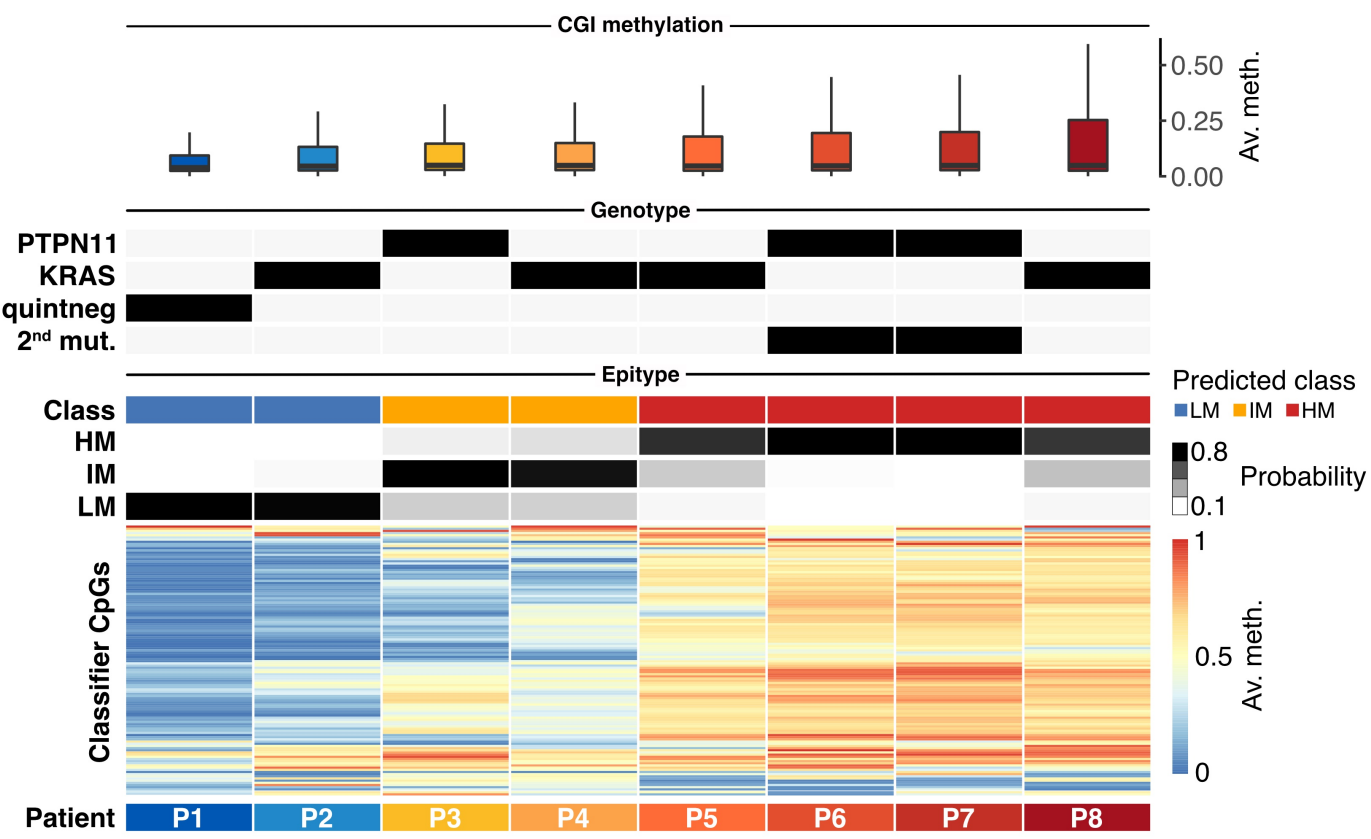


1    **Extended data figures and tables**

2    Hartmann & Schönung et al. 2023

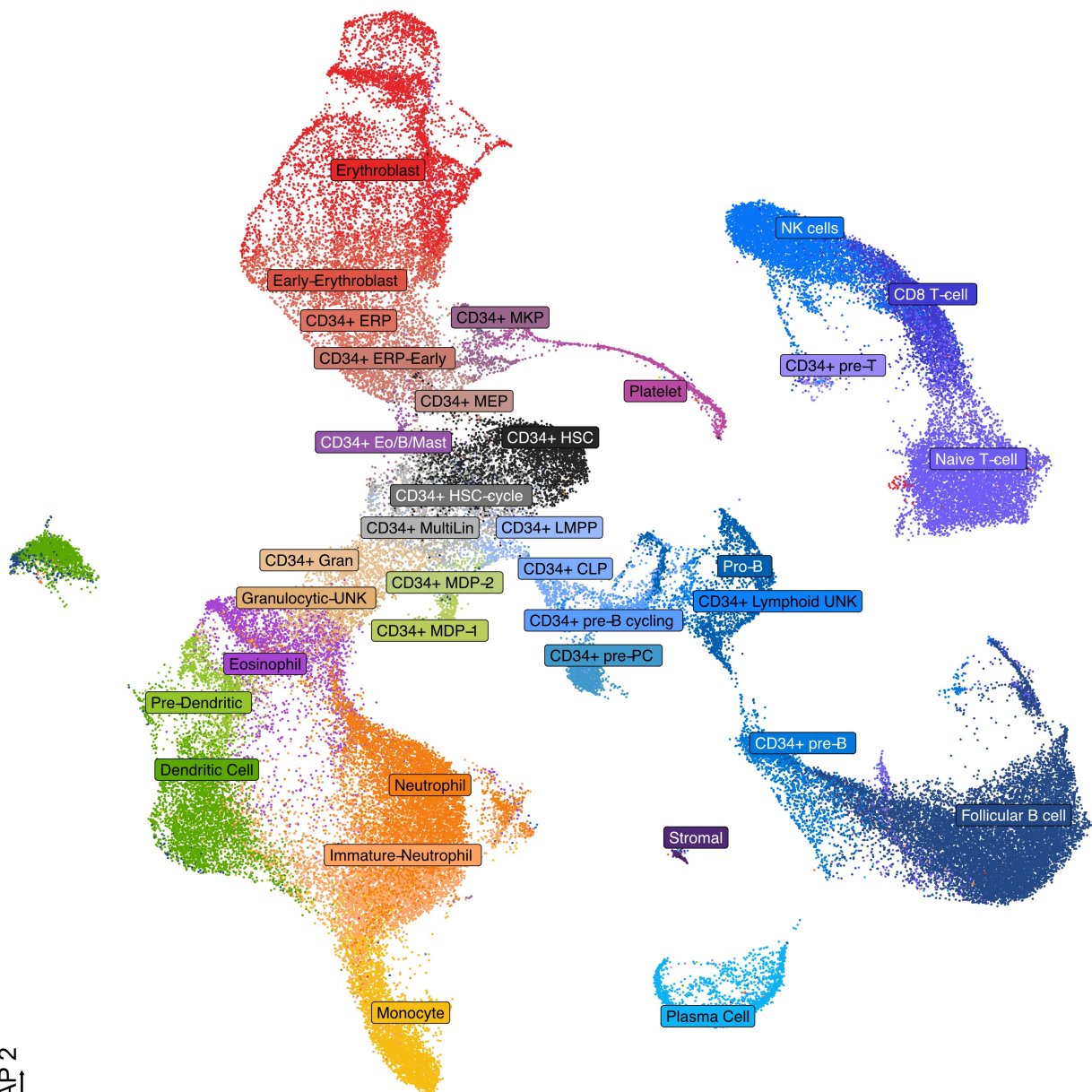
3    **Supplementary information**

4    Excel file 'Supplementary\_Tables" contains Supplementary Tables 1 to 27, including scRNA-  
5    seq datasets, WGBS datasets, flow cytometry samples, murine Ptpn11<sup>E76K</sup> scRNA-seq  
6    datasets, PDX samples, used transcription signatures from literature, and antibodies used  
7    for cell sorting or flow cytometry analyses. Extended Data Figures 1 to 12 can be found  
8    below.



