nature biotechnology

SUPPLEMENTARY INFORMATION

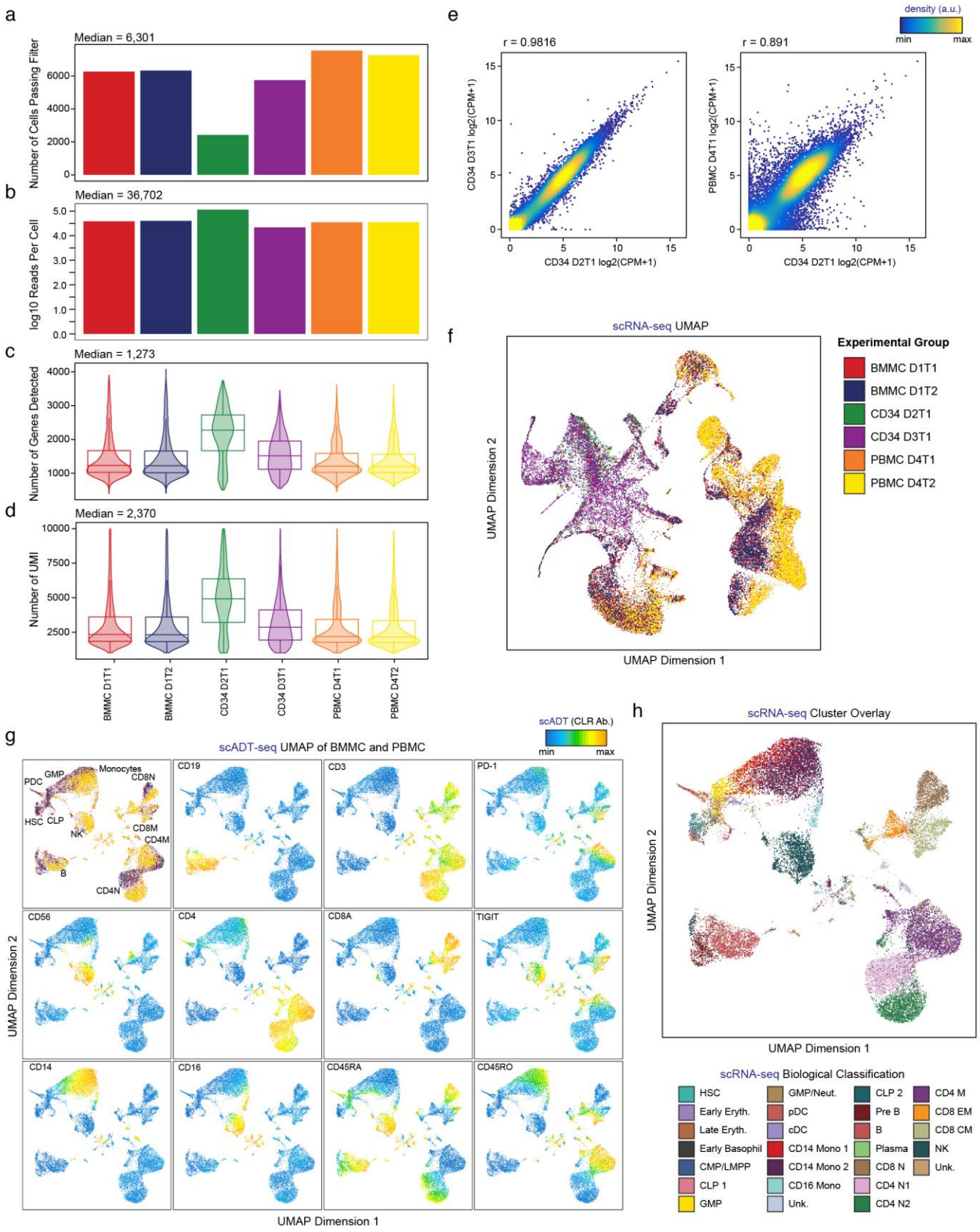
https://doi.org/10.1038/s41587-019-0332-7

In the format provided by the authors and unedited.

Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia

Jeffrey M. Granja[1,2,3,13], Sandy Klemm[3,13]*, Lisa M. McGinnis[3,4,13]*, Arwa S. Kathiria[3], Anja Mezger[3,5], M. Ryan Corces[1,4], Benjamin Parks[3,6], Eric Gars[4], Michaela Liedtke[7], Grace X. Y. Zheng[8], Howard Y. Chang[1,3,9,10], Ravindra Majeti[7] and William J. Greenleaf[1,3,11,12]*

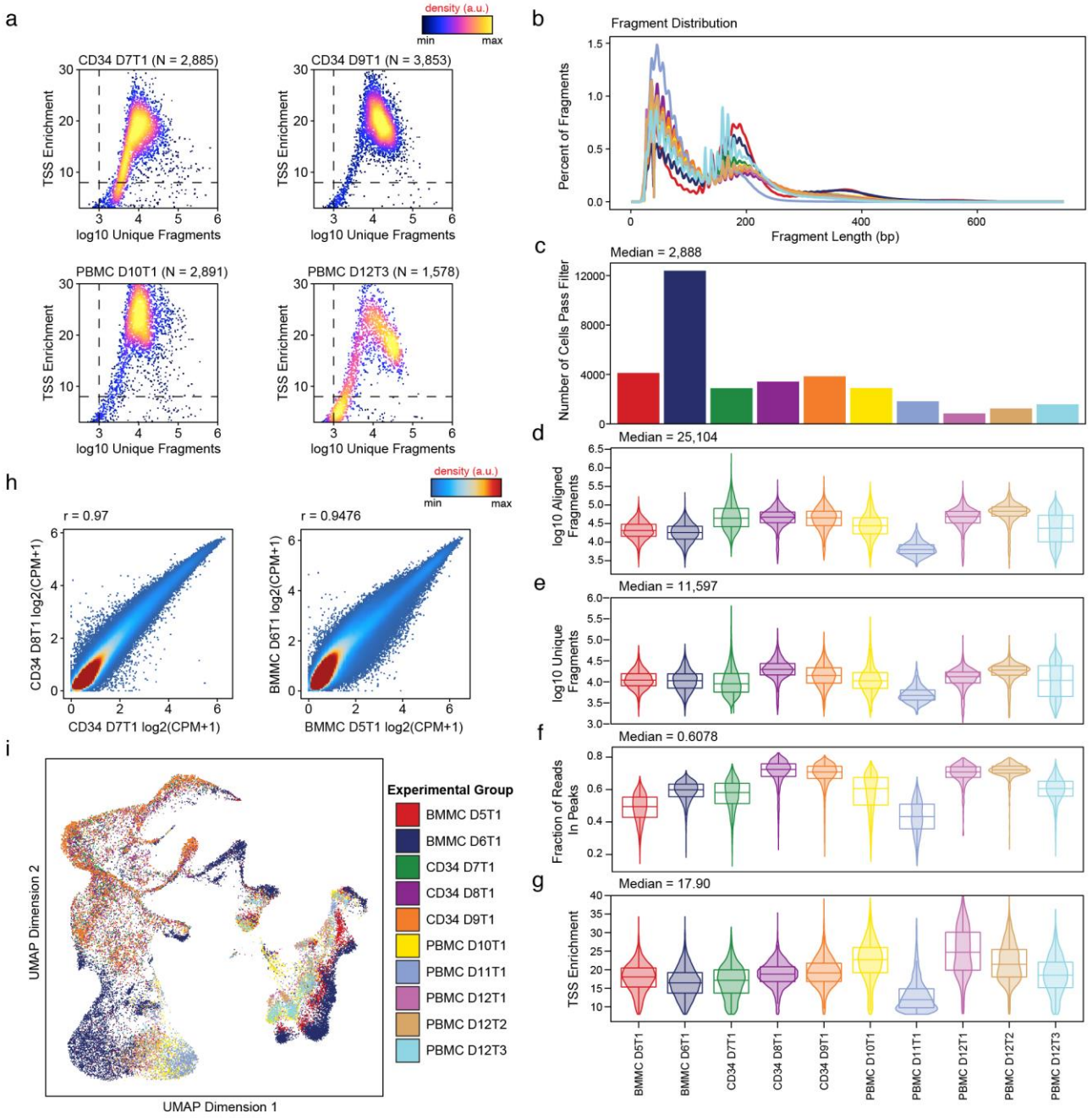
[1]Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA, USA. [2]Biophysics Program, Stanford University School of Medicine, Stanford, CA, USA. [3]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. [4]Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA. [5]Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden. [6]Department of Computer Science, Stanford University School of Engineering, Stanford, CA, USA. [7]Department of Medicine, Division of Hematology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA. [8]10x Genomics, Pleasanton, CA, USA. [9]Department of Dermatology, Stanford University School of Medicine, Redwood City, CA, USA. [10]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA. [11]Department of Applied Physics, Stanford University, Stanford, CA, USA. [12]Chan–Zuckerberg Biohub, San Francisco, CA, USA. [13]These authors contributed equally: Jeffrey M. Granja, Sandy Klemm, Lisa M. McGinnis. *e-mail: klemm@stanford.edu; lisa.mcginnis@stanford.edu; wjg@stanford.edu



Supplementary Figure 1

Quality control of CITE-seq data for hematopoiesis samples.

(a) Number of cells passing filter for each experimental replicate (number of informative genes > 400 and number of unique molecular identifiers (UMI) > 1000). **(b)** Number of aligned reads on average per cell passing filter for each experimental sample. **(c)** Violin and box-whisker plot of the number of informative genes detected per single cell passing filter per experimental sample (n = 2,424 – 7,544). (**d**) Violin and box-whisker plot of the number of unique molecular identified (UMI) transcripts per cell passing filter per experimental sample (n = 2,424 – 7,544). **(e)** Aggregated scRNA-seq (n = 20,287) one to one reproducibility plots for biological replicates (left) and across sample types (right) colored by the density. The correlation (r) represents the Pearson correlation across all genes. **(f)** scRNA-seq experimental sample labels overlay on UMAP of hematopoiesis (n = 35,582). **(g)** scADT-seq UMAP of BMMC and PBMC samples (n = 4) across 14 antibodies. scADT overlay of experimental sample labels, CD19, CD3, CD56, CD4, CD8A, CD14, CD16, CD45RA, CD45RO, TIGIT and PD-1 respectively. Color represents experimental labels or scADT-seq values after CLR transformation. **(h)** scADT-seq UMAP of BMMC and PBMC samples (n = 4) across 14 antibodies colored by scRNA-seq clusters with biological classification. Box-whisker plot; lower whisker is the lowest value greater than the 25% quantile minus 1.5 times the interquartile range (IQR), the lower hinge is the 25% quantile, the middle is the median, the upper hinge is the 75% quantile and the upper whisker is the largest value less than the 75% quantile plus 1.5 times the IQR.

