

# Single-cell multiomics of human fetal hematopoiesis define a developmental-specific population and a fetal signature

Mikael N. E. Sommarin, Rasmus Olofzon, Sara Palo, Parashar Dhapola, Shamit Soneji, Göran Karlsson, and Charlotta Böiers

Division of Molecular Hematology, Department of Laboratory Medicine, Lund Stem Cell Center, Faculty of Medicine, Lund University, Lund, Sweden

## Key Points

- An immunophenotypic and transcriptional single-cell map of first trimester fetal liver hematopoietic stem and progenitor cells is presented.
- A fetal-specific cluster was observed, and universal differences in gene expression between fetal and adult hematopoietic cells were identified.

Knowledge of human fetal blood development and how it differs from adult blood is highly relevant to our understanding of congenital blood and immune disorders and childhood leukemia, of which the latter can originate in utero. Blood formation occurs in waves that overlap in time and space, adding to heterogeneity, which necessitates single-cell approaches. Here, a combined single-cell immunophenotypic and transcriptional map of first trimester primitive blood development is presented. Using CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing), the molecular profile of established immunophenotype-gated progenitors was analyzed in the fetal liver (FL). Classical markers for hematopoietic stem cells (HSCs), such as CD90 and CD49F, were largely preserved, whereas CD135 (FLT3) and CD123 (IL3R) had a ubiquitous expression pattern capturing heterogeneous populations. Direct molecular comparison with an adult bone marrow data set revealed that the HSC state was less frequent in FL, whereas cells with a lymphomyeloid signature were more abundant. An erythromyeloid-primed multipotent progenitor cluster was identified, potentially representing a transient, fetal-specific population. Furthermore, differentially expressed genes between fetal and adult counterparts were specifically analyzed, and a fetal core signature was identified. The core gene set could separate subgroups of acute lymphoblastic leukemia by age, suggesting that a fetal program may be partially retained in specific subgroups of pediatric leukemia. Our detailed single-cell map presented herein emphasizes molecular and immunophenotypic differences between fetal and adult blood cells, which are of significance for future studies of pediatric leukemia and blood development in general.

## Introduction

Knowledge of our fetal blood system can facilitate the generation of transplantable hematopoietic stem cells (HSCs) derived from pluripotent stem cells (PSCs) for future regenerative medicine<sup>1</sup> and also increase our understanding of congenital blood disorders and pediatric leukemia.<sup>2</sup>

Submitted 24 January 2023; accepted 16 June 2023; prepublished online on *Blood Advances* First Edition 28 June 2023. <https://doi.org/10.1182/bloodadvances.2023009808>.

The data generated within this study have been deposited in the Gene Expression Omnibus (GEO) database under accession number GSE192966. Previously published data used within this study are available at: <https://data.mendeley.com/datasets/phfgms85x2/1><sup>6</sup>; <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-9067/12>; ArrayExpress: E-MTAB-6382<sup>15</sup>; <https://pecan.stjude.cloud/static/hg19/pan-all/BALL-1988S-HTSeq.zip><sup>18</sup>; GEO: GSE173076.<sup>19</sup> The code along with all processed data used to perform this analysis are available on OSF: [https://osf.io/8yxb/?view\\_only=6877d308ecca4879bea2688966ec2226](https://osf.io/8yxb/?view_only=6877d308ecca4879bea2688966ec2226).

The code can also be found in a Github repository: [https://github.com/razofz/CB\\_FL](https://github.com/razofz/CB_FL). The FL CS22 data were also uploaded to an online visualization platform: [https://singlecell.broadinstitute.org/single\\_cell/study/SCP2022/fetal-liver-hematopoiesis#study-visualize](https://singlecell.broadinstitute.org/single_cell/study/SCP2022/fetal-liver-hematopoiesis#study-visualize)

The full-text version of this article contains a data supplement.

© 2023 by The American Society of Hematology. Licensed under [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International \(CC BY-NC-ND 4.0\)](https://creativecommons.org/licenses/by-nc-nd/4.0/), permitting only noncommercial, nonderivative use with attribution. All other rights reserved.

Fetal and adult blood cells differ in cell cycle status, molecular profile, and composition of progenitor cells, and hematopoiesis takes place in different niches depending on the stage of development.<sup>3-6</sup> Moreover, some innate immune cells are formed almost exclusively during fetal life.<sup>7</sup> The blood system emerges in waves and is initiated by HSC-independent hematopoiesis in the yolk sac, where immature erythroid and myeloid cells, critical for the growing embryo, emerge, followed by erythromyeloid and lymphomyeloid progenitors.<sup>5,7-9</sup> From around day 27 (~Carnegie Stage [CS] 13), hematopoietic stem and progenitor cells (HSPCs) are formed in the aorta-gonad-mesonephros region and later migrate to the fetal liver (FL), where they mature and expand. The bone marrow (BM), the site of blood production in adults, starts to be colonized at the end of the first trimester.<sup>5</sup>

The different waves and niches add to heterogeneity, and the sparsity of human fetal samples makes studies in early development demanding. Much of our knowledge regarding embryonic and fetal hematopoiesis has been translated from studies of the murine system, but accumulating reports of the human counterpart complement the current model.<sup>5,10</sup> Of late, single-cell studies have emerged focusing on embryonic/fetal development, increasing our knowledge of blood formation during human ontogeny.<sup>11-13</sup> However, direct comparisons between human fetal and adult hematopoiesis are just beginning to be explored.<sup>6,14,15</sup> Immunophenotypic markers used to purify and prospectively analyze progenitors may not be equally expressed throughout development, making direct comparisons of purified populations less robust. Furthermore, the analysis of transcriptomic data from different niches and stages also poses challenges, as heterogeneity may be lost upon integration.

Here, we set out to directly compare the transcriptome and immunophenotype of embryonic/fetal and adult hematopoiesis using CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing).<sup>16</sup> To better preserve heterogeneity in cells from different niches and stages of development, a projection approach was used.<sup>17</sup> This allowed for a comparison of progenitor composition within the fetal and adult HSPC compartment. The projection approach revealed a multipotent progenitor cluster, which was hardly detectable in the adult BM. Investigating a publicly available data set,<sup>6</sup> these cells were specifically found in the FL niche. Furthermore, by comparing molecular signatures across developmental stages, we could define a fetal core signature, with genes specifically upregulated or downregulated in fetal cells compared to adult. By using a publicly available RNA-seq data set from patients with acute lymphoblastic leukemia (ALL),<sup>18</sup> the fetal core signature was found to define some cases of pediatric leukemias, thus enabling the separation of specific subgroups of ALL.

In summary, our comparison between different stages of development highlights important transcriptomic and immunophenotypic differences in human fetal and adult primitive blood cell development.

## Materials and methods

### Sample preparation

Human FLs were donated from elective medical terminations of pregnancy after informed consent and with the approval of the Regional Ethics Review Board (Lund)/the Swedish Ethical Review Authority and the Swedish National Board of Health and Welfare.

Embryos/fetuses were staged according to CS and postconceptional weeks (pcw) and were all from the first trimester of pregnancy (developmental stages CS16-9 pcw) (supplemental Table 1). The FL was mechanically disrupted and dissociated through a 40  $\mu$ m filter. The cells were frozen in StemCellBanker (Amsbio) or in fetal bovine serum with 10% dimethyl sulfoxide. Samples were stored at  $-150^{\circ}\text{C}$ . Cord blood (CB) ( $n = 3$ ) and adult BM (20 to 30-year olds, males;  $n = 2$ ) data sets were from a previous study.<sup>19</sup>

### Sample preparation and cell sorting

FL samples were thawed on the day of experiment and stained with a panel of CITE-seq antibodies (supplemental Table 2), a hashtag oligonucleotide (HTO) antibody to facilitate sample multiplexing, a Fc receptor blocking reagent (Miltenyi Biotec), CD45-Alexa700 (HI30, Biolegend), CD34-FITC (581, Biolegend) and lineage markers (LIN: CD19-BV605 (SJ25C1, BD Bioscience), CD3-PE-Cyanine5 (UCHT1, Biolegend), CD2-PE (RPA-2.10, Biolegend), CD16-BV421 (3G8, Biolegend), CD14-PE-Cyanine7 (M5E2, Biolegend), and CD235a-PE-Cyanine5 (GA-R2, BD Bioscience)). These were incubated at  $4^{\circ}\text{C}$  for 30 min. Samples were then washed and resuspended in phosphate-buffered saline with 2% fetal bovine serum (GE Lifesciences) and 7-aminoactinomycin D (7-AAD) (BD Bioscience). Up to 10 000 LIN<sup>+</sup>CD45<sup>+</sup>CD34<sup>+</sup> cells were sorted from each donor using a BD FACSAria IIu. The CB and adult BM data sets from<sup>19</sup> were generated in a similar way from LIN<sup>+</sup>CD34<sup>+</sup> cells.

### CITE-seq library generation and sequencing

After sorting, the samples were loaded on to the 10 $\times$  genomics 3' version 2 platform (10 $\times$  Genomics) and single-cell RNA-seq was performed according to the manufacturers' instructions with minor changes according to a previous study<sup>16</sup> to allow for CITE-seq. After reverse transcription and cDNA amplification, the resulting libraries were sequenced on a NOVAseq (Illumina). After sequencing, the binary base call files were processed using Cellranger mkfastq to produce FASTQ files, and Cellranger count was used for alignment to the human reference genome (Hg38), filtering, barcode counting, and unique molecular identifier counting. The resulting matrix files were further analyzed in sequential steps using Seurat<sup>20</sup> and Scarf.<sup>17</sup>

### Cell filtering and uniform manifold approximation and projection creation

The output from Cellranger was processed using Seurat (V4).<sup>20</sup> Samples were filtered based on unique molecular identifier, gene, and mitochondrial counts to remove low-quality cells (supplemental Table 1). The cells were demultiplexed using the hashtag antibody, and doublets were removed. Each donor was treated separately. RNA reads were log-normalized and antibody derived tag (ADT) and HTO data were CLR (centered-log ratio) normalized using the NormalizeData function. CS22 FL samples were integrated using the IntegrateData function, and cell cycle effects were regressed out using the ScaleData function. The output data were used to make a unified uniform manifold approximation and projection (UMAP). The cells were clustered using FindClusters (resolution 0.7), which resulted in 11 clusters. These clusters were associated with a hematopoietic progenitor cell state based on the results from the FindAllMarkers function.

## Weighted nearest neighbor analysis

The scaled and normalized RNA and ADT data of FL CS22 were used to perform weighted nearest neighbor analysis (WNN).<sup>20</sup> The FindMultiModalNeighbors function of Seurat (V4) was used with the first 30 PCs of the RNA principal component analysis (PCA) and the first 18 PCs of the ADT PCA with default settings, after which RunUMAP was performed using WNN and default settings. The cells were clustered using FindClusters, with the smart local moving algorithm (resolution of 0.5), resulting in 10 clusters. The metadata and the UMAP coordinates were exported and used in Python for visualization.

