Full title:

Single cell analysis of bone marrow derived CD34+ cells from children with sickle cell disease and thalassemia

Short title:

Analysis of CD34+ cells in SCD and thalassemia

Peng Hua[1,2], Noemi Roy[3,6*], Josu de la Fuente[4*], Guanlin Wang[1], Supat Thongjuea[1], Kevin Clark[1], Anindita Roy[5], Bethan Psaila[2], Neil Ashley[1], Yvonne Harrington[4], Claus Nerlov[1], Suzanne M. Watt[2], Irene Roberts[1,5,6] and James O. J. Davies[1,3,6]

1. MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, John Radcliffe Hospital, Oxford, UK.
2. Nuffield Division of Clinical Laboratory Medicine, Radcliffe Department of Medicine, John Radcliffe Hospital, Oxford, UK.
3. Department of Haematology, Oxford University NHS Foundation Trust, Oxford, UK
4. Department of Paediatrics, St. Mary's Hospital, Imperial College Healthcare NHS Trust, London, UK.
5. Department of Paediatrics, University of Oxford, Children's Hospital, John Radcliffe Hospital, Oxford, UK.
6. Oxford Biomedical Research Centre, University of Oxford

* These authors contributed equally.

Supplemental methods

Human CD34+ HSPC isolation from Bone Marrow

Bone marrow harvests were performed at St. Mary's Hospital, London, under general anesthetic using standard procedures, following written informed consent for collection for autologous pediatric transplantation by responsible adults/parents (HTA research license no. 11118). Informed consent was also obtained for use of the cells for research should they become surplus to transplant requirements, after Oxford South Central C Research Ethics Board and WIMM R&D committee approval (ref. 17/SC/0111) for this project. Viable cells were thawed via dropwise addition of 100% fetal bovine serum (FBS) with DNase I (Roche Applied Science, 10104159001, 200ug/mL). Mononuclear cells were enriched using Ficoll-Paque (<1.077g/mL; GE Healthcare Life Sciences, cat. no. 17-5442-03). Adult bone marrow MNCs were purchased from StemCell Technologies (cat. no. 70001). CD34+ selection was

performed using the CD34 Microbead kit and MACS system (Miltenyi Biotech, cat. no. 130-046-703).

Flow cytometry and cell sorting

Multicolour flow cytometry characterization of CD34+ human bone marrow-derived HSPC was performed as previously reported[21]. Multiple combinations of antibodies were used to analyse the cells as outlined in supplementary Table 3. Briefly, cells were incubated in human FcR blocking reagent diluted in MACS buffer (Miltenyi Biotec. cat. no. 130-059-901 and 130-117-336) and incubated for 10 minutes on ice. Cells were then incubated with a mixture of fluorescently labelled antibodies diluted in MACS buffer for 20 minutes on ice. Cells were washed once, resuspended in MACS buffer and analysed immediately on an FACS Aria II (BD Biosciences). 7AAD (Biolegend) or eF506 (eBioscience) was added at 100ng/ml directly before acquisition to distinguish live and dead cells. Gating was set based on Fluorescence Minus One controls (FMO) and data were further analysed with FlowJo software (TreeStar Inc.). The following antibodies were used: CD34 AF700/PerCP-Cy5.5 (581) or APC (8G12), CD45RA APC-H7/BV650/FITC (HI100), CD38 PE-TxR/BB515/AF700 (HIT2), CD90 PE/BV421 (5E10), CD123 PerCP-Cy5.5/PE-Cy7 (6H6) , CD10 PE-Cy7 (GoH3), CD19 BV711 (HIB19), CD133 BV711 (W6B3) and a lineage cocktail from BD (CD2, CD3, CD14, CD16, CD19, CD56, CD235a).

Colony forming assays

200 cells from CMP, GMP, MEP as defined by the immunophenotype shown in supplemental Table 4. These were sorted (single-cell mode) into 1mL of methylcellulose (MethoCult Enriched, cat. no. H4435, Stem cell technologies), mixed and then plated into 35 mm dishes in triplicate. The cells were allowed to differentiate for 14 days and were assessed morphologically for their colony forming potential.

Single-cell RNA sequencing and bioinformatic analysis

Single-cell RNA sequencing libraries were prepared using Chromium Single Cell 3' Library & Gel Bead Kit v2 (10XGenomics, cat. no. 120237). 10,000 CD34+ cells from each donor were sorted from AriaII (BD Biosciences) on 4-way purity mode using 70-micron nozzle and then loaded on a GemCode Single-Cell Instrument (10x Genomics, Pleasanton, CA, USA) to generate single-cell Gel Bead in emulsion (GEMs) respectively. Reverse transcription (RT) was performed according to the manufacturer's protocol (55 °C for 2 h then 85 °C for 5 min). After RT, GEMs were broken and the single-strand cDNA was purified with DynaBeads MyOne Silane Beads and SPRIselect Reagent (0.6 × SPRI). cDNA was amplified by PCR (98 °C for 3 min; 98 °C for 15 s; 67 °C for 20 s; and 72 °C for 1 min) for 14 cycles then 72 °C for 1 min. The amplified cDNA product was cleaned up using SPRIselect beads (0.6 × SPRI). The cDNA was enzymatically fragmented and end-repair, dA-tailing and adaptor ligation were performed. The patient samples were differentially indexed using Chromium i7 sample indexing. The barcoded and indexed libraries were quantified by quantitative PCR (KAPA Biosystems Library Quantification Kit for Illumina platforms, cat. no. KK4824) and normalised. The library pool was sequenced on an Illumina NovaSeq6000 (S2 kits, PE100bp) (Novogene), using the following settings: 26 cycles Read1, 8 cycles I7 Index, 0 cycle I5 Index and 91 cycles Read2.

Approximately 3,000-5,000 cells were captured for each donor and at least 100,000 reads per cell were obtained (QC30 shown in supplemental Table 2).

The Cell Ranger Single-Cell Software (1.2.0, 10x Genomics) was used to perform sequencing data demultiplexing, barcode processing and gene counting against the reference genome (GRCh38-1.2.0). Gene reads counted by Cell Ranger were then loaded into R (3.5.1) and analyzed using Seurat (3.0.1). Cells were excluded if the number of genes detected was below 500 or the percentage of mitochondrial genes above 5% (supplemental Table 2). "merge" function was then used for pooling data from different biological donors, followed by log-normalization using "NormalizeData" on a scale factor as 10000. All variable genes were detected by "FindVariableFeatures" and genes passing the "vst" filter were included in the following analysis. Biological variation within the same group was removed by regressing the

expression value based on total UMI counts, proportion of mitochondrial genes and batches by “ScaleData” function. Principal component analysis was performed using “RunPCA” function and report by “Elbowplot”. The top 20 principle components (PCs) were used in downstream analysis. Louvain graph-based clustering and dimensionality reduction methods included tSNE using the R package Seurat.

