Cell Reports, Volume *36*

Supplemental information

TooManyPeaks identifies drug-resistant-specific regulatory elements from single-cell leukemic epigenomes

Gregory W. Schwartz, Yeqiao Zhou, Jelena Petrovic, Warren S. Pear, and Robert B. Faryabi

Figure S1

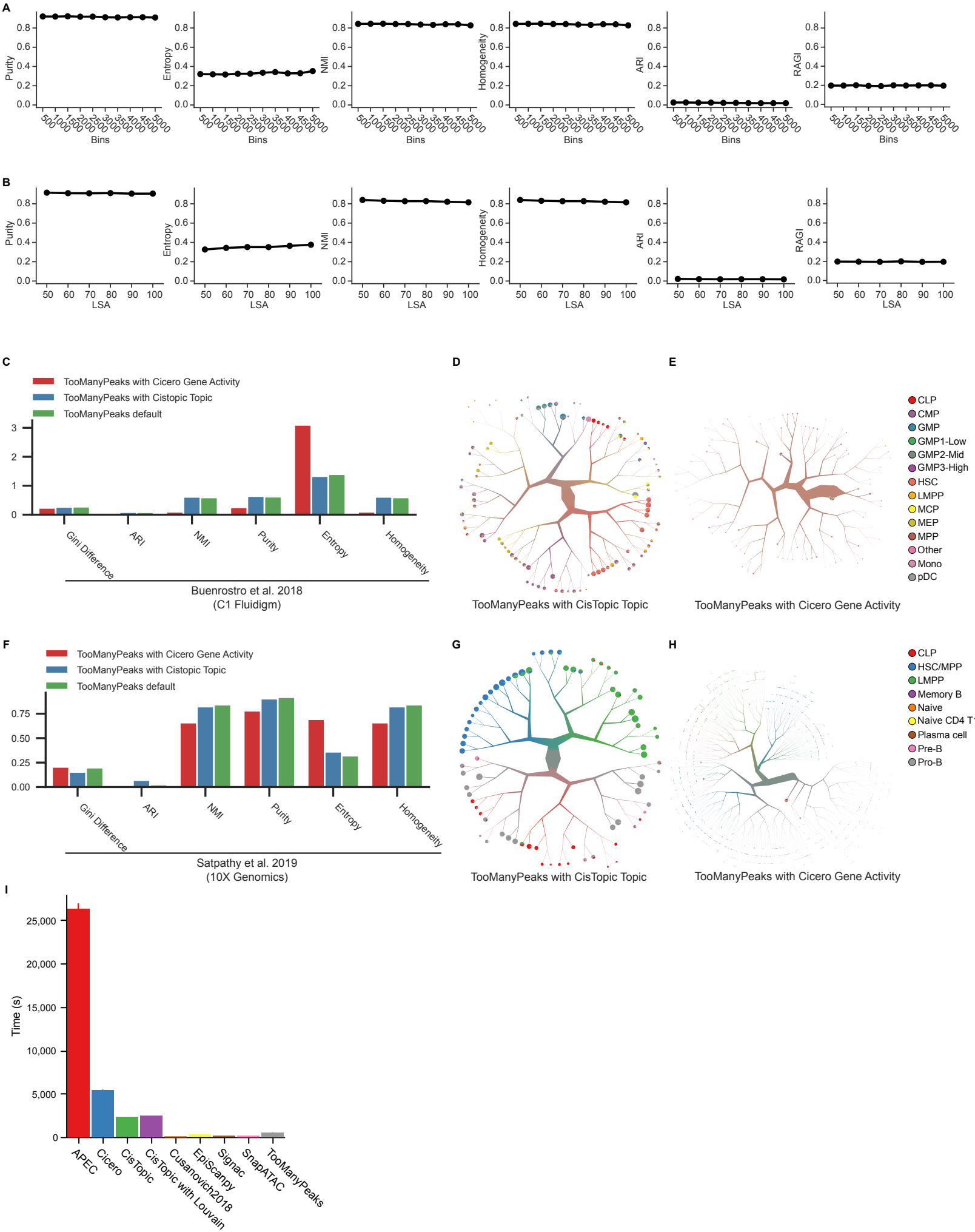

A
Purity
Entropy
NMI
Homogeneity
ARI
RAGI
Bins
B
LSA
C
TooManyPeaks with Cicero Gene Activity
TooManyPeaks with Cistopic Topic
TooManyPeaks default
Gini Difference
ARI
NMI
Purity
Entropy
Homogeneity
Buenrostro et al. 2018
(C1 Fluidigm)
D
TooManyPeaks with CisTopic Topic
E
TooManyPeaks with Cicero Gene Activity
CLP
CMP
GMP
GMP1-Low
GMP2-Mid
GMP3-High
HSC
LMPP
MCP
MEP
MPP
Other
Mono
pDC
F
Satpathy et al. 2019
(10X Genomics)
G
H
HSC/MPP
Memory B
Naive
Naive CD4 T1
Plasma cell
Pre-B
Pro-B
I
Time (s)
APEC
Cicero
CisTopic
CisTopic with Louvain
Cusanovich2018
EpiScanpy
Signac
SnapATAC
TooManyPeaks

