Supplementary Information

Functional and Transcriptional Heterogeneity of Human Hemopoietic Lympho-Myeloid Progenitors at the Single Cell Level

Contents

Supplementary Figure Legends
Supplementary Figures
List of Supplementary Tables
Supplementary Table 1

Not included: Supplementary Tables 2-9 (Separate Excel files)
Supplementary Figure Legends:

Figure S1. Related to Figure 1. Immunophenotypic purification of distinct human hemopoietic stem and progenitor cells (HSPCs).

a) Representative FMO controls used for setting of gates for human CB HSPC sorting.
b) Representative sorting purity plots and average CB HSPC sorting purities (N=50, ± SD) for the functional and the transcriptional analyses.

Figure S2. Related to Figure 2. In vitro functional lympho-myeloid potential of LMPP, MLP and GMP.

a) Representative flow cytometry analysis of the B-cell, NK, monocytic and granulocytic output of CB LMPP, MLP and GMP cultured for 2 weeks on MS-5 stroma with SCF, G-CSF, FLT3L, IL15, IL2 and Dup-697. Frequencies shown are mean from 3 biological replicates calculated as a percentage of human CD45+ cells. Antibodies used are indicated.
b) Representative flow cytometry analysis of T-cell output of CB LMPP, MLP and GMP cultured for 5 weeks on OP9-hDL1 stroma with SCF, FLT3L and IL7. Frequencies shown are average from 5 biological replicates calculated as a percentage of human CD45+ cells. Antibodies used are indicated.

Figure S3. Related to Figure 3. LMPP and GMP are lympho-myeloid progenitors, while MLP is mainly a lymphoid progenitor in liquid culture quantitative in vitro assays.

a) LDA and single cell in vitro culture conditions used in the study.
b) Representative FACS profiles of the outputs from the quantitative in vitro assays. All plots are gated on human CD45+ cells.
c) LDA plots showing the frequency (f - 1 in X cells can give rise to) of LMPP, MLP and GMP cells with B-cell, NK cell, monocytic and granulocytic potential. Plots are generated using R and the lines represent the estimates calculated using ELDA software.
d) Cloning efficiency of the single cell LMPP, MLP and GMP in S7T2GM/G/M condition. Significance is defined using Fisher’s exact test.
e) Cloning efficiency of lymphoid lineages (ei) and myeloid lineages (eii) of single cell LMPP, MLP and GMP cultured in S7T2GM/G/M condition. Bars indicate total cloning efficiency; filled portion indicates the proportion of lymphoid potential (lymphoid plus mixed clones) or myeloid potential (myeloid plus mixed clones). Mean ± SD is shown. Significance was defined using two-tailed students unpaired t-test.

f) Single-, g) bi- and h) multi-lineage outputs from single cells in S7T2GM/G/M condition, presented as a percentage of the positive plated wells.

i) Lymphoid (Ly), myeloid (My) and lympho-myeloid (Ly-My) outputs presented as a percentage of all plated single cell LMPP, MLP and GMP wells in S7T2GM/G/M condition.

For the LDA analysis data come from 4 biological replicates, for the single cell assay in S7T2GM/G/M condition data come from 2 biological replicates.

**Figure S4. Related to Figure 4. Human LMPP, MLP and GMP progenitors have distinct differentiation potential in vivo.**

a) Gating strategy for the sorting of human progenitors for in vivo transplantation.

b) Table summarizing sorting purities of all transplanted CB units used for transplantation assays.

c) Number of cells of LMPP, MLP and GMP populations injected per ossicle.

d) Correlation between the number of transplanted progenitor cells and the myeloid/lymphoid ratio of the output cells.

**Figure S5. Related to Figure 5. Transcriptional profiling of human HSPCs in bulk shows the distinct transcriptional patterns of human progenitor populations.**

a) Heatmap showing the expression of all genes by each HSPC population with hierarchical clustering on genes and populations. Expression values are normalized per gene (by row).

b) Plot comparing the eigenvalues for each principle component (PC) for the PC analysis (PCA) of human HSPC using the top 300 ANOVA genes (black) and 300 randomly selected genes from a randomized expression matrix (red).

c) 3D PCA plot showing the position of HSPC populations using the top 300 ANOVA genes and the first 3 PCs. Percentage variance for each PC is shown.
d) PCA plots showing the position of HSPC populations using 10,000-1,000 genes with the highest variance across all HSPC populations. Percentage variance for each PC is shown.
e) Table of differentially expressed genes in one versus one comparisons of HSPCs. Genes are upregulated in a population column versus row.
f) Heatmap showing the expression of genes recently identified as being expressed by lineage primed hemopoietic cells\textsuperscript{25}. Genes are color-coded according to their classification by Velten et al. Expression values are normalized per gene (by row) and are shown as a minimum-maximum scale.
g-h) Heatmap showing the expression of (f) hemopoietic-related transcription factors that were differentially expressed between the MLP and GMP and (g) cytokine and chemokine related genes by the LMPP, GMP and MLP. Genes affiliated with the lymphoid or myeloid lineages are color-coded (lymphoid: orange, myeloid: green) and genes associated with immune function are labeled in black. Expression values are normalized per gene (by row) and are shown as a minimum-maximum scale.