## ADT gating analysis

The CLR normalized ADT values from Seurat were loaded into Python, transformed by their antilog (base  $e$ ), and then multiplied by 1000 to enable gating in FlowJo. The values were exported using the write\_FCS function and loaded into FlowJo V10 (BD), where conventional gating was performed using internal negative controls to set the gates. The gated populations were exported as csv, and using Python, the ADT values from cells in the exported gates were matched to the original data to identify the cell identities.

The ADTs were used to immunophenotypically define the following populations (gated from LIN<sup>-</sup>CD34<sup>+</sup>CD45<sup>+</sup> cells): HSC<sup>CD49F</sup>: LIN<sup>-</sup>CD34<sup>+</sup>CD38<sup>-/low</sup>CD45RA<sup>-</sup>CD90<sup>+</sup>CD49F<sup>+</sup>; HSC<sup>CD90</sup>: LIN<sup>-</sup>CD34<sup>+</sup>CD38<sup>-/low</sup>CD45RA<sup>-</sup>CD90<sup>+</sup>; MPP: LIN<sup>-</sup>CD34<sup>+</sup>CD38<sup>-/low</sup>CD45RA<sup>-</sup>CD90<sup>-</sup>; LMPP: LIN<sup>-</sup>CD34<sup>+</sup>CD38<sup>-/low</sup>CD45RA<sup>+</sup>CD90<sup>-</sup>; CMP<sup>CD123</sup>: LIN<sup>-</sup>CD34<sup>+</sup>CD38<sup>+</sup>CD10<sup>-</sup>CD45RA<sup>-</sup>CD123<sup>+</sup>; CMP<sup>CD135</sup>: LIN<sup>-</sup>CD34<sup>+</sup>CD38<sup>+</sup>CD10<sup>-</sup>CD45RA<sup>-</sup>CD135<sup>+</sup>; GMP<sup>CD123</sup>: LIN<sup>-</sup>CD34<sup>+</sup>CD38<sup>+</sup>CD10<sup>-</sup>CD45RA<sup>+</sup>CD123<sup>+</sup>; GMP<sup>CD135</sup>: LIN<sup>-</sup>CD34<sup>+</sup>CD38<sup>+</sup>CD10<sup>-</sup>CD45RA<sup>+</sup>CD135<sup>+</sup>; MEP<sup>CD123</sup>: LIN<sup>-</sup>CD34<sup>+</sup>CD38<sup>+</sup>CD10<sup>-</sup>CD45RA<sup>-</sup>CD123<sup>+</sup>; MEP<sup>CD135</sup>: LIN<sup>-</sup>CD34<sup>+</sup>CD38<sup>+</sup>CD10<sup>-</sup>CD45RA<sup>-</sup>CD135<sup>+</sup>; CLP<sup>CD10</sup>: LIN<sup>-</sup>CD34<sup>+</sup>CD38<sup>+</sup>CD45RA<sup>+</sup>CD10<sup>+</sup>; CLP<sup>IL7R</sup>: LIN<sup>-</sup>CD34<sup>+</sup>CD38<sup>+</sup>CD45RA<sup>+</sup>IL7R<sup>+</sup>; CLP<sup>CD7</sup>: LIN<sup>-</sup>CD34<sup>+</sup>CD38<sup>+</sup>CD45RA<sup>+</sup>CD7<sup>+</sup>.

## Projection of samples onto a reference UMAP

The projection of cells on the FL CS22 reference UMAP was done in Scarf.<sup>17</sup> In brief, the UMAP coordinates, HVGs, and cluster information of the FL CS22 analysis in Seurat were loaded into Scarf, and by using the run\_mapping function (with  $k = 11$ ), the nearest neighbors of each cell projected were calculated. This was performed for FL CS16, FL CS22, FL 9 pcw, CB, and adult BM, the latter 2 from Sommarin et al,<sup>19</sup> for immunophenotypically defined populations from Ranzoni et al,<sup>12</sup> and with data from Roy et al.<sup>6</sup> For the data from Ranzoni et al,<sup>12</sup> CORAL was set to "True" in the run\_mapping function to account for the difference in technology used to produce the data. With the adult BM as a reference, the same settings as for the FL CS22 were used, projecting FL CS16, FL CS22, FL 9pcw, CB, and adult BM cells. The projected cells were classified using the get\_target\_classes function (threshold = 0.4). The mapping score was calculated using get\_mapping\_score, and the size of each cell in the reference map was set proportional to the mapping score.

## Harmony integration

The FL CS22 and adult BM data sets were integrated with Harmony (version 0.1.0).<sup>21</sup> The FL and BM Seurat objects were merged (merge function). The data were log-normalized, and highly

variable genes were calculated with the FindVariableFeatures function. Cell cycle effects were regressed out using the ScaleData function. PCA was calculated using the RunPCA function, after which the RunHarmony function was used. The Harmony reduction was then used in FindNeighbors (dimensions 1:15), FindClusters (resolution = 0.8), and RunUMAP (dimensions 1:15, 15 neighbors, minimum distance = 0.005, spread = 6). The metadata and the UMAP coordinates were exported and used in Python for visualization.

## Investigation of molecular differences in primitive FL clusters

To analyze differentially expressed genes (DEGs) in the CS22 MPP-I cluster compared with HSC and MPP-II, respectively, the FindMarkers function of Seurat was used. DEGs (adjusted [adj]  $P$  value < .001; log<sub>2</sub> fold change (FC) > |0.5|) and cells were hierarchically clustered using the clustermap function of Seaborn (Euclidean distances and Ward linkage).

CellRadar (using data from HemaExplorer<sup>22</sup>) (available at <https://karlssong.github.io/cellradar/>) was used to investigate lineage affiliation and to create a radar plot using the min-max scaled median value of marker genes in each cluster (adj  $P$  value < .001; log<sub>2</sub> FC > 0.25). Furthermore, the CellCycleScoring function in Seurat was used for cell cycle phase classification for the CS22 FL cells.

## Differential gene expression analysis

To perform DEG analysis, each cluster of the FL CS22 and the adult BM cells projected to a FL cluster were pseudobulked into 2 replicates per donor sample using the make\_bulk function of Scarf.<sup>17</sup> These pseudobulked populations were then exported as csv and loaded into DESeq2<sup>23</sup> using DESeqDataSetFromMatrix (design = ~sample, with "sample" referring to FL CS22 or adult BM), after which DEG analysis was performed (DESeq function with default settings) between FL CS22 and projected adult BM for each of the CS22 clusters. Clusters DC-I and Ly-III were excluded from the analysis because of their low cell count (<60 cells). The resulting files were then analyzed for upregulated (fetal signature) and downregulated (adult signature) genes using the adjusted  $P$  value < .05 and log<sub>2</sub> FC > |1| to make gene lists for each cluster. For validation purposes, DEG analysis was also performed in DESeq2 using pseudobulked samples without pseudoreplicates ( $n = 2$  per developmental stage).

## Gene set enrichment analysis

The DESeq2 normalized counts for FL CS22 and projected adult BM clusters (including pseudoreplicates;  $n = 4$  per developmental stage) were exported for Gene Set Enrichment Analysis (GSEA 4.3.2)<sup>24</sup> to analyze hallmark gene sets (h.all.v7.5.1).<sup>25</sup> Gene sets with a false discovery rate  $q$ -value of <0.05 were considered significantly enriched/depleted (nr of permutations: 1000, permutation type: gene set). The normalized values for all the CS22 FL clusters (with pseudoreplicates, excluding DC-I and Ly-III) were additionally exported for GSEA, now focused on defining differences between MPP-I/HSC and the MPP-I/MPP-II clusters.

## Principal component analysis

Pseudobulked data of the clusters and the ADT-defined immunophenotypic populations (pseudobulking done as above) were loaded into DESeq2 (DESeqDataSetFromMatrix, design = ~1). After running the DESeq function, the data were transformed using the vst function, and the top 500 most variable genes were used in the prcomp function of R to perform PCA.

## Definition of the fetal derived gene signature

The significant genes (adj *P* value < .05; log<sub>2</sub> FC > |1|) for each cluster of the DESeq2 differential testing (described above) were visualized using the Venn tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). Upregulated and downregulated genes were displayed separately. Clusters DC-I and Ly-III were excluded. Genes found to be differentially expressed in all primitive, lymphoid, or myeloid clusters were included in a gene set named 'fetal core signature.' DEGs were also generated without pseudoreplicates (adj *P* value < .05; log<sub>2</sub> FC > |1|) as well as using FindMarkers in Seurat (adj *P* value < .05; log<sub>2</sub> FC > |0.5|), and these were analyzed as above. The fetal core signature as well as the gene sets generated without pseudoreplicates and with Seurat were further validated using a publicly available bulk RNA-seq data set.<sup>15</sup> The fragments per kilobase of transcript per million mapped reads (FPKM) values were log<sub>2</sub>-transformed (log<sub>2</sub>(FPKM + 1)), and PCA was performed using the gene sets described above and for validation purposes with the top 500 variable genes and the top variable genes excluding genes in the fetal core signature.

## Analysis of the fetal derived gene signature in FL and adult BM

The single-cell FL and adult BM data sets were combined by using the ZarrMerge function of Scarf. The genes specific for FL (upregulated in primitive, Ly, or My) were used to subset the combined FL-BM data set. Using the clustermap function of seaborn, the cells and genes were clustered using hierarchical clustering (Euclidean distances and Ward linkage). In addition, all HOX genes expressed in >10 cells were used to create a heatmap using the plot\_marker\_heatmap function of scarf. Furthermore, using Seurat, the data from Roy et al<sup>6</sup> were log-normalized and scaled, after which the AddModuleScore function was used to calculate scores for the fetal and adult signature genes for each cell.

## Investigating the fetal core signature in pediatric ALL

The fetal gene signature was investigated in an *ETV6::RUNX1* iPSC model using publicly available data from a previous study.<sup>15</sup> FPKM values were log<sub>2</sub>-transformed (log<sub>2</sub>(FPKM + 1)), scaled, and centered using the scale function in R and a heatmap created from the identified universal upregulated fetal core genes (14 out of 15 genes were identified). The z-scored mean expression of the same genes was also calculated for each sample and visualized in a box plot.

