaagggctga
Cdk1	gggcctatggtgttccagta	cgacagcccacatgtcaat
Evi1/Mecom	agttttccccgatctgcaa	ccttgggacactgatcacact
Gprc5c	gagatggccctgatgcac	cagggttgagttggcactg
h19	cggtgtgatggagaggaca	agacggcttctacgacaagg
Hk1	tcccagatgggactgagc	ggactcggaaattcgttcct
Hmga1	gcagacccaagaaactggag	ggcactgcgagtggtgat
Hmga2	cgttcagaagaagcctgctc	ccaactgatgctgaggtagaaa
Ifitm1	tgagatctccacgcctgac	ccaccatcttcctgtcccta
Igf2bp2	gggaaaatcatggaagttgacta	cgggatgttccgaatctg
Ldha	tccgttacctgatgggagag	gcaacattcacaccactcca
Ldhb	acaagtgggtatggcatgtg	acatccaccagggcaagtt
Lin28b	ctctggagtttgaagctgagg	gcacgttgaaccatttacagtg
Mecp2	tggtagctgggatgttaggg	ttgtagtggctcatgcttgc
Meg3	cgaggacttcacgcacaac	attccagatgatggctttgg
Neo1	tgaaacttttgaaagcgacct	agcggacggacatgaaga
Ngp	gcctaaagactgcgacttcc	tgaagaatttccctgtgcaa
Notch1	actatctcggcggcttttc	ctcctcggagcagttagacc
Nusap1	gattgcagaacgcgatgac	aaggcttttaacaacttgtctgc
Pbk	ccttaatcacccaaacattatagga	agacttttcacctccatactcca
Plk1	ttgtagttttggagctctgtcg	agtgccttcctcctcttgtg
Pygb	cgggtggaagatgtcgag	aggcaggcgctcatagaat
Pygl	cagaagatccgagagggatg	aagggtttccatgcctgag
Pygm	agtggaggacgtggaaagg	gctcaggaattcggtcgtag
s100a9	gacaccctgacaccctgag	tgagggcttcatttctcttctc
Serpinb6a	ggaagagctggactttcagg	tggagacagcacctctttga
Tpx2	ctgagatgtggaagcaccag	ggaagatgacggtgtttgga
Vwf	caggggctgcagttatc	ctcattctcttgccatcttgg

Characterization of MPP5 cells



Sup 1

Supplemental Figure 1: Lineage Potential of MPP5 Cells and HSC/MPP Ontogeny Analysis

(A) Representative FACS dot plots of HSC/MPP gating approach.

(B) Comparison of HSC/MPP gating strategies. Respective populations were gated as previously described by Eric Pietras et al.[21] (EP) or Hideyuki Oguro et al.[22] (HO) and the presence of HSC and MPP1-MPP6 cells within those compartments was determined. Uncharacterized cells (UC) did not fall into one of the HSC/MPP gates as defined in this paper, e.g. gaps between gates (see Supplemental Figure 1A). n=3

(C) Endpoint analysis of primary recipient animals. HSPC engraftment in the BM is shown. n=6

(D) Endpoint analysis of primary recipient animals. Engraftment of differentiated cells in the BM is shown. n=6

(E) Endpoint analysis of primary recipient animals. Engraftment of differentiated cells in the Spleen is shown. n=6

