Single-cell proteo-transcriptomic profiling reveals altered characteristics of stem and progenitor cells in patients receiving cytoreductive hydroxyurea in early-phase chronic myeloid leukemia

Hana Komic,[1,2*] Malin S. Nilsson,[1,3*] Lovisa Wennström,[4] Tagore Sanketh Bandaru,[2] Pekka Jaako,[5] Kristoffer Hellstrand,[1,6] Fredrik B. Thorén[1,2#] and Anna Martner[1,3#]

[1]TIMM Laboratory at Sahlgrenska Center for Cancer Research, University of Gothenburg; [2]Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg; [3]Department of Microbiology and Immunology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg; [4]Department of Hematology, Sahlgrenska University Hospital; [5]Sahlgrenska Center for Cancer Research, Department of Microbiology and Immunology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg and [6]Department of Infectious Diseases, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

HK and MSN contributed equally as first authors.

#FBT and AM contributed equally as senior authors.

Correspondence: A. Martner
anna.martner@gu.se

Received: January 15, 2024.
Accepted: August 6, 2024.
Early view: August 15, 2024.

https://doi.org/10.3324/haematol.2024.285071

Supplementary Material

Supplementary Methods

Immunophenotypic analysis of blood and BM samples

Peripheral blood and bone marrow mononuclear cells (PBMC and BM MNC) were stained with a panel consisting of anti-CD15-BV786 (HI98), anti-CD33-BV711 (WM53), anti-CD34-BV421 (581) and anti-CD38-BV510 (HIT2) antibodies (BD Biosciences) and acquired on a five laser BD LSR Fortessa or a BD FACS Aria Fusion (BD Biosciences). Analyses were performed using FlowJo (v10 or later, BD Biosciences).

Cell cycle analysis

Blood samples obtained from three patients before and after HU treatment were thawed and stained for cell surface markers using anti-CD14-PE-Cy7 and anti-CD34-PE (clones M5E2 and 8G12, respectively; BD Biosciences). Cells were washed and fixed with 2% paraformaldehyde (ThermoFisher), and permeabilized with permeabilization buffer (PBS with 0.1% EDTA, 0.5% BSA, 0.1% saponin and 3% FCS). Thereafter, cells were stained with 1µg/ml DAPI (Invitrogen) and anti-Ki67-Alexa Fluor 647 (B56, BD Biosciences) for 30 minutes at room temperature. Cells were acquired on a five laser BD LSR Fortessa (BD Biosciences) at a low flow rate (100-200 events/second) and DAPI fluorescence was analyzed on the linear scale. Analyses were performed using FlowJo (v10.8.1, BD Biosciences).

CITE-seq library preparation

PBMC and BM MNC obtained before or after HU were thawed and labeled with sample-specific barcoded cell hashing antibodies ('Sample Tags', BD Biosciences). MACS MicroBead (Miltenyi Biotec) $CD34^+$ cell selection was performed on samples that had not undergone this enrichment prior to cryopreservation. Cells were pooled to achieve 1:1 ratio between cells from

different samples, and thereafter co-labeled with unconjugated, fluorescent, and oligo-conjugated BD AbSeq antibodies (BD Biosciences) (Tables S3 and S4). $CD14^{-}CD34^{+}$ MNC were isolated from the sample pool using a BD FACSAria Fusion Cell sorter (BD Biosciences). CITE-seq sequencing libraries were generated from the isolated $CD14^{-}CD34^{+}$ cells using the BD Rhapsody Single-Cell Analysis System (BD Biosciences) as described in Nilsson *et al*.[1]

Sequencing

mRNA, AbSeq and Sample tag libraries were sequenced at SNP&SEQ Technology Platform (Uppsala, Sweden; part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory). Sequencing was performed on a NovaSeq 6000 instrument (Illumina) using a S4-200 v1.5 flow cell (Illumina), a custom read setup of 64-8-0-74 (R1-i7-i5-R2) and 20% PhiX control v3 (Illumina).

