

Supplemental Information

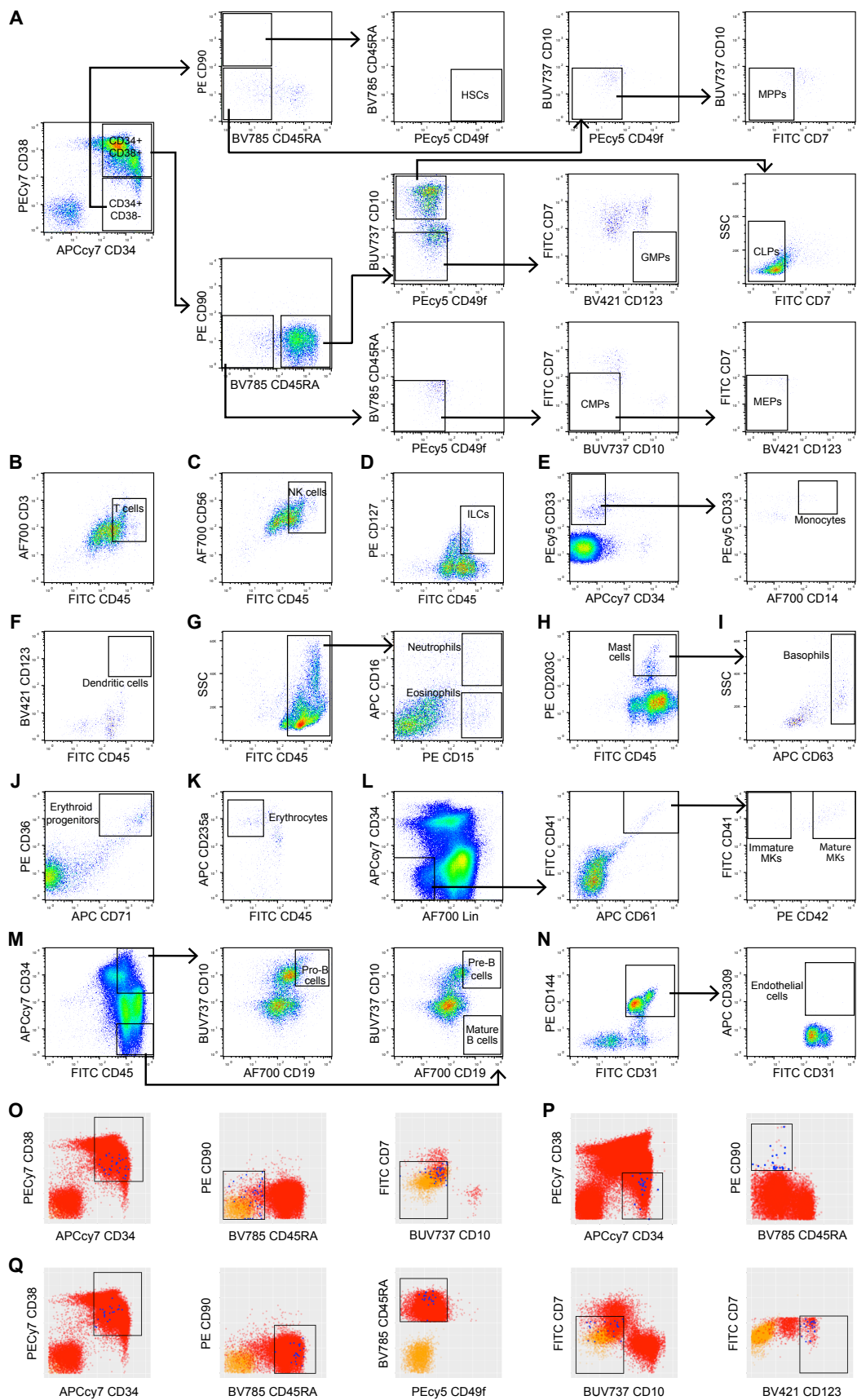
Integrative Single-Cell RNA-Seq and ATAC-Seq