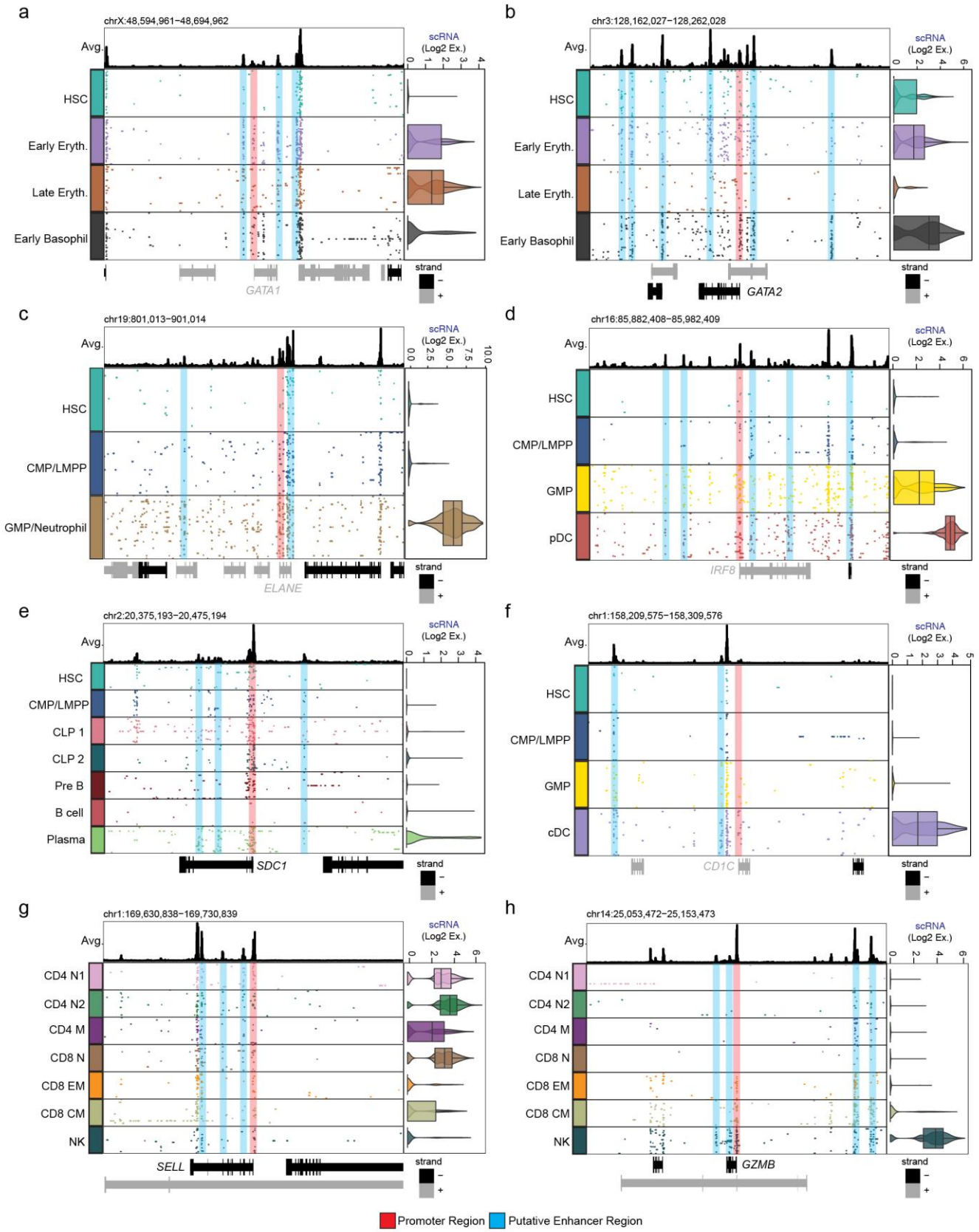


Supplementary Figure 2

Quality control of scATAC-seq data for hematopoiesis samples.

(a) scATAC-seq cell filtering plot of 4 representative scATAC-seq hematopoietic samples. The x-axis is the number of unique accessible fragments and the y-axis is the enrichment of Tn5 insertions at transcription start sites, representing the robust signal to background for each single cell. **(b)** Aggregated scATAC-seq fragment size distributions across individual experiments demonstrating sub-, mono- and multi nucleosome spanning ATAC-seq fragments. **(c)** Number of cells passing filter for each experimental replicate (Unique nuclear fragments > 1000 and TSS enrichment > 8). **(d)** Violin and box-whisker plot of the number of total aligned fragments for each single cell passing filter per experimental sample (n = 836 – 12,394). **(e)** Violin and box-whisker plot of the number of unique aligned nuclear fragments for each single cell passing filter per experimental sample (n = 836 – 12,394). **(f)** Violin and box-whisker plot of the fraction of the total number Tn5 insertions (Reads) that are within the healthy hematopoietic union peak set (n = 452,004) for each single cell passing filter. **(g)** Violin and box-whisker plot of the normalized transcription start site (TSS) enrichment for each single cell passing filter per experimental sample. **(h)** Aggregated scATAC-seq (n = 452,004) one to one reproducibility plots for biological

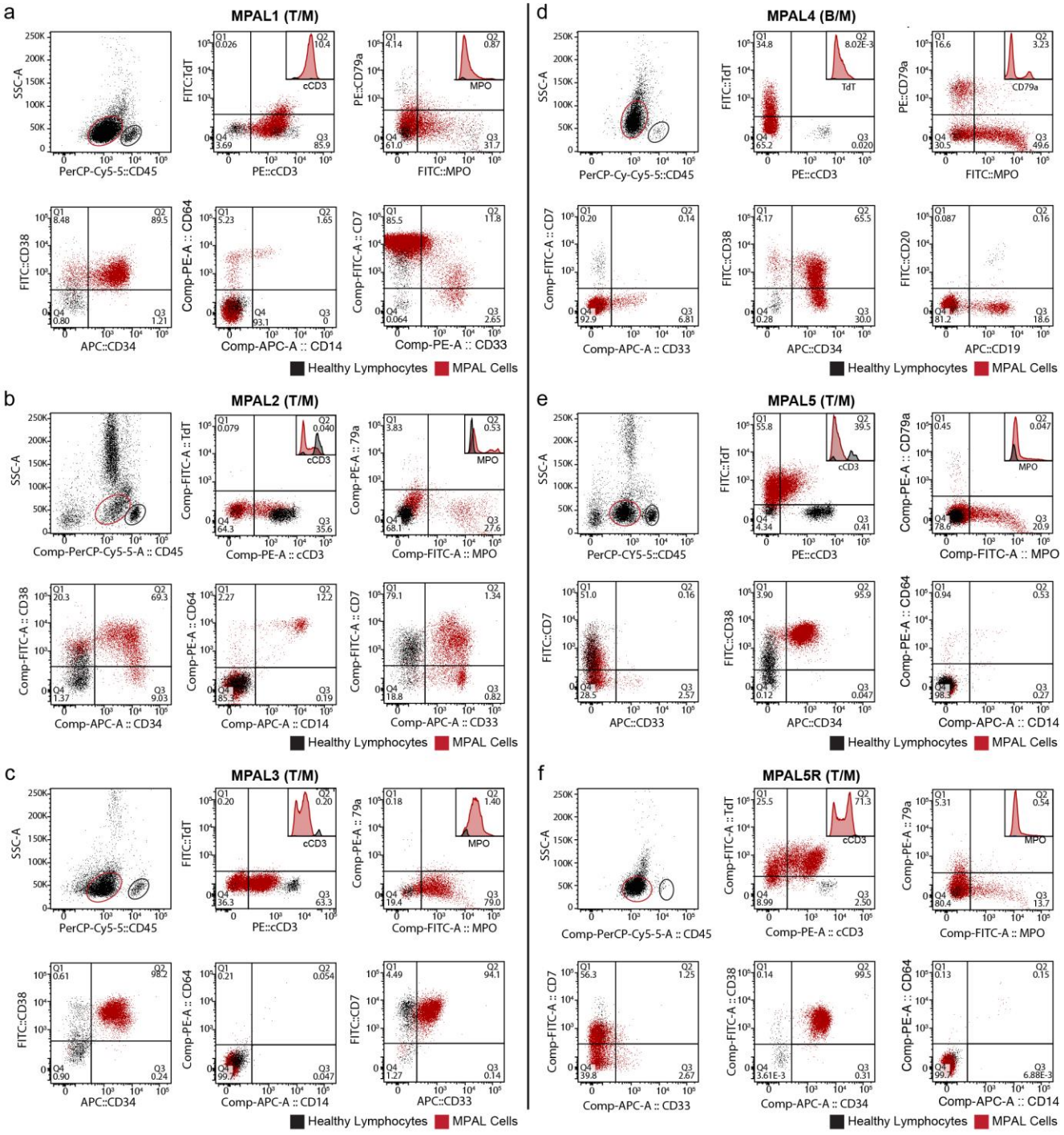
replicates colored by the density. The correlation (r) represents the Pearson correlation across all genes. **(i)** scATAC-seq experimental sample labels overlay on UMAP of hematopoiesis (n = 35,038). Box-whisker plot; lower whisker is the lowest value greater than the 25% quantile minus 1.5 times the interquartile range (IQR), the lower hinge is the 25% quantile, the middle is the median, the upper hinge is the 75% quantile and the upper whisker is the largest value less than the 75% quantile plus 1.5 times the IQR.


a
chrX:48,594,961−48,694,962
Avg.
scRNA (Log2 Ex.)
HSC
Early Eryth.
Late Eryth.
Early Basophil
strand
GATA1
b
chr3:128,162,027−128,262,028
GATA2
c
chr19:801,013−901,014
CMP/LMPP
GMP/Neutrophil
ELANE
d
chr16:85,882,408−85,982,409
GMP
pDC
IRF8
e
chr2:20,375,193−20,475,194
CLP 1
CLP 2
Pre B
B cell
Plasma
SDC1
f
chr1:158,209,575−158,309,576
cDC
CD1C
g
chr1:169,630,838−169,730,839
CD4 N1
CD4 N2
CD4 M
CD8 N
CD8 EM
CD8 CM
NK
SELL
h
chr14:25,053,472−25,153,473
GZMB
Promoter Region
Putative Enhancer Region

Supplementary Figure 3

Validation of key marker genes for both scRNA-seq and scATAC-seq for hematopoiesis.