Data analysis

After initial quality control, mRNA expression data were normalized using the SCTransform function[2], and differences between proliferating cells (i.e. G2/M vs S phase) based on Seurat's CellCycleScoring function were regressed out of the data. To remove patient-specific clustering from the analysis presented in Figure 3, mRNA data were integrated by patient using reciprocal PCA (rPCA) on the top 22 PCs (k.anchor = 5). SCT or integrated assays were used for principal component analysis (PCA) followed by K-nearest neighbor (KNN) graph construction using the top 22-23 PCs, and Louvain algorithm-based clustering. Uniform Manifold Approximation and Projection (UMAP) was employed for clustering visualization.

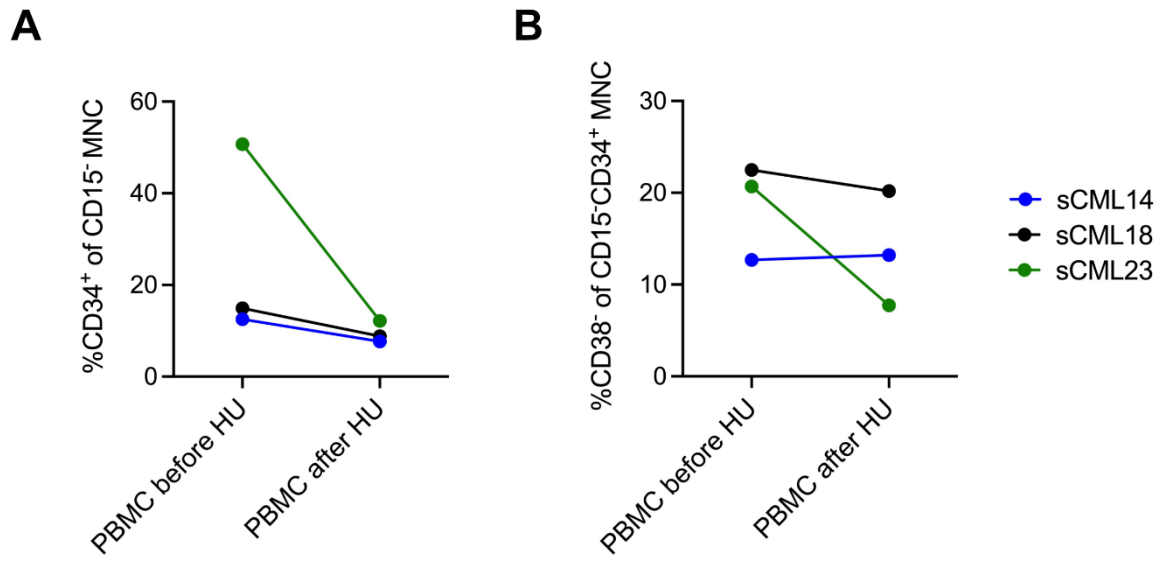
Selection of $CD38^{-}$ cells for reclustering was performed based on centered log ratio (CLR) normalized CD38 protein expression data using Seurat's CellSelector function. A previously

annotated $CD14^{-}CD34^{+}CD38^{-}$ CML dataset (Nilsson *et al.*[1]) was used as a reference to annotate $CD14^{-}CD34^{+}CD38^{-}$ cells by cell label transfer using Seurat's FindTransferAnchors and TransferData functions.

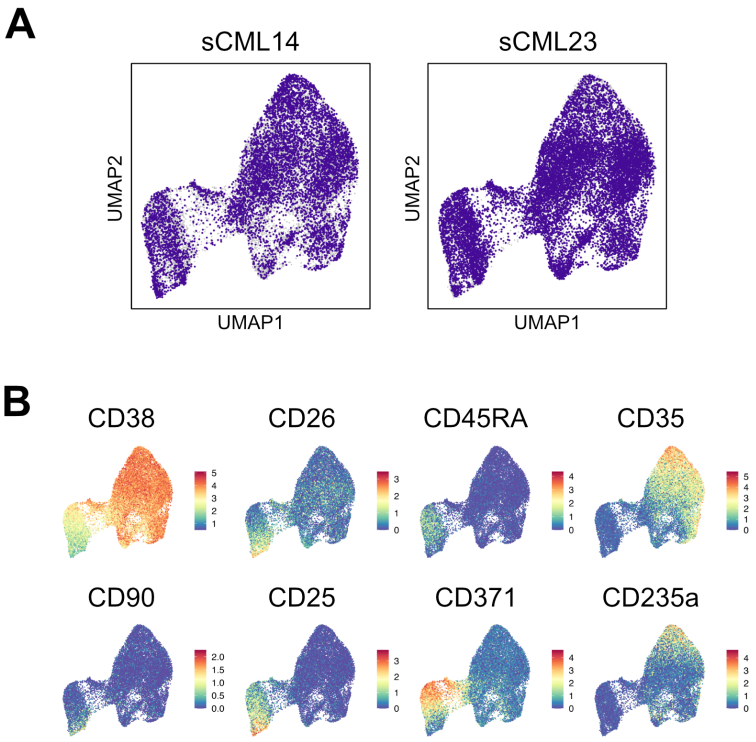
Visualization and differential expression analyses utilized log-normalized RNA or CLR normalized AbSeq counts. UMAP cluster annotations were manually assigned based on differentially expressed genes and proteins from Seurat's FindAllMarkers function as well as feature and dot plot visualizations of expression patterns across the UMAPs. Visualizations were created using ggplot2 (v. 3.4.0), SCPubr (v. 1.0.1), RColorBrewer (v. 1.1.3), GraphPad Prism (v. 9.4.1) and BioRender.com.