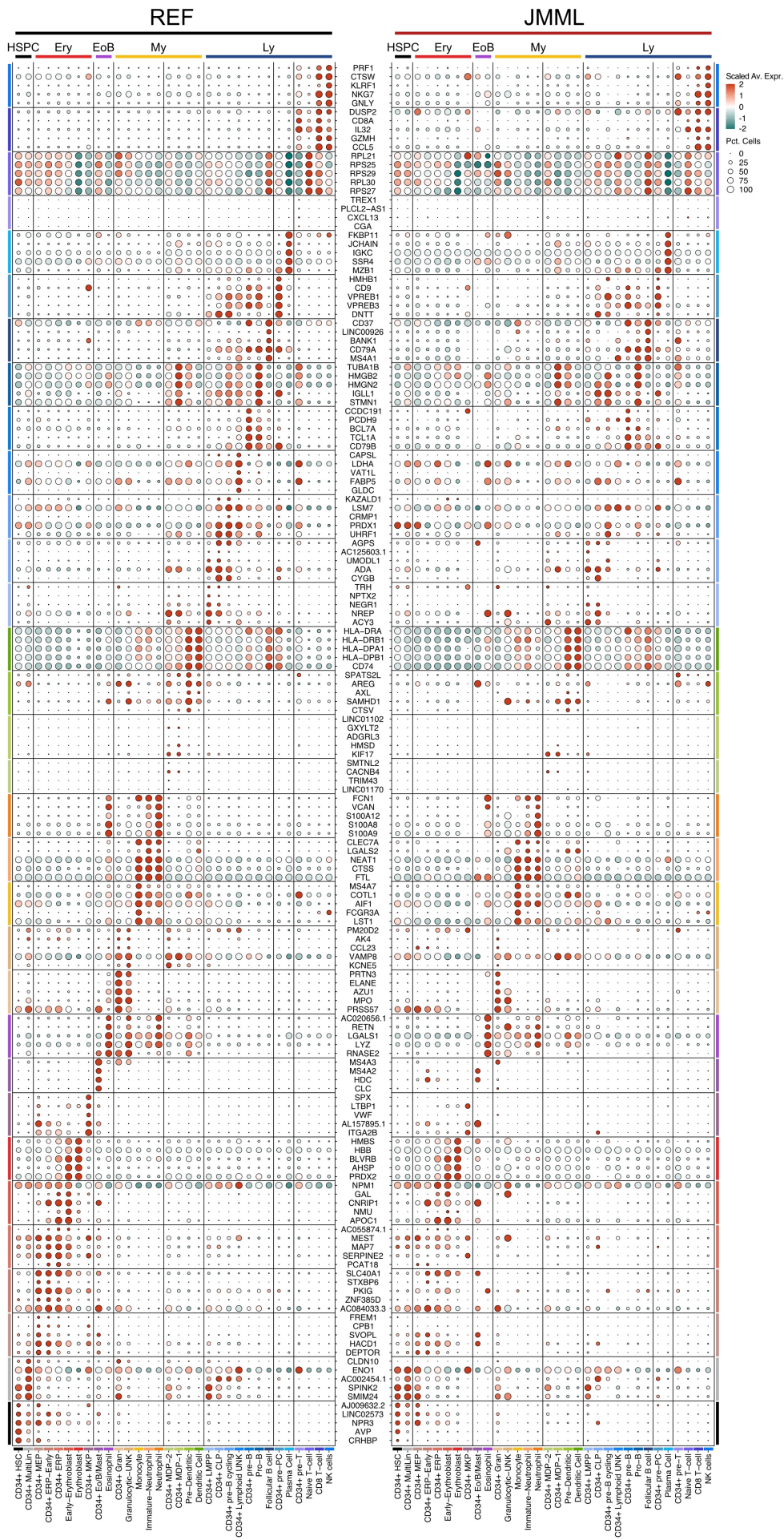
**Extended Data Figure 1: Multi-omics cohort: annotation of patient cohort to JMML DNA methylation subgroups ('epitypes').** Patient stratification was conducted using our published JMML DNA methylation classifier<sup>1</sup>. Heatmap shows average methylation values across 124 classifier CpGs. Based on the probability per JMML subgroup, patients were assigned to one of three subgroups, namely LM (low methylation, blue), IM (intermediate methylation, yellow), or HM (high methylation, red). The annotation tracks specify the genetic background of the patients, which was determined by JMML panel sequencing. As boxplots, average methylation of genome-wide CpG island (CGI) methylation values are summarized. Based on those values and in a semi-quantitative manner, patients were annotated to a patient-specific color throughout the manuscript.

UMAP 2  
UMAP 1





**Extended Data Figure 2: UMAP summarizing the scRNA-seq data of total hematopoiesis in JMML, neonatal cord blood, as well as juvenile and adult bone marrow.** Single-cells are colored by cell types, which were annotated by automated cell type annotation using Seurat's 'label transfer' function with adult bone marrow as a reference. JMML (n=7) and juvenile bone marrow (n=2) data were generated for this study. Neonatal cord blood and adult bone marrow data were taken from the Human Cell Atlas project<sup>2</sup>. The same colors are used for the corresponding cell types throughout the manuscript.



**Extended Data Figure 3: Dot plot summarizing the expression of the top 4 to 5 hematopoietic cell type marker genes as published in the human cell atlas project<sup>2</sup>.**  
Average expression is scaled per gene. Dot sizes represent the percentage of cells per cell type expressing the corresponding gene. Expression values are separately analyzed for JMML and healthy references ('REF'), including neonatal cord blood, as well as juvenile and adult bone marrow.