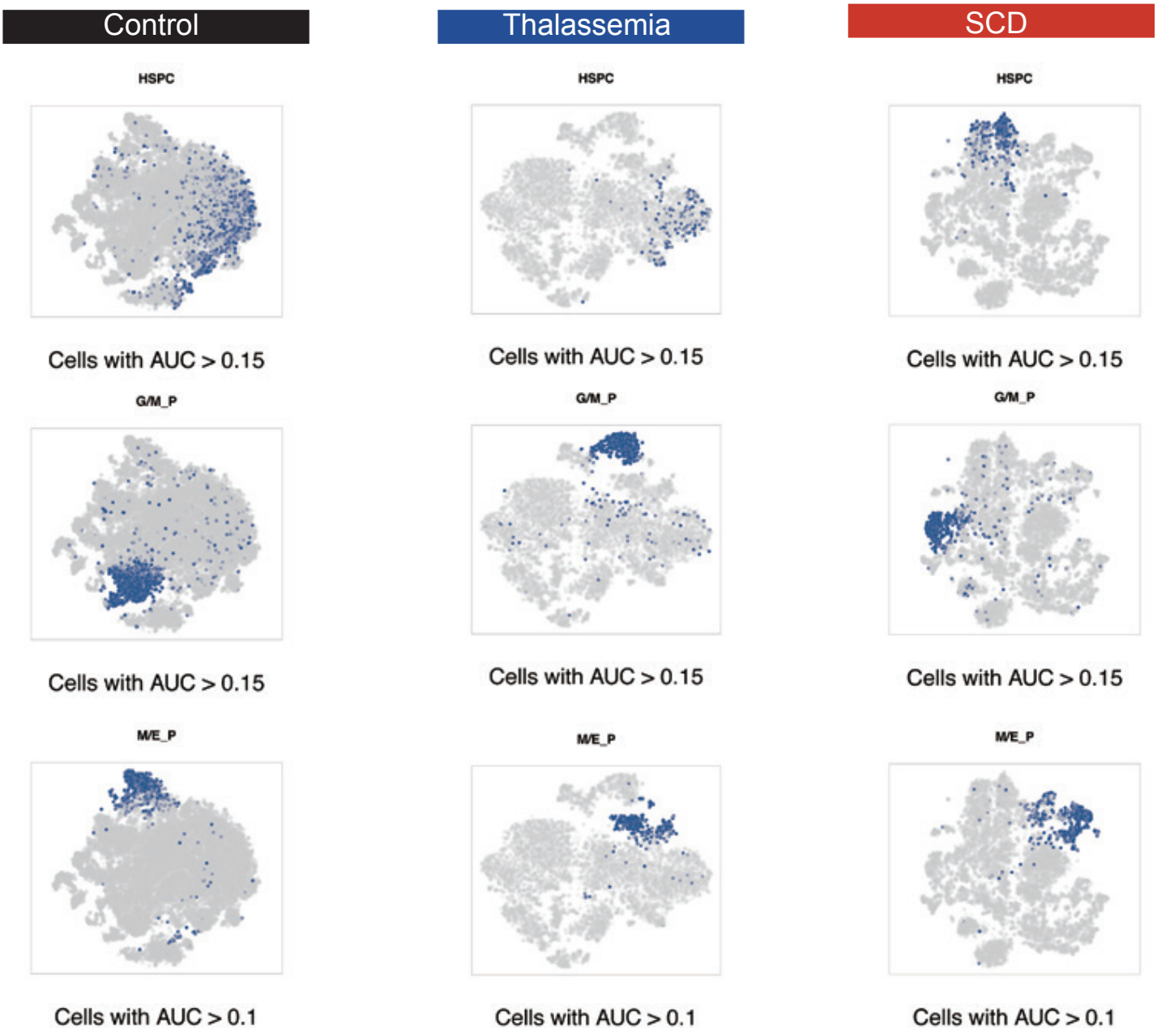
The identity of cells was assigned iteratively based on the current known markers and transcription factors[20,21] listed in supplemental Table 3. Each cell was scored based on the UMI filtered expression of these genes using SCENIC (a computational method for cell-state identification)[19] and the same threshold for each type of cells was set for analysing all samples, providing consistency of cell type identification between the different patient samples. Clusters passing the AUcell filtering were annotated correspondingly and the expression level of markers or transcription factors between the clusters was plotted using violin plot function from the R package scater (v 1.10.1)[25]. The data are available at GEO: GSE133181.

The further details of the code for trajectory analysis can be requested from supat.thongjuea@ndcls.ox.ac.uk.

Statistical analysis

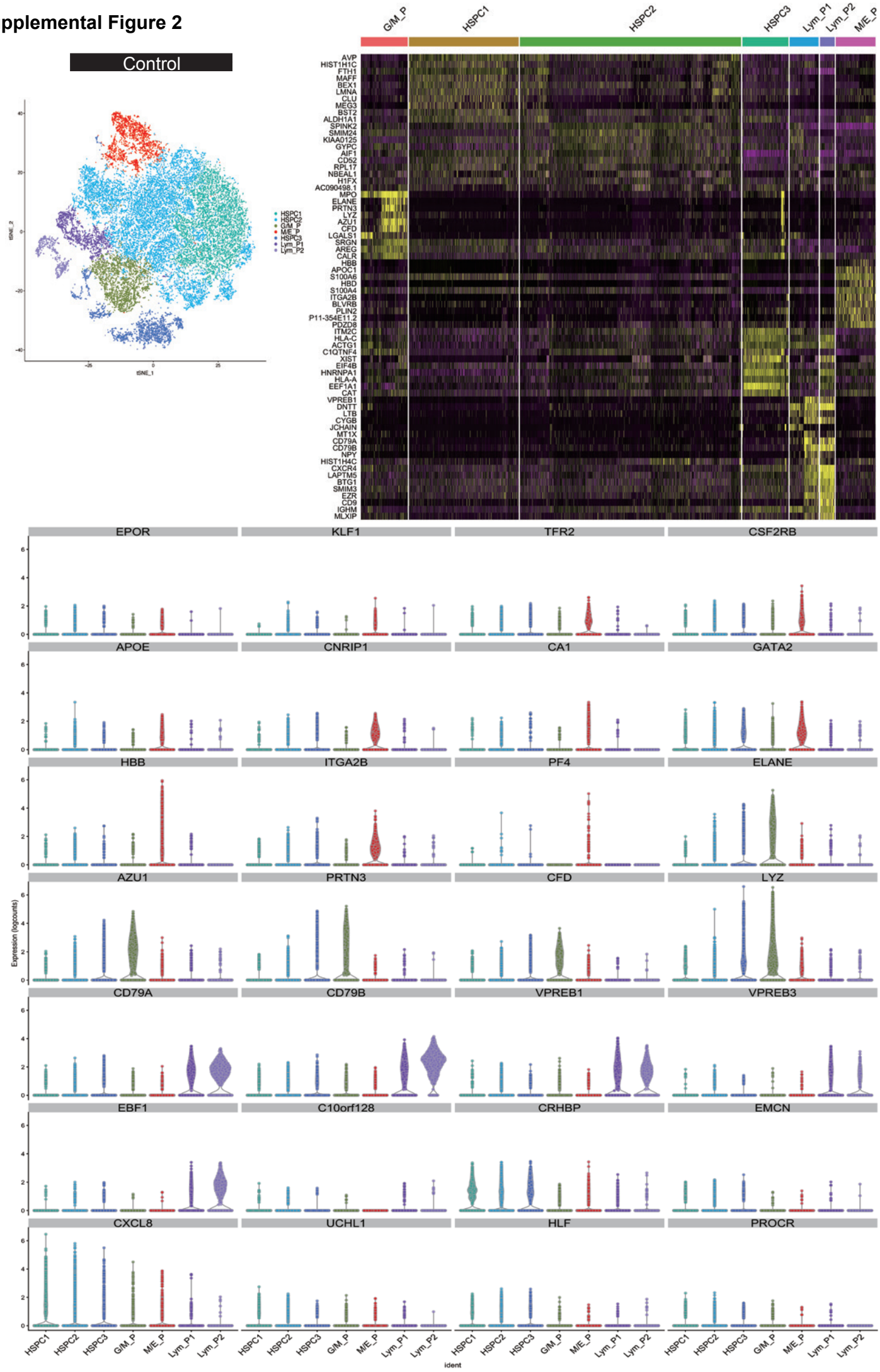
Statistical analysis was performed on GraphPad Prism 7.0. Data are shown as mean ± sem. Two-tailed t-test was used to compare differences in means between controls and patients. N.S $p>0.05$; * $p<0.05$; **, $p<0.005$; ***, $p<0.001$.