The pySCENIC (v. 0.12.1) package in Python (v. 3.1.0) was utilized to address genes regulated by the GATA1 transcription factor. The org was set to hgnc and linked to the RcisTarget databases. A database containing gene promoters (hg38_500bp_up_100bp_down_full_tx_v10_clust.genes_vs_motifs.rankings.feather) was used and only motifs with a Normalized Enrichment Score (NES) > 3.0 were considered as significantly enriched in the transcription factors (TF) module. Downstream analyses were performed using *Grnboost2*, followed by using *pyscenic ctx* and *pyscenic aucell*, using mask dropouts (cells in which expression of either TF or target gene is 0). As input data matrices had been subjected to quality control (QC) in the R analysis pipeline, this aspect was not addressed during SCENIC analysis.

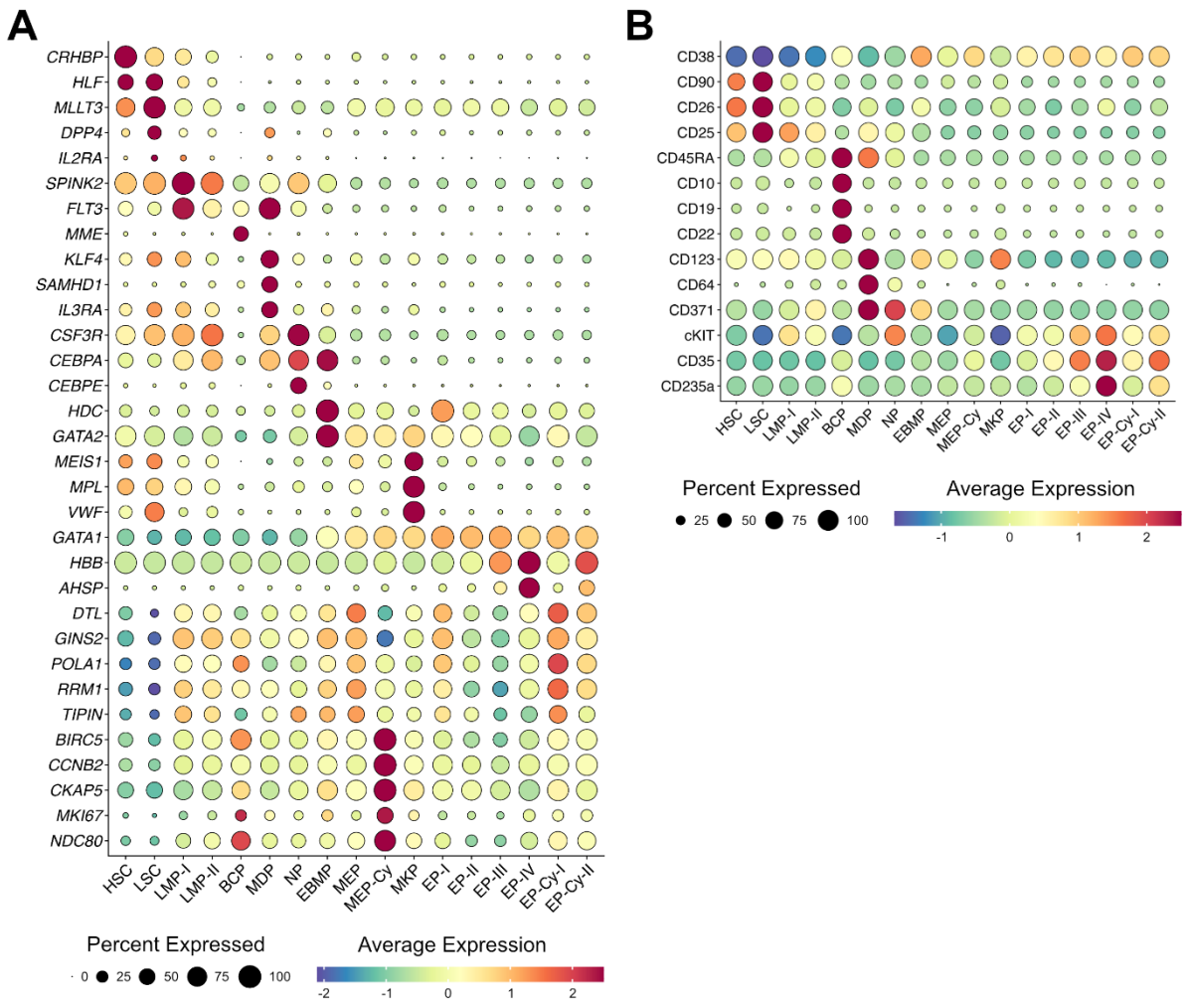
Supplementary Figures



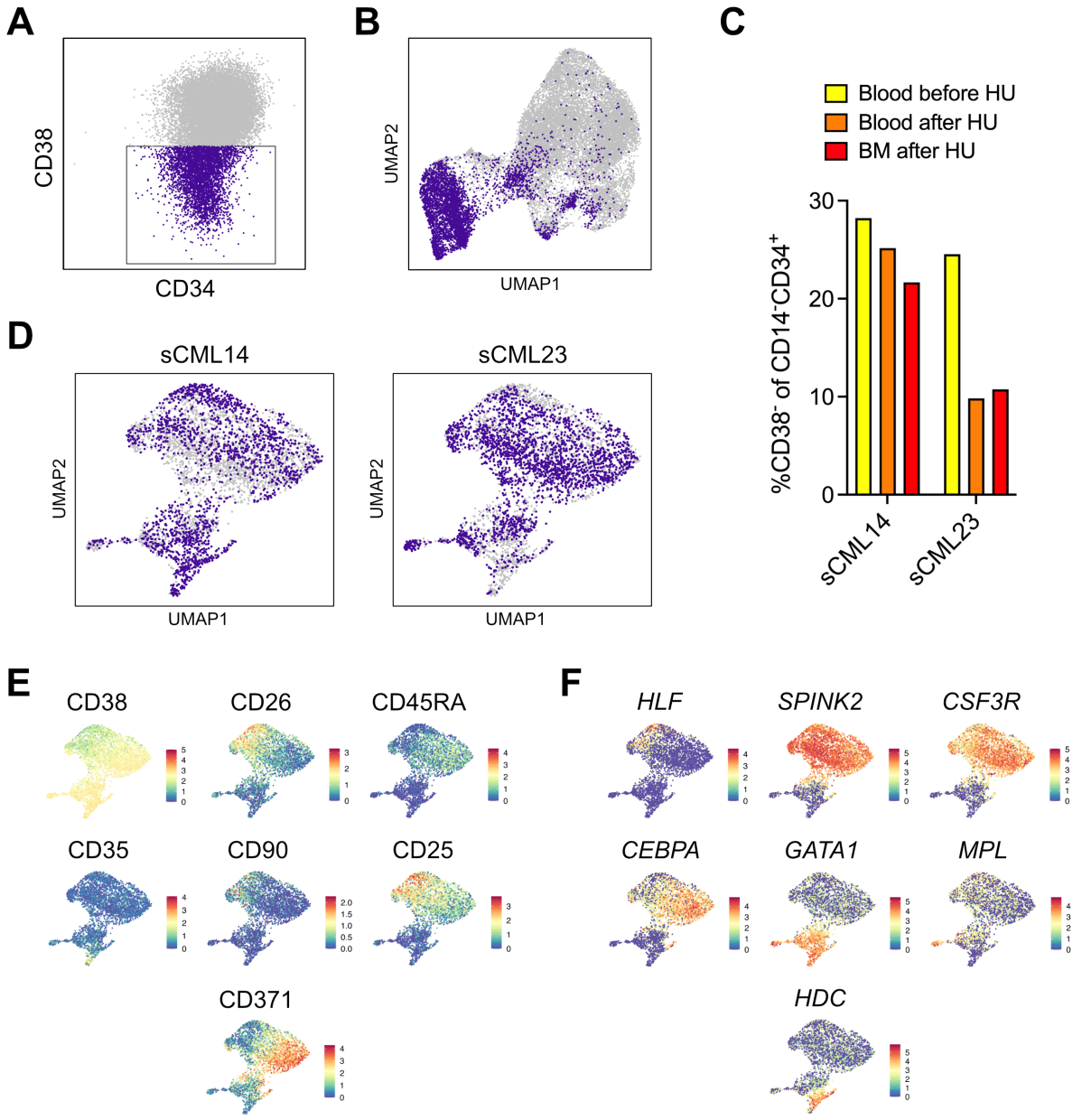
Supplementary Figure 1. Effect of hydroxyurea (HU) treatment on the frequency of **(A)** $CD34^+$ and **(B)** $CD34^+CD38^-$ cells in paired CML patient peripheral blood (PBMC) samples obtained before and after HU treatment. Patients sCML14, sCML18 and sCML23 had received HU treatment for 7, 4, and 9 days respectively, prior to 'after HU' sampling. Samples highlighted in blue (sCML14) and green (sCML23) were subsequently subjected to CITE-seq analysis. MNC, mononuclear cells.