Figure S1: Comparison of TooManyPeaks for varying parameters, feature definition algorithms, and timing. (**A-B**) Comparison of TooManyPeaks across varying parameters on a human peripheral blood cell population ($n = 44{,}814$ cells) (Satpathy et al., 2019). Performance was measured using a clustering benchmark with, from left to right, lower entropy, higher purity, higher normalized mutual information (NMI), higher adjusted Rand index (ARI), and higher homogeneity showing more accurate clustering by varying either the number of bins (A) or the number of LSA dimensions (B). (**C-H**) Running TooManyPeaks using features defined from other scATAC-seq algorithms. (C, F) Clustering benchmark quantification from either the tree generated using CisTopic (D, G) or Cicero (E, H) on $CD34^+$ hematopoietic progenitor cells profiled using 10x Genomics ($n = 7{,}771$ cells) (Satpathy et al., 2019) (C-E), or Fluidigm C1 ($n = 2{,}954$ cells) (Buenrostro et al., 2018) (F-H). (**I**) Timing benchmark of scATAC-seq algorithms. Running time of algorithms were compared using a human bone marrow profiled with Fluidigm C1 ($n = 2{,}954$ cells) (Buenrostro et al., 2018). Bar plots (mean plus / minus standard error) quantifies timing for 3 independent runs using a machine with Ubuntu 20.04, 512GiB Memory, Intel® Xeon® CPU E5-2670 v3 @ 2.30GHz, 2 physical processors 24 cores, and 48 threads. For an unbiased comparison, default or suggested filterings and parameters were used for all algorithms unless otherwise noted (see STAR Methods). Related to Figure 1.

Figure S2

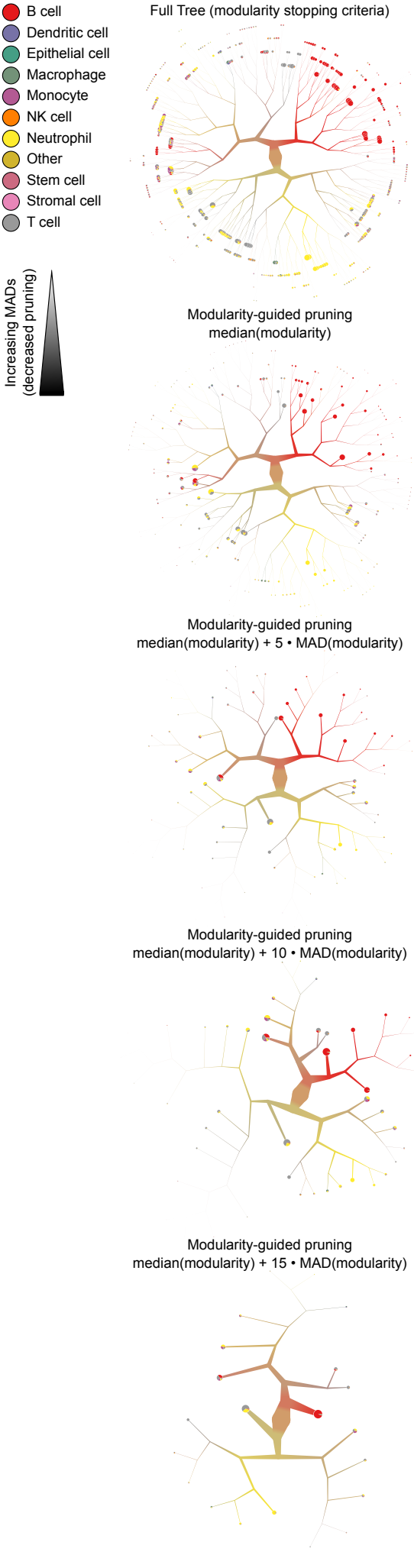
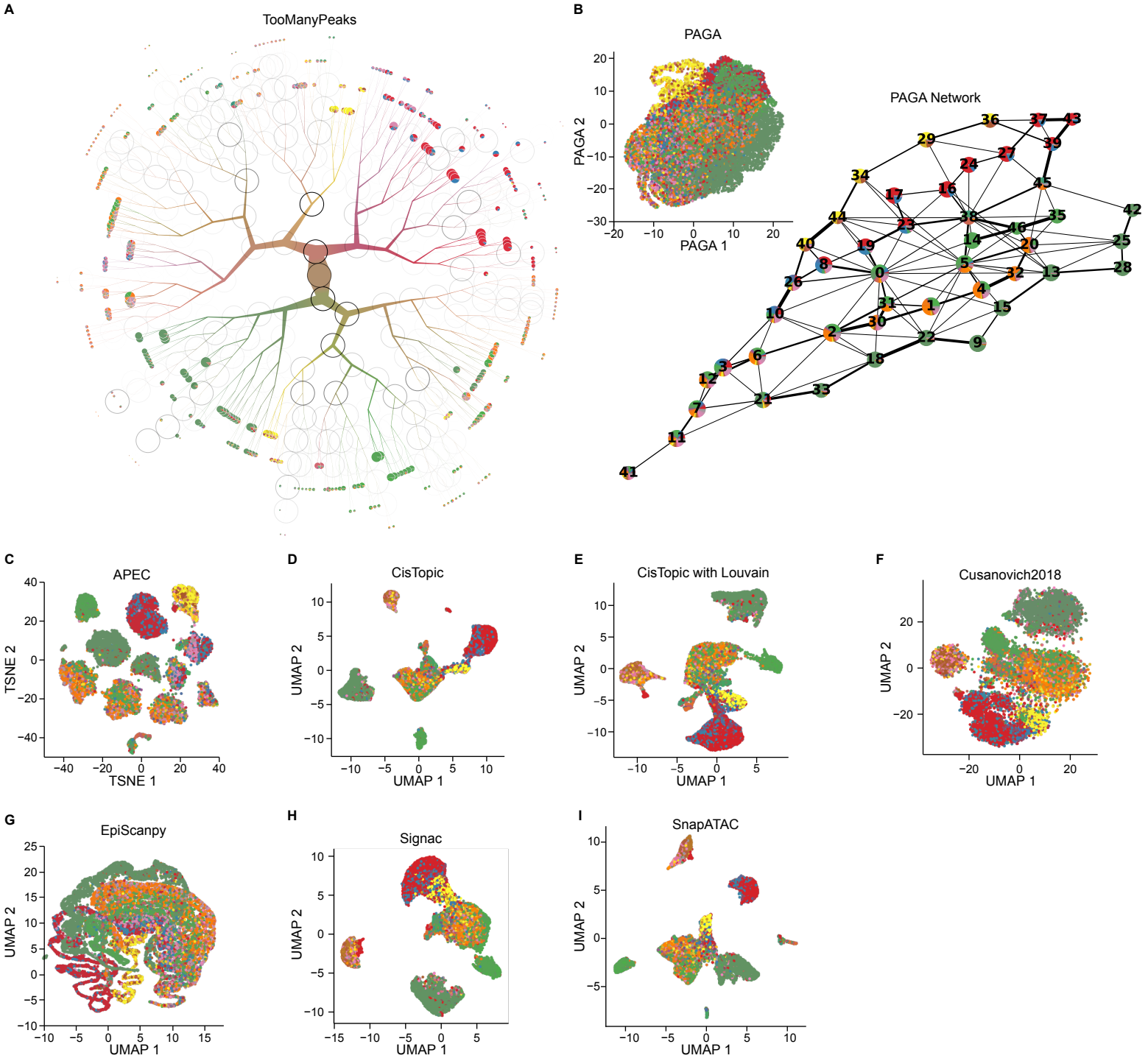