Supplemental Figure 1. A) Single cell 10X chromium analysis showing tSNE plots of Lymphoid, Myeloid, Erythoid and Multipotential progenitor cells from control and patients. Cell type annotation was performed using AUcell based on the expression level of cell type specific genes (supplemental methods; supplemental Table 3).

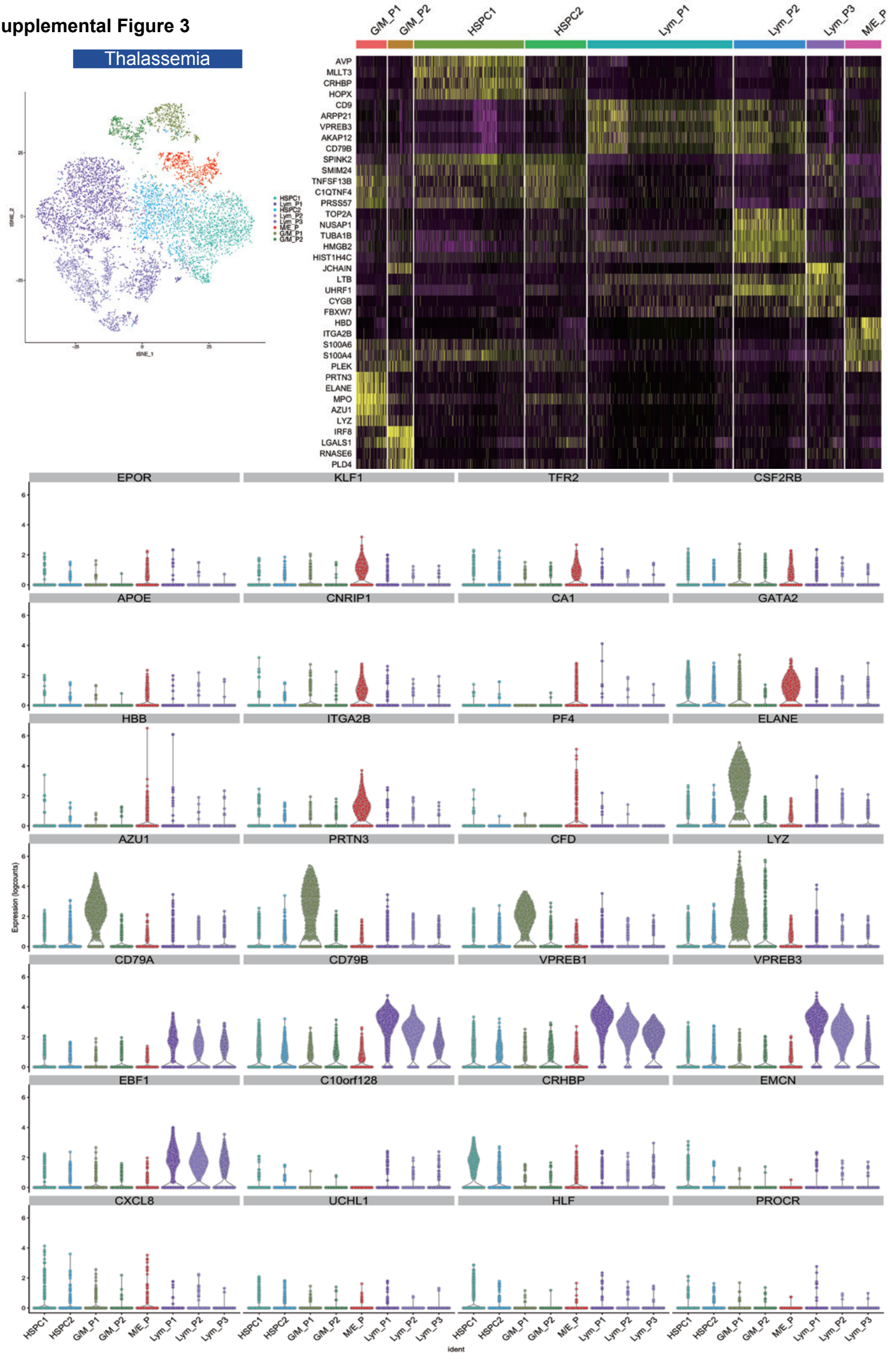


Supplemental Figure 2, 3 and 4 show tSNE plots of the unsupervised clustering of pooled Chromium 10x data from adult controls (supplemental Figure 2), children with Thalassemia (supplemental Figure 3) and children with Sickle cell disease (supplemental Figure 4). The heatmaps show the top 10 differentially expressed genes resulting from unsupervised clustering. These clusters could be annotated as G/M progenitors; HSPC1 (the earliest undifferentiated cluster); less-immature HSPC2 and an HSPC3 was defined in adult patients only. In addition, clusters could be linked to Lymphiod progenitors and M/E progenitors in patient and control samples. The heat maps show the expression level from low (purple) to high (yellow) of these genes. The violin plots show the expression level of the genes that were used to annotate the clusters in each of the clusters.