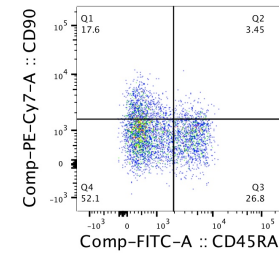
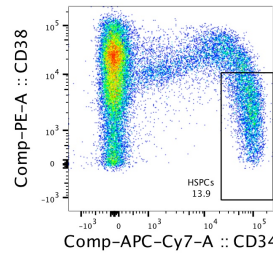
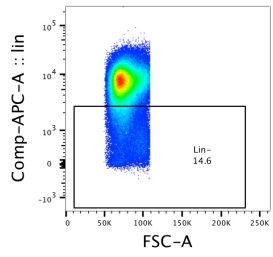
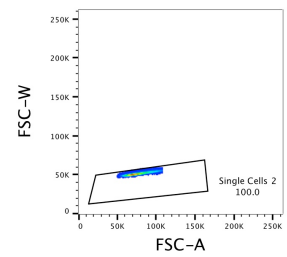
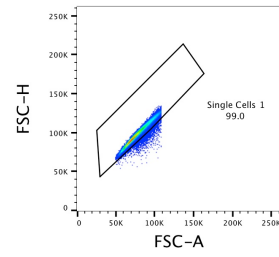
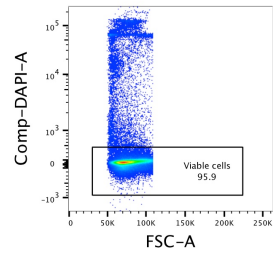
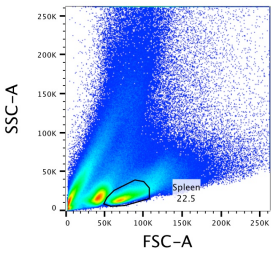


35 **Extended Data Figure 4: Density plots depicting the cell type-specific expression of**  
36 **selected marker genes. a, Hematopoietic stem and progenitor cells (HSPCs). b,**  
37 **Erythrocytes. c, Megakaryocytes. d, Dendritic cells. e, Monocytes and granulocytic**  
38 **neutrophils. f, B cells. g, Plasma cells. h, NK and T cells. i, Eosinophils.**  
39



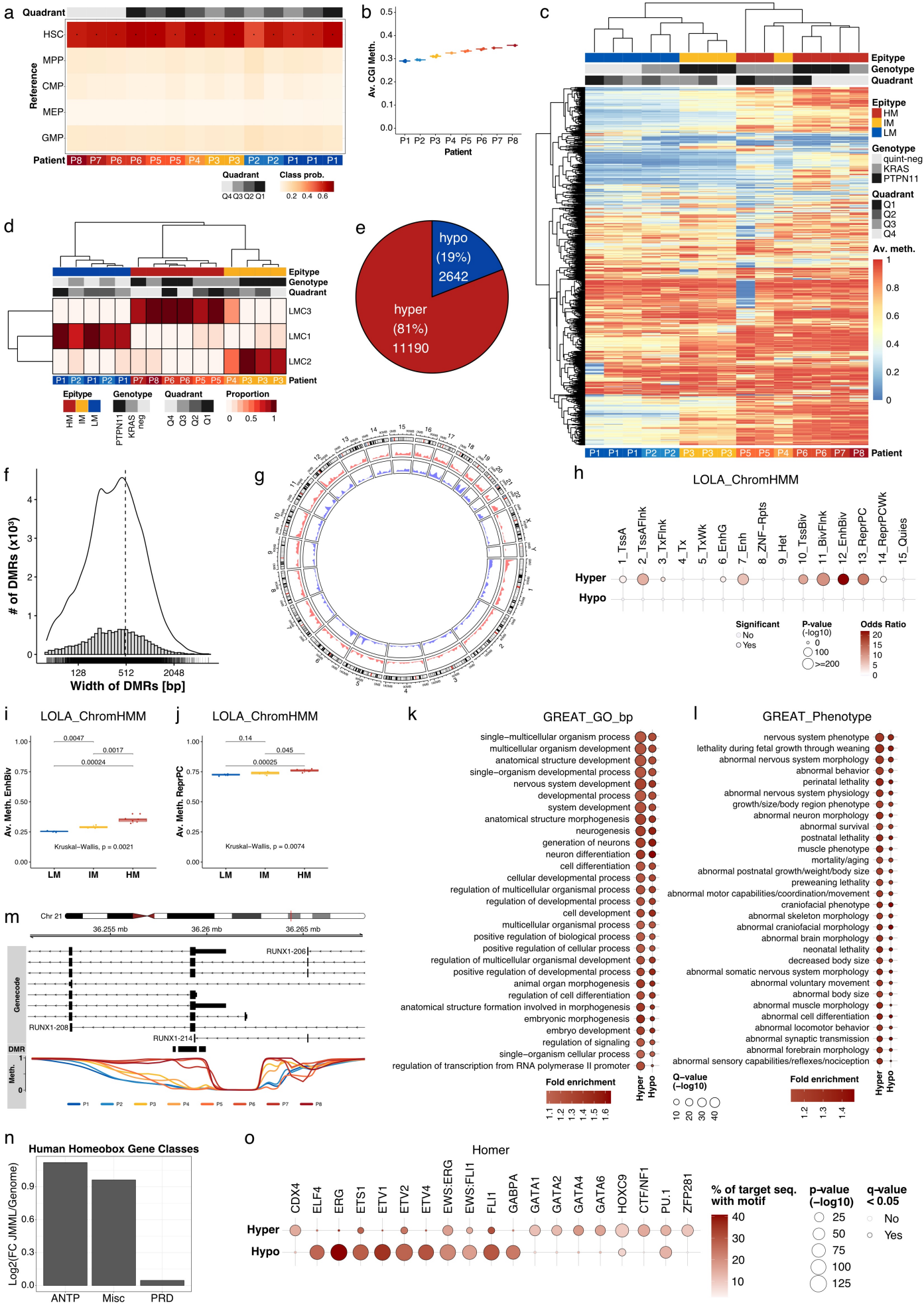


**Extended Data Figure 5: Principal component analyses (scRNA-seq data) of major hematopoietic lineages stratified by healthy references and JMML.** All analyses are ordered in columns per lineage as labeled in b. **a**, Dotplot of principal components. Depicted is the summary of the correlation of principal components with technical and biological features. nCount: number of transcripts/cell; nFeature: number of genes/cell; Cell cycle: cell cycle phase score; Patient: donor of primary material; Subgroup: JMML epitype; Genotype: type of RAS-pathway mutation; Disease: disease status, i.e. healthy or JMML; Cell type: annotated hematopoietic cell type. For every lineage, PC 1 was representative for differentiation (colored per lineage). Additionally, at least one PC was representative for disease-specific features (red) among the first six PCs. Those PCs were used for 2-D PCA plot visualizations (Figures 1g & ED5c,d). **b**, Detailed analysis of differentiation-specific PCs per lineage. From top to bottom: stacked bar plot of cell type contributions over the course of PC1; PC loadings and scaled normalized expression of top genes from PC1. **c**, PCA plot colored by cell types. **d**, PCA plot colored by disease state (JMML in red vs. healthy references in grey). Density plots of cell type distribution across PC1 or disease state for the second dimension. PCA plot for neutrophil lineage in Figure 1g.