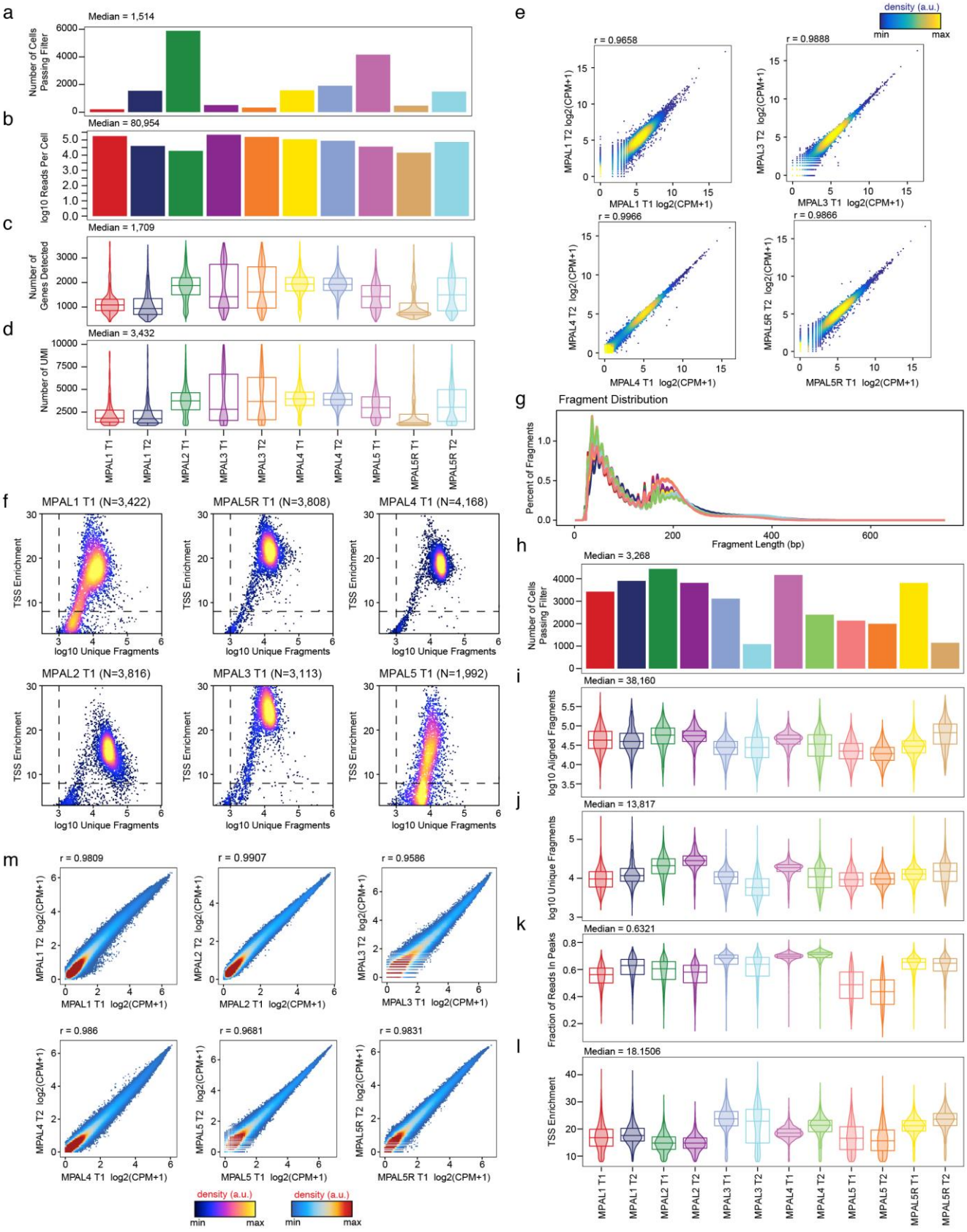
(a-h) Multi-omic tracks; (Top) average track of all clusters displayed, (Middle) binarized 100 random scATAC-seq tracks for each locus at 100bp resolution and (right) violin and box-whisker plot of the scRNA-seq log2 normalized expression for each cluster. Box-whisker plot; lower whisker is the lowest value greater than the 25% quantile minus 1.5 times the interquartile range (IQR), the lower hinge is the 25% quantile, the middle is the median, the upper hinge is the 75% quantile and the upper whisker is the largest value less than the 75% quantile plus 1.5 times the IQR. **(a)** Multi-omic track of *GATA1* (specific in these clusters for Erythroid) for erythroid development from HSC progenitor cells (n = 111 - 1,653). **(b)** Multi-omic track of *GATA2* (specific in these clusters for Basophil) for erythroid development from HSC progenitor cells (n = 111 - 1,653). **(c)** Multi-omic track of *ELANE* (specific in these clusters for GMP/Neutrophil) for neutrophil development from HSC progenitor cells (n = 1,050 – 2,260). **(d)** Multi-omic track of *IRF8* (specific in these clusters for pDC) across pDC development from HSC progenitor cells (n = 544 – 2,260). **(e)** Multi-omic track of *SDC1* (specific in these clusters for Plasma cells) across B cell development and plasma cells (n = 62 – 2,260). **(f)** Multi-omic track of *CD1C* (specific in these clusters for cDC) across cDC development from HSC progenitor cells (n = 325 – 2,260). **(g)** Multi-omic track of *SELL* (specific in these clusters for Naive T cells vs memory, and CD8 central memory vs CD8 effector memory) across NK and T cells (n = 796 – 3,539). **(h)** Multi-omic track of *GZMB* (specific in these clusters for NK cells) across NK and T cells (n = 796 – 3,539).


a
MPAL1 (T/M)
SSC-A
PerCP-Cy5-5::CD45
FITC::TdT
PE::cCD3
cCD3
PE::CD79a
FITC::MPO
MPO
FITC::CD38
APC::CD34
Comp-PE-A :: CD64
Comp-APC-A :: CD14
Comp-FITC-A :: CD7
Comp-PE-A :: CD33
Healthy Lymphocytes
MPAL Cells
b
MPAL2 (T/M)
Comp-PerCP-Cy5-5-A :: CD45
Comp-FITC-A :: TdT
Comp-PE-A :: cCD3
Comp-PE-A :: 79a
Comp-FITC-A :: MPO
Comp-FITC-A :: CD38
Comp-APC-A :: CD34
Comp-PE-A :: CD64
Comp-APC-A :: CD14
Comp-FITC-A :: CD7
Comp-APC-A :: CD33
c
MPAL3 (T/M)
FITC::TdT
PE::cCD3
Comp-PE-A :: 79a
Comp-FITC-A :: MPO
FITC::CD38
APC::CD34
Comp-PE-A :: CD64
Comp-APC-A :: CD14
FITC::CD7
APC::CD33
d
MPAL4 (B/M)
FITC::TdT
PE::cCD3
TdT
PE::CD79a
CD79a
FITC::MPO
Comp-FITC-A :: CD7
Comp-APC-A :: CD33
FITC::CD38
APC::CD34
FITC::CD20
APC::CD19
e
MPAL5 (T/M)
PerCP-CY5-5::CD45
FITC::TdT
PE::cCD3
Comp-PE-A :: CD79a
Comp-FITC-A :: MPO
FITC::CD7
APC::CD33
FITC::CD38
APC::CD34
Comp-PE-A :: CD64
Comp-APC-A :: CD14
f
MPAL5R (T/M)
Comp-PerCP-Cy5-5-A :: CD45
Comp-FITC-A :: TdT
Comp-PE-A :: cCD3
Comp-PE-A :: 79a
Comp-FITC-A :: MPO
Comp-FITC-A :: CD7
Comp-APC-A :: CD33
Comp-FITC-A :: CD38
Comp-APC-A :: CD34
Comp-PE-A :: CD64
Comp-APC-A :: CD14
Healthy Lymphocytes
MPAL Cells
g
Whole Exome Sequencing
MPAL1
MPAL2
MPAL3
MPAL4
MPAL5
ABCC9 ATRX ATRX BCR BRIP1 CCND3 CEBPA CSF2RA CSF3R CTCF CTCF CYLD DNMT3A ECT2L EED EGR2 EP300 ETV6 FAM46C FANCB FANCD2 FANCI FAT1 FAT3 FBXW7 IDH1 IKZF2 IKZF3 JAK3 LAMB4 LRP1B MED12 MGA NCOR2 NOTCH1 NOTCH1 NOTCH2 NRAS PIK3CB POU2AF1 RAD21 RIPK1 RUNX1 RUNX2 SETD2 SH2B3 SPEN TET2 TET2 TLR2 TNFAIP3 TP53 USP29 WHSC1L1 WT1
Missense
Frameshift deletion
Stopgain
Frameshift insertion
Nonframeshift deletion

Supplementary Figure 4

Diagnostic flow cytometry plots for MPALs 1-5R.

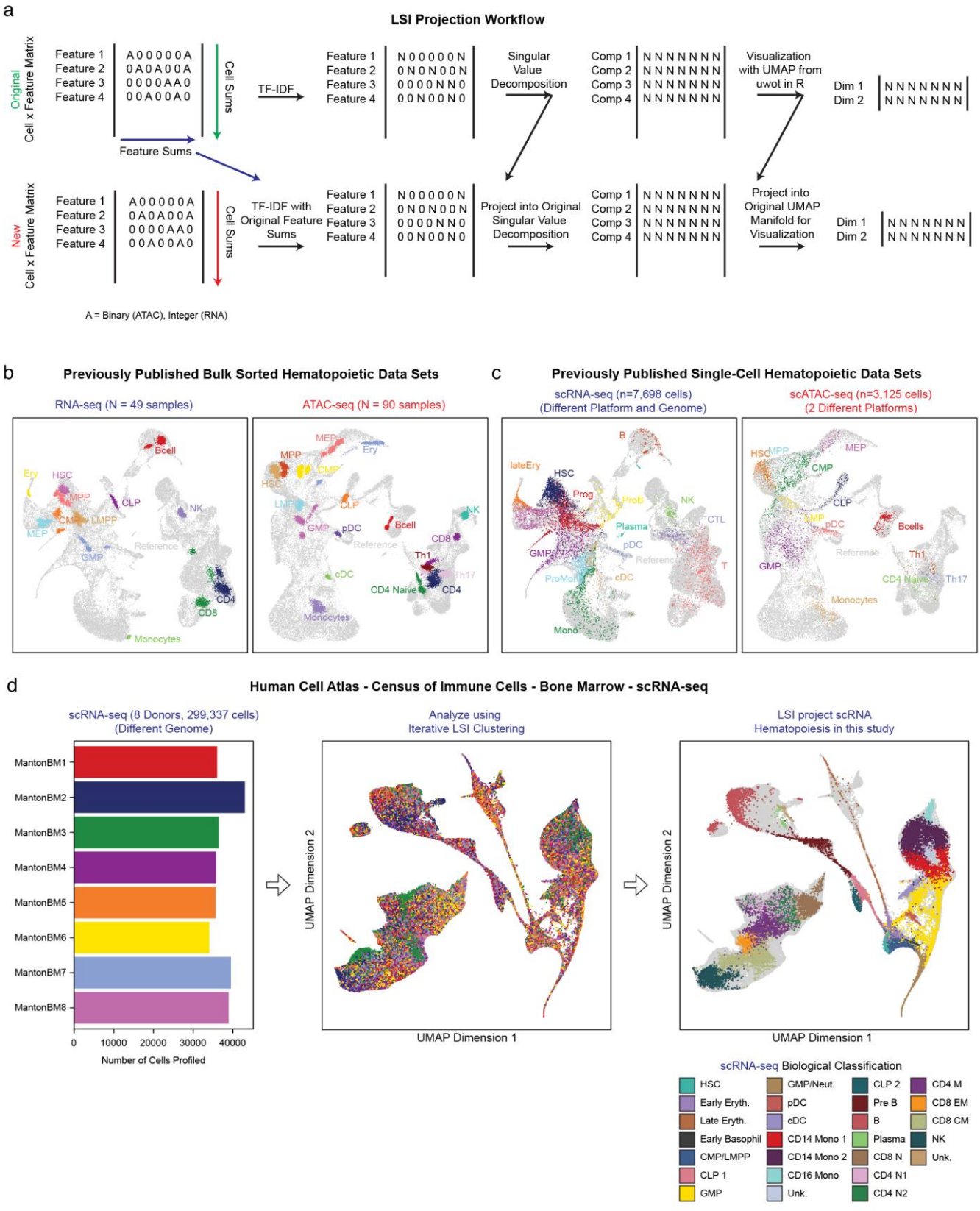
(a-f) Diagnostic flow cytometry plots from the 5 different MPAL cases (MPAL1-5R) gated on blasts area (highlighted in red) and lymphocytes (highlighted in black) from CD45 and side scatter area (SSC-A). **(a)** MPAL 1 shows classic bilineal phenotype with both T-lymphoblasts (cCD3-positive and CD7-positve) and myeloid blasts (MPO-positive and CD33-positive). **(b)** MPAL 2 demonstrates a more complex phenotype with both biphenotypic (single population expressing lymphoid marker CD7 and myeloid marker CD33) and bilineal T-Myeloid patterns (subpopulation expressing monocytic markers CD64, CD33, and CD14). **(c)** MPAL 3 demonstrates a classic biphenotypic case with coexpression of both T-lineage markers (cCD3-positive) and myeloid markers (MPO-positive). **(d)** MPAL4 demonstrates a classic bilineal B/M phenotype expressing B-lineage markers (CD79a and CD19-positive) and myeloid markers (MPO-positive and CD33-positive). **(e)** MPAL5 demonstrates a more complicated phenotype with a subpopulation of blasts expressing T-lineage markers (cCD3-positive and CD7-positive) and a subpopulation expressing myeloid marker MPO. **(f)** MPAL5R post-treatment relapse of MPAL5. Flow cytometry reveals expansion of the T-lymphoblastic subpopulation (cCD3-positive, TdT-positive population) following chemotherapy. **(g)** High-confidence mutations detected in 5 MPAL cases by whole exome sequencing. Missense mutations are shown in blue, frameshift deletions are shown in yellow, stop-gain mutations are shown in purple, frameshift insertions are shown in orange, and nonframeshift deletions are shown in dark gray.