Supplementary Figure 2. Characteristics of cells in the UMAP comprising $CD14^{-}CD34^{+}$ cells from blood and BM samples obtained from patients sCML14 and sCML23 before and after 7-9 days of hydroxyurea (HU) treatment. **(A)** Distribution of cells from each patient across the UMAP. Cells from samples obtained from patient sCML14 are highlighted in purple in the left panel, and those from patient sCML23 in the right panel. **(B)** Feature plots of expression of indicated marker proteins among cells in the UMAP. Red indicates maximum expression and blue minimum expression.

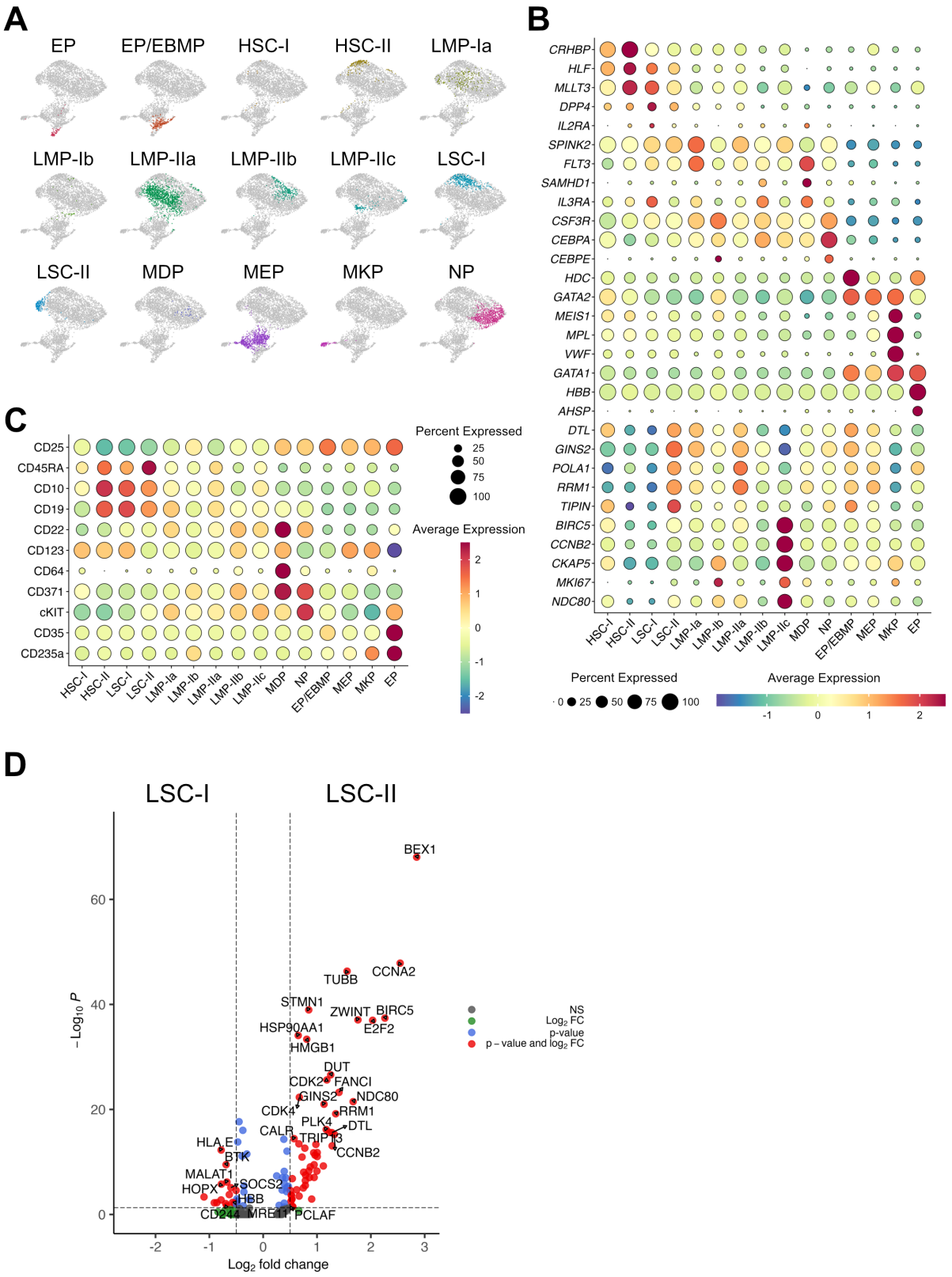


Supplementary Figure 3. Scaled expression of **(A)** genes and **(B)** proteins used to confirm the cell type annotation in Figure 5. Cells in the analysis derive from a previously published dataset of $CD14^{-}CD34^{+}$ bone marrow cells obtained before (n=5) or after (n=7) hydroxyurea (HU) treatment (unpaired samples). Colors represent average expression level and dot size percentage of cells expressing the specific marker. For further detail on the original annotation of these cell types, the reader is referred to Nilsson *et al.