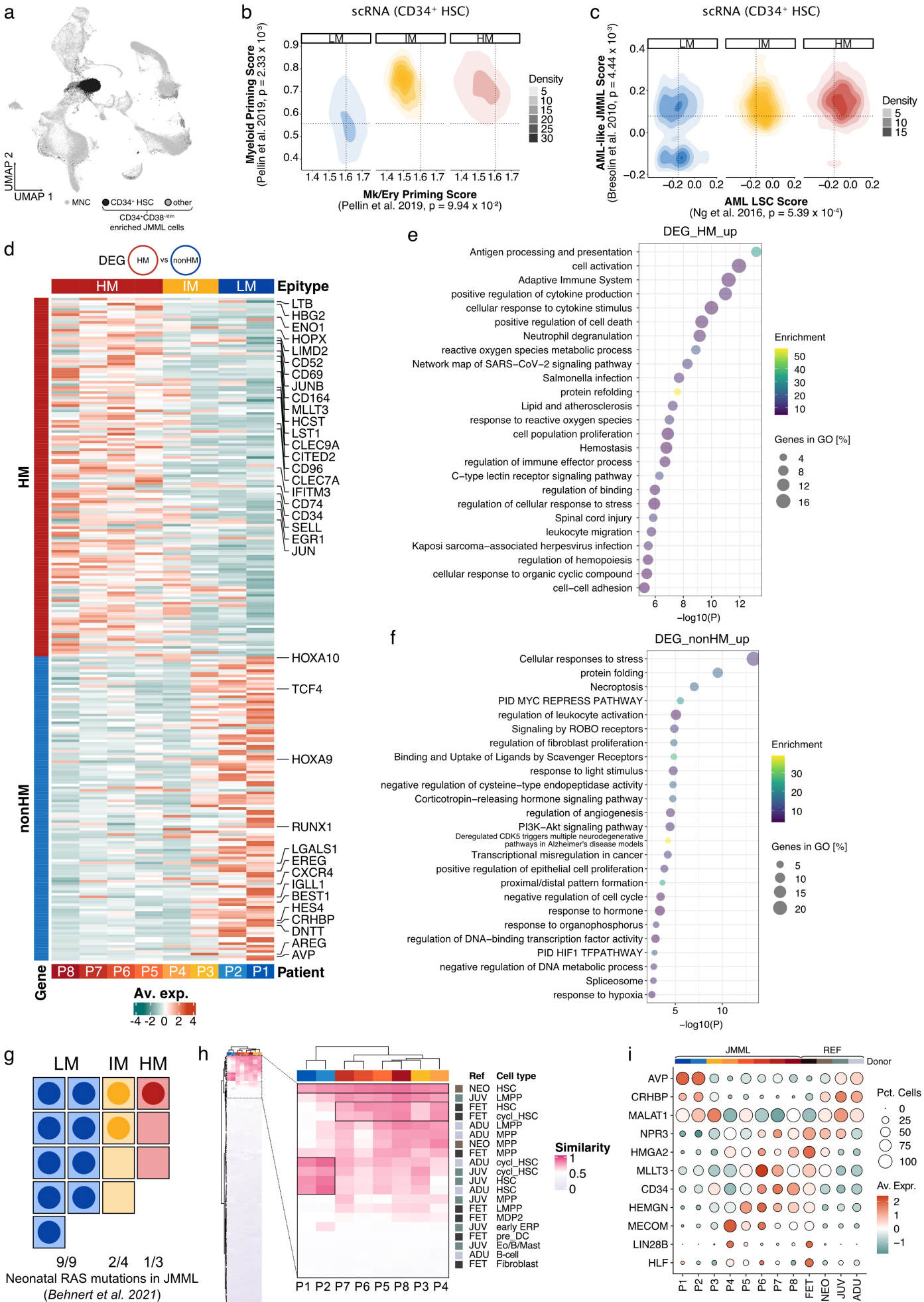




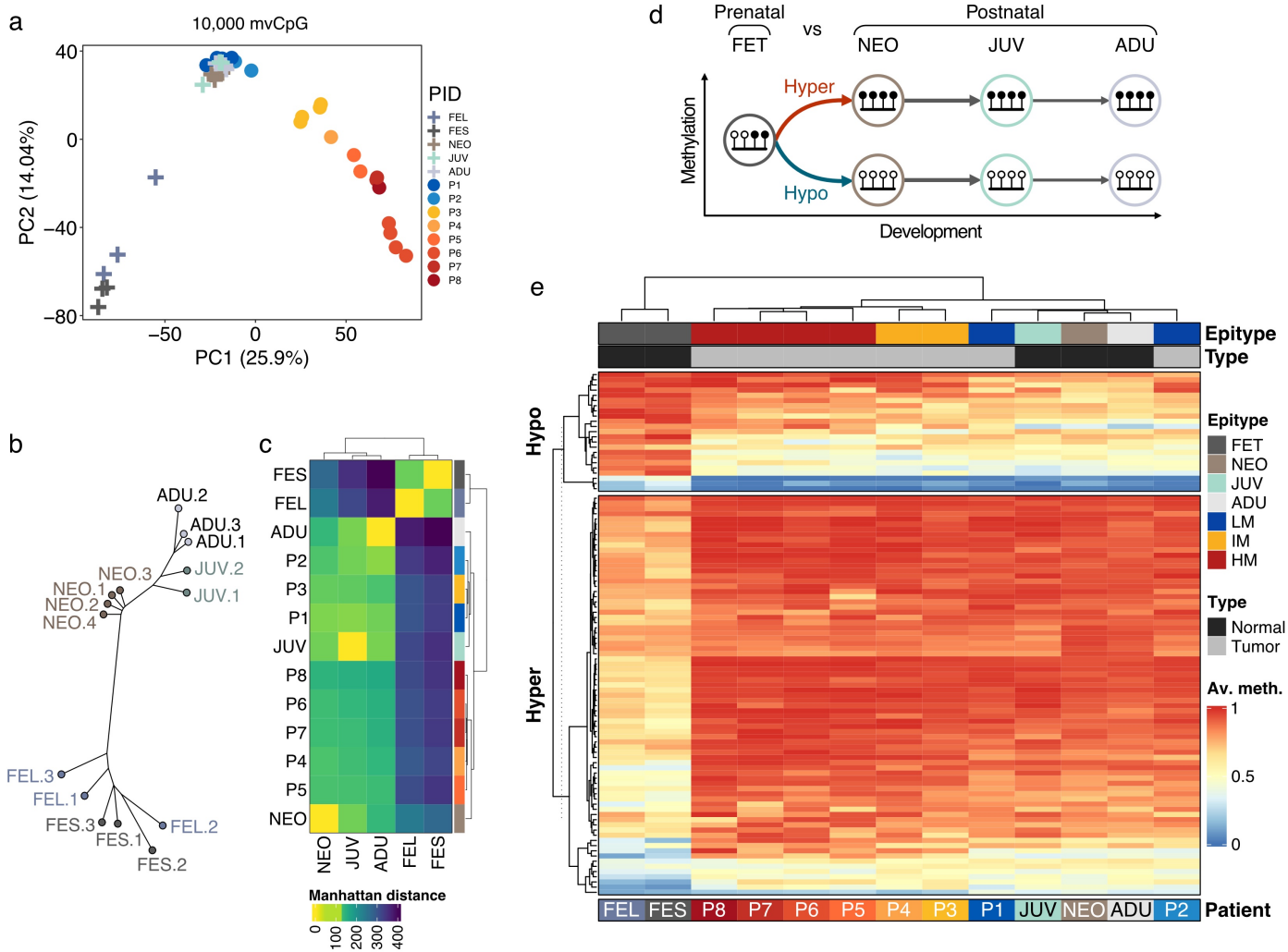
57 **Extended Data Figure 6: Fluorescence activated cell sorting (FACS) strategy for in-**  
58 **depth analysis of JMML HSPCs.** Lin<sup>-</sup>CD34<sup>+</sup>CD38<sup>lo/-</sup> cells were selected for scRNA-seq  
59 analyses. Lin<sup>-</sup>CD34<sup>+</sup>CD38<sup>lo/-</sup>CD45RA/CD90 quadrants were selected for ultra-low input  
60 WGBS analyses.  
61