To investigate the fetal gene signature in childhood ALL, a publicly available bulk RNA-seq data set of almost 2000 patients was used.<sup>18</sup> The following cytogenetic abnormalities were selected: *KMT2A::AFF1*, *KMT2A::MLLT1*, *KMT2A::MLLT3*, *BCR::ABL1*, *ETV6::RUNX1*, and high or low hyperdiploidy. The HTSeq files of these were loaded into DESeq2 using the DESeqDataSetFromHTSeqCount function (design = ~RNA\_lib, to account for

batch effects from different sequencing methods). Next, the DESeq function was run, followed by variance stabilizing transformation by the vst function, and batch effects were removed with removeBatchEffect from the Limma package. The fetal- and adult-specific gene lists were translated into hg37 Ensembl IDs using the getBM function of Biomart (557 Ensembl IDs were detected from the 629 fetal core genes). The genes were used for PCA on all leukemia subtypes above, pooled as well as individually, using the prcomp function with default settings. To validate whether the PC1 variation seen was due to a random effect, a number of Ensembl IDs matching the number of fetal core Ensembl IDs found in the data set were randomly selected after removing genes for which the average expression across the samples was lower than for the lowest-expressed fetal core gene, and a comparison PCA was performed (5 iterations). In addition, for leukemia subtypes where adult counterparts were represented, the fetal core genes were z-scored, and the mean values for upregulated and downregulated fetal core genes were plotted for each age group. For *KMT2A::AFF1*, the top 25 genes and the lowest 25 genes from PC1 were used to generate a heatmap, and a second heatmap was generated looking specifically at expression of different *HOX* genes.

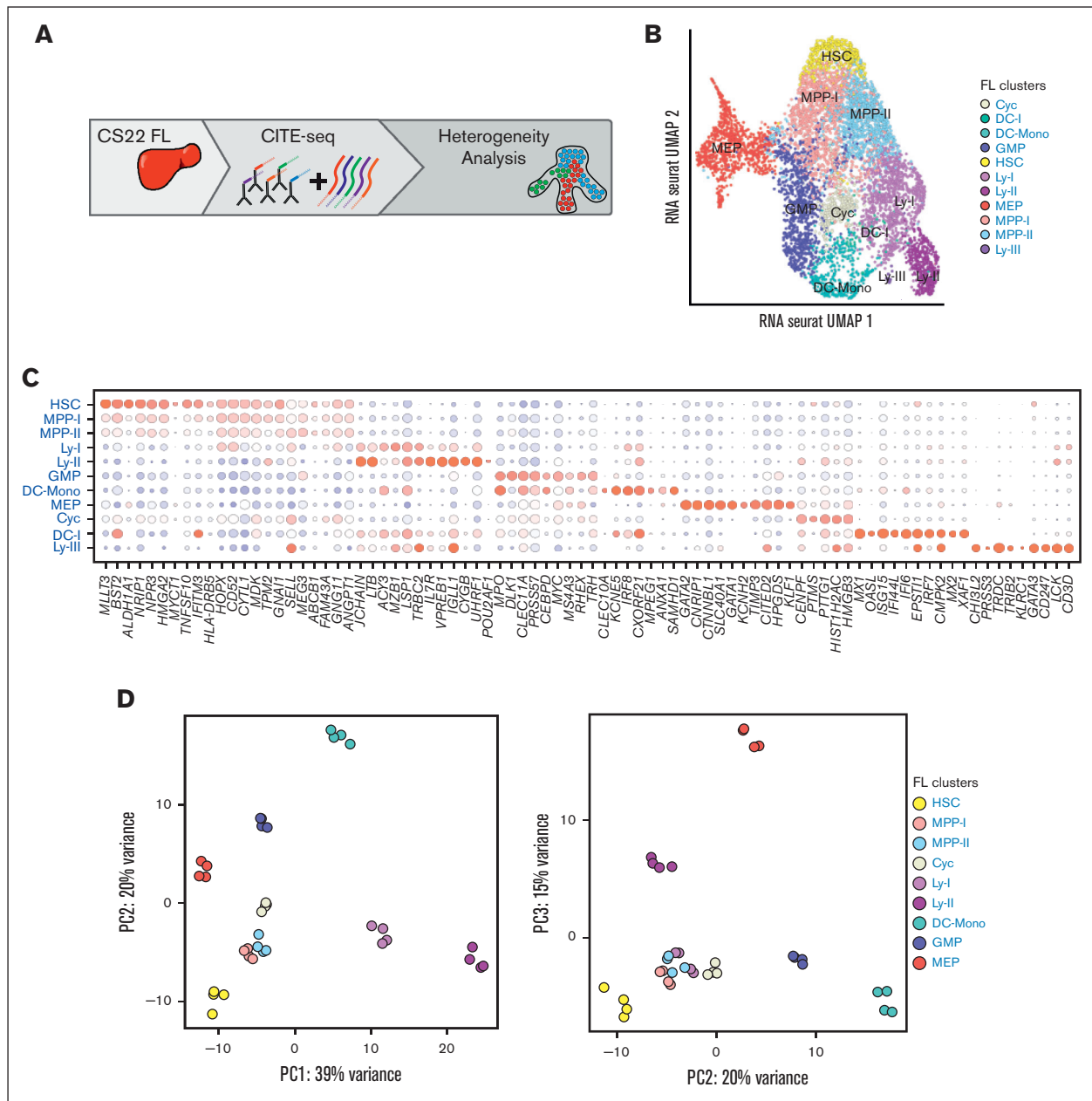
## Statistical analysis

To generate statistics for the contribution of the fetal core signature to the PCA, a bootstrapping approach was used. Here, random sampling of all genes was performed with 10 000 iterations. The variance explained by PC1 of each iteration was used to generate a histogram. The variance described by PC1 of the fetal core signature was compared with the distribution to generate a *P* value. Statistical analysis for the z-scored fetal core was calculated with ANOVA function of R (aov), followed by the TukeyHSD function to generate the adjusted *P* values.

## Results

### A transcriptional map of primitive cells from first-trimester FL

The emerging blood system is a complex mixture of cells from different niches and waves acting simultaneously, generating a heterogeneity that necessitates the use of single-cell assays. In addition, immunophenotypic markers classically used to purify progenitors in CB and adult BM may not necessarily be conserved throughout development. To address these issues, we analyzed FL HSPCs using CITE-seq, a high-throughput single-cell RNA-seq method wherein information on immunophenotype is sampled simultaneously with the transcriptome, through oligo-barcode-tagged antibodies, ADTs.<sup>16</sup> FL samples were obtained from medical terminations of pregnancy in the first trimester, ranging in age from CS16 to 9 pcw (supplemental Table 1). HSPCs were selected based on the expression of CD45 and CD34 and for lack of mature lineage markers (LIN) (Figure 1A; supplemental Figure 1A). The analysis focused on CS22, where 6732 cells from 2 embryos were captured for further analysis after quality control (supplemental Table 1). Eleven clusters were identified and assigned to a hematopoietic state based on the expression of lineage-associated genes and transcription factors (Figure 1B-C; supplemental Figure 1B). Accordingly, the UMAP showed an HSC

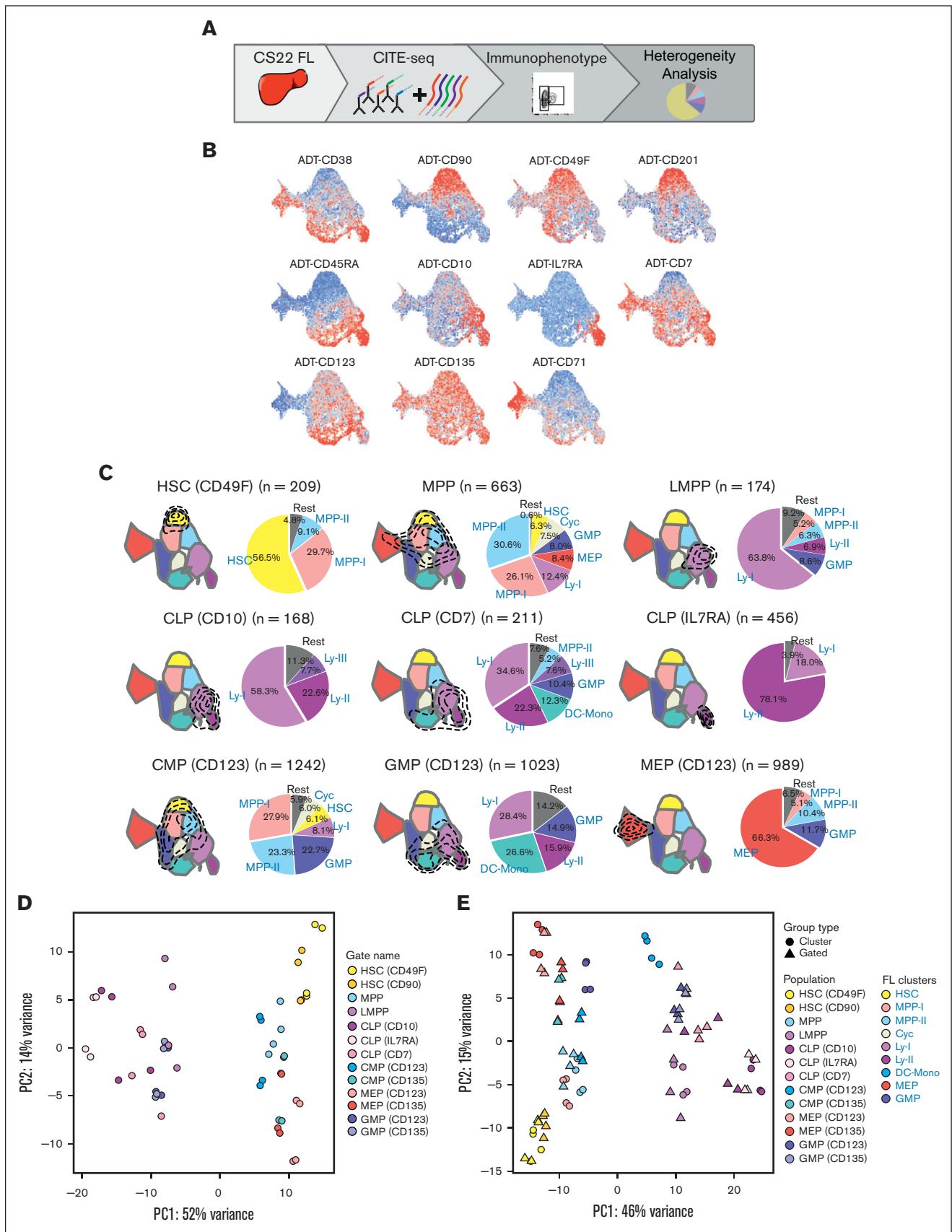


**Figure 1. A transcriptional map of primitive cells from the first trimester FL.** (A) Schematic presentation of the data in the figure. (B) UMAP of cellular states within LIN<sup>+</sup>CD45<sup>+</sup>CD34<sup>+</sup> FL cells at CS22 (2 donors) (C) Bead plot of DEGs between clusters, where the size of the circle is the fraction of cells with expression and the color represents the z-scored mean expression. (D) Cells from the different cell states in (B) were pseudobulked, and PCA was performed using the top 500 variably expressed genes. PC1 vs PC2 (left) and PC2 vs PC3 (right).

cluster at the top, expressing genes like *MLLT3* (protein AF9) and *BST2*.<sup>26</sup> Subsequent to HSCs, cells with a multipotent progenitor (MPP)-like signature were located, followed by a distinct separation into a megakaryocyte-erythroid progenitor axis (MEP), a granulocyte-monocyte progenitor axis (GMP), a dendritic cell-monocyte cluster (DC-Mono), and a lymphoid progenitor axis (Ly-I/Ly-II). The lymphoid axis showed a gradual maturation from Ly-I to Ly-II, and the Ly-II cluster was molecularly more primed toward the B lineage (Figure 1C). The Ly-III was T lineage primed but contained less than 60 cells, as did the DC-I (dendritic cell precursor)

cluster. A cluster located below the MPP clusters had cell cycle-associated genes among the DEGs and was named cycling (Cyc) (Figure 1C).