Analysis of Human Developmental Hematopoiesis

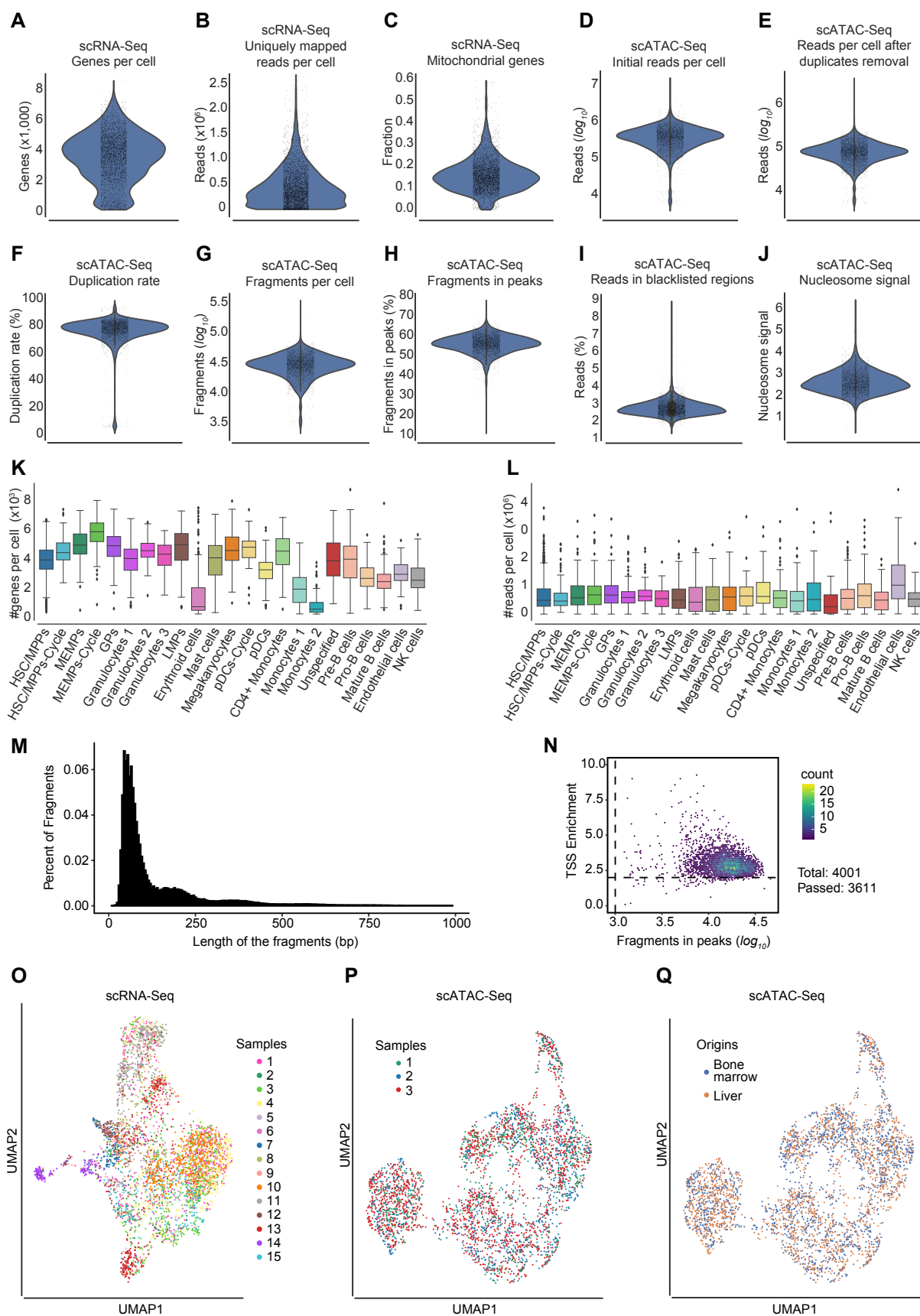
Anna Maria Ranzoni, Andrea Tangherloni, Ivan Berest, Simone Giovanni Riva, Brynelle Myers, Paulina M. Strzelecka, Jiarui Xu, Elisa Panada, Irina Mohorianu, Judith B. Zaugg, and Ana Cvejic

Integrative Single-cell RNA-Seq and ATAC-Seq Analysis of Human Developmental Haematopoiesis

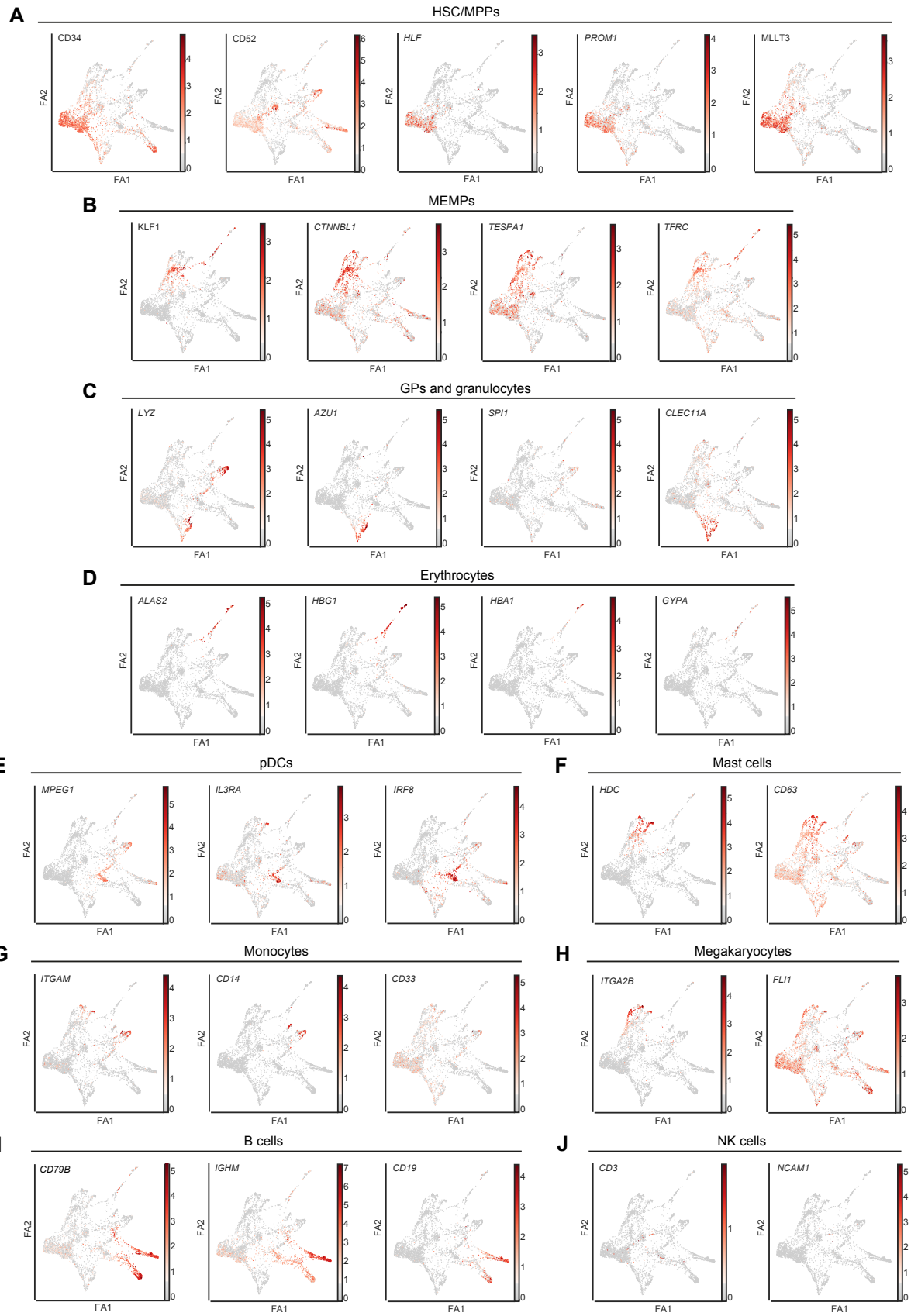
Anna Maria Ranzoni, Andrea Tangherloni, Ivan Berest,
Simone Giovanni Riva, Brynelle Myers, Paulina M. Strzelecka,
Jiarui Xu, Elisa Panada, Irina Mohorianu,
Judith B. Zaugg, Ana Cvejic