B cell
Dendritic cell
Epithelial cell
Macrophage
Monocyte
NK cell
Neutrophil
Other
Stem cell
Stromal cell
T cell
Increasing MADs
(decreased pruning)
Full Tree (modularity stopping criteria)
Modularity-guided pruning
median(modularity)
Modularity-guided pruning
median(modularity) + 5 • MAD(modularity)
Modularity-guided pruning
median(modularity) + 10 • MAD(modularity)
Modularity-guided pruning
median(modularity) + 15 • MAD(modularity)

Figure S2: TooManyPeaks provides a number of pruning options, including a modularity-guided pruning, to control the tree depth (Schwartz et al., 2020). Example of controlling the depth of the TooManyPeaks tree with modularity stopping criteria of mouse spleen and bone marrow cells with modularity-guided pruning when the threshold is set to no pruning or median(modularity) plus 0, 5, 10, and 15 $\times$ MAD(modularity) (from top to bottom in that order). MAD: median absolute deviations. Related to Figure 2.

Figure S3



- B.Fem.Sp
- B.Fo.Sp
- B.FrE.BM
- B.GC.CB.Sp
- B.GC.CC.Sp
- B.mem.Sp
- B.MZ.Sp
- B.PB.Sp
- B.PC.BM
- B.PC.Sp
- B.Sp
- B.T1.Sp
- B.T2.Sp
- B.T3.Sp
- B1b.PC
- BEC.SLN
- DC.103+11b−.SI
- DC.103+11b+.SI
- DC.4+.Sp
- DC.8+.Sp
- DC.pDC.Sp
- Ep.MEChi.Th
- FRC.CD140a+.Madcam−.CD35−.SLN
- GN.BM
- GN.Sp
- GN.Thio.PC
- IAP.SLN
- ILC2.SI
- ILC3.CCR6+.SI
- ILC3.NKp46−CCR6−.SI
- ILC3.NKp46+.SI
- LEC.SLN
- LTHSC.34−.BM
- LTHSC.34+.BM
- MF.226+II+480lo.PC
- MF.Alv.Lu
- MF.Fem.PC
- MF.ICAM+480hi.PC
- MF.LP.SI
- MF.microglia.CNS
- MF.PC
- MF.pIC.Alv.Lu
- MF.RP.Sp
- Mo.6C−II−.Bl
- Mo.6C+II−.Bl
- MPP3.48+.BM
- MPP4.135+.BM
- NK.27−11b+.BM
- NK.27−11b+.Sp
- NK.27+11b−.BM
- NK.27+11b−.Sp
- NK.27+11b+.BM
- NK.27+11b+.Sp
- NKT.Sp.LPS.18hr
- NKT.Sp.LPS.3d
- NKT.Sp.LPS.3hr
- NKT.Sp
- preB.FrD.BM
- preT.DN1.Th
- preT.DN2a.Th
- preT.DN2b.Th
- preT.DN3.Th
- proB.CLP.BM
- proB.FrA.BM
- proB.FrBC.BM
- STHSC.150−.BM
- T.4.Nve.Fem.Sp
- T.4.Nve.Sp
- T.4.Sp.aCD3+CD40.18hr
- T.4.Th
- T.8.Nve.Sp
- T.8.Th
- T.DN4.Th
- T.DP.Th.Rorc(gt)+ .GFP
- T.DP.Th.Rorc(gt)GFP .GFP
- T.DP.Th
- T.ISP.Th
- T8.IEL.LCMV.d7.SI
- T8.MP.LCMV.d7.Sp
- T8.Tcm.LCMV.d180.Sp
- T8.TE.LCMV.d7.Sp
- T8.Tem.LCMV.d180.Sp
- T8.TN.P14.Sp
- Tgd.g1.1+d1.24a+.Th
- Tgd.g1.1+d1.LN
- Tgd.g2+d1.24a+.Th
- Tgd.g2+d1.LN
- Tgd.g2+d17.24a+.Th
- Tgd.g2+d17.LN
- Tgd.Sp
- Treg.4.25hi.Sp
- Treg.4.FP3+.Nrplo.Co