**Extended Data Figure 7: Ultra-low input whole-genome bisulfite sequencing data analysis of Lin<sup>-</sup>CD34<sup>+</sup>CD38<sup>lo/-</sup> HSPCs from primary JMML patient samples.** **a**, DNA methylation-based cell type classifier to determine the cell type-specific DNA methylation profiles across Lin<sup>-</sup>CD34<sup>+</sup>CD38<sup>lo/-</sup>CD45RA/CD90 quadrants<sup>3</sup>. Normal bone marrow from adult donors was taken as a reference. HSC, Lin<sup>-</sup>CD34<sup>+</sup>CD38<sup>-</sup>CD45RA<sup>-</sup>CD90<sup>+</sup>; MPP, Lin<sup>-</sup>CD34<sup>+</sup>CD38<sup>-</sup>CD45RA<sup>-</sup>CD90<sup>-</sup>; CMP, Lin<sup>-</sup>CD34<sup>+</sup>CD38<sup>+</sup>CD10<sup>-</sup>CD45RA<sup>-</sup>CD135<sup>+</sup>; GMP, Lin<sup>-</sup>CD34<sup>+</sup>CD38<sup>+</sup>CD10<sup>-</sup>CD45RA<sup>+</sup>CD135<sup>+</sup>; MEP, Lin<sup>-</sup>CD34<sup>+</sup>CD38<sup>+</sup>CD10<sup>-</sup>CD45RA<sup>-</sup>CD135<sup>-</sup>. **b**, Boxplot of the average genome-wide DNA methylation of CpG islands (CGI) across patients. **c**, Heatmap of the average methylation of the 1,000 most variable promoters. Epitype, JMML DNA methylation subgroup; Genotype, RAS mutation; Celltype, Lin<sup>-</sup>CD34<sup>+</sup>CD38<sup>lo/-</sup>CD45RA/CD90 quadrant. **d**, Reference-free decomposition of DNA methylation data using MeDeCom<sup>4</sup> with the following settings: k = 3, lambda = 0.01. **e-g**, HM vs non-HM DMR calling. **e**, Piechart stratifying DMRs into hyper- (11,190) and hypo-methylated (2,642). **f**, Histogram summarizing the size distribution of JMML DMRs with a median of 512 bp. **g**, Circos plot with histograms summarizing genomic DMR locations. Red represents hyper- and blue represents hypomethylated regions in HM vs. non-HM DMRs. Genome-wide distribution of JMML DMRs. **h-j**, Locus overlap analysis for enrichment of genomic ranges (LOLA) using the 15-state ChromHMM chromatin feature annotation. **h**, Bubble plot summarizing genome-wide enrichment analysis of DMRs in chromatin features (ChromHMM) using locus overlap analysis (LOLA). **i**, Boxplot of average methylation in polycomb-repressed (ReprPC) domains across JMML epitypes. **j**, Boxplot of average methylation in bivalent enhancers (EnhBiv) across JMML epitypes. **k,l**, Genomic Regions Enrichment of Annotations Tool (GREAT)<sup>5</sup> analysis for gene ontology analysis of (K.) biological processes (GO\_bp) and (L.) murine phenotypes. **m**, Locus line plot of DNA methylation at the RUNX1 locus across JMML patients. Smoothed DNA methylation; 'DMR' indicates HM vs. non-HM JMML DMRs. **n**, Bar plot depicting the logarithmic fold change of the number of homeobox genes associated to HM vs. non-HM DMRs relative to the genome-wide abundance. Homeobox gene classes were stratified as previously published<sup>6</sup>: ANTP, *Antennapedia* (incl. HOXL and NKL genes); PRD, *paired* (incl. PAX and PAXL genes); Misc incl. LIM, POU, HNF, SINE, CUT, PROS, ZF, CERS class genes. **o**, Bubble plot of the genome-wide analysis of enriched motifs across DMRs using Hypergeometric Optimization of Motif EnRichment (HOMER) to summarize the enrichment of TF binding motifs across HM vs non-HM DMRs.