a
Median = 1,514
Number of Cells Passing Filter
b
Median = 80,954
log10 Reads Per Cell
c
Median = 1,709
Number of Genes Detected
d
Median = 3,432
Number of UMI
MPAL1 T1
MPAL1 T2
MPAL2 T1
MPAL3 T1
MPAL3 T2
MPAL4 T1
MPAL4 T2
MPAL5 T1
MPAL5R T1
MPAL5R T2
e
density (a.u.)
min
max
r = 0.9658
r = 0.9888
r = 0.9966
r = 0.9866
MPAL1 T2 log2(CPM+1)
MPAL1 T1 log2(CPM+1)
MPAL3 T2 log2(CPM+1)
MPAL3 T1 log2(CPM+1)
MPAL4 T2 log2(CPM+1)
MPAL4 T1 log2(CPM+1)
MPAL5R T2 log2(CPM+1)
MPAL5R T1 log2(CPM+1)
f
MPAL1 T1 (N=3,422)
MPAL5R T1 (N=3,808)
MPAL4 T1 (N=4,168)
MPAL2 T1 (N=3,816)
MPAL3 T1 (N=3,113)
MPAL5 T1 (N=1,992)
TSS Enrichment
log10 Unique Fragments
g
Fragment Distribution
Percent of Fragments
Fragment Length (bp)
h
Median = 3,268
Number of Cells Passing Filter
i
Median = 38,160
log10 Aligned Fragments
j
Median = 13,817
log10 Unique Fragments
k
Median = 0.6321
Fraction of Reads In Peaks
l
Median = 18.1506
TSS Enrichment
m
r = 0.9809
r = 0.9907
r = 0.9586
r = 0.986
r = 0.9681
r = 0.9831
MPAL1 T2 log2(CPM+1)
MPAL1 T1 log2(CPM+1)
MPAL2 T2 log2(CPM+1)
MPAL2 T1 log2(CPM+1)
MPAL3 T2 log2(CPM+1)
MPAL3 T1 log2(CPM+1)
MPAL4 T2 log2(CPM+1)
MPAL4 T1 log2(CPM+1)
MPAL5 T2 log2(CPM+1)
MPAL5 T1 log2(CPM+1)
MPAL5R T2 log2(CPM+1)
MPAL5R T1 log2(CPM+1)
density (a.u.)
min
max
density (a.u.)
min
max

Supplementary Figure 5

Quality control of scRNA-seq and scATAC-seq data for MPAL samples.

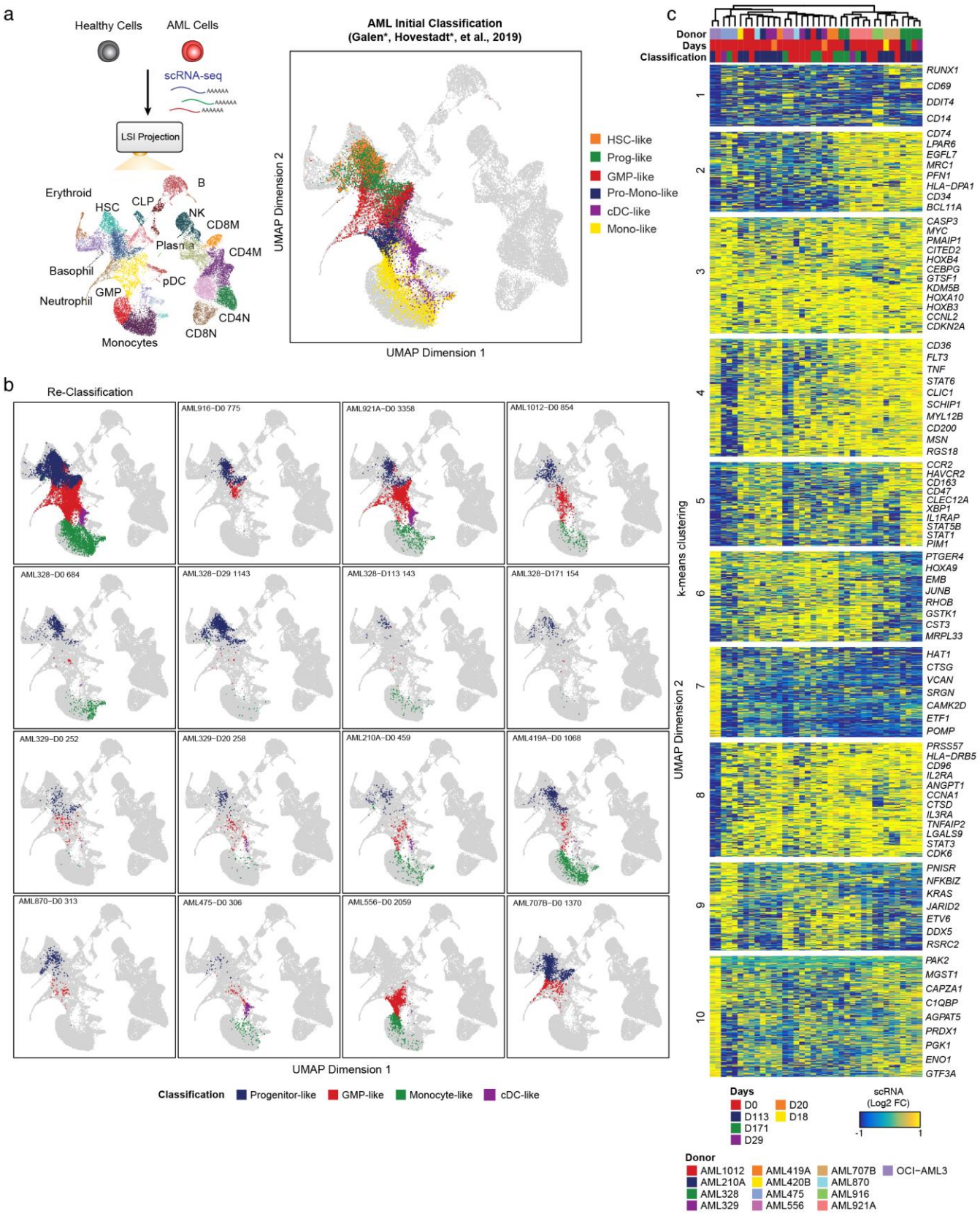
(a) Number of cells passing filter for each experimental replicate (number of informative genes > 400 and number of unique molecular identifiers (UMI) > 1000). **(b)** Number of aligned reads on average per cell passing filter for each experimental sample. **(c)** Violin and box-whisker plot of the number of informative genes detected per single cell passing filter per experimental sample. **(d)** Violin and box-whisker plot of the number of unique molecular identified (UMI) transcripts per cell passing filter per experimental sample. **(e)** Aggregated scRNA-seq (n = 20,287) one to one reproducibility plots for technical replicates colored by the density. The correlation (r) represents the Pearson correlation across all genes. **(f)** scATAC-seq cell filtering plot of 6 representative scATAC-seq MPAL samples. The x-axis is the number of unique accessible fragments and the y-axis is the enrichment of Tn5 insertions at transcription start sites, representing the robust signal to background for each single cell. **(g)** Aggregated scATAC-seq fragment size distributions across individual experiments demonstrating sub-, mono- and multi nucleosome spanning ATAC-seq fragments. **(h)** Number of cells passing filter for each experimental replicate (Unique nuclear fragments > 1000 and TSS enrichment > 8). **(i)** Violin and box-whisker plot of the number of total aligned fragments for each single cell passing filter per experimental sample. **(j)** Violin and box-whisker plot of the number of unique aligned nuclear fragments for each single cell passing filter per experimental sample. **(k)** Violin and box-whisker plot of the fraction of the total number Tn5 insertions (Reads) that are within the MPAL union peak set (n = 346,274) for each single cell passing filter. **(l)** Violin and box-whisker plot of the normalized transcription start site (TSS) enrichment for each single cell passing filter per experimental sample. **(m)** Aggregated scATAC-seq (n = 346,274) one to one reproducibility plots for technical replicates colored by the density. The correlation (r) represents the Pearson correlation across all genes. Box-whisker plot; lower whisker is the lowest value greater than the 25% quantile minus 1.5 times the interquartile range (IQR), the lower hinge is the 25% quantile, the middle is the median, the upper hinge is the 75% quantile and the upper whisker is the largest value less than the 75% quantile plus 1.5 times the IQR.


a
LSI Projection Workflow
Original
Cell x Feature Matrix
Feature 1
Feature 2
Feature 3
Feature 4
A 0 0 0 0 0 A
0 A 0 A 0 0 A
0 0 0 0 A A 0
0 0 A 0 0 A 0
Cell Sums
Feature Sums
TF-IDF
N 0 0 0 0 0 N
0 N 0 N 0 0 N
0 0 0 0 N N 0
0 0 N 0 0 N 0
Singular Value Decomposition
Comp 1
Comp 2
Comp 3
Comp 4
N N N N N N N
Visualization with UMAP from uwot in R
Dim 1
Dim 2
N N N N N N N
N N N N N N N
New
Cell x Feature Matrix
TF-IDF with Original Feature Sums
Project into Original Singular Value Decomposition
Project into Original UMAP Manifold for Visualization
A = Binary (ATAC), Integer (RNA)
b
Previously Published Bulk Sorted Hematopoietic Data Sets
RNA-seq (N = 49 samples)
ATAC-seq (N = 90 samples)
Bcell
Ery
HSC
MPP
CLP
NK
CMP
LMPP
MEP
GMP
Reference
CD4
CD8
Monocytes
MEP
Ery
MPP
CMP
HSC
LMPP
CLP
NK
GMP
pDC
Bcell
CD8
Th1
Th17
cDC
CD4 Naive
CD4
Monocytes
c
Previously Published Single-Cell Hematopoietic Data Sets
scRNA-seq (n=7,698 cells)
(Different Platform and Genome)
scATAC-seq (n=3,125 cells)
(2 Different Platforms)
B
lateEry
HSC
Prog
ProB
NK
CTL
Plasma
pDC
GMP
Reference
T
ProMono
cDC
Mono
HSC
MPP
MEP
CMP
CLP
LMPP
pDC
Bcells
GMP
Reference
Th1
CD4 Naive
Th17
Monocytes
d
Human Cell Atlas - Census of Immune Cells - Bone Marrow - scRNA-seq
scRNA-seq (8 Donors, 299,337 cells)
(Different Genome)
MantonBM1
MantonBM2
MantonBM3
MantonBM4
MantonBM5
MantonBM6
MantonBM7
MantonBM8
0
10000
20000
30000
40000
Number of Cells Profiled
Analyze using
Iterative LSI Clustering
UMAP Dimension 2
UMAP Dimension 1
LSI project scRNA
Hematopoiesis in this study
UMAP Dimension 2
UMAP Dimension 1
scRNA-seq Biological Classification
HSC
Early Eryth.
Late Eryth.
Early Basophil
CMP/LMPP
CLP 1
GMP
GMP/Neut.
pDC
cDC
CD14 Mono 1
CD14 Mono 2
CD16 Mono
Unk.
CLP 2
Pre B
B
Plasma
CD8 N
CD4 N1
CD4 N2
CD4 M
CD8 EM
CD8 CM
NK
Unk.