Figure S3: Stratification and annotation of murine marrow and spleen cells. The TooManyPeaks algorithm for cell-type annotation based on reference cis-regulatory elements was used to predict cell lineages in mouse bone marrow and spleen ($n = 16{,}749$ cells) (Cusanovich et al., 2018). Reference cis-regulatory elements of 92 phenotypically defined progenitor and differentiated hematopoietic cell types are generated from the analyses of bulk ATAC-seq in FACS-sorted cells (Yoshida et al., 2019). **(A)** TooManyPeaks with modularity stopping criteria and no pruning shows all 92 progenitor and differentiated hematopoietic cell types. At each bipartitioining, darker circle circumference represents higher modularity. **(B-I)** PAGA-initiated UMAP (B, left panel) or PAGA network (B, right panel), t-SNE output of APEC (C), as well as UMAP outputs of CisTopic (D), CisTopic with Louvain (E), Cusanovich2018 (F), EpiScanpy (G), Signac (H), and SnapATAC (I) are colored by assigned cell-type labels. For an unbiased comparison, each projection used the corresponding package's UMAP or t-SNE implementation. Related to Figure 2.

Figure S4

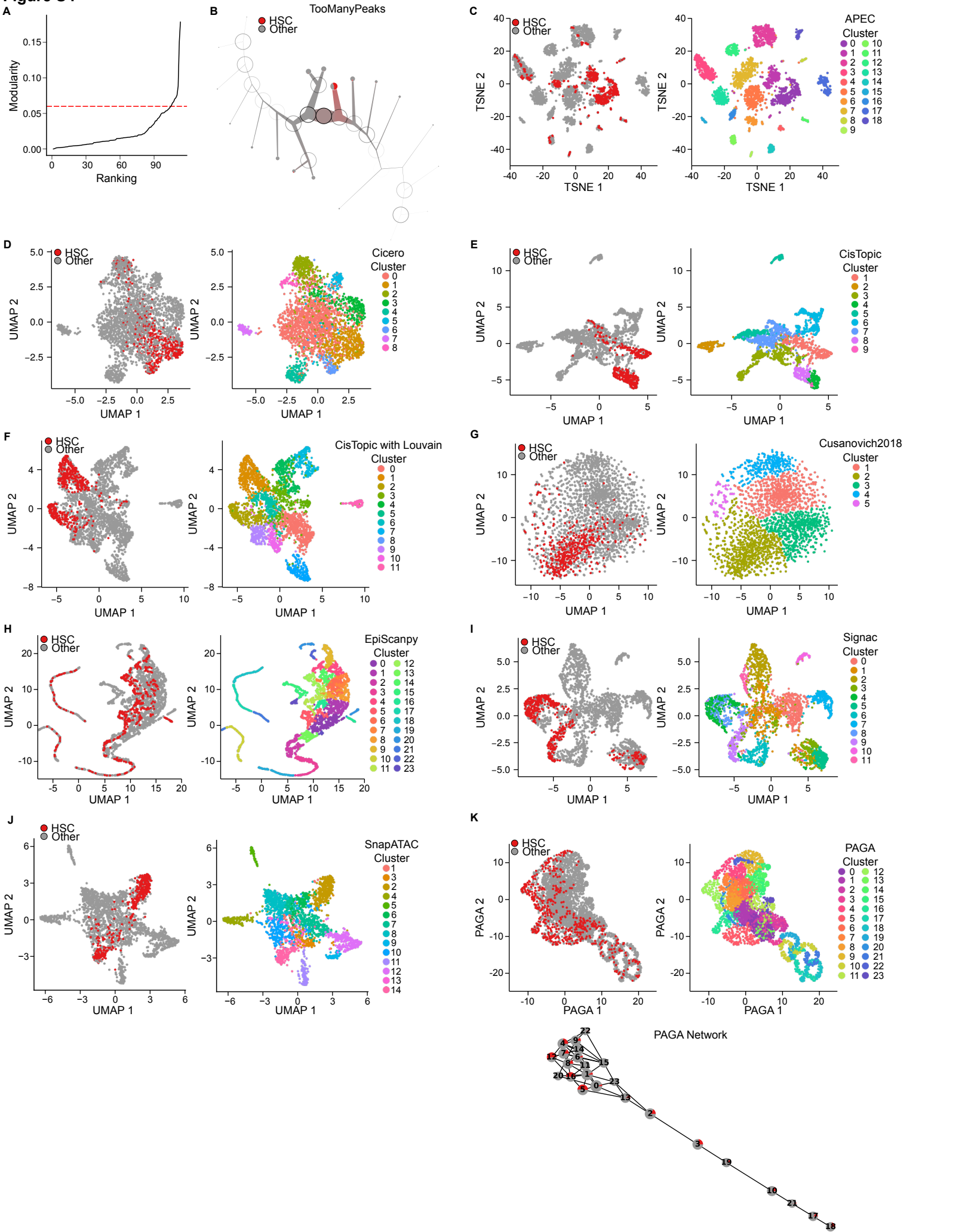

A
Modularity
Ranking
B
TooManyPeaks
HSC
Other
C
HSC
Other
TSNE 1
TSNE 2
APEC
Cluster
D
HSC
Other
UMAP 1
UMAP 2
Cicero
Cluster
E
HSC
Other
UMAP 1
UMAP 2
CisTopic
Cluster
F
HSC
Other
UMAP 1
UMAP 2
CisTopic with Louvain
Cluster
G
HSC
Other
UMAP 1
UMAP 2
Cusanovich2018
Cluster
H
HSC
Other
UMAP 1
UMAP 2
EpiScanpy
Cluster
I
HSC
Other
UMAP 1
UMAP 2
Signac
Cluster
J
HSC
Other
UMAP 1
UMAP 2
SnapATAC
Cluster
K
HSC
Other
PAGA 1
PAGA 2
PAGA
Cluster
PAGA Network

Figure S4: Comparing localization of hematopoietic stem cells (HSC) in the TooManyPeaks tree and UMAP or t-SNE outputs of human bone marrow profiled with Fluidigm C1 (Buenrostro et al., 2018). **(A)** The modularity-guided pruning threshold is set based on the beginning of the left plateau (marked by dashed red line) in the ranked modularity curve of the human bone marrow TooManyPeaks tree nodes ($n = 2{,}954$ cells). **(B)** HSC cells are colored on the pruned TooManyPeaks tree per (A). At each bipartitioining, darker circle circumference represents higher modularity. **(C-K)** HSC cells are colored on UMAP or t-SNE outputs (red dots, left panel) generated by APEC (C), Cicero (D), CisTopic (E), CisTopic with Louvain (F), Cusanovich2018 (G), EpiScanpy (H), Signac (I), SnapATAC (J), and PAGA (K) initiated UMAP (top two panels) or PAGA network (bottom panels). Coordinates from the UMAP or t-SNE outputs (C-K, left panels) colored by each algorithm cluster label (C-K, right panels) fails to clearly localize HSC cells. For an unbiased comparison, each projection used the corresponding package's UMAP or t-SNE implementation. Moreover, default or suggested filters and parameters were used for all algorithms unless otherwise noted (see STAR Methods). Related to Figure 2.