To further investigate relatedness between the clusters, samples were pseudobulked (2 replicates per donor), and PCA was performed. As expected, the more primitive populations clustered together (HSC, MPP-I, and MPP-II), and the lymphoid and myeloid clusters were clearly separated, with the Ly-II cluster positioned further away from the primitive clusters than Ly-I (Figure 1D).



To validate the hematopoietic state assigned to each cluster, a publicly available data set with immunophenotypically defined fetal HSPCs was projected on the CS22 UMAP.<sup>12</sup> Immunophenotypic HSCs, MPPs, and CLPs mapped to associated clusters as expected. However, the purified CMPs and MEPs mapped to both MEP and GMP clusters, whereas the purified GMPs projected to the Ly-I as well as the DC-Mono clusters (supplemental Figure 1C).

Thus, CITE-seq captured the cellular heterogeneity in early blood development, and different progenitor populations could be defined based on the expression of lineage-associated genes. The data set can be viewed in an interactive format on the following platform: [https://singlecell.broadinstitute.org/single\\_cell/study/SCP2022/fetal-liver-hematopoiesis#study-visualize](https://singlecell.broadinstitute.org/single_cell/study/SCP2022/fetal-liver-hematopoiesis#study-visualize).

## Molecular heterogeneity of immunophenotypically defined progenitor compartments

We next sought to further investigate the heterogeneity observed in some of the immunophenotypically defined progenitors using the ADTs (Figure 2A). First, the transcriptome and the ADT data were integrated. The WNN<sup>20</sup> UMAP generated showed high similarity with the transcriptionally defined cell states (supplemental Figure 2A). Next, the abundance of the individual ADTs was investigated within each cluster (supplemental Figure 2B). The HSC cluster expressed CD90 and CD49F and lacked expression of CD38, all in agreement with markers used in conventional purifications of the HSC population (Figure 2B).<sup>27,28</sup> Furthermore, CD201 (EPCR), a marker shown to define fetal HSCs,<sup>29</sup> was also distinctly expressed in the molecular HSC cluster. The comparatively immature lymphoid cluster (Ly-I) expressed CD10 and CD45RA, but was low in CD38 expression, indicative of lymphomyeloid primed progenitors (LMPPs),<sup>30</sup> whereas the relatively more mature lymphoid cluster (Ly-II), in addition to CD10 and CD45RA, expressed CD38, corresponding to a common lymphoid progenitor (CLP) phenotype.<sup>30</sup> This cluster also had a distinct expression of the interleukin 7 receptor (IL7R), known to mark CLPs in the murine system.<sup>31</sup> The MEP cluster was positive for CD71, as expected (Figure 2B).<sup>14,32</sup> However, CD7, CD123 (IL3R), and CD135 (FLT3), the latter 2 markers traditionally used to define GMPs and common myeloid progenitors (CMPs),<sup>30,33</sup> showed a broader expression pattern (Figure 2B).

Next, the ADT information was used to perform conventional immunophenotypic gating (supplemental Figure 2C; supplemental Table 3). Molecularly defined HSC, MPP, LMPP, and MEP clusters were well captured by the conventional adult immunophenotype (Figure 2C; supplemental Figure 2D).

Conventional CLPs are typically defined based on the expression of CD10<sup>30,34</sup> and using this gating strategy, most of the captured cells were indeed part of a lymphoid transcriptional cluster. However, almost 60% of the cells belonged to the lymphomyeloid Ly-I

cell state, likely because of issues discriminating the cells based on CD38 expression. Therefore, other gating strategies were investigated to identify a transcriptionally purer lymphoid cell state. In CB, CD7, when combined with a lack of CD38 expression, has been shown to enrich for lymphoid progenitors.<sup>35,36</sup> However, in the FL, CD7 surface expression measured with the ADT antibody was virtually absent in the CD38-negative fraction (Figure 2B).

Using CD7 in combination with CD38 and CD45RA, we defined a heterogeneous CD7<sup>+</sup> CLP-like population with cells from both lymphoid and myeloid clusters, in agreement with earlier observations.<sup>35-37</sup> Using IL7R<sup>31</sup> in combination with CD45RA and CD38, almost 80% of cells captured belonged to the Ly-II cell state, and the remaining cells were defined as Ly-I. Thus, IL7R was by far the best surface marker to capture the Ly-II cell state during early development (Figure 2C).

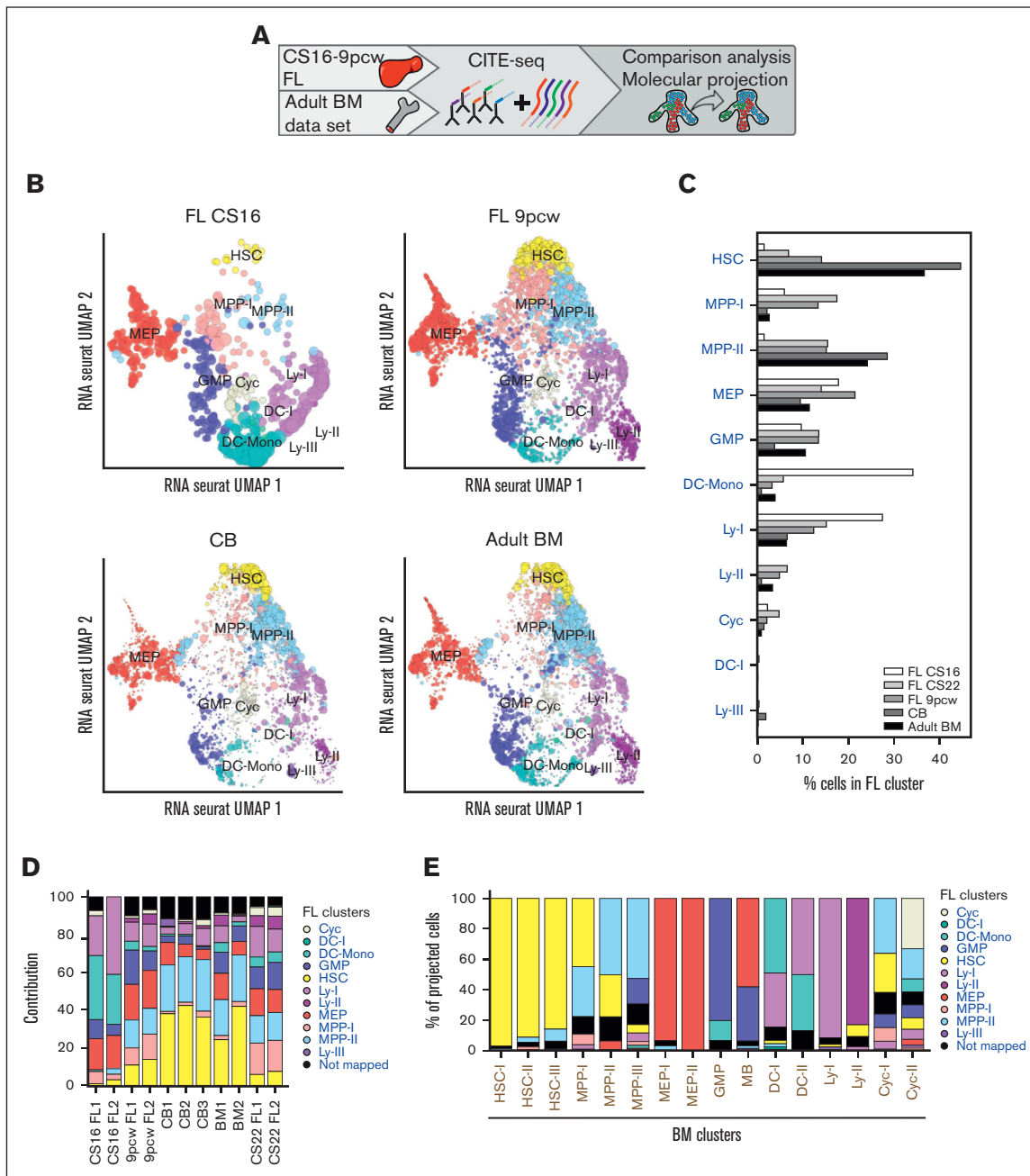
Immunophenotypic CMPs and GMPs showed high heterogeneity as measured by the number of molecular clusters and their relative proportions captured within each immunophenotype (Figure 2C; supplemental Figure 2D). Thus, in immunophenotype-gated GMPs, 16% to 21% of the cells molecularly belonged to the Ly-II lymphoid cell state (Figure 2C; supplemental Figure 2D).<sup>30,33</sup> A PCA of pseudobulked immunophenotypically defined populations showed GMPs located in the vicinity of lymphoid populations, in agreement with the previous analysis (Figure 2D; supplemental Figure 2E). When combining the immunophenotypic populations with the molecularly defined clusters from Figure 1D, heterogeneity was seen for the myeloid GMP and CMP populations, whereas for most other populations, the immunophenotypic population correlated with the corresponding molecular cluster (Figure 2E; supplemental Figure 2F).

Overall, our analysis showed that immunophenotypic gating of HSCs and LMPPs captured populations that were molecularly fairly homogenous, whereas CMPs and GMPs were heterogeneous.