Supplementary figure 1 – Sorting panels, Related to Figure 1, related to Figure 1. A-N. FACS sorting panel and gating strategy for the isolation of phenotypically defined cell types: **A.** Committed and non-committed haematopoietic progenitors, HSCs, MPPs, CMPs, GMPs, MEPs, and CLPs. **B.** T cells. **C.** NK cells. **D.** ILCs. **E.** Monocytes. **F.** Dendritic cells. **G.** Neutrophils and eosinophils. **H.** Mast cells. **I.** Basophils. **J.** Erythroid progenitors. **K.** Erythrocytes. **L.** Immature and mature MKs. **M.** Pro-B cells, pre-B cells, and mature B cells. **N.** Endothelial cells. **O-Q.** Index sorting data FACS plots showing sorted cells (blue), total gated population (red), and unstained population (yellow) for defined cell types: **O.** CMPs. **P.** HSCs. **Q.** GMPs. HSCs - haematopoietic stem cells, MPPs - multipotent progenitors, CMPs - common myeloid progenitors, GMPs - granulocyte-monocyte progenitors, MEPs - megakaryocyte-erythroid progenitors and CLPs - common lymphoid progenitors, NK cells - natural killer cells, ILCs - innate lymphoid cells, MKs - megakaryocytes.

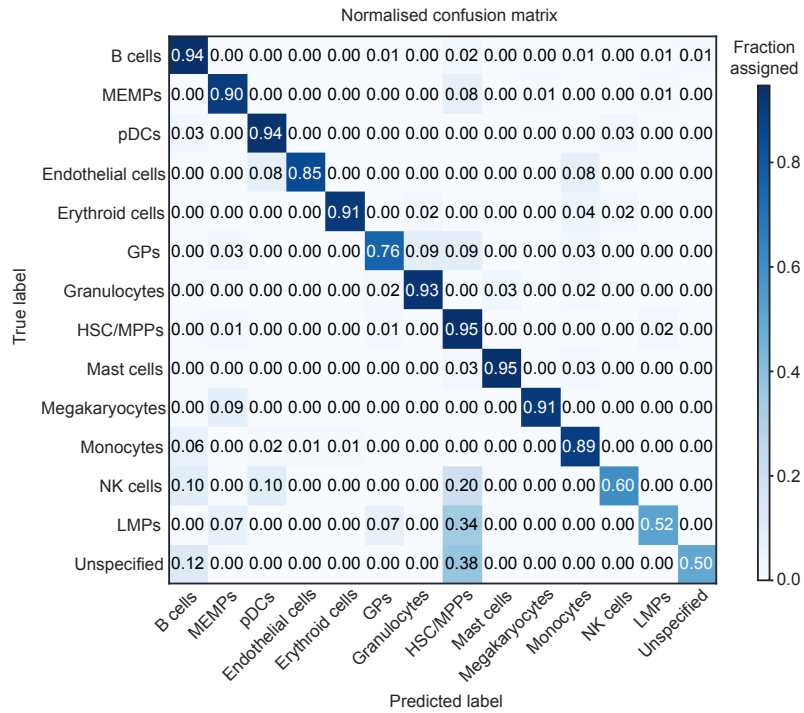


Supplementary figure 2 – Quality control and batch effects correction in scRNA-Seq and scATAC-Seq data, related to Figures 1 and 3. **A.** Violin plots showing the number of expressed genes per cell in scRNA-Seq data. **B.** Violin plots showing the number of uniquely mapped reads against the reference genome per cell in scRNA-Seq data. **C.** Violin plots showing the fraction of mitochondrial genes compared to all genes per cell in scRNA-Seq data. **D.** Violin plots showing the number of reads per cell, prior to duplicates removal, in scATAC-Seq data. The y-axis is in $\log_{10}$ scale. **E.** Violin plots showing the number of reads per cell after duplicates removal in scATAC-Seq data. The y-axis is in $\log_{10}$ scale. **F.** Violin plots showing the duplicate rate in scATAC-Seq data. **G.** Violin plots showing the number of fragments per cell in scATAC-Seq data. The y-axis is in $\log_{10}$ scale. **H.** Violin plots showing the percentage of fragments per peak in scATAC-Seq data. **I.** Violin plots showing the percentage of reads mapping to the blacklist regions in scATAC-Seq data. **J.** Violin plots showing the nucleosome signal per cell in scATAC-Seq data. **K.** Box plot showing the number of genes per cell in each identified cell type in scRNA-Seq data. **L.** Box plot showing the number of uniquely mapped reads per cell in each identified cell type in scRNA-Seq data. **M.** Histogram showing the length of the fragments in terms of base pairs (200 bins). **N.** Scatterplot showing the fragments in peaks with respect to TSS enrichment. The colour intensity represents the number of counts. The x-axis is in $\log_{10}$ scale. **O.** UMAP visualization of the scRNA-Seq samples ($n = 15$) after the batch effect correction with BBKNN. Each colour represents a different sample. **P.** UMAP visualization of the scATAC-Seq samples ($n = 3$) after the batch effect correction with Harmony. Each colour represents a different sample. **Q.** UMAP visualization of the scATAC-Seq bone marrow (blue) and liver (orange) CD34+ CD38- cells after the batch effect correction with Harmony.