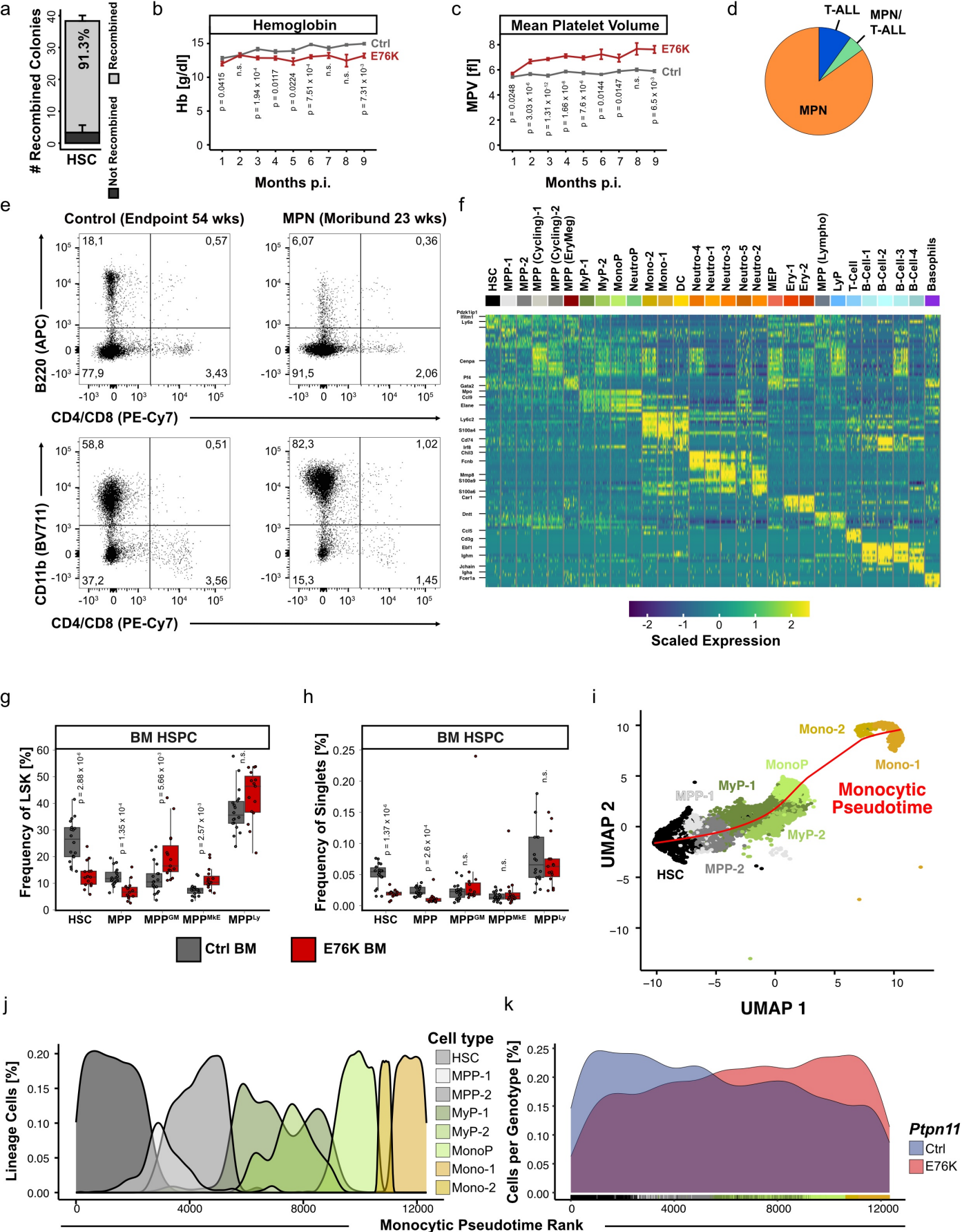


**Extended Data Figure 8: scRNA-seq data analysis of Lin<sup>-</sup>CD34<sup>+</sup>CD38<sup>lo/-</sup> HSPCs from primary JMML patient samples.** **a**, Projection of scRNA-seq data of 20,163 Lin<sup>-</sup>CD34<sup>+</sup>CD38<sup>lo/-</sup> enriched HSPCs onto the UMAP from Figure 1.B (total hematopoiesis, small light grey dots). Additionally sequenced JMML HSPCs are represented as larger dots (CD34<sup>+</sup> HSC in black and others in grey). **b,c**, Density plots of single-cell transcription scores of transcriptionally defined JMML HSCs (CD34<sup>+</sup> HSC) stratified by JMML epitype. **b**, Megakaryocytic/erythroid (Mk/Ery) vs. lympho/myeloid priming. These gene signatures reflect the first branching point in lineage commitment of human HSCs according to Pellin et al. (2019)<sup>7</sup>. **c**, Transcription scores for clinically aggressive JMML (AML-like JMML) as defined by Bresolin et al. (2010)<sup>8</sup> and leukemia stem cell programs (17-gene LSC) as defined by Ng et al. (2016)<sup>9</sup>. **d**, Heatmap of 255 differentially expressed genes (DEGs) identified between CD34<sup>+</sup> HSCs from HM and non-HM JMML patients. Depicted are normalized average expression levels aggregated across all HSCs within each of the patients. **e,f**, Bubble plots summarizing the top 25 significant terms of functional enrichment analysis of HM vs non-HM DEGs using Metascape<sup>10</sup>. Data was stratified by upregulated genes in HM or non-HM HSCs, respectively. Bubble size represents percentage of genes in GO and color the enrichment per term. **g**, Re-analysis of neonatal blood spot genotyping by Behnert et al. (2022)<sup>11</sup>. The frequency of RAS driver mutations in neonatal blood spots of individuals that developed JMML later in life is stratified by JMML epitype. Each square represents a single patient, dark dots represent RAS-positive blood spots and unfilled squares RAS-negative cases. **h**, Reference-based cell type similarity analysis (logistic regression) of JMML HSCs to a reference map of 116 hematopoietic reference cell types across developmental stages (rows). Values above 0.5 (pink) and below 0.5 (grey) refer to similarity and dissimilarity, respectively, to the reference cell type; 0.5 represents randomness (white). Left panel shows the entire results. Right panel shows a zoom into the top part of the heatmap which includes the references with positive similarity values. Row labeling refers to the reference stage (Ref: FET, fetal; NEO, neonatal; JUV, juvenile; ADU, adult) and the cell type, respectively. Reference cell types shown in the main figure 3b are marked by black boxes. **i**, Dot plot of manually curated HSC marker from pre- and postnatal references. The gene list is manually curated based on previously published data<sup>2,7,12,13</sup>. Average expression is scaled per gene. Dot sizes represent the percentage of cells expressing the corresponding gene.



**Extended Data Figure 9: Ultra-low input whole-genome bisulfite sequencing data analysis of Lin<sup>-</sup>CD34<sup>+</sup>CD38<sup>low</sup> HSPCs from primary JMML patient samples in the context of normal HSC development.** Ultra-low input WGBS analysis of normal and JMML HSCs from primary samples. FET corresponds to both fetal spleen (FES) and fetal liver (FEL) samples. NEO is neonatal blood. JUV and ADU are juvenile and adult bone marrow. **a**, Principal component analysis (PCA) of the 10k most-variable CpGs (mvCpGs) across the entire WGBS dataset including reference and JMML HSCs. Low-risk HSCs cluster with postnatal HSCs. **b,c**, Methylation-based phylogenetic relationships between JMML and healthy reference HSCs across development using WGBS data. **b**, Phylogenetic tree analysis of healthy HSCs based on genomic regions defined in 3.F. Selected methylation dynamics reflect the expected order of normal HSC development. **c**, Heatmap showing the distances of samples in the phylogenetic tree. Manhattan distances were calculated based on the branch length per subgroup. **d,e**, 'Epigenetic scars' across normal and malignant HSCs. Epigenetic scars are per definition DNA methylation changes that occur at a single step during development but do not change downstream of this event. 'Hypo' and 'hyper' correspond to scars whose methylation de- or increase during developmental state transition. **d**, Schematic of epigenetic scars of the transition between fetal and postnatal stages **e**, Heatmap showing methylation beta values of epigenetic scars from the transition of fetal to postnatal HSCs. Hypo: loss of DNA methylation; hyper: gain of DNA methylation.