## Projection analysis defines a fetal-specific multipotent cluster with erythromyeloid signature

Next, we set out to directly compare the fetal cells with our adult BM data set of HSPCs, analyzed with CITE-seq in a similar way<sup>19</sup> (Figure 3A; supplemental Figure 3A). Looking at immunophenotypically defined populations using PCA showed that the populations were mainly separated by developmental stage (PC1; 51% of variance) (supplemental Figure 3B). To preserve heterogeneity within the sample, a nearest neighbor projection approach was next used.<sup>17</sup> A reference map was built, and the investigated cells were mapped onto the reference, so only molecular differences within the reference sample were considered. Using FL CS22 HSPCs as a reference, 37% of the cells in adult BM mapped to the fetal HSC cluster, whereas only 7% mapped to this cluster in the FL CS22 sample. In contrast, only 6% of the adult cells mapped to the lymphomyeloid Ly-I cluster, compared

**Figure 2. Molecular heterogeneity of immunophenotypically defined progenitor compartments.** (A) Schematic presentation of the data in the figure. (B) FL CS22 UMAPs of ADT expression of single surface markers; red represents high expression and blue low expression. (C) Transcriptional cell states captured for immunophenotype-gated populations are shown within the contours on the FL CS22 UMAP. Pie charts show percentages of the different molecular clusters within each immunophenotype-gated progenitor. Colors represent the cluster colors, defined in Figure 1B. (D) PCA of pseudobulked ADT-gated cell populations using the top 500 variably expressed genes. (E) Combined PCA of pseudobulked ADT-gated cell populations and clusters (defined in Figure 1B), top 500 variably expressed genes. (For immunophenotypic definition of populations, refer to "Materials and methods").

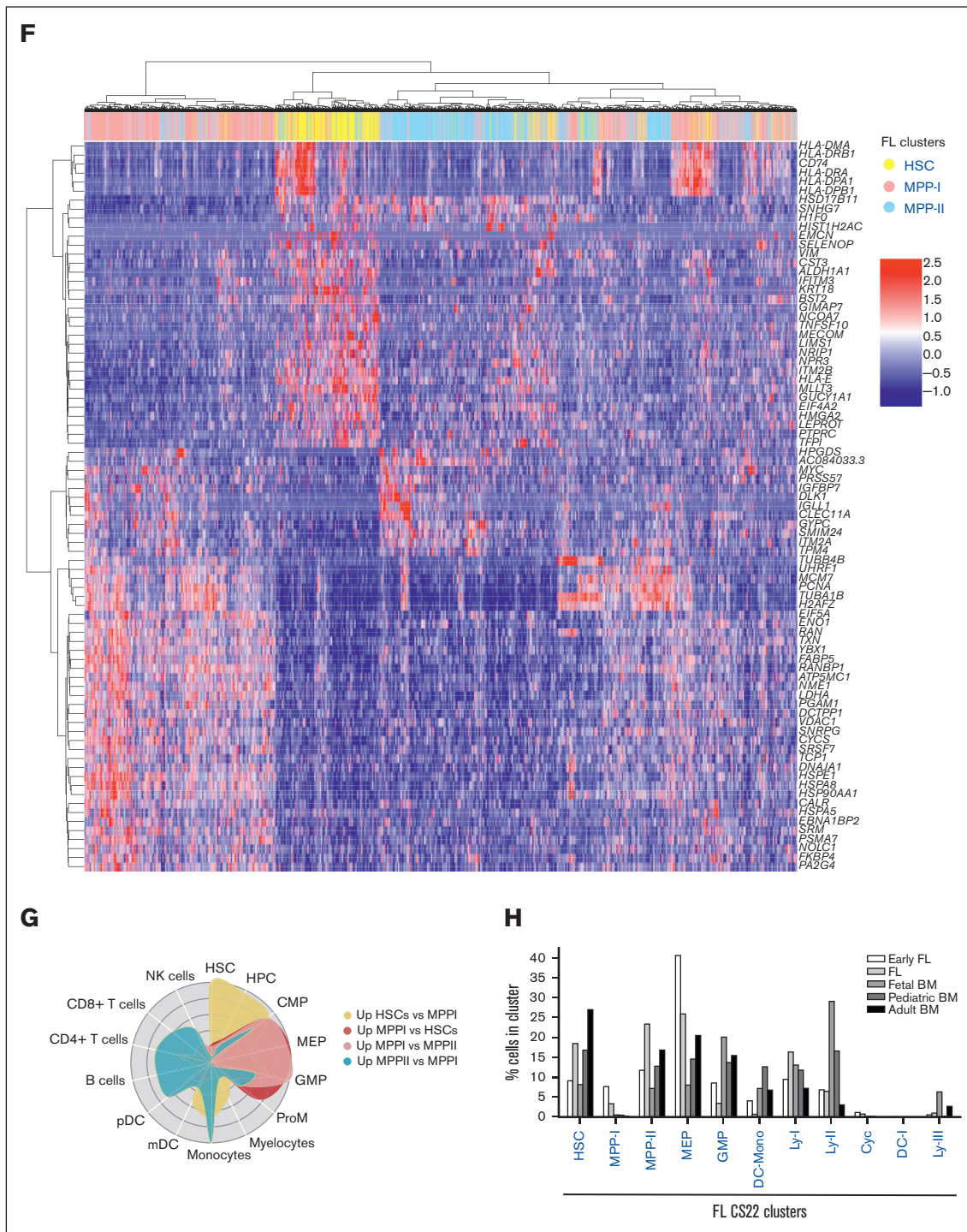


**Figure 3. Projection analysis defines a fetal-specific multipotent cluster with erythromyeloid signature.** (A) Schematic presentation of data in the figure. (B) Projection of FL CS16, FL 9pcw, CB, and adult BM on the FL CS22 UMAP, size of dots represents mapping-score and color of dots represents FL-derived clusters as in Figure 1B. (C) Quantification of projected cells (Figure 3B) in all mapped developmental stages (FL CS16: 143 cells, FL CS22: 6732 cells, FL 9pcw: 1139 cells, CB: 6984 cells, adult BM: 4905 cells). (D) Contribution (percent) of individual donors to the mapping in Figure 3B-C. (E) Classification of adult BM-derived clusters (x-axis) on FL CS22 (also refer to supplemental Figure 3A). (F) Hierarchical clustering of DEGs in primitive CS22 FL cell populations (HSC, MPP-I, and MPP-II). (G) CellRadar plots of DEGs between HSC, MPP-I, and MPP-II. (H) Quantification of classified cells in all mapped developmental stages from Roy et al.<sup>6</sup> In total, 90% to 95% of the cell from the different stages mapped. (Early FL; first trimester, FL, and fetal BM; second trimester).

with 15% in CS22 (Figure 3B-E). Furthermore, the MPP-I cluster, which represented about 17% of the CS22 FL cells, was hardly detectable in the BM sample (3% of total cells). To narrow down the time window in development when these changes in cell states occurred, CB-derived cells were analyzed, generating

similar results as those from adult BM (Figure 3B-D).<sup>19</sup> Investigating 2 more fetal time points, CS16 and 9 pcw (supplemental Table 1), 6% and 13%, respectively, of the fetal cells mapped to MPP-I, identifying MPP-I as a fetal-specific cell state (Figure 3B-D).





**Figure 3 (continued)**

Intriguingly, when using adult BM as a reference, almost no FL cells mapped to the adult HSC clusters and few to the more primitive MPP clusters, whereas the adult MPP-III was enriched in the CS22 and 9 pcw developmental stages. Furthermore, adult MEP-I was enriched in all stages of the FL (supplemental Figure 3C-E). The fetal-associated MPP-I mainly mapped to MPP-III on the adult BM map (supplemental Figure 3F). Thus, transcriptionally defined adult HSCs were virtually absent in first-trimester FLs.

We next asked if the differences in cellular states observed could be detected with an integration approach. Integration of fetal CS22 and adult BM samples with Harmony<sup>21</sup> showed that transcriptionally defined HSCs (Harmony cluster 0) clearly increased with developmental stage and a lymphoid cluster (Harmony cluster 8) decreased with age. Harmony clusters 1 and 2 were dominated by fetal cells and contained a large fraction of the fetal MPP-I cluster (supplemental Figure 4A-D).

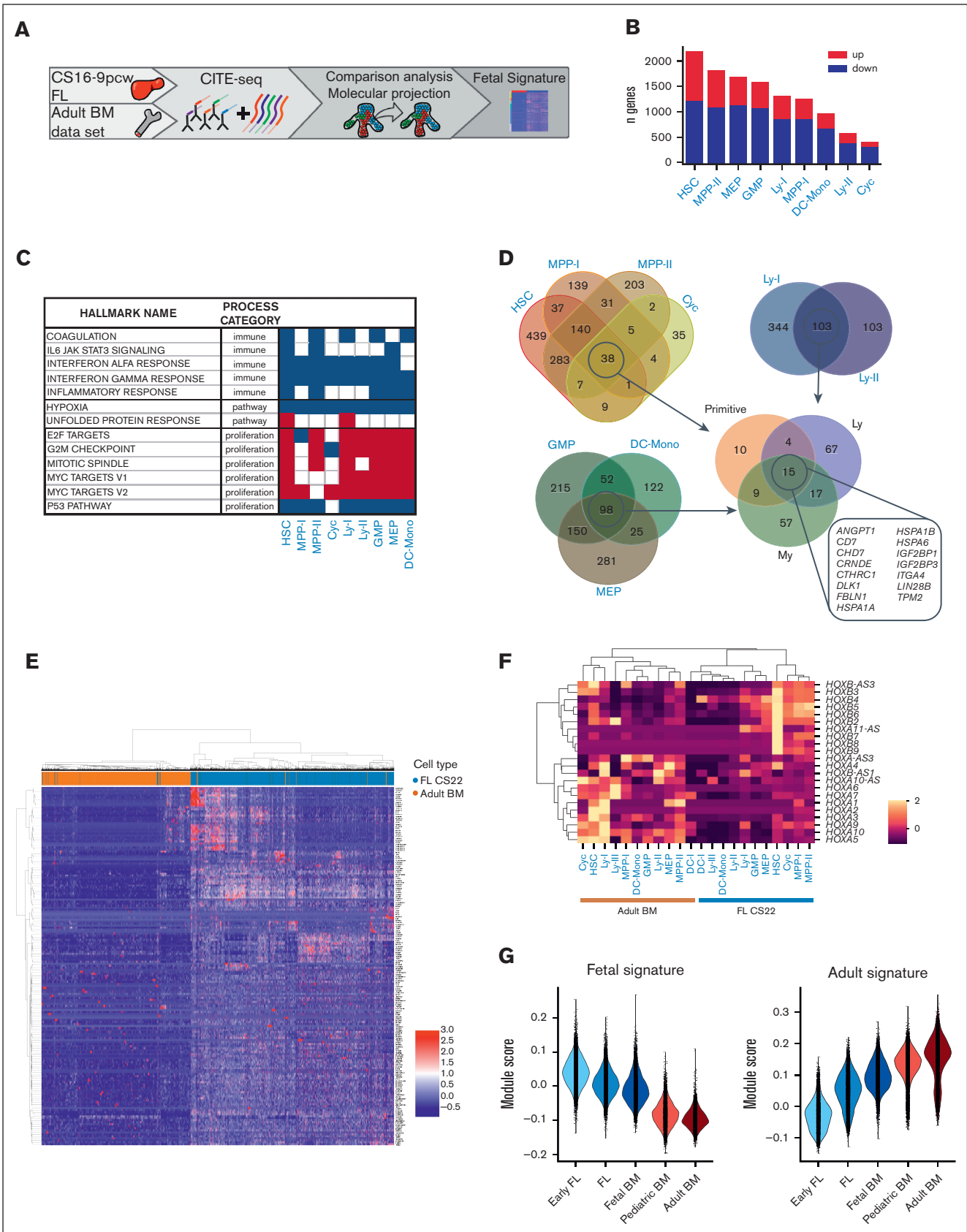


Figure 4.