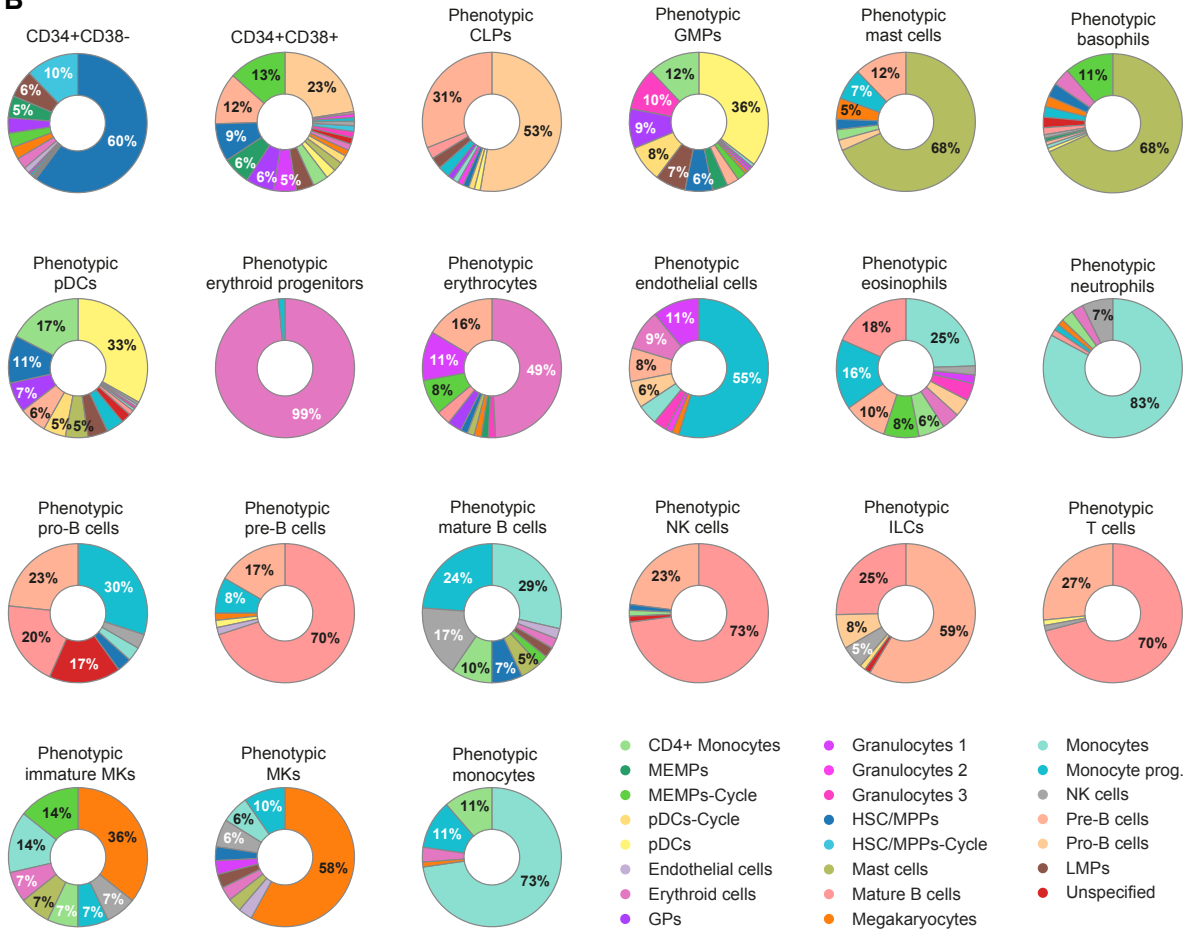


Supplementary figure 3 – Expression of top marker genes along the differentiation trajectory, related to Figure 2. (A-J) FDG visualisation of the *log*-normalised gene expression of marker genes along the differentiation trajectory. **A.** HSC/MPPs (*CD34*, *CD52*, *HLF*, *PROM1*, and *MLLT3*). **B.** MEMPs (*KLF1*, *CTNBL1*, *TESPA1*, and *TFRC*). **C.** GPs and granulocytes (*LYZ*, *AZU1*, *SPI1*, and *CLEC11A*). **D.** Erythrocytes (*ALAS2*, *HBG1*, *HBA1*, and *GYPA*). **E.** pDCs (*MPEG1*, *IL3RA*, and *IRF8*). **F.** Mast cells (*HDC* and *CD63*). **G.** Monocytes (*ITGAM*, *CD14*, and *CD33*). **H.** Megakaryocytes (*ITGA2B* and *FLI1*). **I.** B cells (*CD79B*, *IGHM*, and *CD19*). **J.** NK cells (*CD3* and *NCAM1*). Force-Directed Graph - FDG; ForceAtlas2 - FA2; HSC/MPPs - haematopoietic stem cells/multipotent progenitors; MEMPs - megakaryocyte-erythroid-mast progenitors; GPs - granulocytic progenitors; pDCs-Cycle - cycling plasmacytoid dendritic cells.