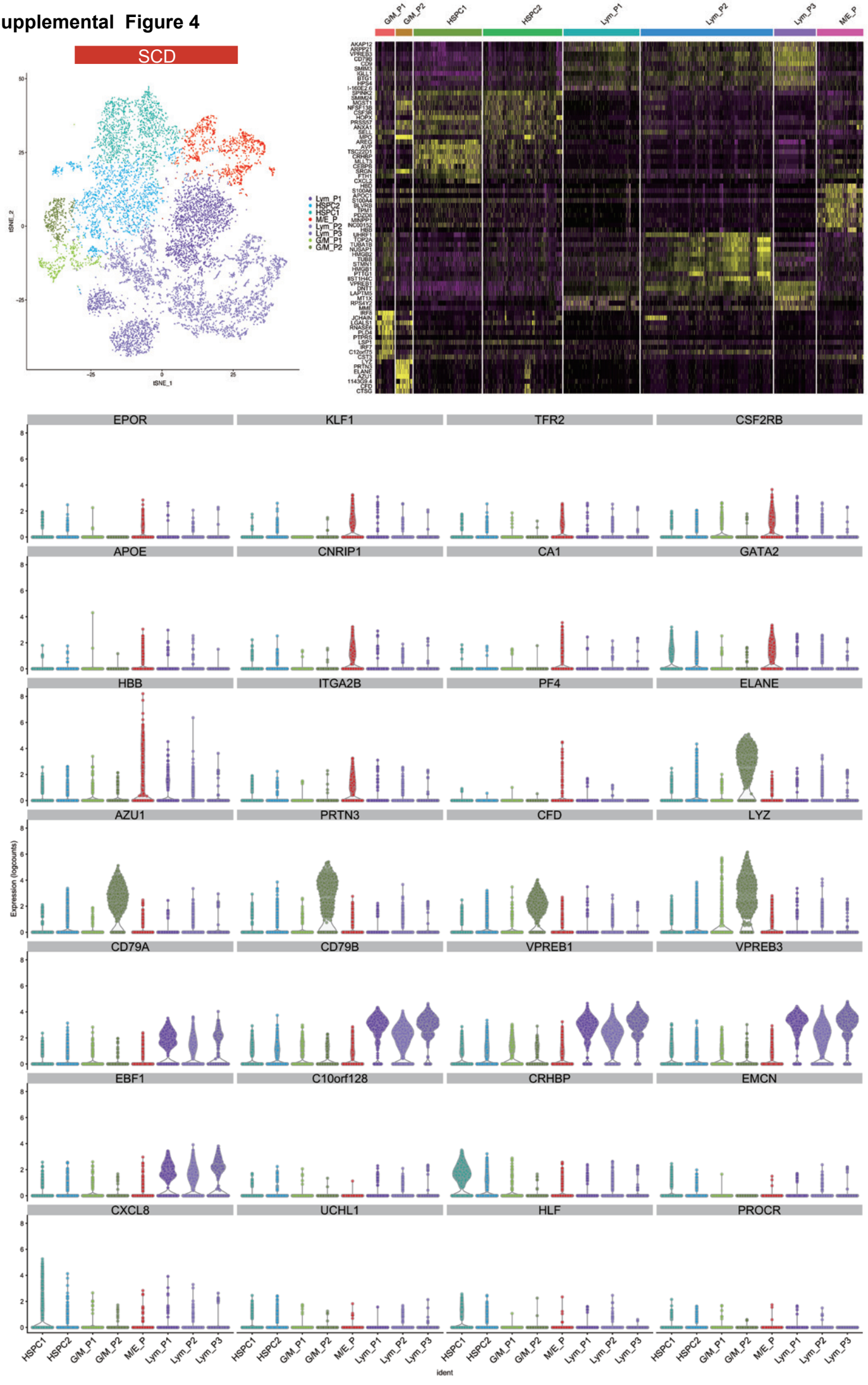
Supplemental Figure 2



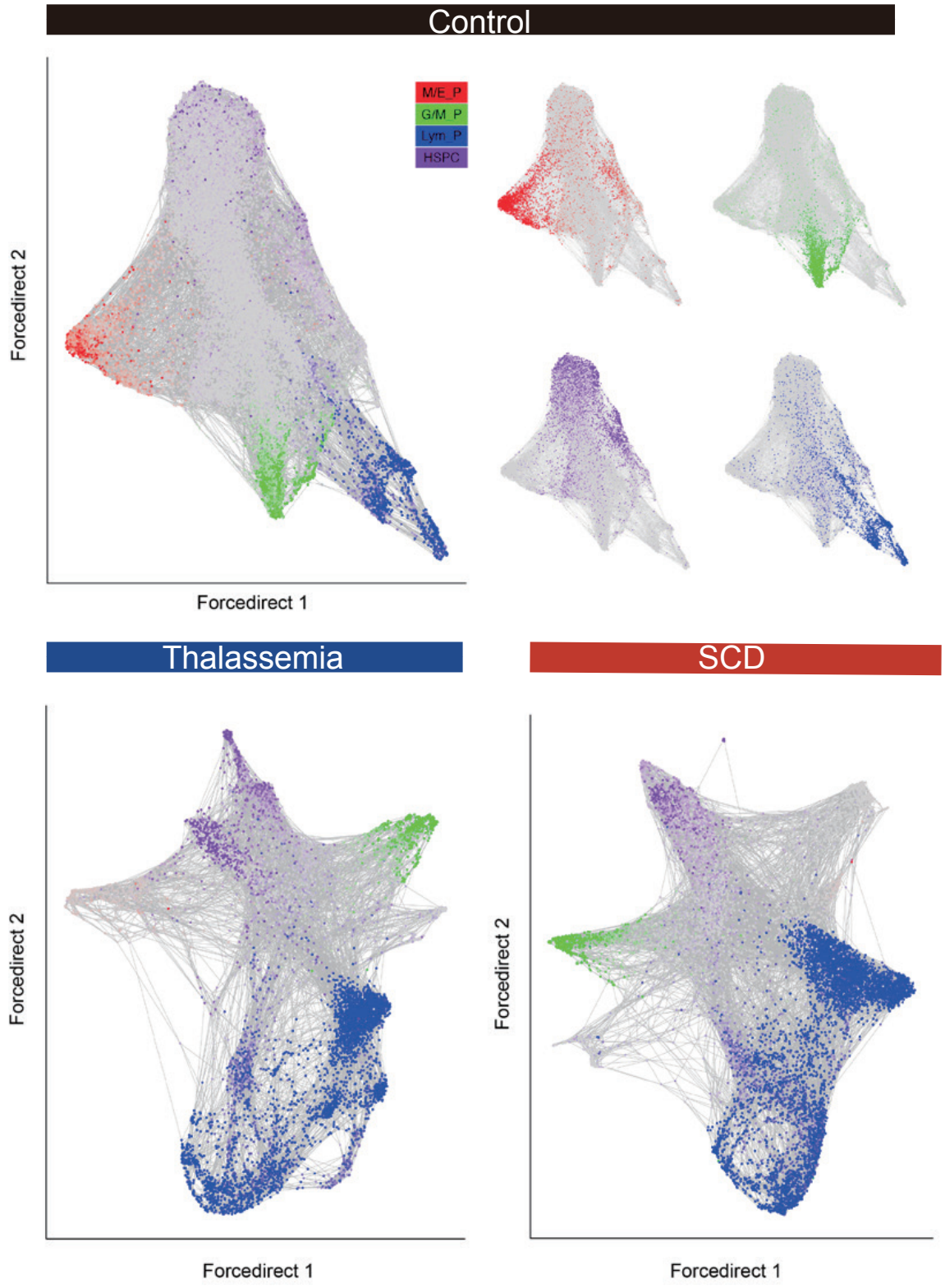
Supplemental Figure 3



Supplemental Figure 4

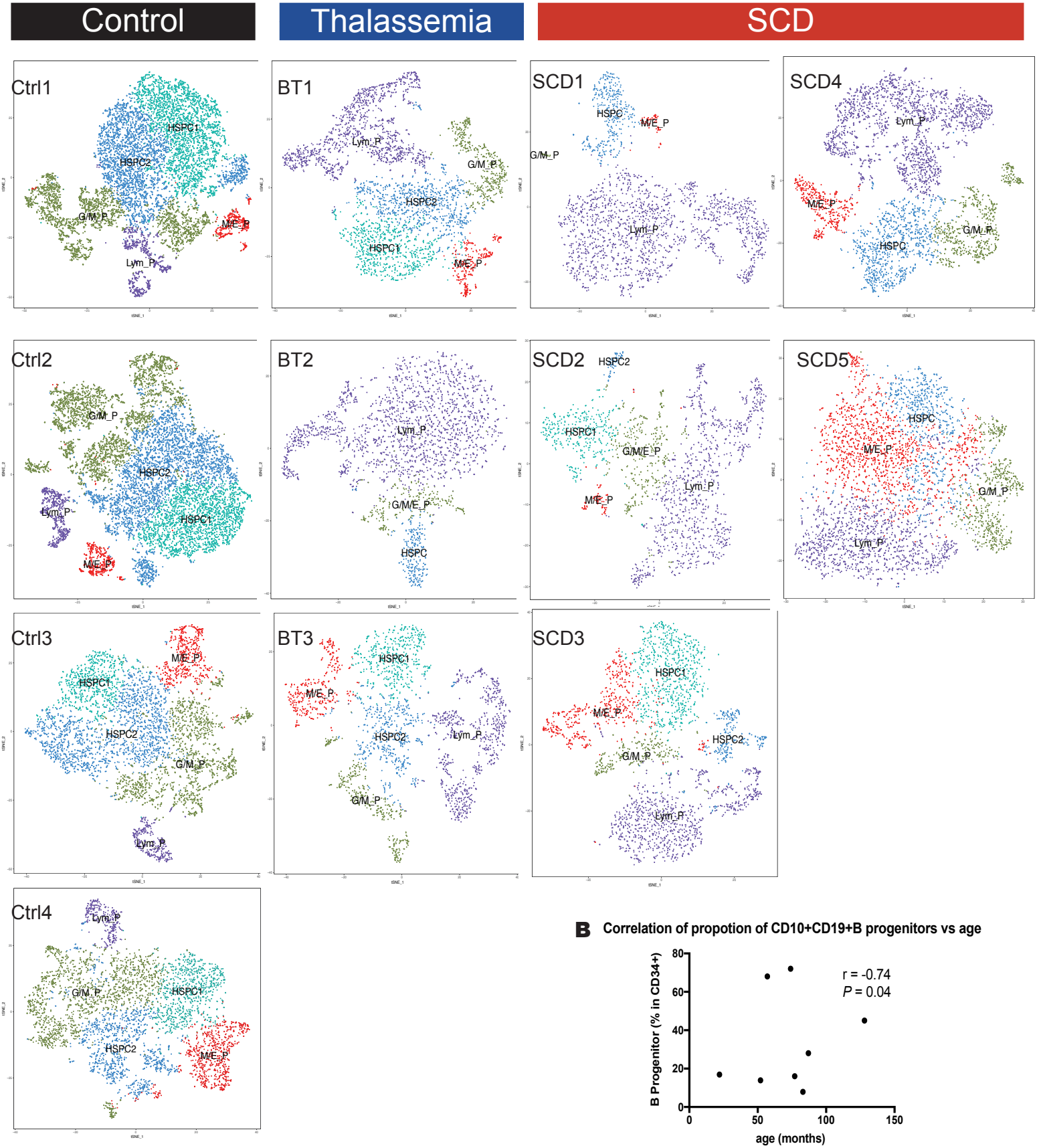


Supplemental Figure 5. Trajectory analysis showing confirming marked expansion of lymphoid progenitors in the patient samples compared to controls. This