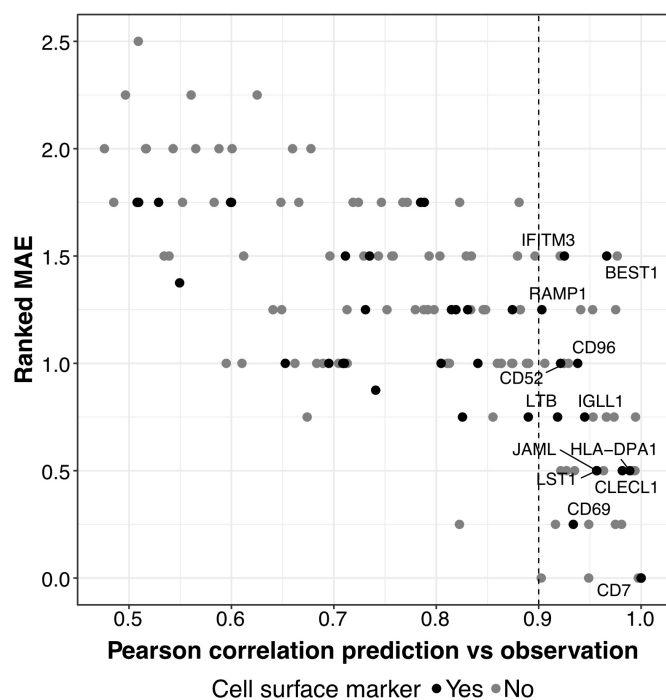




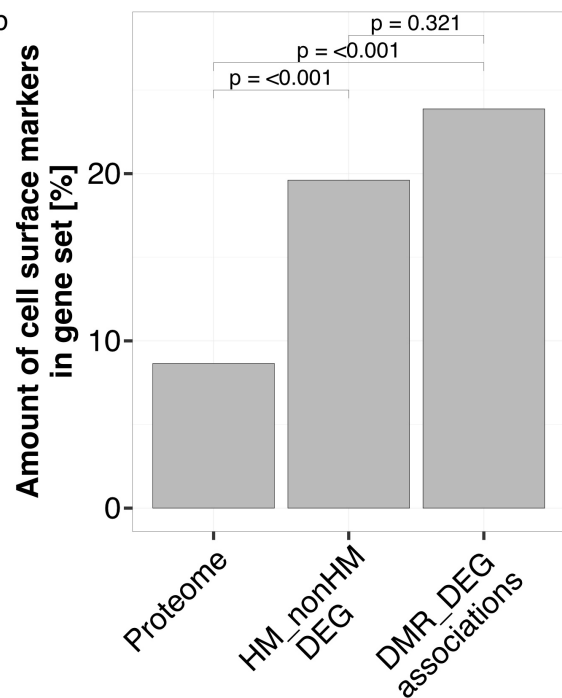


**Extended Data Figure 10: HSC-specific *Ptpn11*<sup>E76K/+</sup> JMML mouse model recapitulates human disease.** **a**, Recombination efficacy of the loxP-flanked PGK-Neo cassette in HSCs. *Ptpn11*<sup>E76K/+</sup> *Scl*-CreER<sup>T</sup> mice were injected at the age of 6 weeks on 5 consecutive days with tamoxifen and recombination of the loxP-flanked PGK-Neo cassette was analyzed after 72h in HSCs by PCR. **b-c**, Differential blood counts in *Ptpn11*<sup>E76K/+</sup> *Scl*-CreER<sup>T</sup> mice (E76K, n = 29) and *Ptpn11*<sup>+/+</sup> *Scl*-CreER<sup>T</sup> controls (Ctrl, n = 23). The results for hemoglobin levels (Hb; **b**) and mean platelet volume (MPV; **c**) are shown over a time course of 9 months. Mean and SEM are depicted. Statistical significances was determined using Student's t-test. **d**, Disease phenotype in moribund E76K mice (n = 20). **e**, Flow cytometry measurement of bone marrow cells from a control mouse and a E76K mouse with MPN. The percentage of cells in each gate is depicted. Axis labels indicate the respective surface markers and fluorophores. **f**, Expression of hematopoietic marker genes in scRNA-seq. The scaled expression of the top 5 marker genes for each of the 29 clusters identified in the scRNA-seq data is depicted as a heatmap. Each cluster was downsampled to 25 cells for visualization. Cluster-specific marker genes were empirically selected and labeled. **g-h**, Hematopoietic stem and progenitor compartment remodeling in E76K (n = 15) and Ctrl (n = 18) mice 18-weeks after tamoxifen induction. HSPC frequencies were determined using flow cytometry and are plotted as frequencies of LSK cells (**g**) and singlets (**h**). Whiskers are defined as 1.5 times interquartile range. Statistical significance was calculated using Student's t-test. **i**, UMAP embedding showing cells from the monocytic differentiation lineage and the calculated monocytic pseudotime. **j-k**, Density plots display the distribution of 3000 randomly selected bone marrow cells from E76K and Ctrl mice along the monocytic pseudotime. The distribution is colored by cell type (**j**) or genotype (**k**), respectively. The annotated cell types are visualized as bars below the density curve per genotype (**k**).