*[1] BCP, B cell progenitors; EBMP, eosinophil/basophil/mast cell progenitors; EP, erythroid progenitors; HSC, hematopoietic stem cells; LMP, lympho-myeloid progenitors; LSC, leukemic stem cells; MDP, monocyte/dendritic cell progenitors; MEP, megakaryocytic/erythroid progenitors; MKP, megakaryocytic progenitors; NP, neutrophil progenitors.



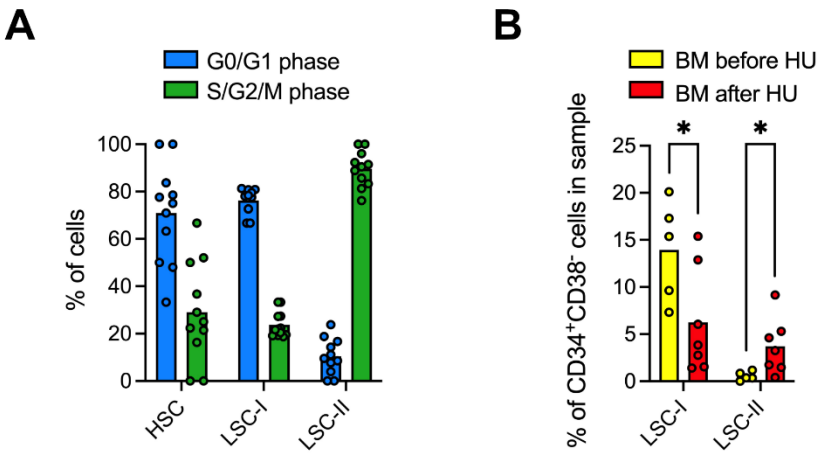
Supplementary Figure 4. Generation and characterization of the UMAP comprising $CD34^{+}CD38^{-}$ cells from blood and bone marrow (BM) samples obtained from patients sCML14 and sCML23 before and after 7-9 days of hydroxyurea (HU) treatment. **(A)** Gating strategy for CITE-seq protein expression-based subsetting of immature $CD34^{+}CD38^{-}$ cells. **(B)** Distribution of subsetted $CD34^{+}CD38^{-}$ cells across the $CD14^{-}CD34^{+}$ UMAP. Subsetted cells are highlighted in purple. **(C)** Frequency of $CD38^{-}$ among $CD14^{-}CD34^{+}$ cells based on CITE-seq protein expression-based gating in blood samples obtained before HU treatment and blood and BM samples obtained after HU treatment for each patient. **(D)** Distribution of cells from