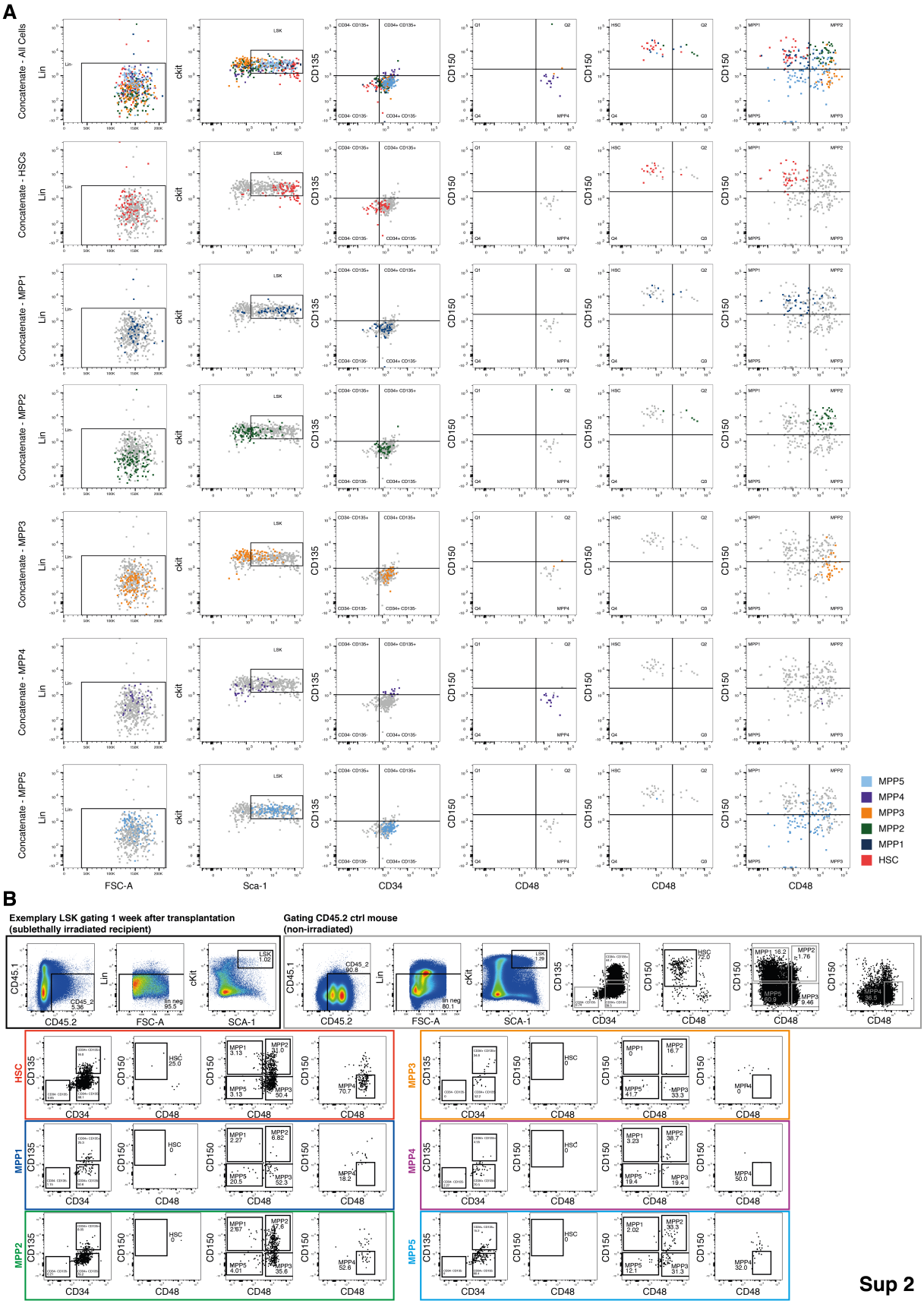
(F) Analysis of secondary recipients of HSC and MPP5 transplantations. The relative percentage of donor contribution to the peripheral blood, myeloid, B cell and T cell lineage is shown. Mean is shown. n=5-6 (16 weeks 2nd Transplantation MPP5 n=4)

(G) Representative FACS dot plots of *in vitro* ontogeny analysis.

(H) Model showing potential HSC/MPP ontogeny.

For all panels mean +SD is shown. Panel (C,D,E) unpaired student's t-test. (F) two-way ANOVA. *p <0.05; **p <0.01; ***p <0.001. n.s., not significant. n indicates number of biological replicates. (B-F) 1 independent experiment.