analysis allowed two different trajectories of CD34+ Lymphoid progenitors to be discerned.



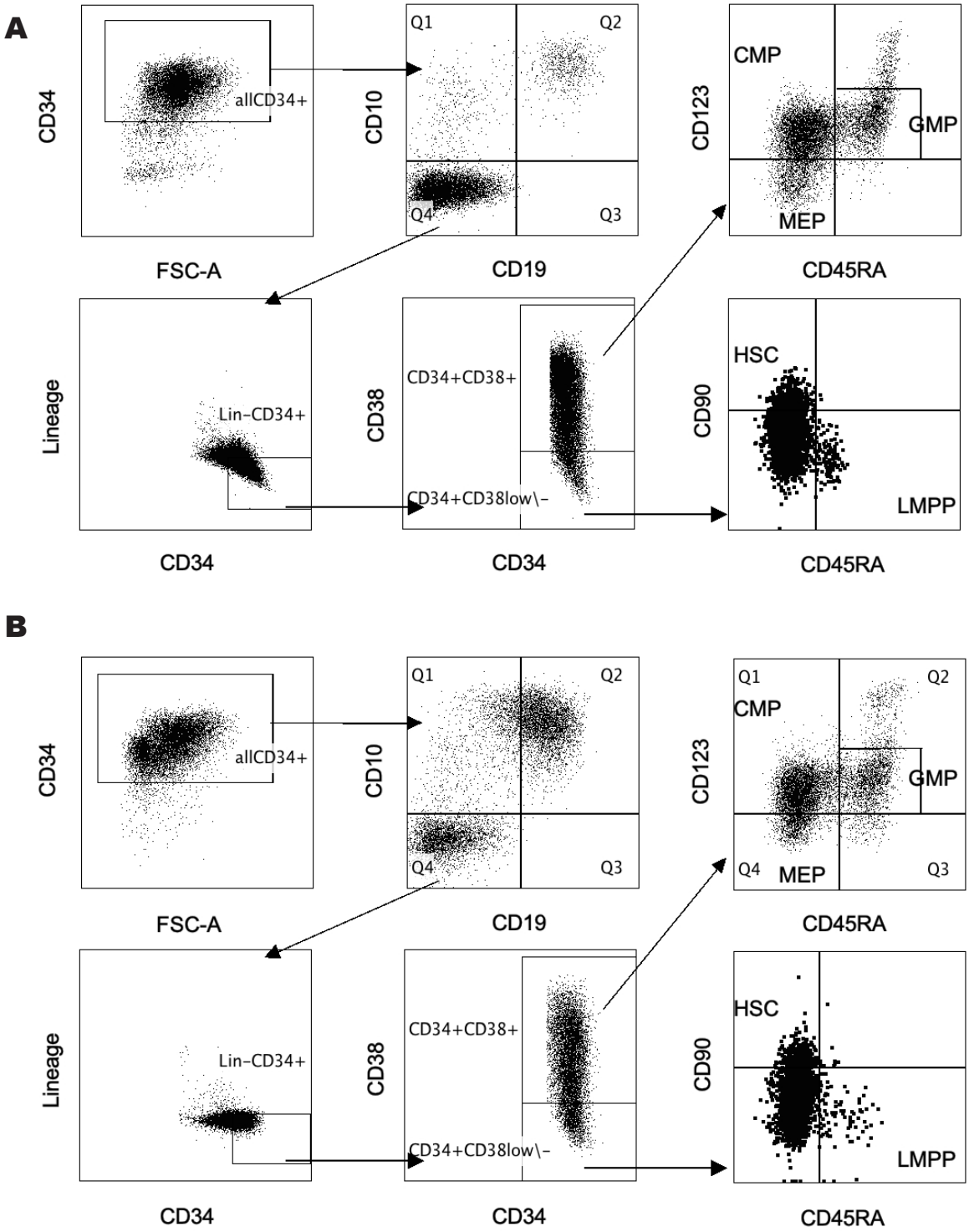
Supplemental Figure 6. A) show tSNE plots of the unsupervised clustering of individual donor's Chromium 10x data. B) shows correlation analysis between proportion of CD10+CD19+ cells against the age of patients.