each patient in the reclustered $CD34^{+}CD38^{-}$ UMAP. Cells from samples obtained from patient sCML14 are highlighted in purple in the left panel, and those from patient sCML23 in the right panel. **(E-F)** Feature plots highlighting the expression of selected marker **(E)** proteins and **(F)** genes across the UMAP. Red indicates maximum expression and blue minimum expression.



Supplementary Figure 5. Characteristics of CD34^{+}CD38^{-} cell subset annotations generated through cell label transfer from a previously published reference dataset comprising CD14^{-} CD34^{+}CD38^{-} CML/healthy bone marrow cells (Nilsson *et al.*[1]). **(A)** Plots highlighting cells

given each individual cell label in the transfer. EBMP, eosinophil/basophil/mast cell progenitors; EP, erythroid progenitors; HSC, hematopoietic stem cells; LMP, lympho-myeloid progenitors; LSC, leukemic stem cells; MDP, monocyte/dendritic cell progenitors; MEP, megakaryocytic/erythroid progenitors; MKP, megakaryocytic progenitors; NP, neutrophil progenitors. **(B-C)** Scaled expression of **(B)** genes and **(C)** proteins used to confirm the validity of the cell label transfer annotation in Figure 7. Colors represent average expression level and dot size percentage of cells expressing the specific marker. For further detail on the original reference dataset annotation, see Nilsson *et al*.[1] **(D)** Volcano plot of differentially expressed genes between cells annotated as LSC-I and LSC-II. Green dots mark non-significantly regulated genes with $|Log_2FC| > 0.5$, blue dots significantly regulated genes (adjusted p values < 0.05) with $|Log_2FC| \leq 0.5$ and red dots significantly regulated genes with $|Log_2FC| > 0.5$. Statistics by Wilcoxon Rank Sum test with Bonferroni correction.



Supplementary Figure 6. Characterization of stem cell subsets in the confirmatory unpaired cohort, comprising bone marrow (BM) cells obtained from CML patients before (n=5) or after 4-19 (median 9) days of hydroxyurea (HU) treatment (n=7). **(A)** Proportion of cells in G0/G1 vs S/G2/M phase in each stem cell subset in the 12 samples. HSC, hematopoietic stem cells; LSC, leukemic stem cells. As HSC and LSC-II subsets were each missing in one patient, 11 data points are displayed for these subsets. **(B)** Frequency of LSC-I and LSC-II out of all $CD34^{+}CD38^{-}$ cells in samples obtained before or after HU treatment. Statistics by Mann-Whitney. * $p<0.05$.

Supplementary Tables

Supplementary Tables are available in a separate excel file.

Supplementary Table 1. Chronic phase CML patient samples included in flow cytometry and CITE-seq analyses. Data include patient ID, age, gender, and specification of peripheral blood/bone marrow samples included in flow cytometry and CITE-seq analyses (with cell numbers analyzed). In addition, hydroxyurea (HU) treatment duration and dosing is indicated for samples obtained after HU treatment.

Supplementary Table 2. Genes included in the CITE-seq analysis.

Supplementary Table 3. Oligo-conjugated AbSeq antibodies used in the CITE-seq analysis.

Supplementary table 4. Antibodies used for fluorescence-activated cell sorting of $CD14^-$ $CD34^+$ cells and AbSeq signal muting. The table additionally specifies the purpose of inclusion of each antibody.