**Figure S6. Related to Figure 6. Transcriptional single lympho-myeloid progenitor cell heterogeneity.**

a) Heatmap showing hierarchical clustering on genes (rows) and single cells (columns).
b) Top, bi), Cell type composition of the 3 clusters identified in the dendrogram in (a). Middle, bii), differentially expressed genes between cluster pairs. Bottom, biii) expected number of cells per cluster, based on the functional assays.
c) Diffusion map dimensionality reductions colored in by cell type (ci) and cluster membership (cii).
d) Heatmap showing clustering of single LMPP, GMP and MLP using the 55 most highly and variably expressed genes between clusters. Data from single cell RNA-sequencing, from two donors was used. The heat map shows clustering from one of two. Data from the other donor is in **Fig. 6b**. Log-normalised gene expression (rows) for each single cell (columns) is shown.
e) Cell type composition of the 3 clusters identified into heatmap in (d).
f) PCA plot colored in cell type (fi) or by cluster membership (fii).
For single cell qRT-PCR gene expression analysis: data from 4 biological replicates. For single cell RNA sequencing analysis: data from 2 biological replicates (replicate 1 in Fig. 6, replicate 2 in Supplementary Fig. 6).

**Figure S7. Related to Figure 7. Further functional purification of the lympho-myeloid progenitor populations.**

a) CD10 level in LMPPs in the 3 clusters identified from clustering of single cell RNA Seq data presented Fig. 6b. Wilcoxon rank sum test was used to define significance.

b) Histograms showing expression of CD10 and CD45RA for LMPP with different functional output. The percentage cut-offs used to define the new sorting strategy for LMPP\textsuperscript{my} and LMPP\textsuperscript{mix} are shown. Thresholds were defined based on maximum CD10 and CD45RA expression of LMPPs with myeloid output.

c) Cloning efficiency of single MLP, LMPP, LMPP\textsubscript{ly}, LMPP\textsubscript{mix} and GMP in SF7b condition. Significance was defined using Fisher’s exact test.

d) Cloning efficiency of lymphoid lineages (d\textsubscript{i}) and myeloid lineages (d\textsubscript{ii}) of single cell MLP, LMPP, LMPP\textsubscript{ly}, LMPP\textsubscript{mix} and GMP in SF7b condition. Bars indicate total cloning efficiency; filled portion indicates the proportion of lymphoid potential (lymphoid plus mixed clones) or myeloid potential (myeloid plus mixed clones). Mean ± SD is shown. Significance was defined using two-tailed students unpaired t-test.

e) Single-\textsubscript{a}, f) bi-\textsubscript{-} and g) multi-lineage outputs from single cells in SF7b condition, presented as % of the plated wells with output.

h) CD38 level in GMPs in the 3 clusters identified from clustering of single cell RNA Seq data presented Fig. 6b. Wilcoxon rank sum test was used to define significance.

i) Correlation between expression of selected genes and cell surface marker expression in GMPs. \(\rho\) indicates Spearman's rank correlation coefficient values and \(p\) is corresponding p-value from cor.test function in R.
j) Histograms showing expression of CD38 for GMP with different functional output. The percentage cut-off used to define the new sorting strategy for GMP CD38$^{\text{hi}}$ and CD38$^{\text{mid}}$ are shown. This cutoff corresponds to the highest CD38 level of GMPs with lympho-myeloid functional output.

For functional assay based on the revised LMPP sorting strategy (SF7b condition) data are from 6 biological replicates; for LMPP, MLP and GMP controls data from 9 biological replicates (including 3 biological replicates used for Fig. 3i-n).

List of Supplementary Tables:

**Supplementary Table 1**: List of antibodies used for FACS sorting and analysis.

**Supplementary Table 2**: List of Taqman probes, related to Figure 2.

**Supplementary Table 3**: LDA and single cell liquid culture well outputs, related to Figure 3 and Supplementary Figure 3.