Figure S5

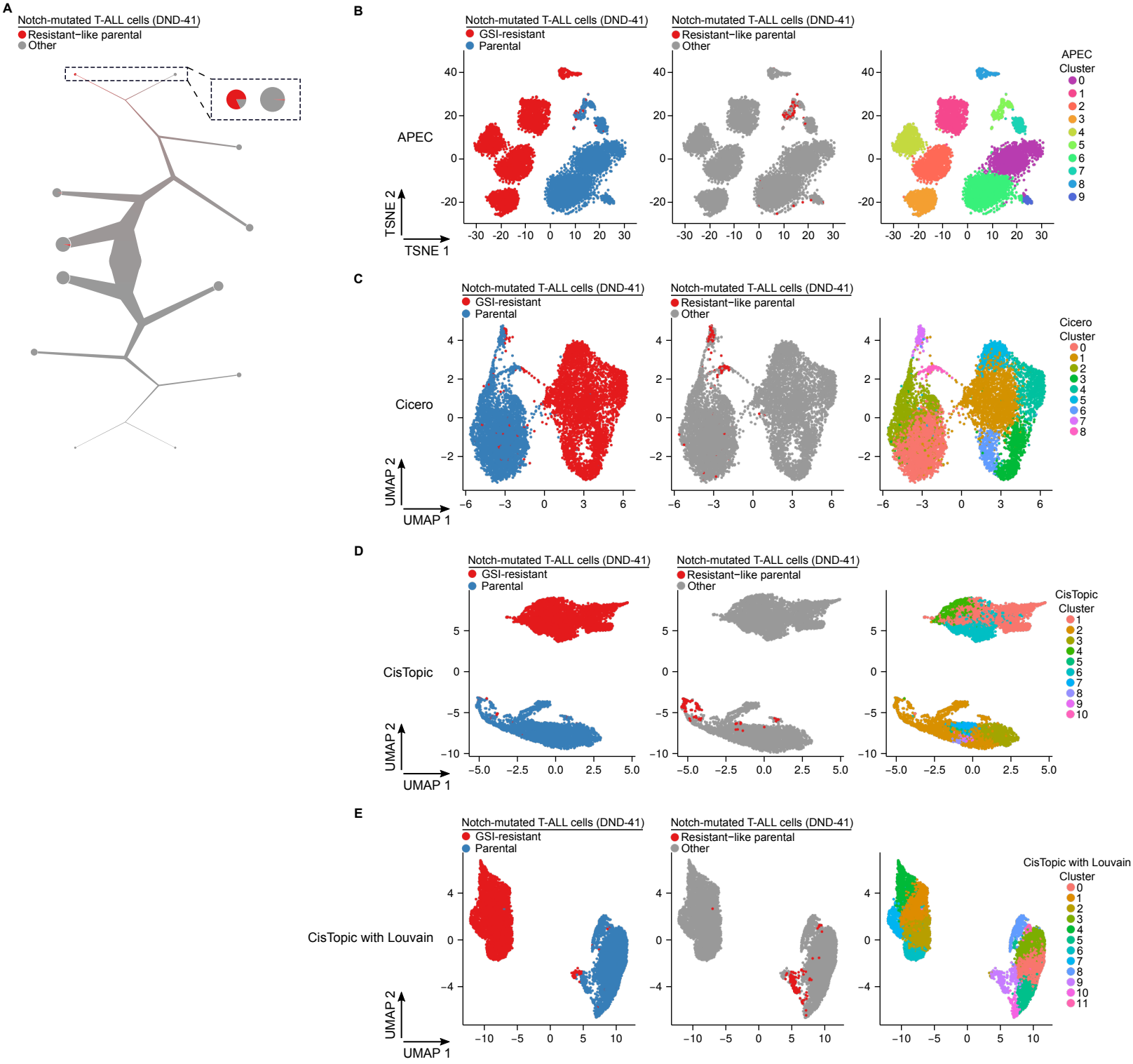

A
Notch-mutated T-ALL cells (DND-41)
Resistant−like parental
Other
B
Notch-mutated T-ALL cells (DND-41)
GSI-resistant
Parental
Notch-mutated T-ALL cells (DND-41)
Resistant−like parental
Other
APEC
Cluster
APEC
TSNE 2
TSNE 1
C
Notch-mutated T-ALL cells (DND-41)
GSI-resistant
Parental
Notch-mutated T-ALL cells (DND-41)
Resistant−like parental
Other
Cicero
Cluster
Cicero
UMAP 2
UMAP 1
D
Notch-mutated T-ALL cells (DND-41)
GSI-resistant
Parental
Notch-mutated T-ALL cells (DND-41)
Resistant−like parental
Other
CisTopic
Cluster
CisTopic
UMAP 2
UMAP 1
E
Notch-mutated T-ALL cells (DND-41)
GSI-resistant
Parental
Notch-mutated T-ALL cells (DND-41)
Resistant−like parental
Other
CisTopic with Louvain
Cluster
CisTopic with Louvain
UMAP 2
UMAP 1

Figure S5: Comparisons of DND-41 T-ALL scATAC-seq data visualization. **(A)** The TooManyPeaks tree colored by resistant-like parental cells. **(B-E)** projection outputs of APEC (B), Cicero (C), CisTopic (D), and CisTopic with Louvain (E) colored by resistance status (left), resistant-like parental cells as defined by TooManyPeaks (middle), or algorithm cluster assignment (right) ($n = 7{,}989$ cells). For an unbiased comparison, each projection used the corresponding package's UMAP or t-SNE implementation. Moreover, default or suggested filters and parameters were used for all algorithms unless otherwise noted (see STAR Methods). Related to Figure 3.