A



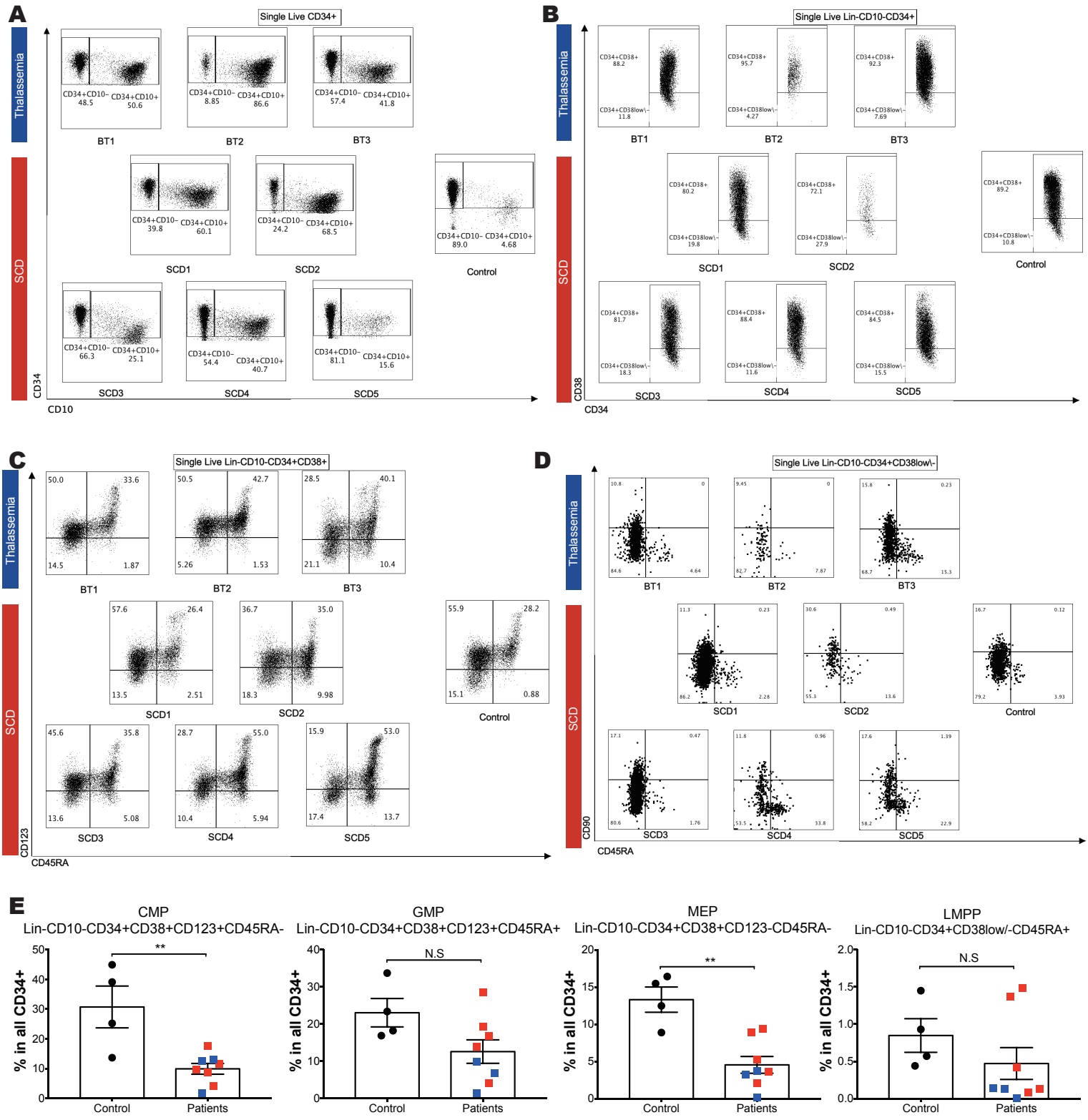
B **Correlation of propotion of CD10+CD19+B progenitors vs age**

Supplemental Figure 7. Gating strategy of flow analysis normal donor A) and SCD patient B).