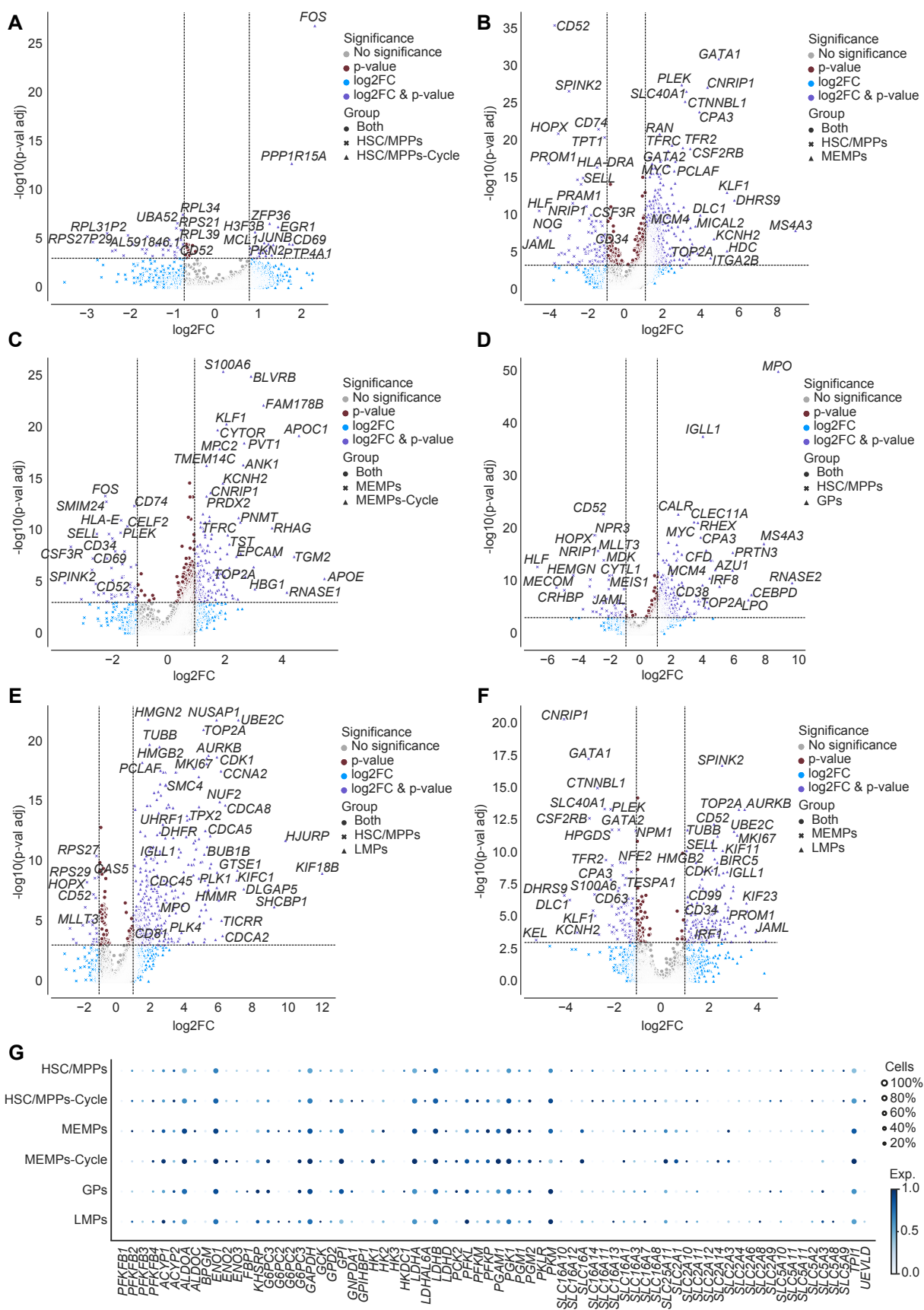
A



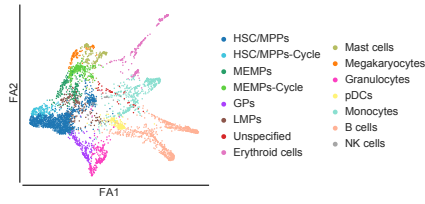
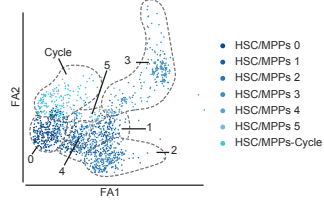
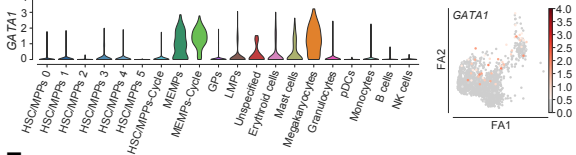
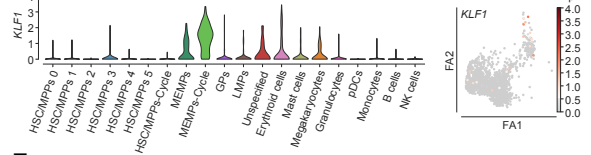
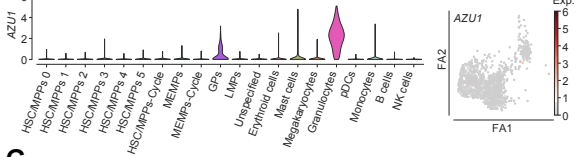
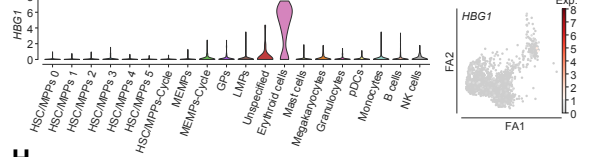
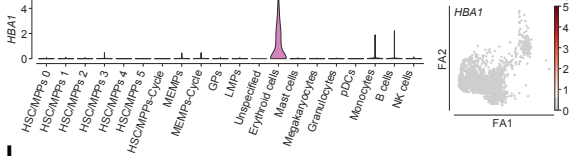
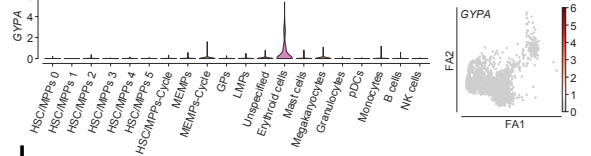