Figure S6

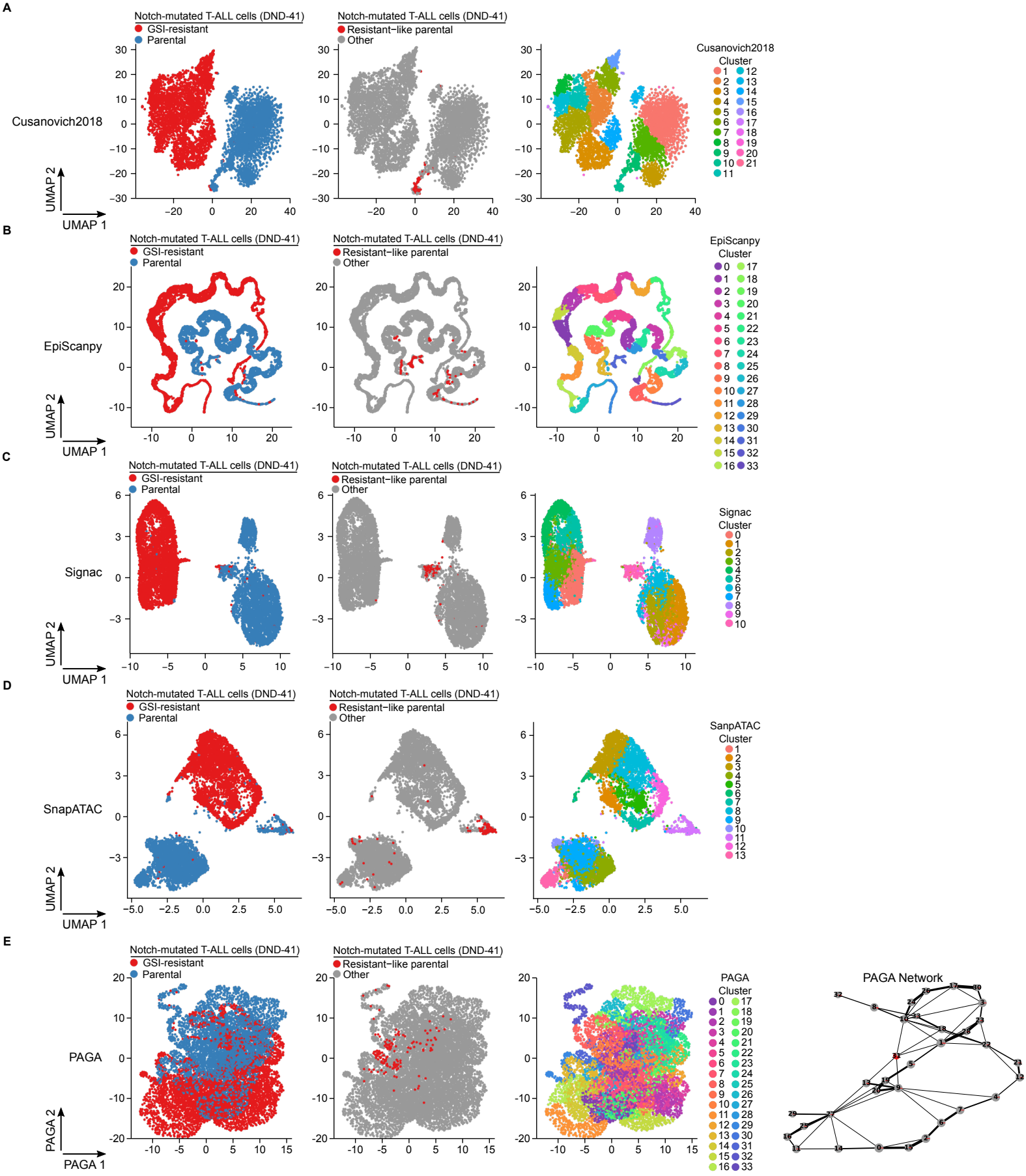