**Supplementary Table 4**: Table of ANOVA genes between HSPC populations, related to Figure 5.

**Supplementary Table 5**: Table of genes from one versus one comparison between HSPC populations, related to Figure 5.

**Supplementary Table 6**: Table of genes from LMPP, MLP and GMP signatures, related to Figure 5.

**Supplementary Table 7**: MetaCore Enrichment analysis of genes upregulated in the LMPP, MLP and GMP.

**Supplementary Table 8**: List of Taqman probes, related to Figure 6.

**Supplementary Table 9**: Single cell liquid culture well outputs from LMPP$^{\text{ly}}$, LMPP$^{\text{mix}}$, GMP CD38$^{\text{hi}}$ and CD38$^{\text{mid}}$ cells, related to Figure 7.
Supplementary Figure 1

(a) Supplementary Figure 1a:

1. FMO Hoechst
2. CD34-PerCP
3. CD38-FITC

(b) Supplementary Figure 1b:

1. Lin-
2. CD34+CD38-
3. CD38-CD10-

HSC 99.5 ± 1.0

MPP 99.6 ± 0.7

LMPP 99.4 ± 1.8

MLP 99.7 ± 0.9

HSC 97.0 ± 3.0

GMP 99.4 ± 1.4

MEP 99.1 ± 1.6
Supplementary Figure 2

(a) LMPP
CD15 CD14 CD56 CD19
14.4 12.8 15.1
10.6 18.7

MLP
CD15 CD14 CD56 CD19
7.9 12.3 34.5
14.1 19.0

GMP
CD15 CD14 CD56 CD19
16.7 36.7 0.2
11.7 4.3

(b) hCD45
CD1a+CD7+
LMPP
11.3 57.6 11.8 21
5.3 27.1 13.7 11

MLP
39.8 57.7 7.2 32.2
7.8 10 5.7 12.5

GMP
CD1a CD7 CD4 CD8
8.3 71.5 3.9 15.1
9.0 15.1 14.6 24.2

5.0 27.6 14.6 24.2
**Supplementary Figure 3**

### Condition 1: SGF15/2
LDA (Fig. 3b, Supplementary Fig. 3c) & single cell (Fig. 3c-h)

<table>
<thead>
<tr>
<th>Cytokines</th>
<th>Time point of analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 ng/ml SCF</td>
<td>2-2.5 weeks</td>
</tr>
<tr>
<td>10 ng/ml G-CSF</td>
<td></td>
</tr>
<tr>
<td>10 ng/ml FLT3L</td>
<td></td>
</tr>
<tr>
<td>10 ng/ml IL15</td>
<td></td>
</tr>
<tr>
<td>10 ng/ml IL2</td>
<td></td>
</tr>
<tr>
<td>10^{-7} M Dup-697</td>
<td></td>
</tr>
</tbody>
</table>

### Condition 2: S7T2GM/G/M (taken from ref 18)
single cell (Supplementary Fig. 3d-i)

<table>
<thead>
<tr>
<th>Cytokines</th>
<th>Time point of analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 ng/ml SCF</td>
<td>4 weeks</td>
</tr>
<tr>
<td>10 ng/ml IL7</td>
<td></td>
</tr>
<tr>
<td>50 ng/ml TPO</td>
<td></td>
</tr>
<tr>
<td>10 ng/ml IL2</td>
<td></td>
</tr>
<tr>
<td>20 ng/ml GM-CSF</td>
<td></td>
</tr>
<tr>
<td>20 ng/ml G-CSF</td>
<td></td>
</tr>
<tr>
<td>10 ng/ml M-CSF</td>
<td></td>
</tr>
</tbody>
</table>

### Condition 3: SF7b/Dox (taken from ref 37)
single cell (Fig. 3i-n)

<table>
<thead>
<tr>
<th>Cytokines</th>
<th>Time point of analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 ng/ml SCF</td>
<td>B/NK/M/G: 3 weeks</td>
</tr>
<tr>
<td>20 ng/ml FLT3L</td>
<td>T-cell: 6 weeks</td>
</tr>
<tr>
<td>10 ng/ml IL7</td>
<td></td>
</tr>
<tr>
<td>20 ng/ml Insulin</td>
<td></td>
</tr>
<tr>
<td>1 μg/ml Doxycyclin</td>
<td>(at week 3)</td>
</tr>
</tbody>
</table>

#### Cytokines
- SCF
- G-CSF
- FLT3L
- IL15
- IL2
- IL7
- TPO
- GM-CSF
- M-CSF
- Insulin
- Doxycyclin

#### Time point of analysis
- 2-2.5 weeks
- 4 weeks
- 4 weeks
- B/NK/M/G: 3 weeks
- T-cell: 6 weeks
- 2-2.5 weeks
- 4 weeks
- 4 weeks
- B/NK/M/G: 3 weeks
- T-cell: 6 weeks