Supplementary Figure 6

Evaluation of LSI projection workflow for previously published bulk and single-cell hematopoietic data sets across different platforms.

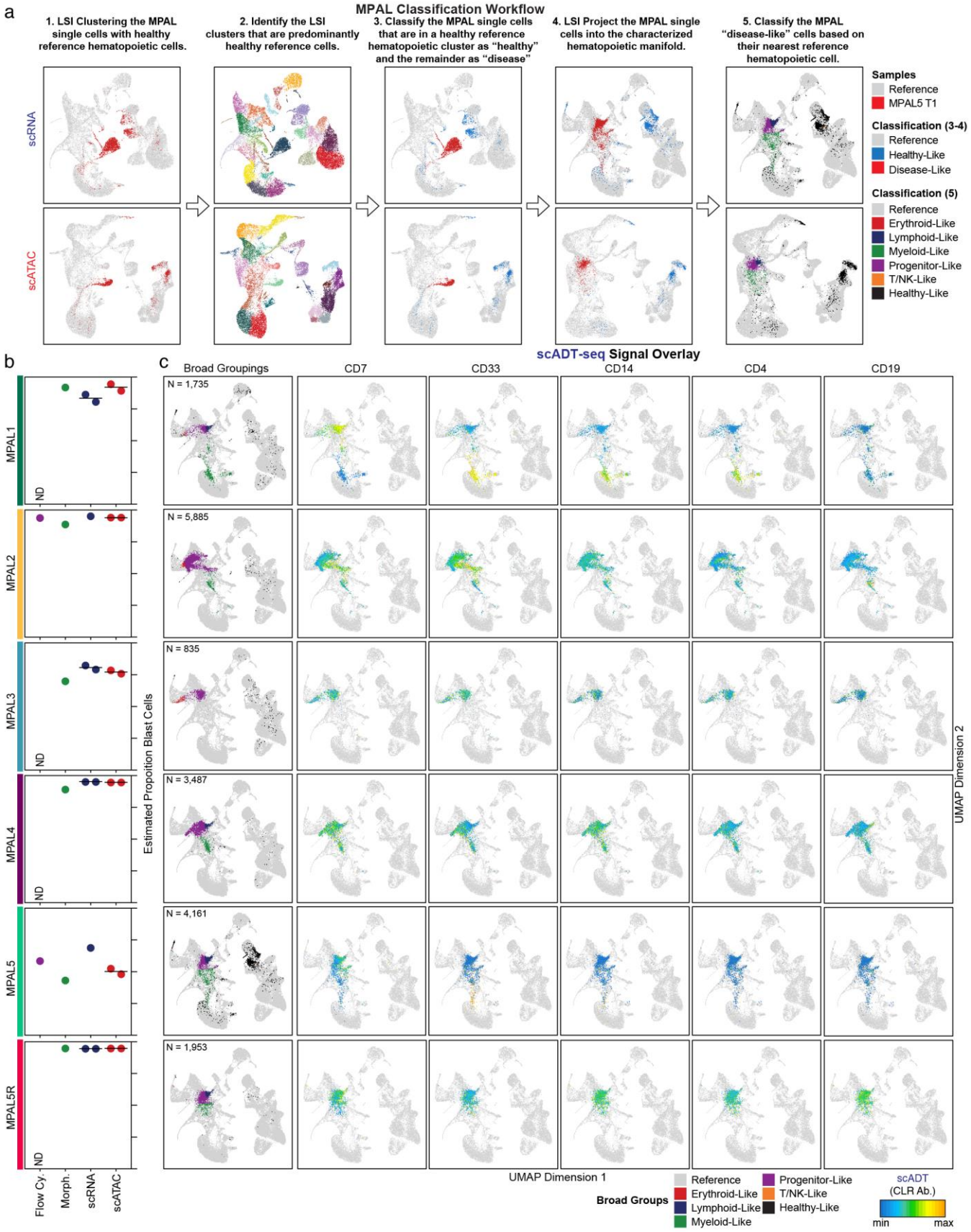
(**a)** Overview of LSI projection workflow. Briefly, when previously computing the TF-IDF transform for the original hematopoietic manifold, we store the feature sums (document frequency) and use this information to compute the TF-IDF transform for the new matrix. We then use the hematopoietic singular value decomposition loadings on the new TF-IDF matrix to project the new matrix into a common subspace. To visualize the hematopoietic subspace, we constructed a UMAP projection using uwot in R. We then used this learned manifold to project the new matrix subspace into the original UMAP projection. **(b)** LSI projection of downsampled previously published bulk sorted hematopoietic data sets[18,20]. (Left) RNA-seq downsampled bulk projections for 49 samples (n = 250 downsampled cells). (Right) ATAC-seq downsampled bulk projections for 90 samples (n = 250 downsampled cells). **(c)** LSI projection of downsampled previously published single-cell hematopoietic data sets labeled by previous classifications[20–22]. (Left) scRNA-seq projections of previous study healthy bone marrow cells (different platform and different aligned genome) colored by previous classifications. (Right) scATAC-seq projections for healthy bone marrow and peripheral blood samples (2 different platforms across 3 studies), colored by ground truth isolated populations. **(d)** Projection of hematopoietic scRNA-seq into of Human Cell Atlas (HCA) Census of Immune Cells. (Left) Number of cells per each of 8 bone marrow donors. (Middle) UMAP projection of LSI iterative clustering of HCA bone marrow scRNA-seq. (Right) LSI projection of our scRNA-seq hematopoietic single cells into HCA bone marrow UMAP colored by cluster definitions.



Supplementary Figure 7

LSI projection of previously published healthy and AML scRNA-seq identifies malignant programs across AML subpopulations.

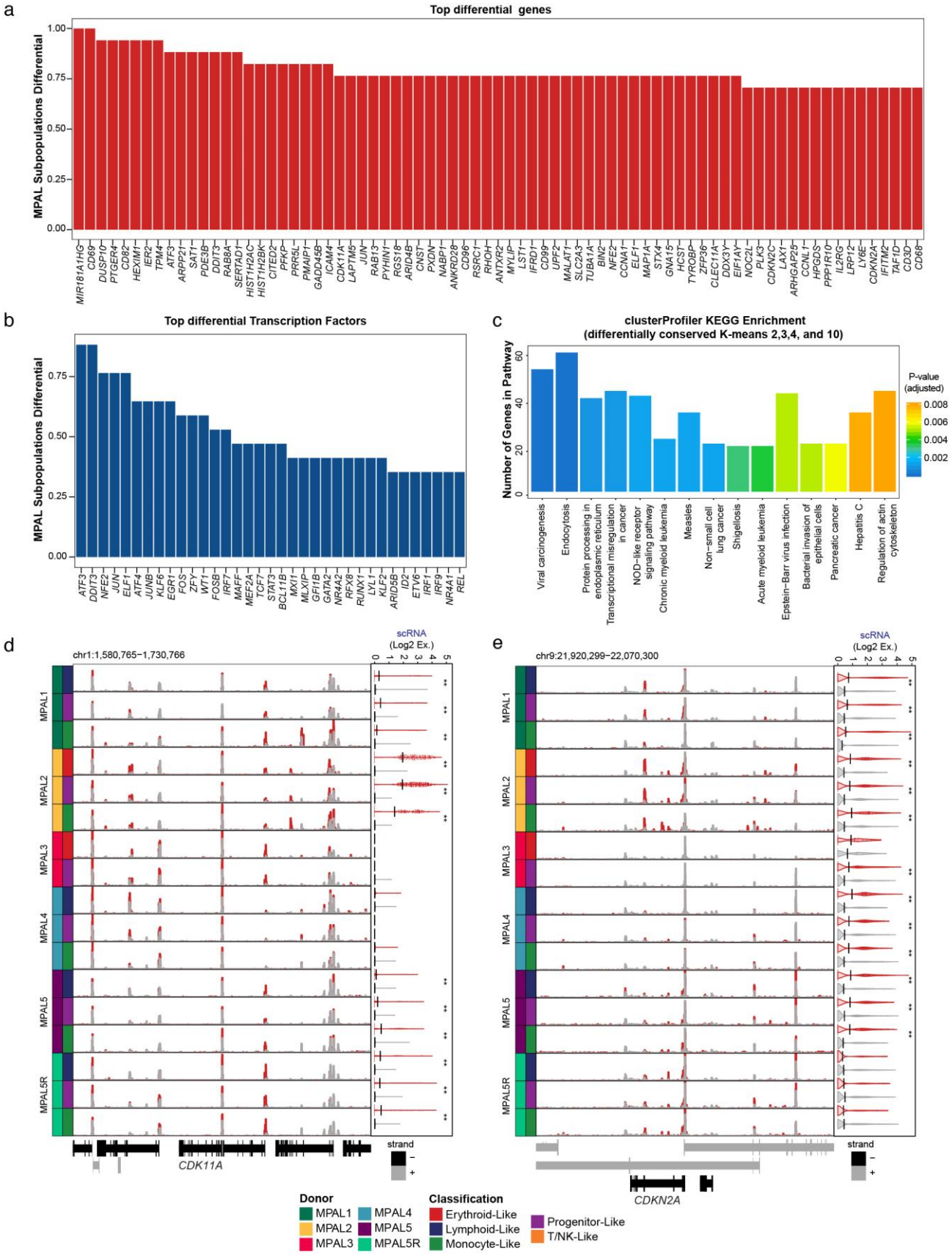
(a) (Left) Schematic of LSI projection. (Right) Initial projection of all AML malignant single-cells colored by previous classifications[19] (n = 31,890). **(b)** Re-classification of scRNA-seq AML single-cells based on closest normal cells in healthy hematopoiesis (See Methods). Broader re-classification increases the number of cells per category for improved power in differential analyses. LSI projection for each individual AML samples onto scRNA-seq healthy hematopoiesis colored by re-classifications (denoted is the sample id and number of cells, n = 143 – 3,358). **(c)** K-means differential scRNA-seq heatmap (k = 10), colored by log2 fold change, comparing each AML sample subpopulations (classifications) vs their closest normal bone marrow cells from the same study[19].


a
MPAL Classification Workflow
1. LSI Clustering the MPAL single cells with healthy reference hematopoietic cells.
2. Identify the LSI clusters that are predominantly healthy reference cells.
3. Classify the MPAL single cells that are in a healthy reference hematopoietic cluster as "healthy" and the remainder as "disease"
4. LSI Project the MPAL single cells into the characterized hematopoietic manifold.
5. Classify the MPAL "disease-like" cells based on their nearest reference hematopoietic cell.
scRNA
scATAC
Samples
Reference
MPAL5 T1
Classification (3-4)
Reference
Healthy-Like
Disease-Like
Classification (5)
Reference
Erythroid-Like
Lymphoid-Like
Myeloid-Like
Progenitor-Like
T/NK-Like
Healthy-Like
b
MPAL1
MPAL2
MPAL3
MPAL4
MPAL5
MPAL5R
ND
Flow Cy.
Morph.
scRNA
scATAC
Estimated Propoition Blast Cells
c
scADT-seq Signal Overlay
Broad Groupings
CD7
CD33
CD14
CD4
CD19
N = 1,735
N = 5,885
N = 835
N = 3,487
N = 4,161
N = 1,953
UMAP Dimension 1
UMAP Dimension 2
Broad Groups
Reference
Erythroid-Like
Lymphoid-Like
Myeloid-Like
Progenitor-Like
T/NK-Like
Healthy-Like
scADT
(CLR Ab.)
min
max

Supplementary Figure 8

Classification of MPALs by projection onto the hematopoietic hierarchy.