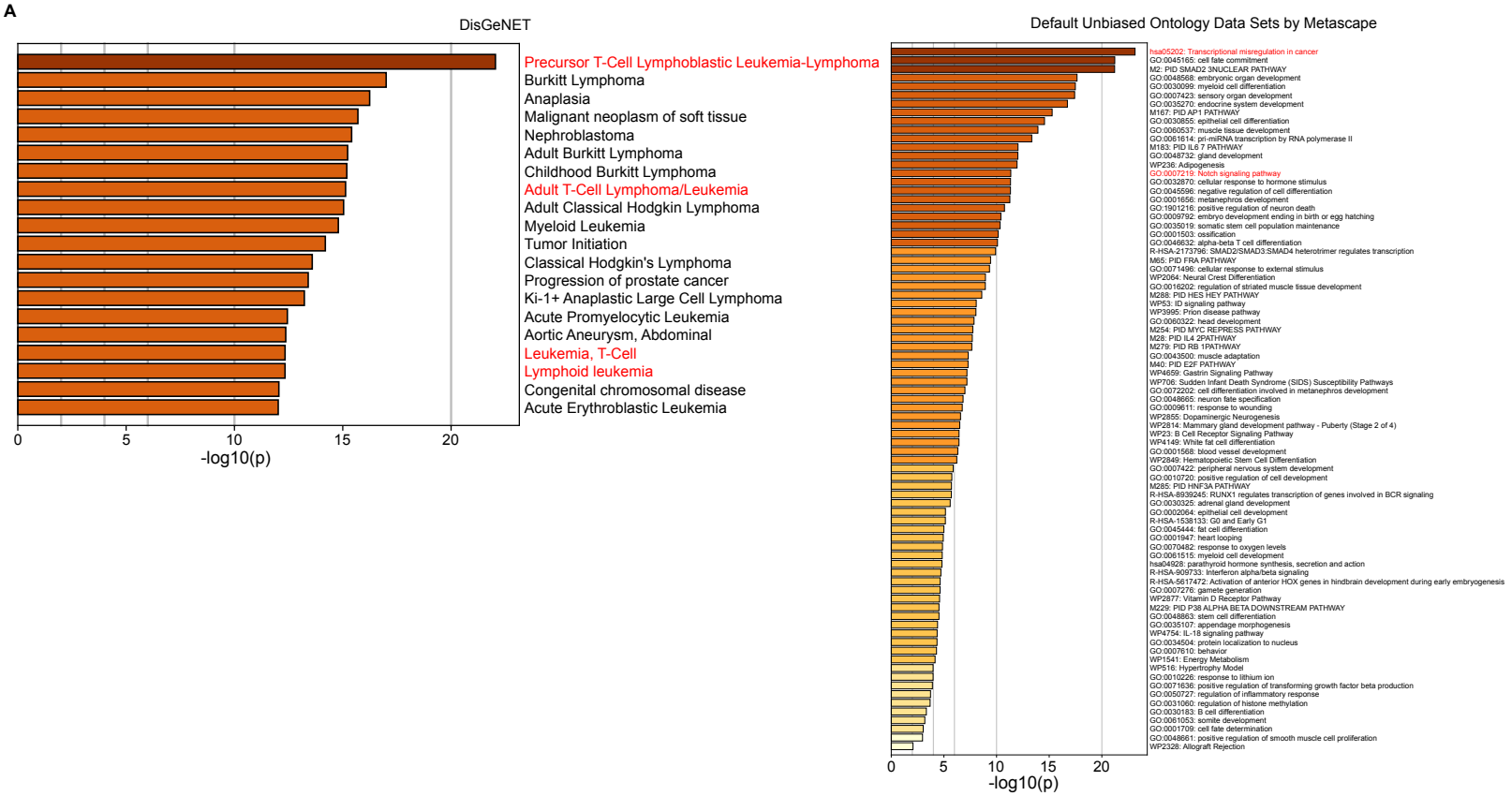