---

**B cell frequency**

**NK cell frequency**

**Monocyte frequency**

**Granulocyte frequency**

---

**Potentiality**

**LMPP**

**MLP**

**GMP**

---

**ei** Percentage lymphoid potential

**ei** Percentage myeloid potential

---

**f** Uni-lineage

**g** Bi-lineage

**h** Multi-lineage

---

**i** Potentiality
Supplementary Figure 4

**a**

Gating strategy

- **CD34**+/**CD38**
- **CD34**+/**CD38**

**b**

<table>
<thead>
<tr>
<th>sample</th>
<th>population</th>
<th>LMPP</th>
<th>MLP</th>
<th>GMP</th>
<th>HSC</th>
<th>MPP</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCB 1</td>
<td></td>
<td>99.5</td>
<td>0.75</td>
<td>4.66</td>
<td>0.17</td>
<td>0.34</td>
</tr>
<tr>
<td>UCB 2</td>
<td></td>
<td>94</td>
<td>0.75</td>
<td>5.24</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>UCB 3</td>
<td></td>
<td>92.5</td>
<td>5.24</td>
<td>1.53</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCB 4</td>
<td></td>
<td>94.5</td>
<td>1.12</td>
<td>1.68</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCB 5</td>
<td></td>
<td>90.4</td>
<td>4.37</td>
<td>2.62</td>
<td>0</td>
<td>0.44</td>
</tr>
<tr>
<td>UCB 6</td>
<td></td>
<td>92.2</td>
<td>6.38</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCB 7</td>
<td></td>
<td>94.6</td>
<td>1.39</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCB 8</td>
<td></td>
<td>93.7</td>
<td>1.15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCB 9</td>
<td></td>
<td>94.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCB 10</td>
<td></td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCB 11</td>
<td></td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCB 12</td>
<td></td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCB 13</td>
<td></td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCB 14</td>
<td></td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCB 15</td>
<td></td>
<td>0.84</td>
<td>0</td>
<td>99.2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCB 16</td>
<td></td>
<td>0.52</td>
<td>0.52</td>
<td>97.4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCB 17</td>
<td></td>
<td>0</td>
<td>0</td>
<td>99.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCB 18</td>
<td></td>
<td>0</td>
<td>0</td>
<td>93.3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCB 19</td>
<td></td>
<td>0</td>
<td>0</td>
<td>98.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCB 20</td>
<td></td>
<td>0</td>
<td>0</td>
<td>97.8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCB 21</td>
<td></td>
<td>0.83</td>
<td>0</td>
<td>94.2</td>
<td>0</td>
<td>0.82</td>
</tr>
<tr>
<td>UCB 22</td>
<td></td>
<td>0.41</td>
<td>0</td>
<td>95.9</td>
<td>0</td>
<td>0.82</td>
</tr>
</tbody>
</table>

**c**

Number of injected cells per ossicle

**d**

Myeloid/lymphoid ratio

- **LMPP** $r = 0.1062$
- **MLP** $r = 0.1082$
Supplementary Figure 5

**Panel a**
- Heatmap of gene expression data, with colors representing row Z-scores.
- Legends indicate different cell types: HSC (red), MPP (green), LMPP (blue), MLP (yellow), CMP (magenta), GMP (cyan), MEP (gray).

**Panel b**
- Eigenvalue plot showing the variance explained by each principal component.
- Two datasets are compared: 300 ANOVA genes (black) and 300 random genes (red).

**Panel c**
- 3D scatter plot of the first three principal components (PC1, PC2, PC3) for 300 ANOVA genes.

**Panel d**
- Scatter plots showing the distribution of 10000 variant genes on PC1 and PC2.

**Panel e**
- Table listing the expression values for different cell types (HSC, MPP, LMPP, MLP, CMP, GMP, MEP) across various cell lines.

**Panel f**
- Gene expression heatmap for different cell types.

**Panel g**
- Heatmap showing gene expression levels for various cell types, with different colors indicating expression levels.

**Panel h**
- Additional gene expression heatmap highlighting specific genes.

**Legend**
- Various cell types and genes are color-coded for easy identification.
Supplementary Figure 6

a) Heatmap showing gene expression across different cell types and clusters.

b) Heatmap showing the expression of high A vs B and high B vs A genes across clusters.

bii) Heatmap showing the expression of Lymphoid Genes and Myeloid Genes.

c) Scatter plots showing the expression of genes across different cell types.

ci) Scatter plot showing the expression of genes across Cluster 1, Cluster 2, and Cluster 3.

cii) Scatter plot showing the expression of genes across Cluster 1 and Cluster 2.

d) Heatmap showing the expression of genes across different cell types and clusters.

de) Heatmap showing the expression of genes across different cell types and clusters.

f) Scatter plots showing the expression of genes across different cell types.

fii) Scatter plots showing the expression of genes across Cluster 1, Cluster 2, and Cluster 3.
Supplementary Figure 7

(a) CD10 level for LMPP

(b) Percentage of positive plated wells for CD38

(c) Uni-lineage

(d) Bi-lineage

(e) Multi-lineage

(f) GMP

(g) MPO expression (Induction)

(h) CD45RA expression

(i) CD38 expression (Index sorting)

(j) GMP