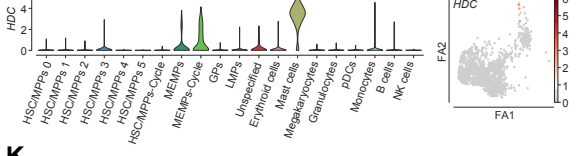
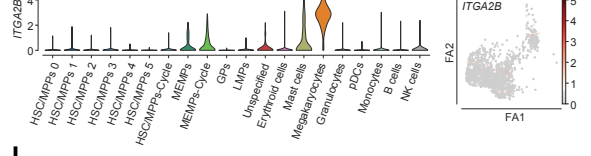
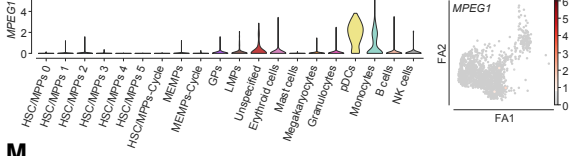
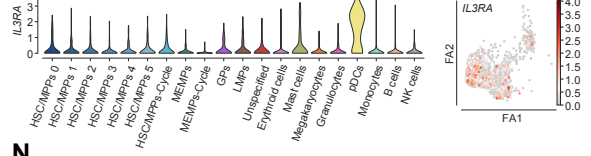
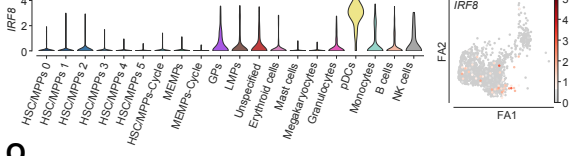
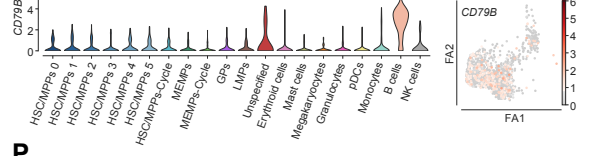
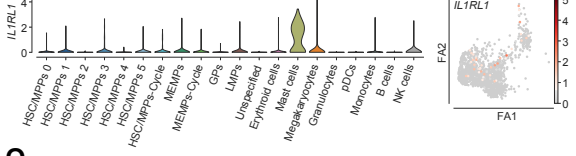
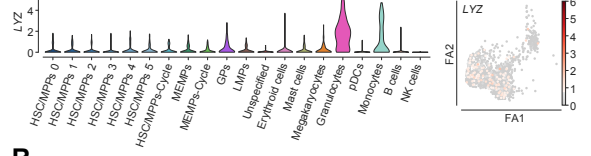
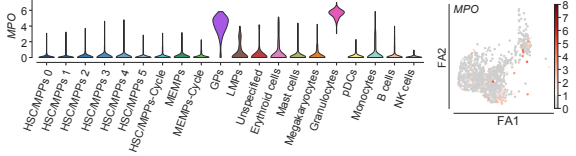
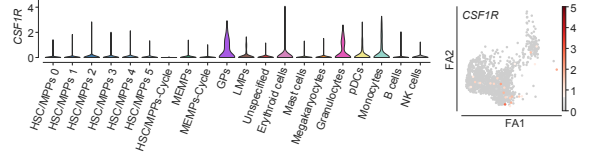
B