Supplementary Table 5. Seurat FindAllMarkers cluster differential gene expression analysis in the $CD14^-CD34^+$ UMAP. Differential gene expression analysis of each cluster in Figure 3A (specified under "cluster") against cells in all other clusters. Columns specify Wilcoxon Rank Sum test p values (p_val), average $\log_2$ fold change (avg_log2FC), fraction of cells expressing the gene in each population (pct.1 and pct.2), and adjusted p values based on Bonferroni correction (p_val_adj). The analysis was run only considering genes expressed in

at least 25% of cells in either population and the table filtered to only contain upregulated genes with $log_2FC > 0.25$.

Supplementary Table 6. Seurat FindAllMarkers cluster differential protein expression analysis in the CD14$^-$CD34$^+$ UMAP. Differential protein expression analysis of each cluster in Figure 3A (specified under "cluster") against cells in all other clusters. Columns specify Wilcoxon Rank Sum test p values (p_val), average log_2 fold change (avg_log2FC), fraction of cells expressing the protein in each population (pct.1 and pct.2), and adjusted p values based on Bonferroni correction (p_val_adj). The analysis was run only considering proteins expressed in at least 25% of cells in either population and the table filtered to only contain upregulated proteins with $log_2FC > 0.25$.

Supplementary Table 7. Differential expression analysis shows that target genes of *GATA1* are upregulated among CD14$^-$CD34$^+$ cells following HU treatment. Genes analyzed were identified as belonging to the *GATA1* regulon in SCENIC analyses of paired blood samples from patients sCML14 and sCML23 obtained before and after (7 and 9 days, respectively) HU. Columns **(B-F)** show results from differential gene expression analysis in blood samples obtained after vs before HU for patient sCML14, and columns **(G-K)** corresponding results for patient sCML23. Subcolumns specify Wilcoxon Rank Sum test p values (p_val), average log_2 fold change (avg_log2FC), fraction of cells expressing the gene in each population (pct.1 and pct.2), and adjusted p values based on Bonferroni correction (p_val_adj). The table was filtered to only include genes significantly differentially expressed (adjusted p value < 0.05) and with $|log2FC| > 0.25$ in at least one of the comparisons, and cells colored according to direction of regulation (upregulated genes in red and downregulated genes in blue).

Supplementary Table 8. Seurat FindAllMarkers cluster differential gene expression analysis in the CD14$^-$CD34$^+$CD38$^-$ UMAP. Differential gene expression analysis of each cluster in Figure 6A (specified under "cluster") against cells in all other clusters. Columns specify Wilcoxon Rank Sum test p values (p_val), average $\log_2$ fold change (avg_log2FC), fraction of cells expressing the gene in each population (pct.1 and pct.2), and adjusted p values based on Bonferroni correction (p_val_adj). The analysis was run only considering genes expressed in at least 25% of cells in either population and the table filtered to only contain upregulated genes with $\log_2$FC > 0.25.

Supplementary Table 9. Seurat FindAllMarkers cluster differential protein expression analysis in the CD14$^-$CD34$^+$CD38$^-$ UMAP. Differential protein expression analysis of each cluster in Figure 6A (specified under "cluster") against cells in all other clusters. Columns specify Wilcoxon Rank Sum test p values (p_val), average $\log_2$ fold change (avg_log2FC), fraction of cells expressing the protein in each population (pct.1 and pct.2), and adjusted p values based on Bonferroni correction (p_val_adj). The analysis was run only considering proteins expressed in at least 25% of cells in either population and the table filtered to only contain upregulated proteins with $\log_2$FC > 0.25.

Supplementary Table 10. Differential gene expression analysis between LSC-II and LSC-I from patients sCML14 and sCML23. Columns specify Wilcoxon Rank Sum test p values (p_val), average $\log_2$ fold change (avg_log2FC), fraction of cells expressing the gene in each population (pct.1 and pct.2), and adjusted p values based on Bonferroni correction (p_val_adj). The analysis was run only considering genes expressed in at least 10% of cells in either population and the table filtered to only contain genes with |log2FC| > 0.25 and p_val_adj < 0.05.

Supplementary References

1. Nilsson MS, Komic H, Gustafsson J, et al. Multiomic single-cell analysis identifies von Willebrand factor and TIM3-expressing BCR-ABL1+ CML stem cells. bioRxiv. 2023 Sep 17. doi:10.1101/2023.09.14.557507 [preprint, not peer-reviewed].

2. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 2019;20(1):296.