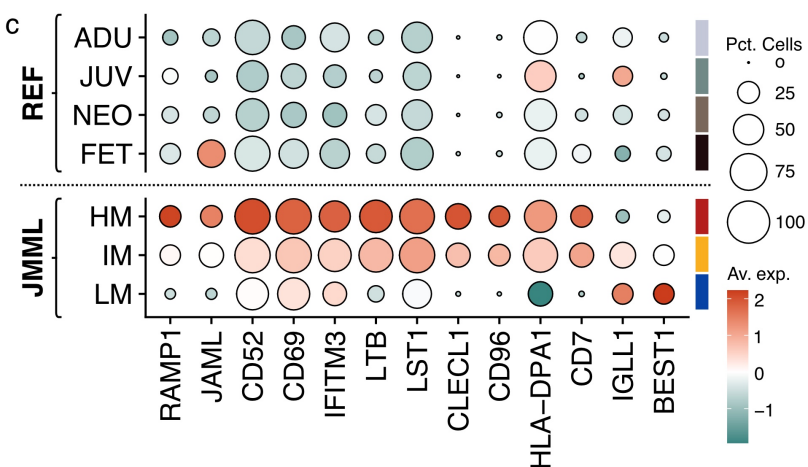
a



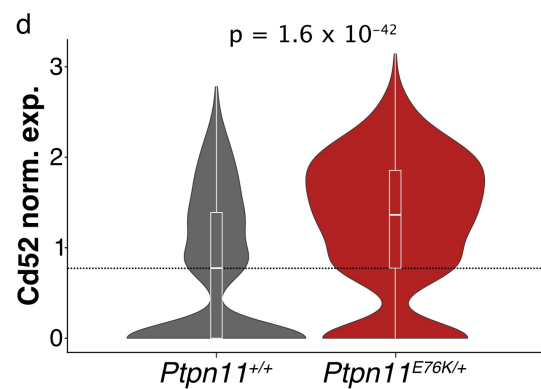
b



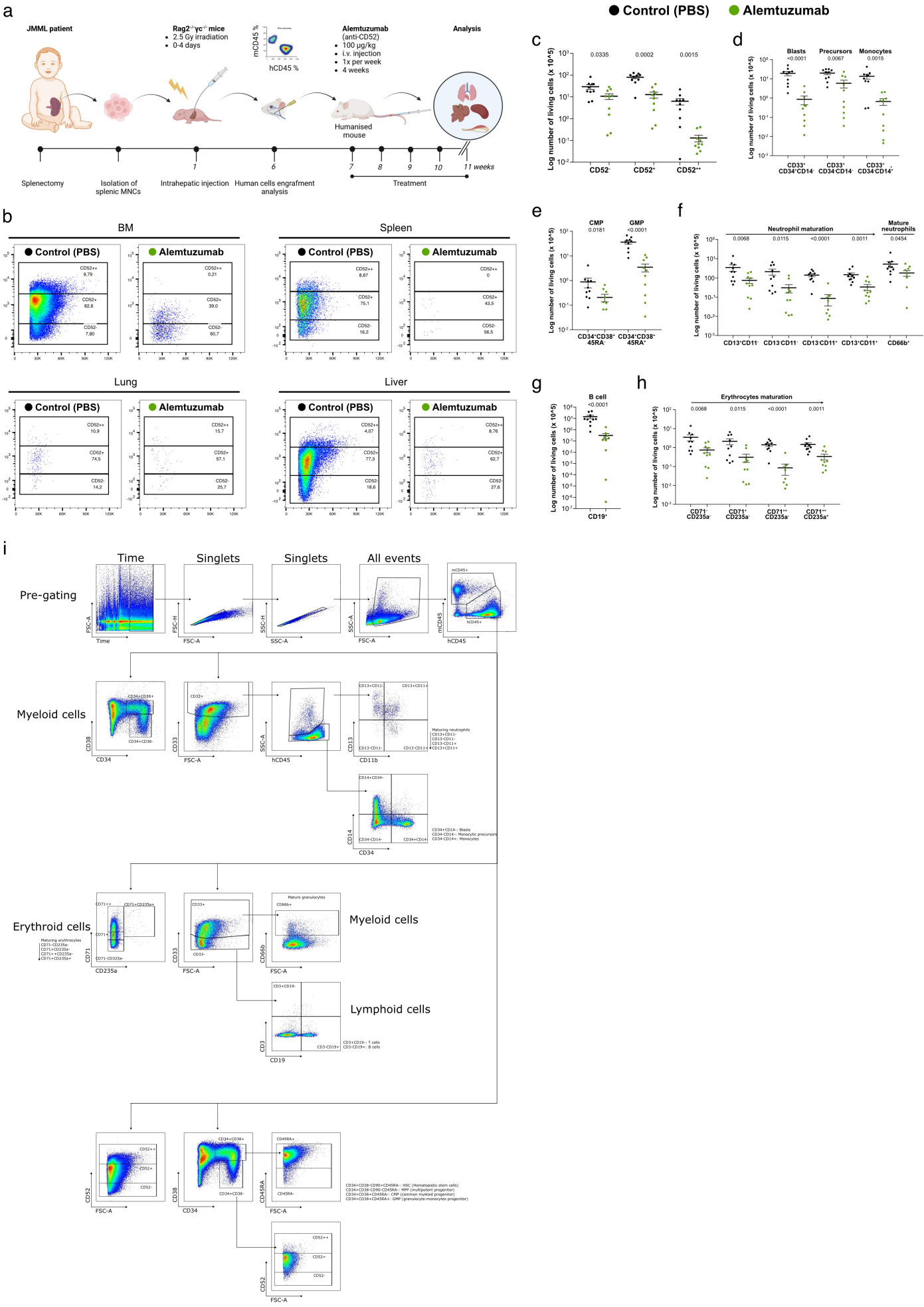
c



d



**Extended Data Figure 11: Integrative analysis of methylation and transcription changes in JMML HSCs reveals novel prognostic biomarkers and therapeutic targets including CD52.** **a-b**, DMR-DEG associations to determine genes which subgroup-specific expression that can be explained by differential methylation. HM vs non-HM DMRs were associated with HM vs non-HM DEGs applying a generalized linear model with stepwise feature elimination using AIC. Per DEG, DMRs in a region 100 kb up- and downstream of the TSS were considered for potential associations. **a**, GLM results including all genes of which expression changes are associated with methylation changes across subgroups. Lower ranked MAE values represent genes that better fit the model. Pearson correlation between prediction and observation was used to calculate the confidence of the DMR-DEG associations. All genes above 0.9 were considered as candidates for further analysis. **b**, Amount of surface markers across gene sets reveals an enrichment of differentially expressed surface markers in JMML stem cells. As a reference, the previously published in silico human surfaceome was used<sup>14</sup>. 'DEG' are all HM vs. nonHM DEGs (Figure 2.I), 'DMR\_DEG' are genes that were identified in glm analysis. **c**, Dot plot summarizing the scaled normalized expression of all DMR-associated high-confidence candidate surface marker genes in JMML and healthy reference HSCs. Circle size represents the percentage of cells expressing the gene. **d**, Violin plot of normalized single-cell RNA-seq expression of Cd52 in murine HSCs of *Ptpn11*<sup>+/+</sup> and *Ptpn11*<sup>E76K</sup> mice.



**Extended Data Figure 12: Anti-CD52 (Alemtuzumab) treatment depletes total hematopoiesis and disrupts disease-propagation in JMML patient-derived xenografts (PDX).** **a**, Schematic of anti-CD52 (Alemtuzumab) treatment of JMML PDX. Created with BioRender.com **b**, FACS plots showing CD52 expression in different organs in anti-CD52 treated (alemtuzumab, green) and untreated (Control, black) mice. **c-h**, Logarithmic scale of absolute cell number per pair of femora of human CD45<sup>+</sup> cells in treated vs. untreated mice. **c**, Total human CD45<sup>+</sup> cells stratified by CD52 expression. **d-h**, Different hematopoietic lineages based on classical cell type surface marker expression: **(d)** monocytes and blasts; **(e)** myeloid precursors; **(f)** neutrophils; **(g)** B cells; **(h)** erythrocytes. **i**, Gating scheme for immunophenotypic characterization of PDX mice.

## References

- 1 Schönung, M. *et al.* International Consensus Definition of DNA Methylation Subgroups in Juvenile Myelomonocytic Leukemia. *Clin Cancer Res* **27**, 158-168, doi:10.1158/1078-0432.CCR-20-3184 (2021).
- 2 Hay, S. B., Ferchen, K., Chetal, K., Grimes, H. L. & Salomonis, N. The Human Cell Atlas bone marrow single-cell interactive web portal. *Exp Hematol* **68**, 51-61, doi:10.1016/j.exphem.2018.09.004 (2018).
- 3 Farlik, M. *et al.* DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation. *Cell Stem Cell* **19**, 808-822, doi:10.1016/j.stem.2016.10.019 (2016).
- 4 Lutsik, P. *et al.* MeDeCom: discovery and quantification of latent components of heterogeneous methylomes. *Genome Biol* **18**, 55, doi:10.1186/s13059-017-1182-6 (2017).
- 5 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501, doi:10.1038/nbt.1630 (2010).
- 6 Holland, P. W., Booth, H. A. & Bruford, E. A. Classification and nomenclature of all human homeobox genes. *BMC Biol* **5**, 47, doi:10.1186/1741-7007-5-47 (2007).
- 7 Pellin, D. *et al.* A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nat Commun* **10**, 2395, doi:10.1038/s41467-019-10291-0 (2019).
- 8 Bresolin, S. *et al.* Gene expression-based classification as an independent predictor of clinical outcome in juvenile myelomonocytic leukemia. *J Clin Oncol* **28**, 1919-1927, doi:10.1200/JCO.2009.24.4426 (2010).
- 9 Ng, S. W. *et al.* A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* **540**, 433-437, doi:10.1038/nature20598 (2016).
- 10 Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523, doi:10.1038/s41467-019-09234-6 (2019).
- 11 Behnert, A. *et al.* Exploring the genetic and epigenetic origins of juvenile myelomonocytic leukemia using newborn screening samples. *Leukemia* **36**, 279-282, doi:10.1038/s41375-021-01331-0 (2022).
- 12 Lehnertz, B. *et al.* HLF expression defines the human hematopoietic stem cell state. *Blood* **138**, 2642-2654, doi:10.1182/blood.2021010745 (2021).
- 13 Calvanese, V. *et al.* Mapping human haematopoietic stem cells from haemogenic endothelium to birth. *Nature* **604**, 534-540, doi:10.1038/s41586-022-04571-x (2022).
- 14 Bausch-Fluck, D. *et al.* The in silico human surfaceome. *Proc Natl Acad Sci U S A* **115**, E10988-E10997, doi:10.1073/pnas.1808790115 (2018).