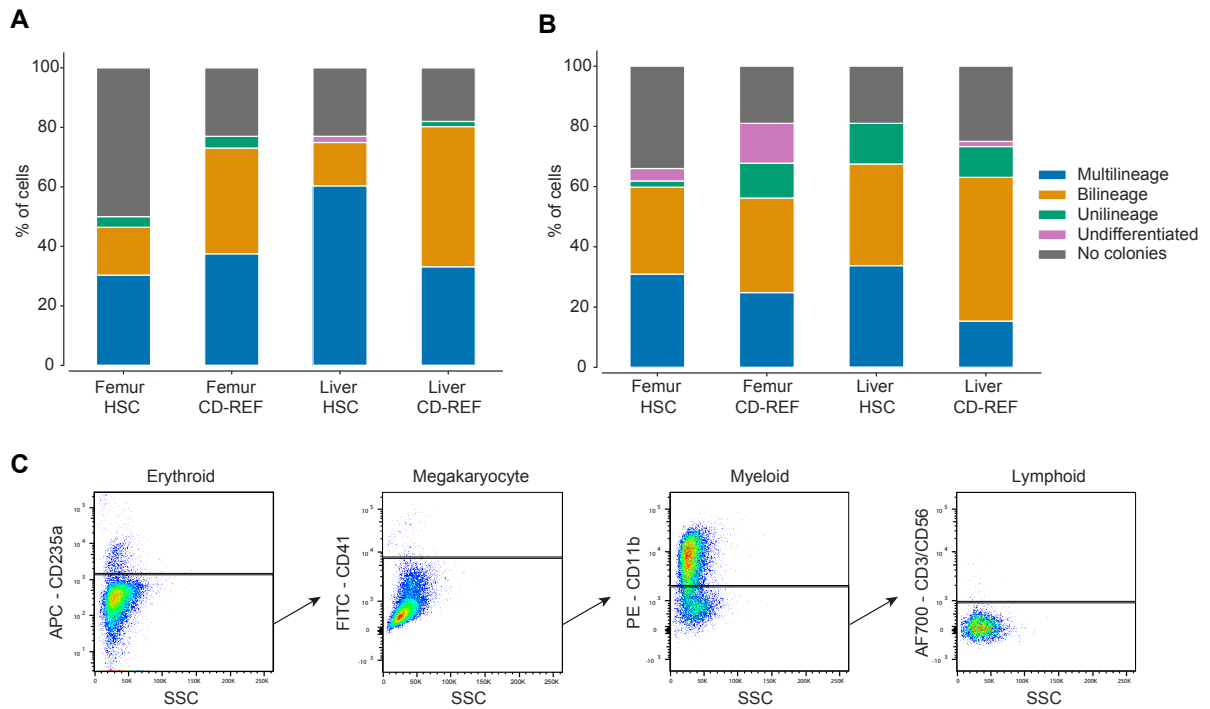
Supplementary figure 4 – Validation of the cell type assignment and transcriptional heterogeneity of phenotypically defined cell populations, related to Figures 1 and 3. **A.** Confusion matrix showing the cell type assignment achieved by the DNN on the test set (901 cells), considering the top 30 marker genes per cell type (14 distinct cell types). The colour intensity represents the fraction of the assigned cells per cell type. **B.** Donut plots showing the percentage of transcriptionally defined (i.e., manually curated) cell populations in each of the phenotypically defined populations (Expanded from Figure 1C). Each colour represents a different cell type. HSC/MPPs-Cycle - cycling haematopoietic stem cells/multipotent progenitors; HSC/MPPs - haematopoietic stem cells/multipotent progenitors; MEMPs - megakaryocyte-erythroid-mast progenitors; MEMPs-Cycle - cycling megakaryocyte-erythroid-mast progenitors; GPs - granulocytic progenitors; LMPs - lymphomyeloid progenitors; pDCs-Cycle - cycling plasmacytoid dendritic cells; pDCs - plasmacytoid dendritic cells.



Supplementary figure 5 – Differential expression analysis of the progenitor compartment, related to the STAR Methods section. (A-F) Volcano plot showing DEGs between two cell types of interest. **A.** HSC/MPPs and HSC/MPPs-Cycle. **B.** HSC/MPPs and MEMPs. **C.** MEMPs and MEMPs-Cycle. **D.** HSC/MPPs and GPs. **E.** HSC/MPPs and LMPs. **F.** MEMPs and LMPs. The x-axes show the $\log_2$ fold-change (magnitude of change), while the y-axes show the $-\log_{10}$ adjusted p-value (statistical significance). We used the Wilcoxon rank-sum with the Benjamini-Hochberg correction. Colours represent the significance of the genes, both in terms of p-value and $\log_2$ fold-change. **G.** Dot plot of the expression of metabolic genes involved in glycolysis in the identified progenitor compartment. The expression of the genes is standardised between 0 and 1. For each gene, the minimum value is subtracted and the result is divided by the maximum. The spot size indicates the percentage of cells that express the gene of interest within each cell type. The colour intensity represents the standardised expression level. HSC/MPPs-Cycle - cycling haematopoietic stem cells/multipotent progenitors; HSC/MPPs - haematopoietic stem cells/multipotent progenitors; MEMPs - megakaryocyte-erythroid-mast progenitors; MEMPs-Cycle - cycling megakaryocyte-erythroid-mast progenitors; GPs - granulocytic progenitors; LMPs - lympho-myeloid progenitors.

A**B****C****D****E****F****G****H****I****J****K****L****M****N****O****P****Q****R**

Supplementary figure 6 – Expression of lineage-specific marker genes in HSC/MPP sub-populations, related to 2. **A.** FDG visualisation of the identified differentiation trajectory. **B.** FDG visualisation of the HSC/MPP sub-populations. **(C-R)** Left panels - violin plots showing the log-normalised median gene expression of lineage-specific marker genes in the HSC/MPP sub-populations and more differentiated haematopoietic cells. Right panels - FDG visualisations of the log-normalised gene expression of the same genes as in violin plots along the differentiation trajectory, considering only the HSC/MPP sub-populations. **C.** *GATA1*. **D.** *KLF1*. **E.** *AZU1*. **F.** *HBG1*. **G.** *HBA1*. **H.** *GYPA*. **I.** *HDC*. **J.** *ITGA2B*. **K.** *MPEG1*. **L.** *IL3RA*. **M.** *IRF8*. **N.** *CD79B*. **O.** *IL1RL1*. **P.** *LYZ*. **Q.** *MPO*. **R.** *CSF1R*. Force-Directed Graph - FDG; ForceAtlas2 - FA2; HSC/MPPs - haematopoietic stem cells/multipotent progenitors; HSC/MPPs-Cycle - cycling haematopoietic stem cells/multipotent progenitors; MEMPs - megakaryocyte-erythroid-mast progenitors; MEMPs-Cycle - cycling megakaryocyte-erythroid-mast progenitors; GPs - granulocytic progenitors; LMPs - lymphomyeloid progenitors; pDCs - plasmacytoid dendritic cells.