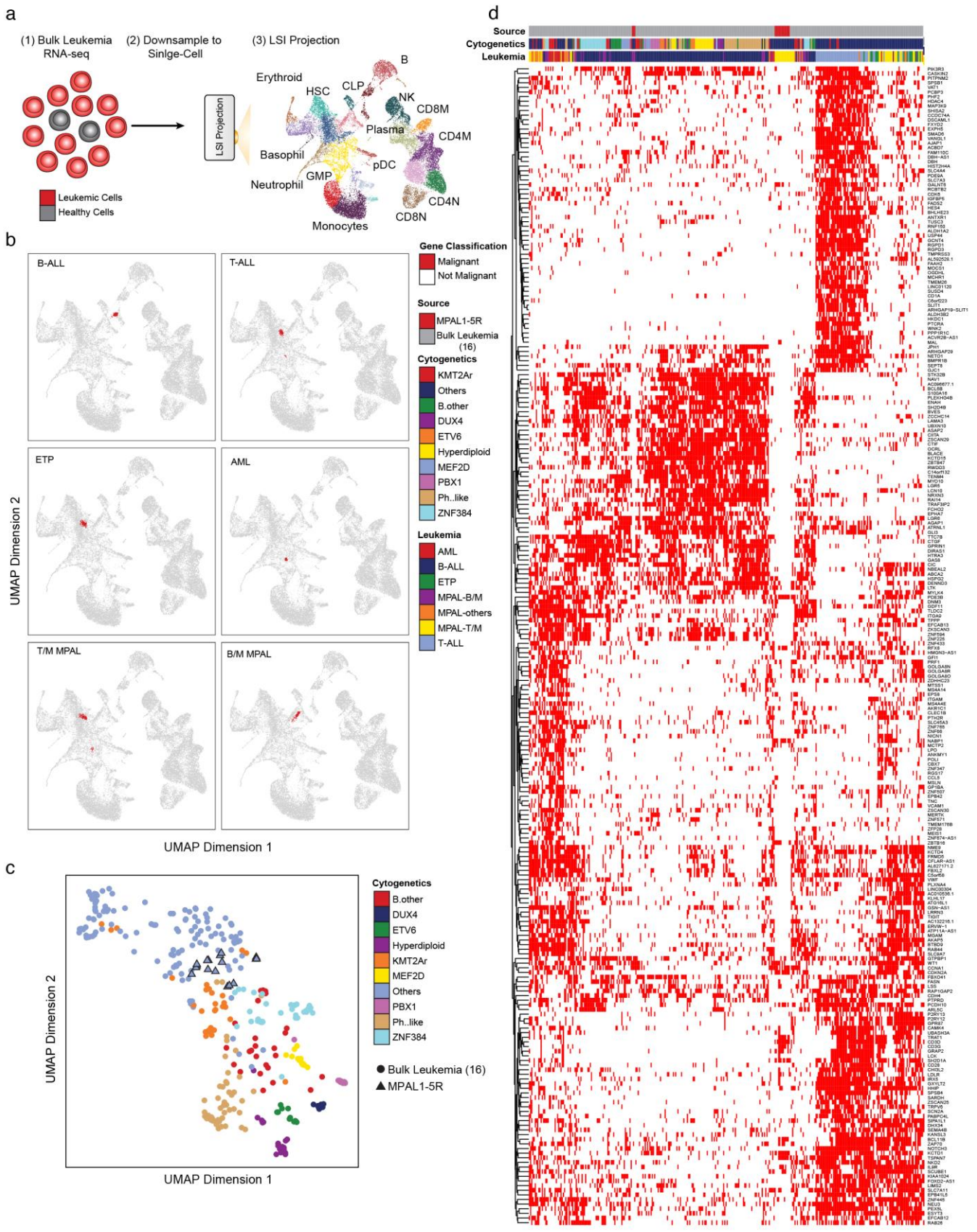
(a) MPAL single cell classification workflow for scRNA and scATAC-seq. First cells were clustered with reference hematopoietic cells (1) and classified based on healthy hematopoietic clusters (2-3). Cells were then LSI projected into the reference hematopoietic manifold (4) and then classified based on the nearest reference cell hematopoietic compartment (5). MPAL 5 replicate 1 is shown as an example for scRNA (Top) and scATAC-seq (Bottom). **(b)** Proportion of estimated blast cells for each MPAL with Flow Cytometry, Morphology, scRNA and scATAC-seq (Range is from 0 to 1). **(c)** (Left) Projected MPALs colored by hematopoietic compartments as described in **a**. (Right) scADT-seq overlay of CD7, CD33, CD14, CD4 and CD19 on MPAL single cells LSI projected onto hematopoiesis.


a
Top differential genes
MPAL Subpopulations Differential
1.00
0.75
0.50
0.25
0.00
MIR181A1HG CD69 DUSP10 PTGER4 CD82 HEXIM1 IER2 TPM4 ATF3 ARPP21 SAT1 PDE3B DDIT3 RAB8A SERTAD1 HIST1H2AC HIST1H2BK CITED2 PFKP PRR5L PMAIP1 GADD45B ICAM4 CDK11A LAPTM5 JUN RAB13 PYHIN1 RGS18 ARID4B CNST PXDN NABP1 ANKRD28 CD96 RSRC1 RHOH ANTXR2 MYLIP LST1 IFRD1 CD99 UPF2 MALAT1 SLC2A3 TUBA1A BIN2 NFE2 CCNA1 ELF1 MAP1A STX4 GNA15 HCST TYROBP ZFP36 CLEC11A DDX3Y EIF1AY NOC2L PLK3 CDKN2C LAX1 ARHGAP25 CCNL1 HPGDS PPP1R10 IL2RG LRP12 LY6E CDKN2A IFITM2 TAF1D CD3D CD68
b
Top differential Transcription Factors
MPAL Subpopulations Differential
0.75
0.50
0.25
0.00
ATF3 DDIT3 NFE2 JUN ELF1 ATF4 JUNB KLF6 EGR1 FOS ZFY WT1 FOSB IRF7 MAFF MEF2A TCF7 STAT3 BCL11B MXI1 MLXIP GFI1B GATA2 NR4A2 RFX8 RUNX1 LYL1 KLF2 ARID5B ID2 ETV6 IRF1 IRF9 NR4A1 REL
c
clusterProfiler KEGG Enrichment
(differentially conserved K-means 2,3,4, and 10)
Number of Genes in Pathway
0
20
40
60
P-value
(adjusted)
0.008
0.006
0.004
0.002
Viral carcinogenesis
Endocytosis
Protein processing in endoplasmic reticulum
Transcriptional misregulation in cancer
NOD−like receptor signaling pathway
Chronic myeloid leukemia
Measles
Non−small cell lung cancer
Shigellosis
Acute myeloid leukemia
Epstein−Barr virus infection
Bacterial invasion of epithelial cells
Pancreatic cancer
Hepatitis C
Regulation of actin cytoskeleton
d
chr1:1,580,765−1,730,766
scRNA
(Log2 Ex.)
0 1 2 3 4 5
MPAL1
MPAL2
MPAL3
MPAL4
MPAL5
MPAL5R
**
strand
−
+
CDK11A
e
chr9:21,920,299−22,070,300
scRNA
(Log2 Ex.)
0 1 2 3 4 5
MPAL1
MPAL2
MPAL3
MPAL4
MPAL5
MPAL5R
**
strand
−
+
CDKN2A
Donor
MPAL1
MPAL2
MPAL3
MPAL4
MPAL5
MPAL5R
Classification
Erythroid-Like
Lymphoid-Like
Monocyte-Like
Progenitor-Like
T/NK-Like

Supplementary Figure 9

Visualization of top differential genes and accessible peak regions.

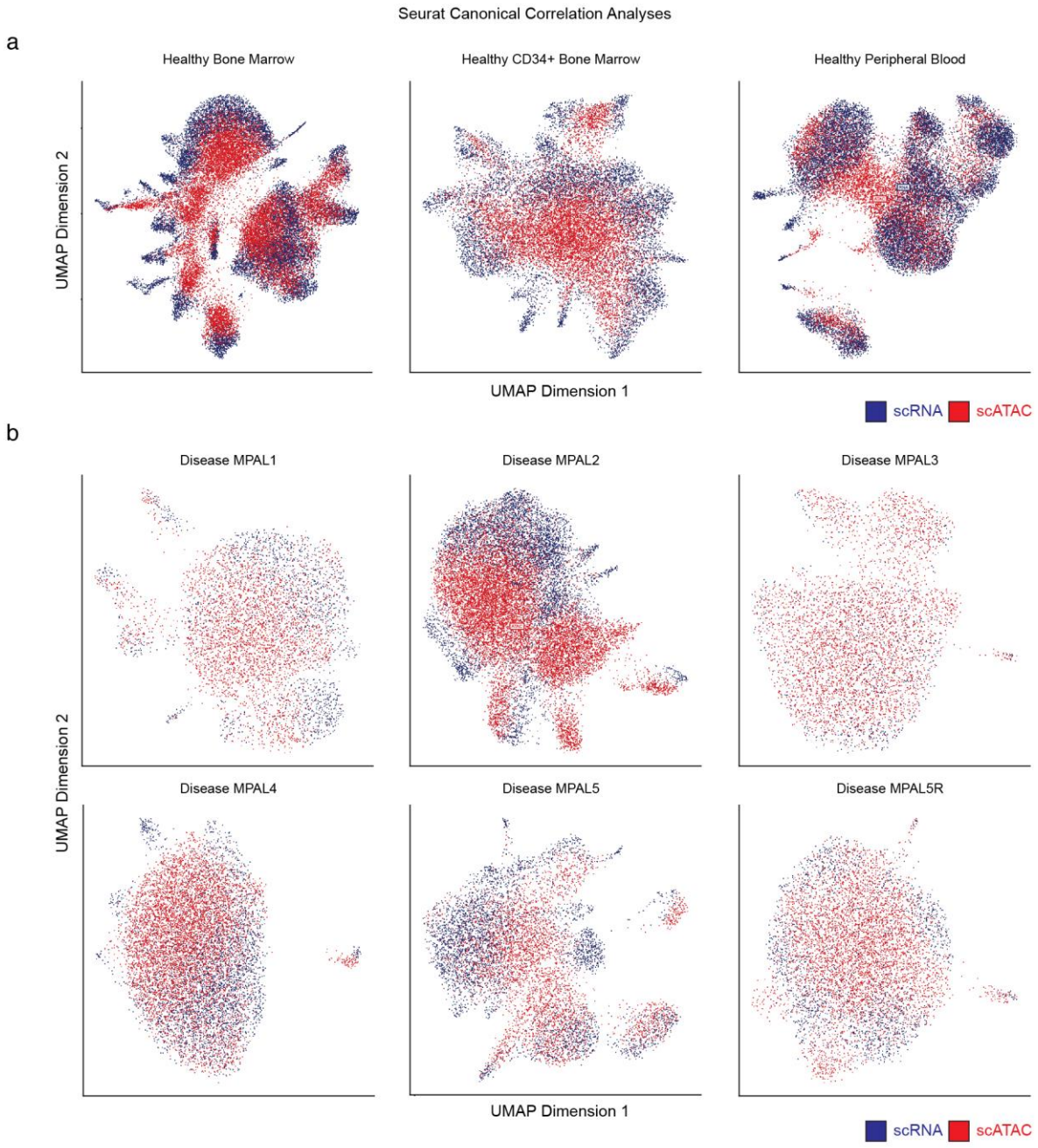
(a) Top conserved differential genes across the MPAL hematopoietic compartments. **(b)** Top conserved differential transcription factors across the MPAL hematopoietic compartments. **(c)** KEGG pathway enrichment of genes (scRNA) differentially conserved k-means 2, 3, 4, and 10 (Figure 2c, n = 2,117 genes). Color represents the significance (hypergeometric test adjusted p-value with the Benjamini-Hochberg correction) of the KEGG path way generated using clusterProfiler (Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16,** 284–287 (2012)). **(d-e)** Multi-omic differential tracks (Left) scATAC tracks showing MPAL disease subpopulations (red) closest normal cells (grey). (Right) Violin plot of the log2 normalized expression for MPAL disease subpopulations (red) and closest normal cells (grey); black line represents the mean and asterisk denote significance (LFC > 0.5 and FDR < 0.01 from Figure 2c). **(d)** Multi-omic differential track of *CDK11A*, up-regulated in MPALs 1, 2, 5 and 5R (n = 89 - 500). **(e)** Multi-omic differential track of *CDKN2A*, up-regulated in MPALs 1, 2, 3, 4, and 5 (n = 89 - 500).