Characterization of MPP5 cells



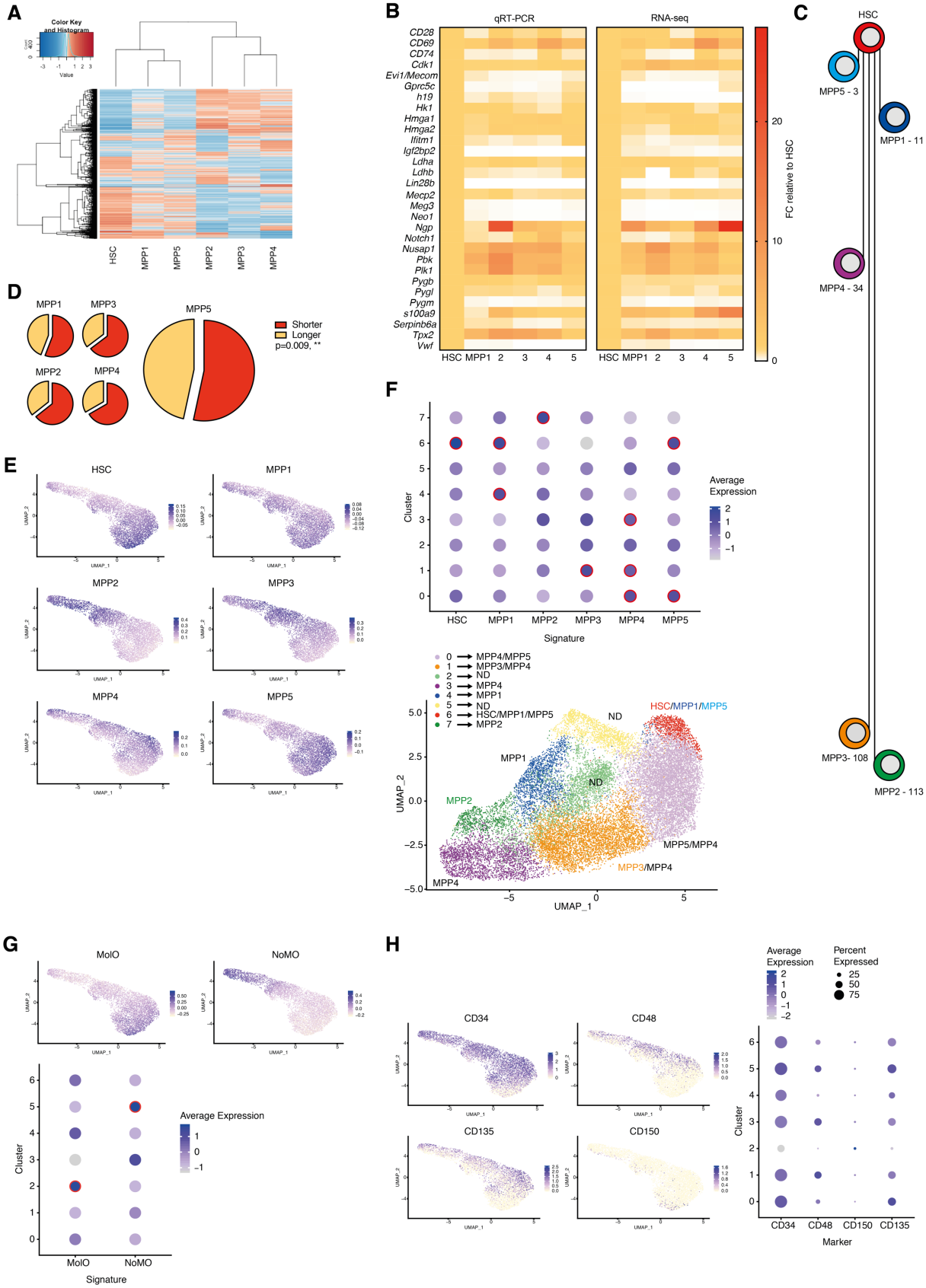
Sup 2

Supplemental Figure 2: Lineage Potential of MPP5 Cells and HSC/MPP Ontogeny Analysis

(A) Representative FACS dot plots of high throughput simultaneous division and differentiation tracking per-ancestor analysis. Analysis of surface marker phenotype 24 h after sort and *in vitro* culture.

(B) Representative FACS dot plots of in vivo short-term transplantation data. Analysis of surface marker phenotype 1 week after transplantation.

Characterization of MPP5 cells



Sup 3

Supplemental Figure 3: Molecular analysis of HSPCs and LSK cells on the population and single-cell level

(A) Clustered heat map. The colors represent the normalized average read count in each of the six cell populations for all differentially expressed genes (FDR = 0.1).

(B) Heatmap representing normalized mean relative expression of genes analyzed by qRT-PCR (left panel) and RNA-seq (right panel).

(C) Identification of differentially used alternative polyadenylation sites. HSC and MPP1-4 data was originally published in[2].