Supplementary figure 7 – Colony formation and the lineage output of CD-REF cells and phenotypic HSCs isolated from the foetal liver and femur, related to Figure 6. A-B. Stacked bar chart showing the differentiation potential and efficiency of colony formation of CD-REF cells compared to the phenotypic HSCs isolated from foetal liver and femur. The y-axis shows the percentage of colonies. The colonies have been divided by their differentiation potential as determined by FACS, namely: multilineage, bilineage, unilineage, and undifferentiated. Data from two different experiments are shown. **C.** Representative FACS plot showing the lineage composition of a quadrilineage colony.

Supplementary table 1 – Cell-surface markers used to isolate cell types, related to Figure 1.

Phenotypic cell type	Cell-surface markers
HSCs	Lin- CD34+ CD38- CD45RA- CD90+ CD49f+/-
MPPs	Lin- CD34+ CD38- CD90- CD45RA- CD49f- CD10- CD7-
CMPs	Lin- CD34+ CD38+ CD90- CD45RA- CD49f- CD10- CD7-
MEPs	Lin- CD34+ CD38+ CD90- CD45RA- CD49f- CD10- CD7- CD123-
GMPs	Lin- CD34+ CD38+ CD90- CD45RA+ CD49f- CD10- CD7- CD123+/-
CLPs	Lin-, CD34+ CD38+ CD90- CD45RA+ CD49f - CD10+ CD7-
T cells	CD45+ CD3+
NK cells	CD45+ CD56+
ILCs	Lin+/- CD45+ CD127+
Monocytes	CD34- CD33+ CD14+
Dendritic cells	Lin- CD45+ CD123+
Mast cells	CD45+ CD203c+
Basophils	CD45+ CD203c+ CD63+
Neutrophils	CD45+ CD15+ CD16+
Eosinophils	CD45+ CD15+ CD16-
Erythroid progenitors	CD36+ CD71+
Erythrocytes	CD235a+
Immature MKs	Lin- CD34- CD41+ CD61+ CD42-
Mature MKs	Lin- CD34- CD41+ CD61+ CD42+
Pro-B cells	CD45+ CD34+ CD19+ CD10+
Pre-B cells	CD45+ CD34- CD19+ CD10+
Mature B cells	CD45+ CD34- CD19+ CD10-
Endothelial cells	CD31+ CD144+ CD309+

Supplementary table 2 – Phenotypic cell types per organ, related to Figure 1.

Phenotypic cell type	Femur	Hip	Liver	Total
Committed progenitors (Lin-CD34+CD38+)	291	128	257	676
Non-committed progenitors (Lin-CD34+CD38-)	262	114	187	563
Basophils	76	0	47	123
CLPs	33	0	41	74
CMPs	130	12	76	218
CD62L-CD52-CD114-CD125-CD117+	79	0	86	165
Dendritic cells	85	0	87	172
Erythroid progenitors	126	0	0	126
Endothelial cells	35	0	29	64
Eosinophils	20	0	29	49
Erythrocytes	22	0	39	61
GMPs	117	7	88	212
HSCs	50	8	67	125
ILCs	75	0	83	158
MEPs	131	22	95	248
Immature MKs	48	0	0	48
Mature MKs	31	0	0	31
Monocytes	45	0	43	88
MPPs	121	38	111	270
Bone marrow cells	137	0	0	137
Mast cells	41	0	0	41
Neutrophils	39	0	31	70
NK-cell	74	0	0	74
CD-REF cells	500	0	0	500
T cells	79	0	0	79
Mature B cells	20	0	22	42
Pre-B cells	34	0	26	60
Pro-B cells	26	0	4	30
Total	2727	329	1448	4504

Supplementary table 3 – scRNA-Seq samples, related to the STAR Methods section.

No.	Gates	Min #counts	Max #counts	Min #genes	Max #genes	Max %mito
1	Non-committed progenitors Committed progenitors	10,000	1,750,000	200	7,500	40
2	Non-committed progenitors Committed progenitors HSCs	10,000	2,000,000	200	7,500	40
3	Non-committed progenitors Committed progenitors HSCs MPPs CMPs GMPs MEPs	10,000	1,500,000	200	7,500	40
4	Non-committed progenitors Committed progenitors HSCs MPPs CMPs GMPs MEPs	10,000	2,750,000	200	9,000	40
5	Immature MKs Mature MKs	10,000	1,750,000	200	9,000	40
6	Non-committed progenitors Committed progenitors HSCs MPPs CMPs GMPs MEPs	10,000	1,500,000	350	7,500	40
7	Bone marrow cells	400	1,750,000	200	7,500	60
8	Endothelial cells Pro-B cells Pre-B cells Mature B	400	1,500,000	200	7,500	60
9	CLPs GMPs Monocytes	1,000	2,000,000	200	7,500	40
10	CD-REF cells	1,000	550,000	200	6,000	40
11	ILCs NK cells T cells	2,000	1,500,000	200	10,000	40
12	Neutrophils Eosinophils Erythrocytes	2,000	1,500,000	200	6,000	40
13	Dendritic cells Mast cells Basophils	2,000	1,500,000	200	7,500	40
14	Erythroid progenitors	500	3,000,000	100	4,500	40
15	CD62L- CD52- CD114- CD125- CD117+	2,000	1,000,000	1,200	7,000	40