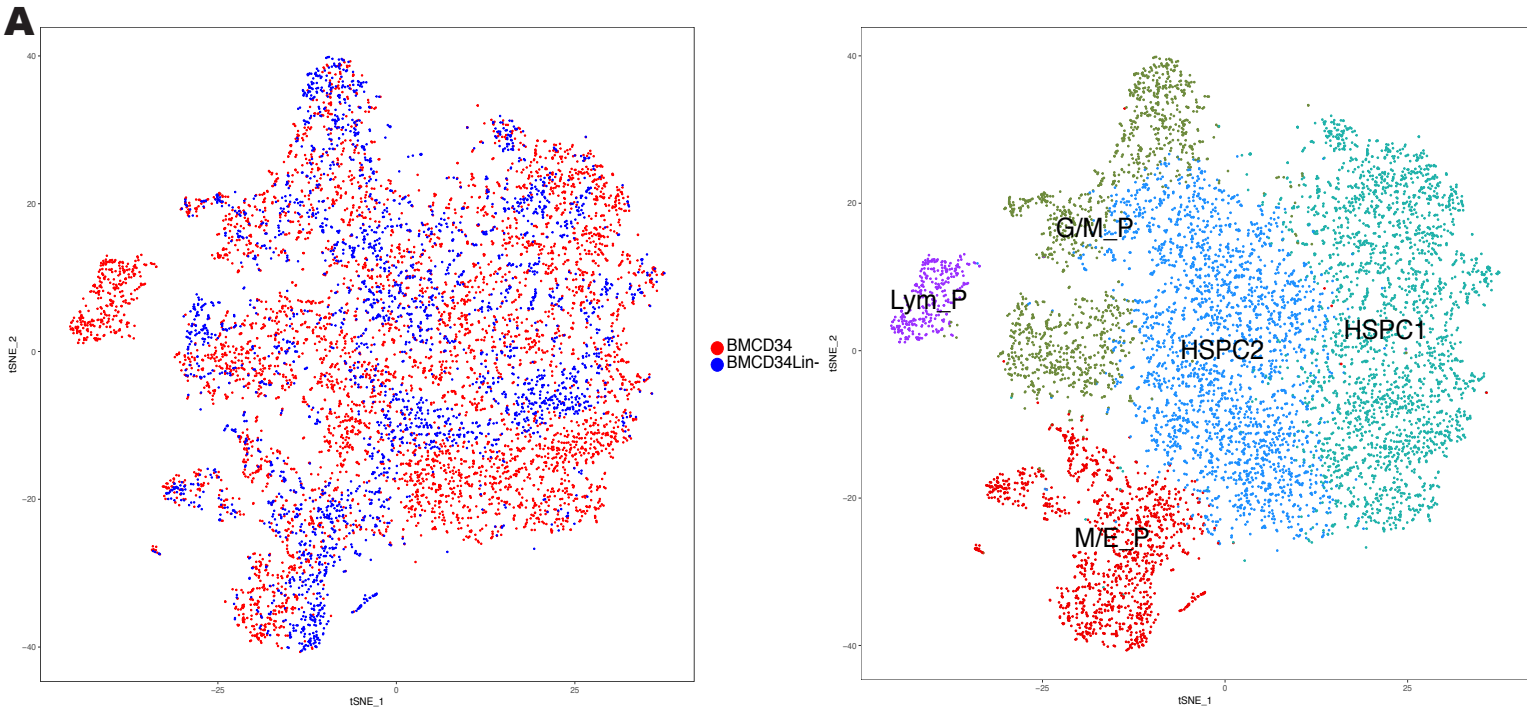


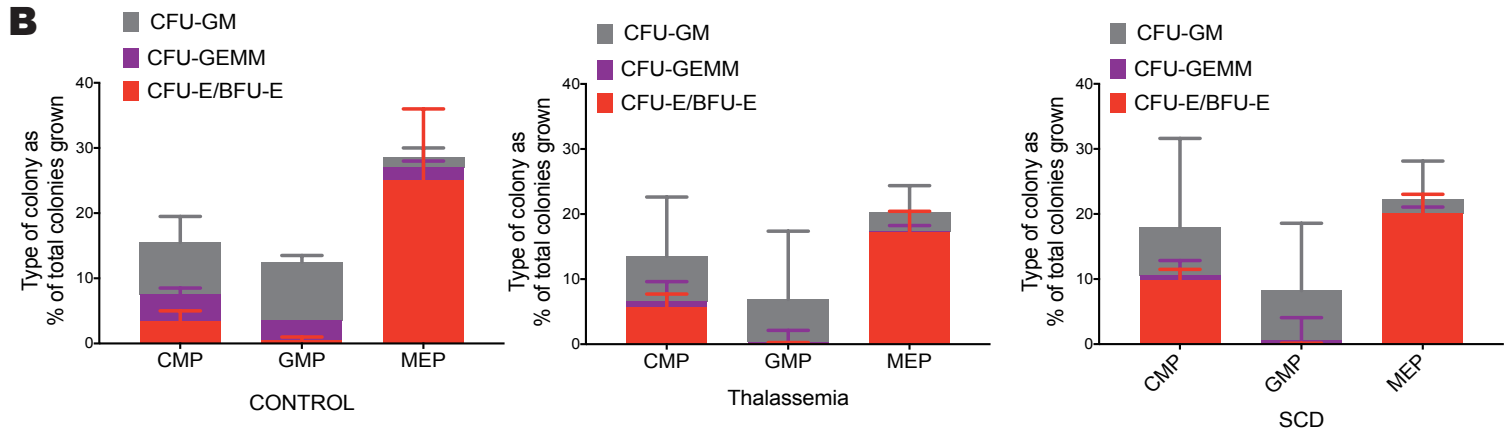
Supplemental Figure 8. A) Flow cytometry plots of CD34 against CD10 gating on total CD34+ cells showing marked expansion of the CD10+CD34+ compartment in children with hemoglobinopathies. **B**) Flow cytometry plots of CD34 against CD38 gating on Lin-CD10-CD34+ cells showing that the CD34+CD38+ fraction was similar in patients and control when the lymphoid progenitor population was excluded from analysis. **C**) Flow plots of Myeloid progenitors (CMP, GMP and MEP compartments) showing these compartments were similar in patients and control when the CD10+ CD34+ lymphoid progenitor population was excluded. **D**) Flow cytometry plots of CD45RA against CD90 gating on Lin-CD10-CD34+CD38low\- cells patients and controls have similar numbers of phenotypic HSCs if the CD34+ lymphoid progenitors are excluded. **E**) Quantitative data of MEP, LMPP, CMP and GMP cells in patients and controls showing significant reductions in the proportion of CMP and MEPs if the CD34+CD10+ lymphoid progenitors are included in the analysis.