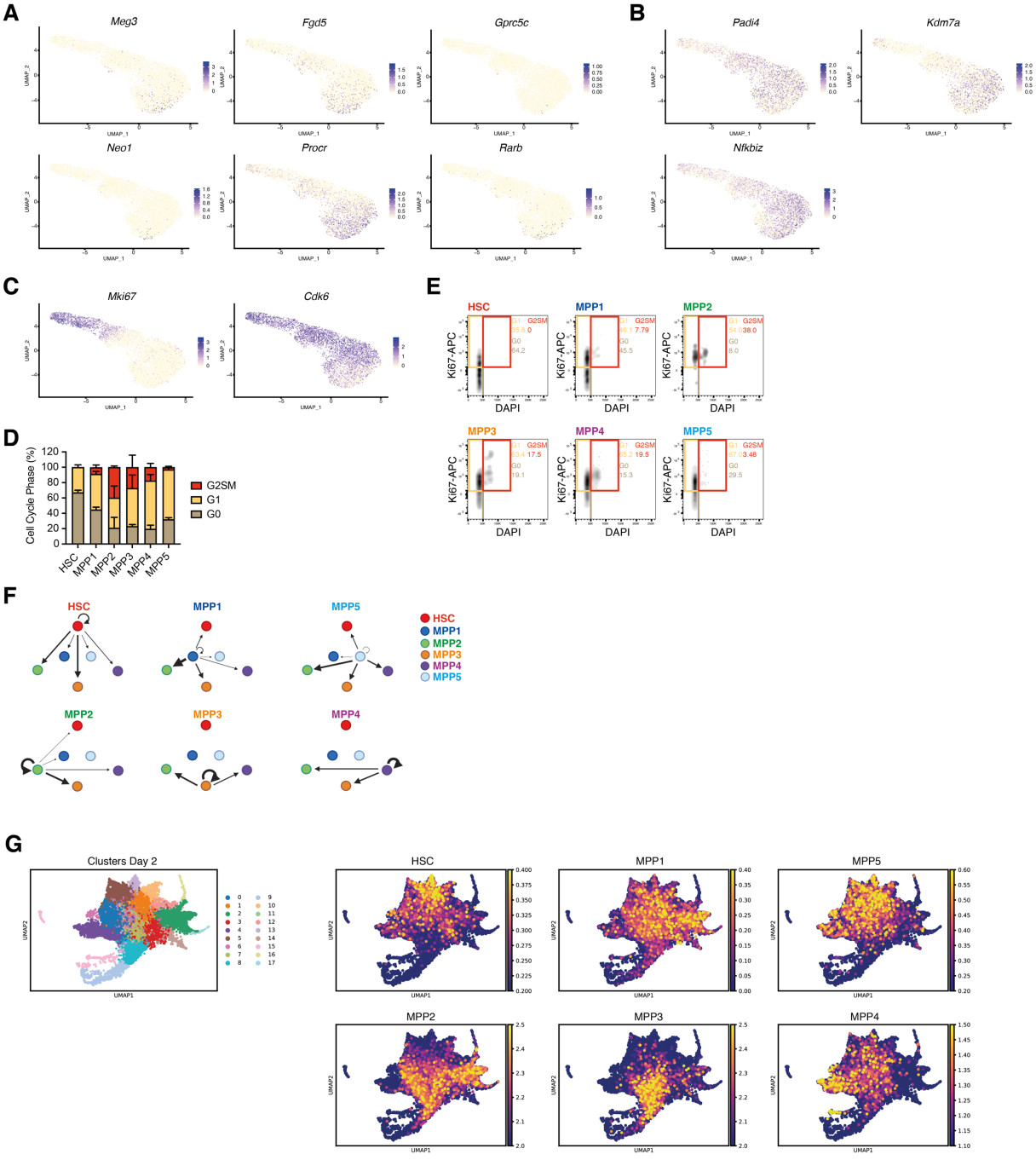
(D) Analysis of 3'-UTR length considering changes in 3'-UTRs belonging to the same exon (UTR-APA only). Binomial test. HSC and MPP1-4 data was originally published in[2].

(E) UMAP projection of HSPC gene signatures derived from the population RNA-seq data analysis. Colors encode signature enrichment.

(F) Upper panel: Expression of HSPC gene signatures in the respective clusters. Red circles indicate maximal enrichment +/- 0.02 considering the signature scoring. Analysis of maximal enrichment was used to assign colors and MPP identity to the respective clusters. Lower panel: UMAP projection. Clusters are color-coded based on enrichment scores and clusters were associated to the respective gene signatures identified in the population RNA-seq analysis. Data was derived from a previously published study[23]. ND: not enriched for any HSPC signature.

(G) UMAP projection of MolO and NoMO gene signatures[15] and expression in the respective single cell clusters. Colors are based on enrichment scores. Red cycles indicate maximal enrichment.

(H) UMAP projection and expression in the respective single cell clusters of HSPC surface markers.


A
Meg3
Fgd5
Gprc5c
Neo1
Procr
Rarb
B
Padi4
Kdm7a
Nfkbiz
C
Mki67
Cdk6
D
Cell Cycle Phase (%)
G2SM
G1
G0
HSC
MPP1
MPP2
MPP3
MPP4
MPP5
E
Ki67-APC
DAPI
F
G
Clusters Day 2
UMAP1
UMAP2

Sup 4

Supplemental Figure 4: Molecular analysis of HSPCs and LSK cells on the population and single-cell level

(A, B, C) UMAP projection and representation of single genes including HSC and MPP5 marker and regulators (A, B) and cell cycle regulators (C). Colors encode standardized gene expression.

(D) Flow cytometry-based cell cycle analysis of HSPCs. Mean +SD is shown. n=5 (HSC, MPP1, MPP3-5), n=3 (MPP2). 1 independent experiment.

(E) Representative FACS dot plots of HSC/MPP cell cycle analysis. Contour plot including outliers is shown.

(F) Representation of clone label transference. Maps showing each population connected by arrows with a width determined by the fraction of day 2 clones transferred to the linked compartment at day 4. Transfers less than 0.05 are not shown.

(G) UMAP projection of day 2 sc profiling. Clusters are color-coded based on enrichment scores of the respective gene signatures identified in the population RNA-seq analysis.