To better characterize the fetal-specific MPP-I, the cells were compared with the HSC and MPP-II clusters. Among DEGs, an upregulation of *MYC* and heat shock–related genes was detected, and an erythromyeloid lineage affiliation was observed with CellRadar (Figure 3F–G). Looking at cell cycle scores, the MPP-I had an increased number of cycling cells compared with the HSC but less than the MEP cluster, thus matching a trajectory from HSC to MPP-I to MEP. Of note, in this analysis, the cluster named Cyc was mainly enriched for G2/M (supplemental Figure 4E). GSEA on pseudobulked clusters<sup>24,25</sup> showed upregulation of *MYC* targets, *MTORC1* signaling, and unfolded protein response and oxidative phosphorylation gene sets in MPP-I compared with the HSC and MPP-II clusters, indicative of a more active metabolic state (supplemental Figure 4F).

To confirm the fetal specificity of the MPP-I cell state, a published data set that included samples from different developmental stages was used.<sup>6</sup> Mapping these data onto our CS22 data showed enrichment of MPP-I in the early FL sample, whereas the fetal, pediatric, and adult BM were almost completely depleted of the MPP-I cell state (Figure 3H).

Taken together, these data show that the MPP-I cluster harbors transient, fetal-specific, erythromyeloid–primed multipotent progenitors.

### Cluster-specific differential gene expression analysis defines a fetal-specific gene signature

To further understand how fetal and adult cell types differ, the projected adult cells were compared with their fetal counterparts using DESeq2 (Figure 4A). The clusters were pseudobulked for the analysis, and as with the immunophenotypic samples, PC1 separated the clusters based on developmental stage (supplemental Figure 5A–B). DEGs were analyzed for each cluster (Figure 4B) and GSEA<sup>24,25</sup> identified a general enrichment of pathways involved in proliferation in fetal clusters, whereas gene sets involved in inflammation and immune system processes were downregulated, in agreement with an earlier study.<sup>6</sup> The fetal-specific MPP-I cells had only about 130 adult counterparts, making the analysis less robust, but showed upregulation of *MYC* targets in the fetal MPP-I, as was also detected in the comparison with the fetal HSC and MPP-II clusters (Figure 4C; supplemental Figure 4F).

To identify common upregulated (fetal signature) and downregulated (adult signature) genes, Venn diagrams were used. A gene set of 179 upregulated and 450 downregulated genes was identified in the primitive, lymphoid (Ly), and myeloid (My) groups (Figure 4D; supplemental Figure 5C; supplemental Tables 4–5). The gene signature retrieved was validated on a data set with fetal, induced (i) PSC-derived and adult BM HSPCs,<sup>15</sup> and could separate samples by developmental stage, showing the biological significance of the derived fetal core signature (supplemental

Figure 5D). Performing the analysis on pseudobulked samples without the pseudoreplicates generated fewer but highly overlapping genes with the fetal core signature. In contrast, no clear separation of the samples by developmental stage was seen using DEGs identified with a single-cell method (Seurat) (supplemental Figure 5E–H).

Among the genes upregulated in the fetal core signature were *LIN28B*, *IGF2BP1*, and *IGF2BP3*, all part of the *LIN28B*-let7 axis and known to be involved in self-renewal of fetal HSCs in the murine system<sup>38</sup>; Delta-like gene 1 (*DLK1*), a gene shown to be a negative regulator of HSC formation in the mouse embryo,<sup>39</sup> and *CHD7*, an epigenetic remodeler known to interact with *RUNX1* and inhibit differentiation<sup>40</sup> (Figure 4D–E). Furthermore, *CD7* was more generally expressed among fetal progenitors, as was also seen at the protein level with the ADT marker (Figure 2B). Among the downregulated genes, the HLA complex of both class I and class II dominated. We also noticed that homeobox (*HOX*) gene clusters were differentially expressed in fetal and adult HSPCs, where *HOXB* genes were in general more expressed in fetal and *HOXA* were more expressed in adult primitive clusters (Figure 4F).

Applying our core to another data set with samples from different developmental stages,<sup>6</sup> the fetal signature was seen to gradually decrease throughout development, whereas the adult signature increased (Figure 4G). Thus, by directly comparing fetal and adult counterparts from our projection analysis, a fetal core signature with genes specifically upregulated or downregulated in fetal compared with adult counterparts was identified.

### The fetal core signature differentiates between pediatric and adult ALL

By investigating neonatal blood spots and through twin studies, translocations giving rise to ALL in children have been backtracked to birth and an in utero origin.<sup>2</sup> One question is whether a remnant of the fetal signature can be detected in the pediatric leukemic cells. Looking at the upregulated universal fetal core genes in a publicly available RNA-seq data set of a preleukemic iPSC model of *ETV6::RUNX1* (*TEL::AML1*),<sup>15</sup> fetal genes were to a large extent expressed in the iPSC-derived normal (known to recapitulate fetal development) and *ETV6::RUNX1*-expressing hematopoietic progenitors (Figure 5A).

Next, using a publicly available RNA-seq data set with almost 2000 samples from patients with B-ALL, the fetal core signature was investigated in B-ALLs of known in utero origin.<sup>18</sup> B-ALLs with different driver mutations of interest were selected, pooled, and batch-corrected (*KMT2A::AFF1*, *KMT2A::MLLT1*, *KMT2A::MLLT3*, *BCR::ABL1*, *ETV6::RUNX1*, and high or low hyperdiploidy, in total, 645 B-ALL samples; supplemental Figure 6A). The selected samples were pooled, and a PCA plot was generated using the fetal core signature. Intriguingly, the PC1 seemed to separate samples

**Figure 4. Cluster-specific differential gene expression analysis defines a fetal-specific gene signature.** (A) Schematic presentation of data in the figure. (B) Number of upregulated and downregulated genes in pseudobulked FL CS22 clusters compared with adult BM counterparts. Red bars; upregulated genes (fetal signature), blue bars; downregulated genes (adult signature), adj *P* value < .05. (C) GSEA of selected hallmark gene sets involved in immune processes, pathways, and proliferation; red; upregulated, blue; downregulated in FL according to NES value. False discovery rate *q*-value < 0.05. (D) Venn diagrams of upregulated genes in FL compared with adult BM defined by log2 fold change > 1 and adj *P* value < .05. (E) Heatmap showing expression of the fetal signature (primitive, My, Ly in Figure 4D) genes in FL CS22 (blue) and adult BM (brown), with cells displayed on the x-axis and genes on the y-axis. (F) Heatmap of mean *HOX* gene expression per cluster in FL CS22 and adult BM. (G) Module scores of the fetal (left) and adult (right) gene signature across development using data from an earlier study.<sup>6</sup>

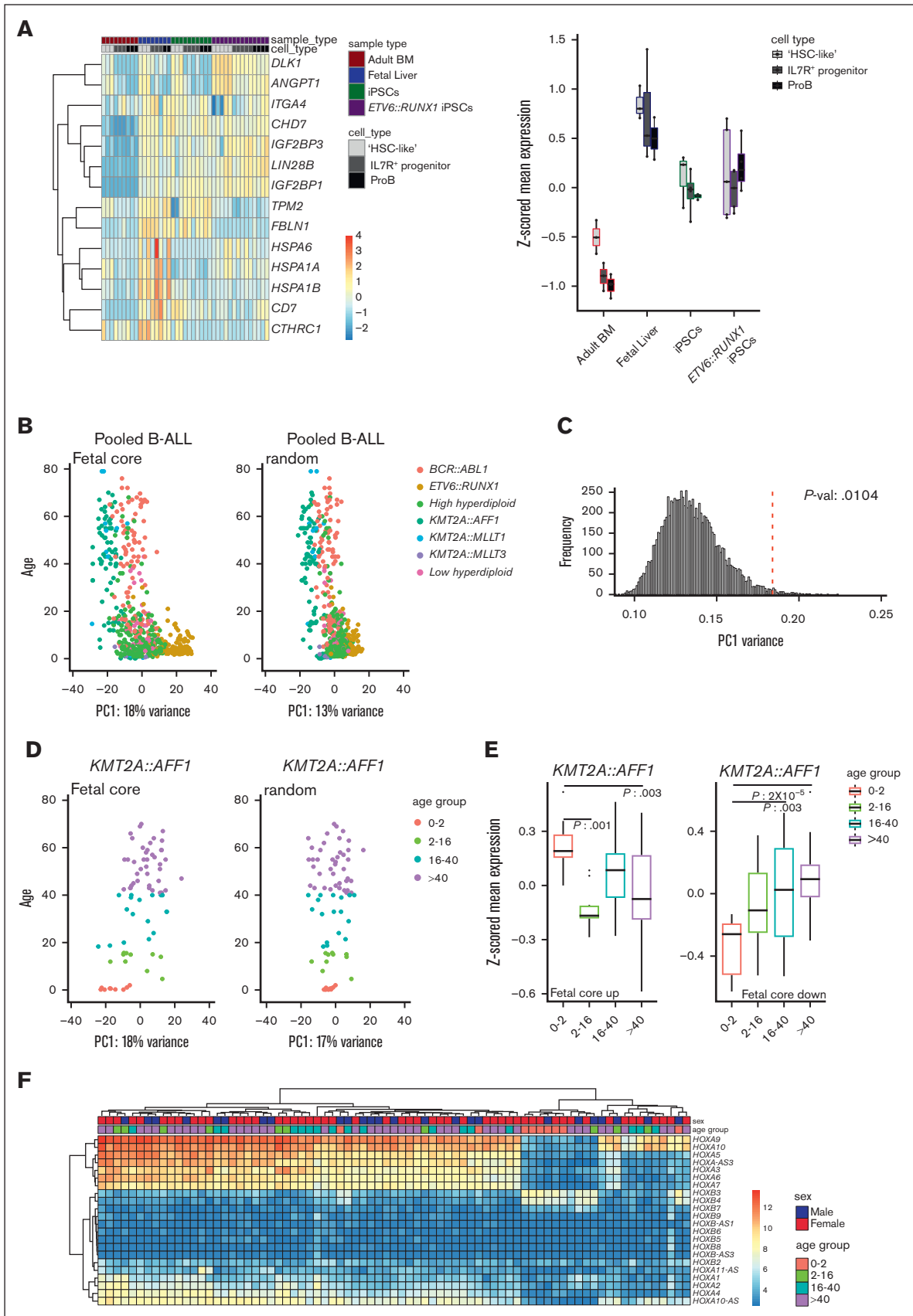


Figure 5.

based on age, thus indicating that the fetal core may discriminate between pediatric and adult B-ALL (Figure 5B-C). Investigating the subtypes of B-ALL individually, an association with *ETV6::RUNX1* and high hyperdiploid ALL, 2 very frequent cytogenetic abnormalities in children, was seen ( $P$  value:  $1 \times 10^{-4}$  and  $3 \times 10^{-4}$ , respectively),<sup>2</sup> whereas PC1 variance in low hyperdiploid, *BCR::ABL1*, and *KMT2A::AFF1* did not significantly differ from random variance (Figure 5D; supplemental Figure 6B-C and data not shown).