Supplemental Figure 8

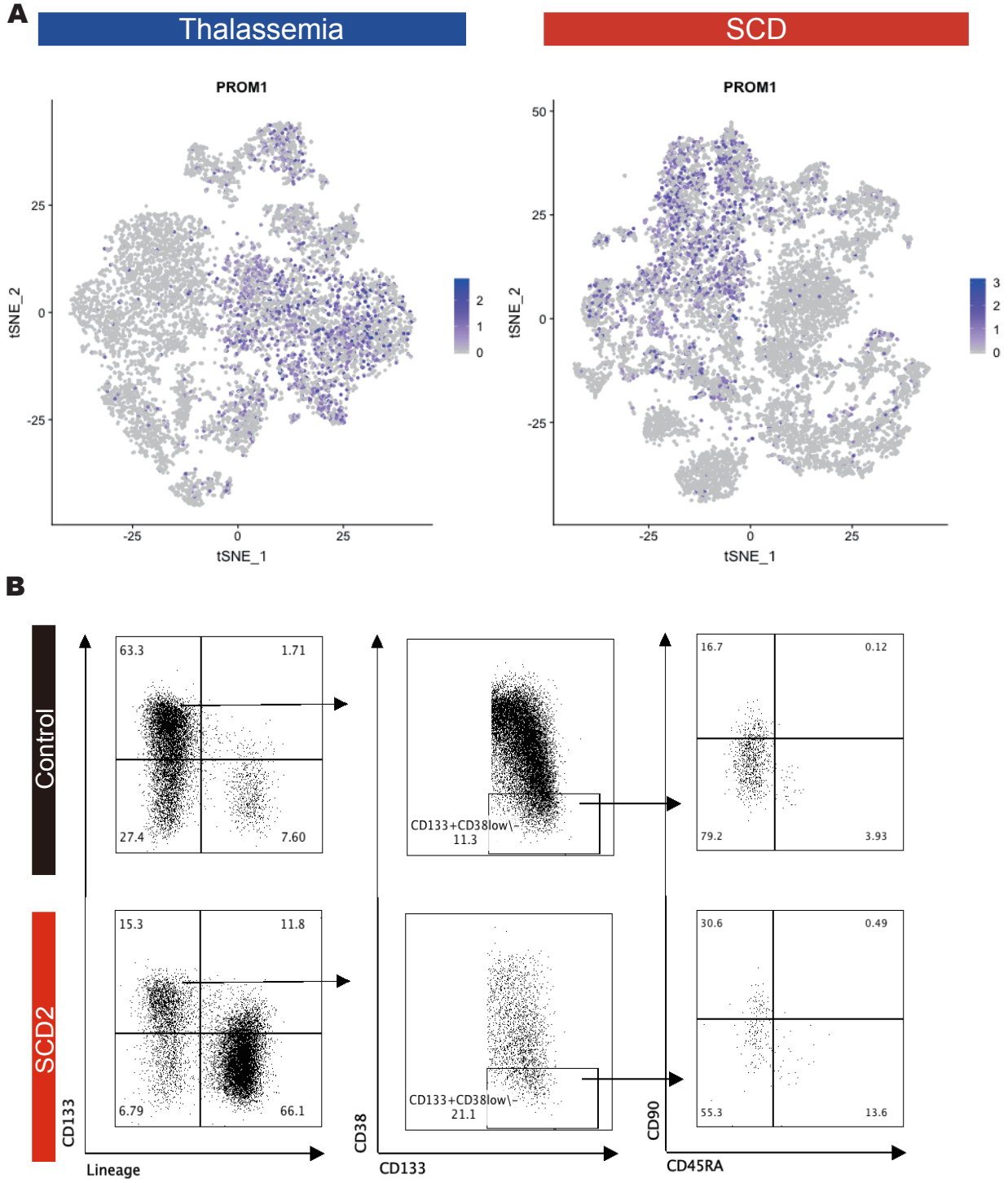


Supplemental Figure 9. A) tSNE plots showing the effects of lineage depletion on 10X data. All CD34+ cells from a control sample are shown in red and data from the same sample that included a lineage depletion are shown in blue. This shows that lymphoid progenitors are specifically excluded by the lineage cocktail. B) Readout of colony forming assays. Cells were FACS sorted based on CD123 and CD45RA and grown on methocult. These data show similar colony forming potential of cells derived from patients and controls.





Supplemental Figure 10. A) tSNE plots of Chromium 10x data showing expression of PROM1 (CD133). This gene is barely detected in the expanded CD34+ lymphoid cells in patients. B) Flow plots showing the gating strategy for defining Lin-CD133+CD34+CD38low\-CD45RA- CD90+ HSCs in control and patient samples.



Supplemental Table 1. Patient information

	Age	Diagnosis	Medication	Hb g/l	Neutrophils $x10^9$/l	Plt $x10^9$/l	Antibody screen	Reason for transplant
BT1	1 year 10 months	Beta thalassaemia major	Desferrioxamine	144	8.7	429	-ve	Quality of life
BT2	4 years 9 months	Beta thalassaemia major	Desferrioxamine	137	5.5	328	Anti-Cw	Quality of life
BT3	4 years 4 months	Beta thalassaemia major	Desferrioxamine	100	4.7	453	-	Quality of life
SCD1	10 years 8 months	HbSS	Hydroxycarbamide	127	2.3	186	Multiple red cell and HLA Abs	Cerebrovascular complications and chest crises
SCD2	6 years 2 months	HbSS	Hydroxycarbamide	98		996	-ve	Frequent VOC and splenic sequestration
SCD3	6 years 11 months	HbSS	Hydroxycarbamide	93	3.1	381	-ve	Recurrent VOC including chest crises
SCD4	7 years 3 months	HbSS		99	1.6	203	Anti-Chido	Recurrent VOC crises
SCD5	6 years 5 months	HbSS	Hydroxycarbamide	105	3.5	265	Anti-HLA	Cerebrovascular complications, chest crises and recurrent VOC
Control 1	53 years							
Control 2	21 years							
Control 3	30 years							
Control 4	41 years							

Supplemental Table 2. QC30 scores of single-cell RNA sequencing:

	Sequencing saturation%	Q30 bases in barcode%	Q30 bases in RNA Read%	Q30 bases in sample Index%	Q30 bases in UMI%	No of cell bar-code sequenced	Mitochondrial genes (%)	Max detected genes	Min detected genes	No of cells passed filter
Ctrl1	69.4	98.5	95.0	97.8	97.0	10792	<5	4000	500	10498
Ctrl2	76.3	98.4	94.6	97.8	97.0	11238	<5	4000	500	10681
Ctrl3	94.0	94.9	89.8	93.2	94.6	5131	<5	4000	500	5014
Ctrl4	93.3	95.1	90.6	92.2	94.8	3013	<5	5000	500	2893
BT1	95.3	95.1	91.2	94.3	94.8	5115	<5	4000	500	4849
BT2	93.7	95.1	90.4	92.9	94.8	4108	<5	4000	500	3760
BT3	89.5	95.1	90.2	93.6	94.7	2788	<5	4000	500	2425
SCD1	94.0	94.8	90.4	92.5	94.5	2966	<5	4000	500	2496
SCD2	92.4	95.1	90.1	87.5	94.7	2337	<5	4000	500	2173
SCD3	95.2	95.1	90.5	91.7	94.8	2321	<5	4000	500	1883
SCD4	93.9	95.1	90.8	92.7	94.7	3192	<5	4000	500	2725
SCD5	97.9	95.1	90.0	91.7	94.7	4874	<5	4000	500	1961

Supplemental Table 3. Genes and transcription factors used to identify cell phenotype in 10X analysis:

M/E_P	EPOR, CA1, HBB, ITGA2B, GATA1, GATA2, APOC1, APOE, TFR2, TFR1, PF4, CNRIP1
G/M_P	ELANE, AZU1, PRTN3, CFD, LYZ
Lym_P	CD79A, CD79B, VPREB1, VPREB3, EBF1
HSPC1	CRHBP, EMCN, CXCL8, UCHL1, HLF, PROCR, C10orf128