a
(1) Bulk Leukemia RNA-seq
(2) Downsample to Sinlge-Cell
(3) LSI Projection
LSI Projection
Leukemic Cells
Healthy Cells
B
Erythroid
HSC
CLP
NK
CD8M
Plasma
CD4M
Basophil
pDC
GMP
Neutrophil
CD4N
CD8N
Monocytes
b
B-ALL
T-ALL
ETP
AML
T/M MPAL
B/M MPAL
UMAP Dimension 2
UMAP Dimension 1
Gene Classification
Malignant
Not Malignant
Source
MPAL1-5R
Bulk Leukemia (16)
Cytogenetics
KMT2Ar
Others
B.other
DUX4
ETV6
Hyperdiploid
MEF2D
PBX1
Ph..like
ZNF384
Leukemia
AML
B-ALL
ETP
MPAL-B/M
MPAL-others
MPAL-T/M
T-ALL
c
UMAP Dimension 2
UMAP Dimension 1
Cytogenetics
B.other
DUX4
ETV6
Hyperdiploid
KMT2Ar
MEF2D
Others
PBX1
Ph..like
ZNF384
Bulk Leukemia (16)
MPAL1-5R
d
Source
Cytogenetics
Leukemia

Supplementary Figure 10

LSI projection of bulk leukemia RNA-seq onto hematopoietic hierarchy.

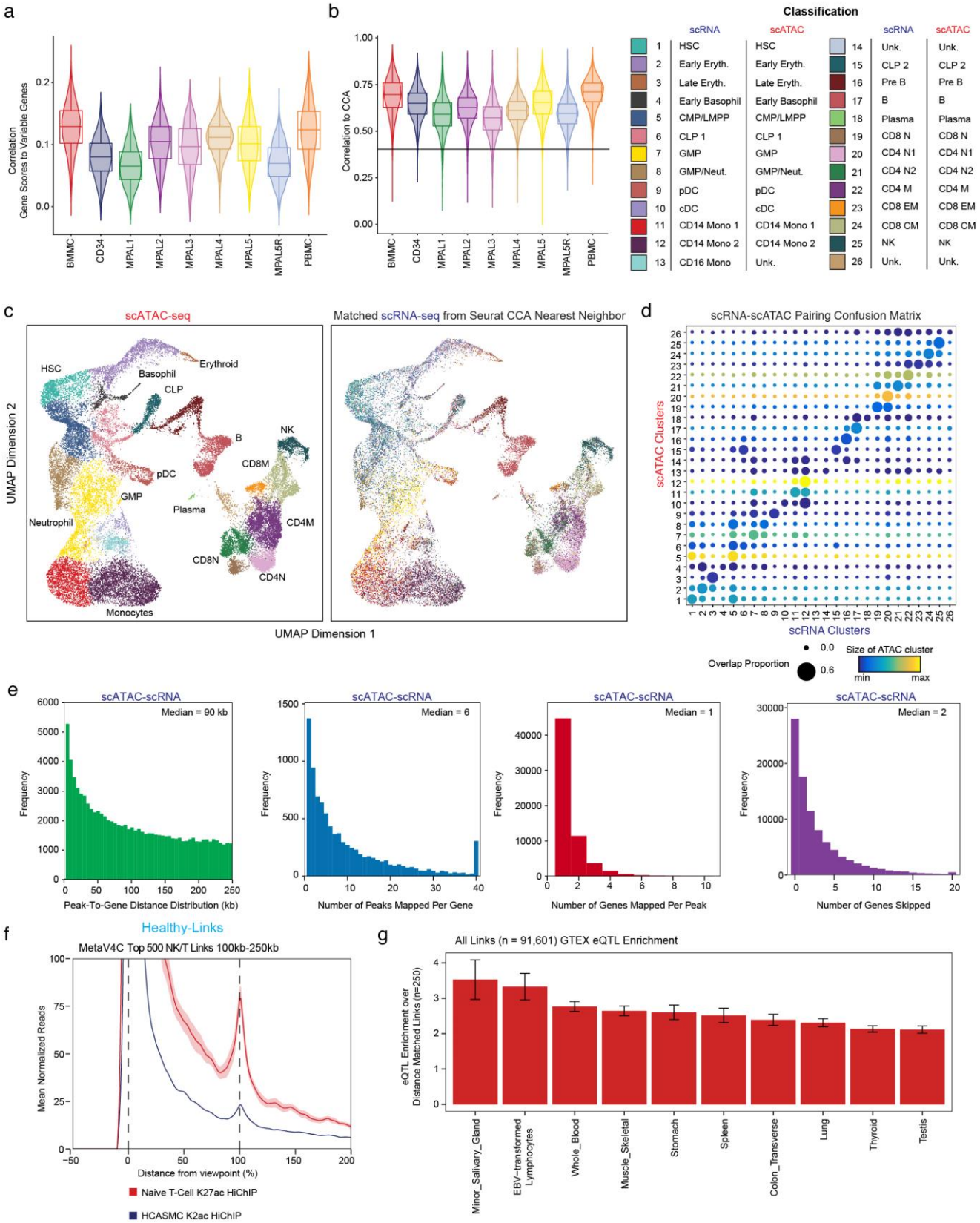
(a) Schematic of LSI projection of downsampled bulk leukemia RNA-seq onto healthy hematopoiesis. **(b)** Representative downsampled LSI projections (n = 250) for B-ALLs, non-ETP T-ALLs, ETP T-ALLs, AMLs, T/M MPALs and B/M MPALs from previous studies[16]. **(c)** LSI UMAP of differentially up-regulated gene expression profiles across bulk leukemias[16] (n = 321) and MPAL subpopulations assayed in this study (n = 17), colored by cytogenetics. **(d)** Binary heatmap of variable malignant genes across leukemia classifications. Each cell in the heatmap is colored whether the gene was identified as malignant for the leukemic sample.



Supplementary Figure 11

Seurat canonical correlation analysis alignment of scRNA and scATAC-seq hematopoietic and MPAL samples.

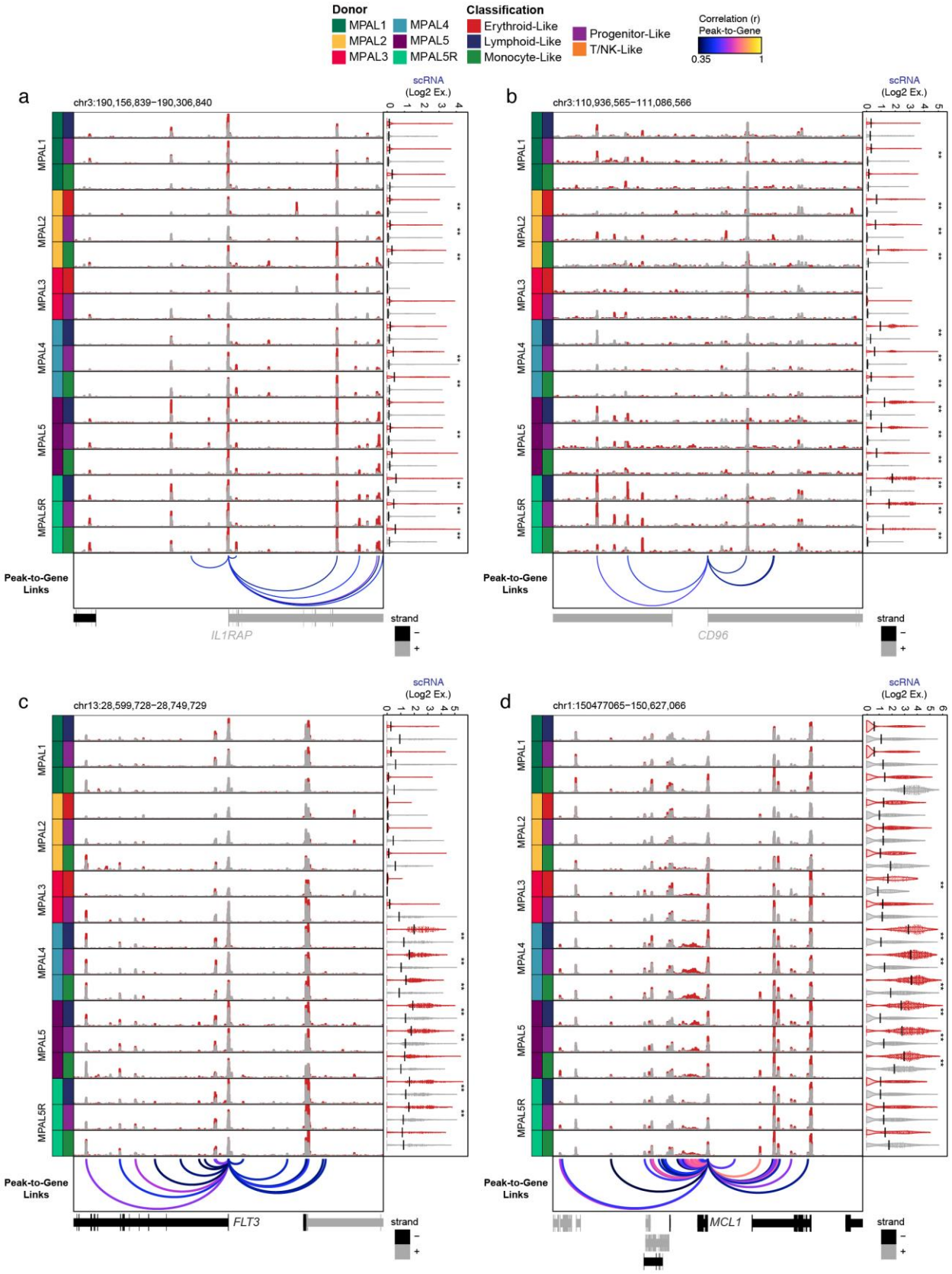
(a) UMAP of CCA alignment of scATAC-seq using Cicero gene activity scores and scRNA-seq for (Left) bone marrow (nATAC = 12,602; nRNA = 16,510), (Middle) CD34+ enriched bone marrow (nATAC = 8,176; nRNA = 10,160), (Right) peripheral blood (nATAC = 14,804; nRNA = 8,368). **(b)** UMAP of CCA alignment of scATAC-seq using Cicero gene activity scores and scRNA-seq for MPAL samples 1-5R (nATAC = 4,127 – 8,255; nRNA = 835 – 5,885).