To investigate if a fetal core program was retained specifically in pediatric B-ALL, a link between our fetal core and age was next examined in leukemia subtypes where both pediatric and adult patients were represented. *KMT2A::AFF1* (*MLL::AF4*) almost exclusively gives rise to B-ALL and is found in many different age groups, including infants, for whom a clear in utero origin has been demonstrated.<sup>41</sup> Furthermore, in infants, few additional mutations are needed in contrast to older children.<sup>42</sup> The fetal core program was specifically maintained in *KMT2A::AFF1* infants compared with older patients. In the other 2 types of leukemia investigated, high hyperdiploid and *BCR::ABL1*, the latter a subtype seen in about 25% of adult B-ALLs but only 3% to 4% of pediatric cases,<sup>43</sup> an association with the downregulated gene set was seen (Figure 5E; supplemental Figure 6D).

To look further into the difference between infant, pediatric, and adult *KMT2A::AFF1* patients, a heatmap of DEGs was made (supplemental Figure 6E). *HOXA3*, *HOXA7*, and *HOXA10* were less expressed in some infants. Looking specifically at the *HOX* genes expressed in our fetal and adult data sets (Figure 4E), one cluster with mostly infants had lower expression of the different *HOXA* genes investigated but higher expression of *HOXB3* and *HOXB4* (Figure 5F). Indeed, a correlation was observed between *HOXA* expression, prognosis, and age in an earlier study.<sup>44</sup>

Thus, in some subtypes of pediatric leukemia associated with an in utero origin, part of a fetal core signature seems to remain in the leukemic cells. The data have the potential to explain differences in pediatric and adult leukemia carrying similar driver mutations, which is potentially important for prognosis and therapy response.

## Discussion

The hematopoietic system is a paradigmatic hierarchical organization of tissues, and careful immunophenotypic isolation and characterization of progenitors have formed the foundation of the hematopoietic hierarchy of today in both mice and human.<sup>45,46</sup> Recent developments in single-cell omics have allowed for further purification and analysis of heterogeneity, and with these findings, the hematopoietic hierarchy can be regarded as a continuum rather than a step-wise hierarchical organization.<sup>47,48</sup> Until recently, these studies have mainly investigated adult hematopoiesis, but more reports looking into human blood development are

rapidly emerging.<sup>6,11-13,49,50</sup> Most of these studies focused on molecular heterogeneity, but recently CITE-seq was used to link the immunophenotype of second-trimester HSCs to engraftment potential.<sup>50</sup> Within this study, we interrogated FL HSPCs using single-cell RNA-seq, relying on CITE-seq to also capture the immunophenotype along with the transcriptome. By focusing on early primitive hematopoiesis, the whole HSPC population could be interrogated and the relative proportions of all CD34-positive populations compared across developmental time points, thereby also capturing the immunophenotype enabling a unique view of the classically defined progenitors. Although most of the conventional markers of HSPCs (CD90, CD45RA, CD71, and IL7R) showed good correlation with the transcriptionally defined clusters, markers of GMPs and CMPs, that is, CD123 and CD135 (FLT3), were more generally expressed. This was even more apparent when using the ADTs to perform immunophenotypic gating, which showed substantial molecular heterogeneity within GMPs and CMPs. This heterogeneity in defining myeloid progenitors was not specific to fetal development but was also seen in adults.<sup>19</sup> Looking at lymphoid surface markers, CD10 mainly captured the lymphomyeloid cluster Ly-I, whereas our data revealed that IL7R in combination with CD38 and CD45RA could capture the Ly-II cell state with high purity in the embryo. This was the opposite of what was seen in adults, where the population captured by CD10 comprised 70% Ly-II cells vs 43% when using IL7R (data not shown and a previous study<sup>19</sup>). Thus, both fetal and adult myeloid immunophenotypically defined progenitors captured molecularly heterogeneous populations, whereas HSCs, LMPPs, and MEPs defined fairly pure molecular populations.

An advantage with high-throughput single-cell RNA-seq methods is that there is no need to preselect populations based on immunophenotype. However, difficulties remain when cells from different stages and niches are compared. Earlier studies have merged different stages onto the same analysis, relying on batch correction to enable sample integration.<sup>6,11,49</sup> However, when merging cell populations from different stages into the same analysis, the developmental differences may take over and heterogeneity within the sample itself may be lost. Another problem with integrating data is that it can be difficult to determine if integration is incomplete or if there is a real difference between the samples. Our integrated UMAP is a good example of this, where without the projection approach, the minimal contribution of adult cells in certain clusters could easily be misinterpreted as incomplete integration (supplemental Figure 4). The projection approach used herein allows for signals responsible for lineage determination to be kept.<sup>17</sup> By mapping the query cells onto the reference map, only molecular differences that separate the reference cells will be investigated. From this analysis, cell state differences over development could be captured, and an increase in molecular HSCs was observed with age, whereas the relative fraction of lymphoid

**Figure 5. The fetal core signature differentiates between pediatric and adult ALL.** (A) Heatmap (left) and box plots (z-scored mean expression, right) displaying expression of fetal universally upregulated genes in an iPSC model expressing *ETV6::RUNX1*.<sup>15</sup> Top rows show sample and cell types ("HSC-like," IL7R<sup>+</sup> progenitor, and proB). (B) PCA of pooled samples from patients with B-ALL using the fetal core signature (left) and a representative plot of random genes (right). PC1 vs age. Color coded according to translocation status. (C) Histogram showing variance of PC1 of the pooled samples with random selected genes (10 000 iterations). The fetal core PC1 variance is indicated by a dotted red line. (D) PCA of *KMT2A::AFF1* (*MLL::AF4*) using the fetal core signature (left) and representative plot of random genes (right). PC1 vs age (color coded per age group). (E) Boxplots showing z-scored mean expression of the fetal core signature in *KMT2A::AFF1* per age group. Upregulated and downregulated gene sets are shown separately, adjusted  $P$  values < 0.05 indicated in the figure. The box includes first to third quartile and the line indicates the median. (F) Heatmap of *HOX* gene expression in *KMT2A::AFF1* B-ALL. Top rows show sex and age group, respectively.

progenitors (Ly-I/Ly-II) decreased, in agreement with an earlier study.<sup>6</sup> In our approach, 2 different reference maps were used: FL CS22 and adult BM. The projection analysis of FL CS22 on the adult BM reference revealed substantial molecular differences between HSCs in FL and adult BM. The molecular FL HSCs mainly projected to adult BM MPP-I and MPP-III, but not to the adult HSC molecular clusters. The discrepancy between fetal and adult HSC states is important to understand as it may facilitate the development of functional HSCs from PSCs for future regenerative medicine.<sup>1</sup>

Furthermore, the MPP-I cluster was found to be almost exclusively a fetal cell state, suggesting that this population could constitute a fetal-specific progenitor population arising from an HSC-independent wave. The MPP-I cells displayed an erythromyeloid gene signature while maintaining a primitive gene program. No suitable statistical method was found to verify the differences in fetal and adult cell states, as the cells were from 2 donors. That said, using another data set,<sup>6</sup> our MPP-I was shown to be virtually absent from both pediatric and adult samples, as well as fetal BM, but not from donor-matched FL. For future characterization, studies such as ours should be performed on enriched yolk sac and early FL progenitors. A combination of immunophenotypic markers presented in our study (supplemental Figure 2B) may allow for prospective isolation and further characterization of the progenitor population and reveal its origin. Such studies may also provide markers and molecular signatures that can separate early from definitive hematopoietic waves.

By using the projection analysis, fetal and adult counterparts could be transcriptionally compared, and a core set of genes linked to fetal or adult identity could be identified. Multiple methods exist to perform DEG analysis on single-cell data sets, and in general, pseudobulking approaches are found to be more reliable.<sup>51,52</sup> Pseudoreplicates were made to account for intradonor variability as well as to increase statistical power in the analysis as donors were limited. The derived core was validated and compared by other methods and shown to be highly biologically relevant, and our data highlights substantial ontogeny-dependent transcriptional differences.

Childhood B-ALL has in many cases been shown to have a fetal origin,<sup>2</sup> and the leukemic cells could potentially retain expression of fetal-specific genes or lack expression of adult-specific genes.<sup>53</sup> Thus, our fetal core signature was used to investigate childhood leukemia from publicly available data of B-ALL samples with different cytogenetic abnormalities.<sup>18</sup> The fetal core signature correlated with the age of the patient in some leukemia subtypes, indicating that a fetal gene expression program may remain in these patients. For *KMT2A::AFF1*, it was clear that *HOXA* genes were downregulated in some infants compared with those in adults, in agreement with an earlier study where a correlation was observed between lower expression of *HOXA* genes, poor prognosis, and

young age.<sup>44</sup> Of note, in our fetal-adult data set *HOXA* genes were less expressed in fetal HSCs compared with adult (Figure 4F).

Our study links immunophenotypes with transcriptomes at the single-cell level, providing a unique map of human fetal blood development. Comparison to adult hematopoiesis gives insight into how the hematopoietic progenitor compartment changes with development. A fetal core signature with universally upregulated or downregulated genes depending on developmental stage could be identified, a gene set that could separate specific subtypes of B-ALL samples based on age. The importance of the fetal core genes in the formation of the preleukemic clone in utero and in leukemia progression and development in general remains to be explored.

## Acknowledgements

The authors acknowledge the personnel of the Lund Stem Cell Center FACS core facility, Anna Fossum and Zhi Ma, and Center for Translational Genomics (CTG) at Lund University for their assistance with the CITE-seq experiments, especially Linda Geirsonson Ulfsson and Eva Erlandsson, and Ariana Calderón for help with the graphical abstract. The authors acknowledge support from the StemTherapy Program, the Crafoord Foundation (C.B.), the Swedish Childhood Cancer Foundation (C.B. and G.K.), the Ragnar Söderberg Foundation (C.B. and G.K.), the Knut and Alice Wallenberg Foundation (G.K.), the Swedish Research Council (C.B. and G.K.), and the Swedish Cancer Society (G.K.).