a
Correlation Gene Scores to Variable Genes
BMMC
CD34
MPAL1
MPAL2
MPAL3
MPAL4
MPAL5
MPAL5R
PBMC
b
Correlation to CCA
Classification
scRNA
scATAC
1 HSC HSC
2 Early Eryth. Early Eryth.
3 Late Eryth. Late Eryth.
4 Early Basophil Early Basophil
5 CMP/LMPP CMP/LMPP
6 CLP 1 CLP 1
7 GMP GMP
8 GMP/Neut. GMP/Neut.
9 pDC pDC
10 cDC cDC
11 CD14 Mono 1 CD14 Mono 1
12 CD14 Mono 2 CD14 Mono 2
13 CD16 Mono Unk.
14 Unk. Unk.
15 CLP 2 CLP 2
16 Pre B Pre B
17 B B
18 Plasma Plasma
19 CD8 N CD8 N
20 CD4 N1 CD4 N1
21 CD4 N2 CD4 N2
22 CD4 M CD4 M
23 CD8 EM CD8 EM
24 CD8 CM CD8 CM
25 NK NK
26 Unk. Unk.
c
scATAC-seq
Matched scRNA-seq from Seurat CCA Nearest Neighbor
HSC
Basophil
Erythroid
CLP
B
NK
CD8M
pDC
GMP
Plasma
CD4M
Neutrophil
CD8N
CD4N
Monocytes
UMAP Dimension 2
UMAP Dimension 1
d
scRNA-scATAC Pairing Confusion Matrix
scATAC Clusters
scRNA Clusters
Overlap Proportion
0.0
0.6
Size of ATAC cluster
min
max
e
scATAC-scRNA
Median = 90 kb
Frequency
Peak-To-Gene Distance Distribution (kb)
Median = 6
Number of Peaks Mapped Per Gene
Median = 1
Number of Genes Mapped Per Peak
Median = 2
Number of Genes Skipped
f
Healthy-Links
MetaV4C Top 500 NK/T Links 100kb-250kb
Mean Normalized Reads
Distance from viewpoint (%)
Naive T-Cell K27ac HiChIP
HCASMC K2ac HiChIP
g
All Links (n = 91,601) GTEX eQTL Enrichment
eQTL Enrichment over Distance Matched Links (n=250)
Minor_Salivary_Gland
EBV-transformed Lymphocytes
Whole_Blood
Muscle_Skeletal
Stomach
Spleen
Colon_Transverse
Lung
Thyroid
Testis

Supplementary Figure 12

Evaluation of scRNA and scATAC-seq alignment and peak-to-gene linkage across hematopoiesis and MPAL samples.

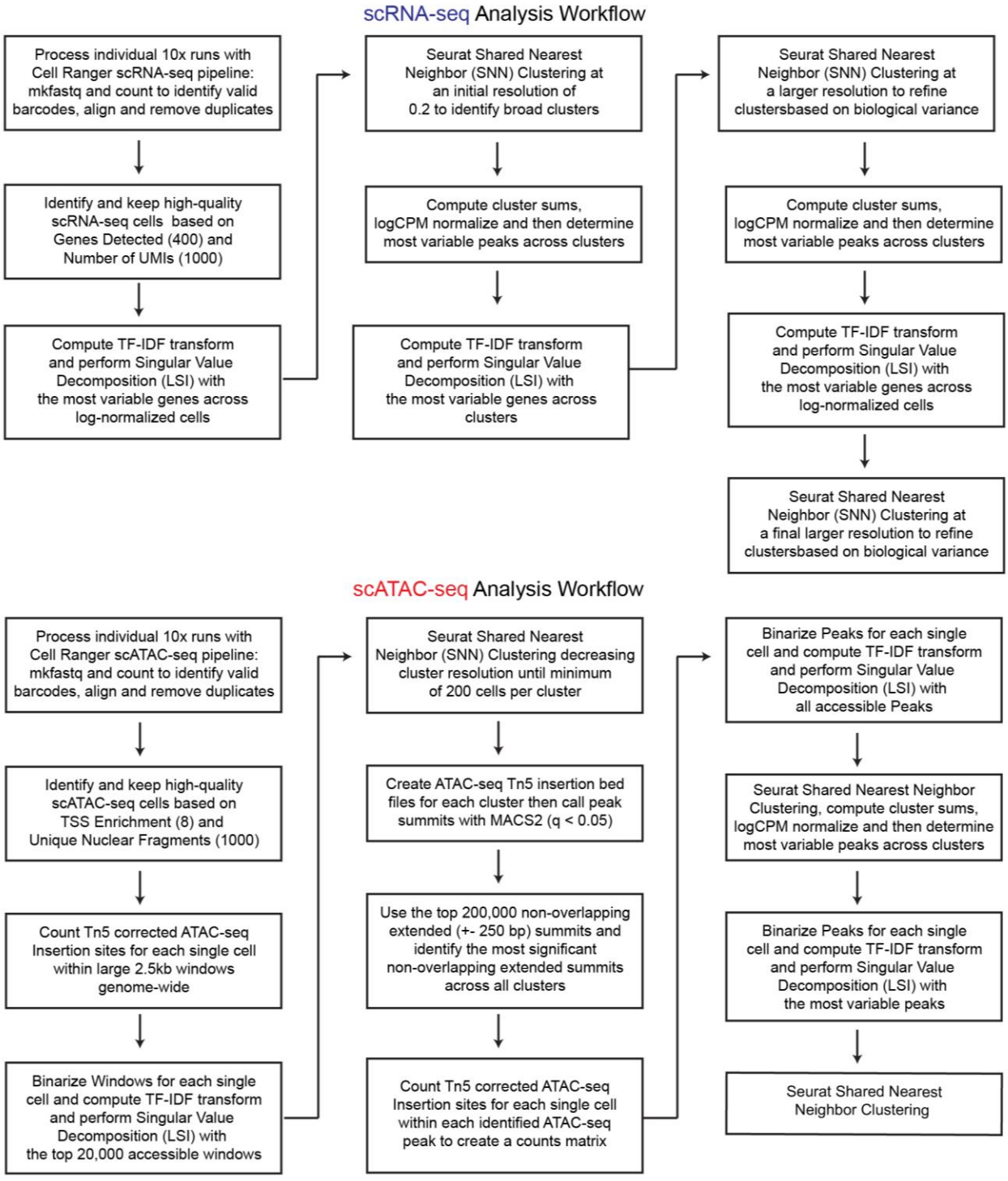
(a) Spearman rank correlation between scATAC-seq Cicero gene activity scores to scRNA-seq for each mapped cell within across all biological experiments (n = 4,127 – 16,510). **(b)** Pearson correlation of CCA scRNA and scATAC-seq nearest-neighbors. The cutoff (R > 0.45) for high quality nearest neighbor mappings is shown (n = 4,127 – 16,510). **(c)** (Left) UMAP of scATAC-seq hematopoiesis colored by scATAC-seq clusters (n = 35,038). (Right) UMAP of scATAC-seq hematopoiesis colored by filtered mapped scRNA-seq clusters (n = 34,507). **(d)** Confusion matrix of initial clusters for mapped scRNA-seq to scATAC-seq clusters for hematopoiesis (Figure 1b-c). Above shows the assigned biological classifications for each cluster across both scRNA and scATAC-seq. **(e)** (Left) Distribution of peak-to-gene distances. (Left-Middle) Distribution of number of peaks mapped per gene (median = 6). (Right-Middle) Distribution of number of genes mapped per peak (median = 1). (Right) Distribution of number of genes skipped for peak-to-gene links (median = 2). **(f)** MetaV4C plots of K27ac HiChIP in Naive T and HCASMC cells for top 500 biased T/NK (broad classification) peak-to-gene links that are identified only in healthy hematopoiesis. The line represents the average signal and the shading represents the range of the signal times the square root of 2 between biological replicates (n = 2). **(g)** Peak-to-gene links (n = 91,601) enrichment in GTEx eQTLs over a permuted background distance-matched set (permutations = 250) for the union set of peak-to-gene links. The mean enrichment is shown, and the error bars indicate 1 standard deviation. Box-whisker plot; lower whisker is the lowest value greater than the 25% quantile minus 1.5 times the interquartile range (IQR), the lower hinge is the 25% quantile, the middle is the median, the upper hinge is the 75% quantile and the upper whisker is the largest value less than the 75% quantile plus 1.5 times the IQR.


Donor
MPAL1
MPAL2
MPAL3
MPAL4
MPAL5
MPAL5R
Classification
Erythroid-Like
Lymphoid-Like
Monocyte-Like
Progenitor-Like
T/NK-Like
Correlation (r)
Peak-to-Gene
0.35
1
a
chr3:190,156,839−190,306,840
scRNA
(Log2 Ex.)
MPAL1
MPAL2
MPAL3
MPAL4
MPAL5
MPAL5R
Peak-to-Gene
Links
IL1RAP
strand
b
chr3:110,936,565−111,086,566
CD96
c
chr13:28,599,728−28,749,729
FLT3
d
chr1:150477065−150,627,066
MCL1

Supplementary Figure 13

Peak-to-gene links nominate putative regulatory regions that nominate key leukemic genes.

(a-d) Multi-omic differential track; (Middle) Aggregated scATAC tracks showing MPAL disease subpopulations (red) and closest normal cells (grey). (Right) Distribution of log2 normalized expression of gene of interest for MPAL disease subpopulations (red) and closest normal cells (grey); black line represents the mean and asterisk denote significance (LFC > 0.5 and FDR < 0.01 from Figure 2c). Violin plot represents the smoothed density of the distribution of the log2 normalized expression and the black line represents the mean log2 normalized expression. (Bottom) Peak-to-gene links for gene of interest colored by Pearson correlation of the peak accessibility and gene expression (see methods). **(a)** Multi-omic differential track for *IL1RAP* (n = 89 - 500). **(b)** Multi-omic differential track for *CD96* (n = 89 - 500). **(c)** Multi-omic differential track for *FLT3* (n = 89 - 500). **(d)** Multi-omic differential track for *MCL1* (n = 89 - 500).



Supplementary Figure 14

Analysis workflows for processing of scRNA-seq and scATAC-seq data.

(Top) scRNA-seq analysis workflow. Briefly cells are aligned using 10x cell ranger, quality filtered, and clustered using a feature optimization approach (see methods). (Bottom) scATAC-seq analysis workflow. Briefly cells are aligned using 10x cell ranger atac, quality filtered, clustered in large windows genome-wide, peak-calling on clusters, creation of a counts matrix and clustered using a feature optimization approach (see methods). (see methods).