## Authorship

Contribution: M.N.E.S., G.K., and C.B. conceptualized the study; M.N.E.S. performed the methodology and conducted the formal analysis; M.N.E.S., P.D., and R.O. curated the data; M.N.E.S., S.P., and R.O. conducted the investigation; M.N.E.S. and C.B. visualized the data and wrote the original draft of the manuscript; M.N.E.S., S.P., G.K., and C.B. reviewed and edited the manuscript; C.B. and G.K. acquired funds; and S.S., G.K., and C.B. supervised the study.

Conflict-of-interest disclosure: P.D. and G.K. are board members and have equity in Nygen Analytics AB, all unrelated to this work. The remaining authors declare no competing financial interests.

ORCID profiles: M.N.E.S., [0000-0002-2581-5543](https://orcid.org/0000-0002-2581-5543); R.O., [0000-0001-8079-8718](https://orcid.org/0000-0001-8079-8718); P.D., [0000-0002-8070-7238](https://orcid.org/0000-0002-8070-7238); S.S., [0000-0002-8007-2260](https://orcid.org/0000-0002-8007-2260); C.B., [0000-0002-4876-1218](https://orcid.org/0000-0002-4876-1218).

Correspondence: Göran Karlsson, Division of Molecular Hematology, Lund Stem Cell Center, Lund University, BMC B12, 221 84 Lund, Sweden; email: [goran.karlsson@med.lu.se](mailto:goran.karlsson@med.lu.se); and Charlotta Böiers, Division of Molecular Hematology, Lund Stem Cell Center, Lund University, BMC B12, 221 84 Lund, Sweden; email: [charlotta.boiers@med.lu.se](mailto:charlotta.boiers@med.lu.se).

## References

1. Wahlster L, Daley GQ. Progress towards generation of human haematopoietic stem cells. *Nat Cell Biol.* 2016;18(11):1111-1117.
2. Greaves M. A causal mechanism for childhood acute lymphoblastic leukaemia. *Nat Rev Cancer.* 2018;18(8):471-484.

3. Kim I, Saunders TL, Morrison SJ. Sox17 dependence distinguishes the transcriptional regulation of fetal from adult hematopoietic stem cells. *Cell*. 2007;130(3):470-483.
4. Yuan J, Nguyen CK, Liu X, Kanellopoulou C, Muljo SA. Lin28b reprograms adult bone marrow hematopoietic progenitors to mediate fetal-like lymphopoiesis. *Science*. 2012;335(6073):1195-1200.
5. Ivanovs A, Rytsov S, Ng ES, Stanley EG, Elefanty AG, Medvinsky A. Human haematopoietic stem cell development: from the embryo to the dish. *Development*. 2017;144(13):2323-2337.
6. Roy A, Wang G, Iskander D, et al. Transitions in lineage specification and gene regulatory networks in hematopoietic stem/progenitor cells over human development. *Cell Rep*. 2021;36(11):109698.
7. Ghosn E, Yoshimoto M, Nakauchi H, Weissman IL, Herzenberg LA. Hematopoietic stem cell-independent hematopoiesis and the origins of innate-like B lymphocytes. *Development*. 2019;146(15):dev170571.
8. Böiers C, Carrelha J, Lutteropp M, et al. Lymphomyeloid contribution of an immune-restricted progenitor emerging prior to definitive hematopoietic stem cells. *Cell Stem Cell*. 2013;13(5):535-548.
9. Palis J. Hematopoietic stem cell-independent hematopoiesis: emergence of erythroid, megakaryocyte, and myeloid potential in the mammalian embryo. *FEBS Lett*. 2016;590(22):3965-3974.
10. Elsaid R, Soares-da-Silva F, Peixoto M, et al. Hematopoiesis: a layered organization across chordate species. *Front Cell Dev Biol*. 2020;8:606642.
11. Popescu DM, Botting RA, Stephenson E, et al. Decoding human fetal liver haematopoiesis. *Nature*. 2019;574(7778):365-371.
12. Ranzoni AM, Tangherloni A, Berest I, et al. Integrative single-cell RNA-seq and ATAC-seq analysis of human developmental hematopoiesis. *Cell Stem Cell*. 2021;28(3):472-487.e7.
13. Calvanese V, Capellera-Garcia S, Ma F, et al. Mapping human haematopoietic stem cells from haemogenic endothelium to birth. *Nature*. 2022;604(7906):534-540.
14. Notta F, Zandi S, Takayama N, et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science*. 2016;351(6269):aab2116.
15. Boiers C, Richardson SE, Laycock E, et al. A human IPS model implicates embryonic B-myeloid fate restriction as developmental susceptibility to B acute lymphoblastic leukemia-associated ETV6-RUNX1. *Dev Cell*. 2018;44(3):362-377.e7.
16. Stoeckius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods*. 2017;14(9):865-868.
17. Dhapola P, Rodhe J, Olofzon R, et al. Scarf enables a highly memory-efficient analysis of large-scale single-cell genomics data. *Nat Commun*. 2022;13(1):4616.
18. Gu Z, Churchman ML, Roberts KG, et al. PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nat Genet*. 2019;51(2):296-307.
19. Sommarin MNE, Dhapola P, Safi F, et al. Single-cell multiomics reveals distinct cell states at the top of the human hematopoietic hierarchy. *bioRxiv*. 2021:2021.2004.2001.437998.
20. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184(13):3573-3587.e29.
21. Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with harmony. *Nat Methods*. 2019;16(12):1289-1296.
22. Bagger FO, Sasivarevic D, Sohi SH, et al. BloodSpot: a database of gene expression profiles and transcriptional programs for healthy and malignant haematopoiesis. *Nucleic Acids Res*. 2016;44(D1):D917-924.
23. 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.
24. 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 U S A*. 2005;102(43):15545-15550.
25. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417-425.
26. Bujanover N, Goldstein O, Greenspan Y, et al. Identification of immune-activated hematopoietic stem cells. *Leukemia*. 2018;32(9):2016-2020.
27. Notta F, Doulatov S, Laurenti E, Poepl A, Jurisica I, Dick JE. Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science*. 2011;333(6039):218-221.
28. Majeti R, Park CY, Weissman IL. Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. *Cell Stem Cell*. 2007;1(6):635-645.
29. Subramaniam A, Talkhonchek MS, Magnusson M, Larsson J. Endothelial protein C receptor (EPCR) expression marks human fetal liver hematopoietic stem cells. *Haematologica*. 2019;104(2):e47-e50.
30. Doulatov S, Notta F, Eppert K, Nguyen LT, Ohashi PS, Dick JE. Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat Immunol*. 2010;11(7):585-593.
31. Kondo M, Weissman IL, Akashi K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell*. 1997;91(5):661-672.
32. Mori Y, Chen JY, Pluvineau JV, Seita J, Weissman IL. Prospective isolation of human erythroid lineage-committed progenitors. *Proc Natl Acad Sci U S A*. 2015;112(31):9638-9643.

33. Manz MG, Miyamoto T, Akashi K, Weissman IL. Prospective isolation of human clonogenic common myeloid progenitors. *Proc Natl Acad Sci U S A*. 2002;99(18):11872-11877.
34. Galy A, Travis M, Cen D, Chen B. Human T, B, natural killer, and dendritic cells arise from a common bone marrow progenitor cell subset. *Immunity*. 1995;3(4):459-473.
35. Hoebeke I, De Smedt M, Stolz F, et al. T-B- and NK-lymphoid, but not myeloid cells arise from human CD34(+)CD38(-)CD7(+) common lymphoid progenitors expressing lymphoid-specific genes. *Leukemia*. 2007;21(2):311-319.
36. Hao QL, Zhu J, Price MA, Payne KJ, Barsky LW, Crooks GM. Identification of a novel, human multilymphoid progenitor in cord blood. *Blood*. 2001; 97(12):3683-3690.
37. Haddad R, Guardiola P, Izac B, et al. Molecular characterization of early human T/NK and B-lymphoid progenitor cells in umbilical cord blood. *Blood*. 2004;104(13):3918-3926.
38. Copley MR, Babovic S, Benz C, et al. The Lin28b-let-7-Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. *Nat Cell Biol*. 2013;15(8):916-925.
39. Mirshekar-Syahkal B, Haak E, Kimber GM, et al. Dlk1 is a negative regulator of emerging hematopoietic stem and progenitor cells. *Haematologica*. 2013;98(2):163-171.
40. Hsu J, Huang HT, Lee CT, et al. CHD7 and Runx1 interaction provides a braking mechanism for hematopoietic differentiation. *Proc Natl Acad Sci U S A*. 2020;117(38):23626-23635.
41. Gale KB, Ford AM, Repp R, et al. Backtracking leukemia to birth: identification of clonotypic gene fusion sequences in neonatal blood spots. *Proc Natl Acad Sci U S A*. 1997;94(25):13950-13954.
42. Andersson AK, Ma J, Wang J, et al. The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat Genet*. 2015; 47(4):330-337.
43. Bernt KM, Hunger SP. Current concepts in pediatric Philadelphia chromosome-positive acute lymphoblastic leukemia. *Front Oncol*. 2014;4:54.
44. Starkova J, Zamostna B, Mejstrikova E, Krejci R, Drabkin HA, Trka J. HOX gene expression in phenotypic and genotypic subgroups and low HOXA gene expression as an adverse prognostic factor in pediatric ALL. *Pediatr Blood Cancer*. 2010;55(6):1072-1082.
45. Jacobsen SEW, Nerlov C. Haematopoiesis in the era of advanced single-cell technologies. *Nat Cell Biol*. 2019;21(1):2-8.
46. Seita J, Weissman IL. Hematopoietic stem cell: self-renewal versus differentiation. *Wiley Interdiscip Rev Syst Biol Med*. 2010;2(6):640-653.
47. Velten L, Haas SF, Raffel S, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol*. 2017;19(4):271-281.
48. Laurenti E, Göttgens B. From haematopoietic stem cells to complex differentiation landscapes. *Nature*. 2018;553(7689):418-426.
49. Jardine L, Webb S, Goh I, et al. Blood and immune development in human fetal bone marrow and Down syndrome. *Nature*. 2021;598(7880):327-331.
50. Vanuytsel K, Villacorta-Martin C, Lindstrom-Vautrin J, et al. Multi-modal profiling of human fetal liver hematopoietic stem cells reveals the molecular signature of engraftment. *Nat Commun*. 2022;13(1):1103.
51. Squair JW, Gautier M, Kathe C, et al. Confronting false discoveries in single-cell differential expression. *Nat Commun*. 2021;12(1):5692.
52. Junttila S, Smolander J, Elo LL. Benchmarking methods for detecting differential states between conditions from multi-subject single-cell RNA-seq data. *Brief Bioinform*. 2022;23(5):bbac286.
53. Symeonidou V, Jakobczyk H, Bashanfer S, et al. Defining the fetal origin of MLL-AF4 infant leukemia highlights specific fatty acid requirements. *Cell Rep*. 2021;37(4):109900.