A
Notch-mutated T-ALL cells (DND-41)
GSI-resistant
Parental
Notch-mutated T-ALL cells (DND-41)
Resistant−like parental
Other
Cusanovich2018
Cluster
Cusanovich2018
UMAP 2
UMAP 1
B
Notch-mutated T-ALL cells (DND-41)
GSI-resistant
Parental
Notch-mutated T-ALL cells (DND-41)
Resistant−like parental
Other
EpiScanpy
Cluster
EpiScanpy
UMAP 2
UMAP 1
C
Notch-mutated T-ALL cells (DND-41)
GSI-resistant
Parental
Notch-mutated T-ALL cells (DND-41)
Resistant−like parental
Other
Signac
Cluster
Signac
UMAP 2
UMAP 1
D
Notch-mutated T-ALL cells (DND-41)
GSI-resistant
Parental
Notch-mutated T-ALL cells (DND-41)
Resistant−like parental
Other
SanpATAC
Cluster
SnapATAC
UMAP 2
UMAP 1
E
Notch-mutated T-ALL cells (DND-41)
GSI-resistant
Parental
Notch-mutated T-ALL cells (DND-41)
Resistant−like parental
Other
PAGA
Cluster
PAGA
PAGA 2
PAGA 1
PAGA Network

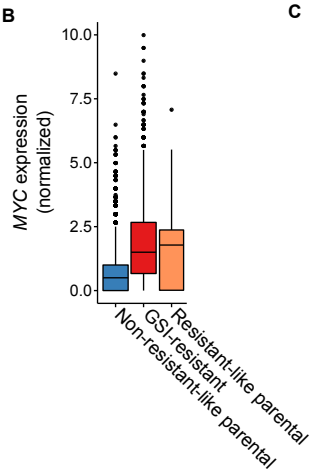
Figure S6: Comparisons of DND-41 T-ALL scATAC-seq data visualization, continued from Figure S5. **(A-D)** projection outputs of Cusanovich2018 (A), EpiScanpy (B), Signac (C), and SnapATAC (D) colored by resistance status (left), resistant-like parental cells as defined by TooManyPeaks (middle), or algorithm cluster assignment (right). (**E**) Projection output of PAGA-initiated UMAP colored by resistant status (left), resistant-like parental (second from left), cluster assignment (second from right). Right panel shows PAGA network colored by resistant-like parental cells ($n = 7{,}989\,\text{cells}$). For an unbiased comparison, each projection used the corresponding package's UMAP or t-SNE implementation. Moreover, default or suggested filters and parameters were used for all algorithms unless otherwise noted (see STAR Methods). Related to Figure 3.

Figure S7

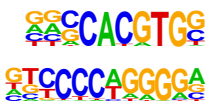
A



B

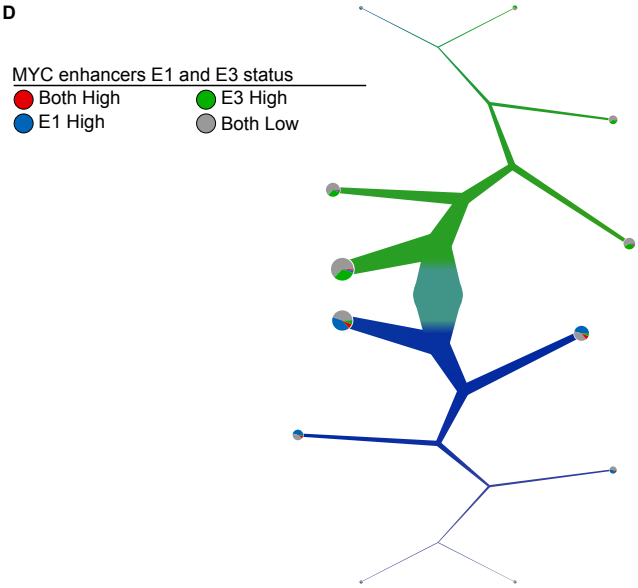


C

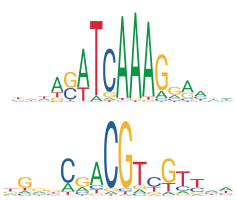


Motif	Motif p-value	Motif q-value	$\log_2$ FoldChange	Expression p-value	Expression q-value
c-Myc	1e-6	0.00	1.08	5.67e-6	1.94e-3
EBF1(EBF)	1e-24	0.00	0.856	0.0228	0.397

D



E



Motif	Motif p-value	Motif q-value	$\log_2$ FoldChange	Expression p-value	Expression q-value
TCF7	1.33e-5	0.0306	-2.63	2.45e-7	9.30e-5
GMEB1	-0.987	0.0283	-0.987	0.0283	0.350

F

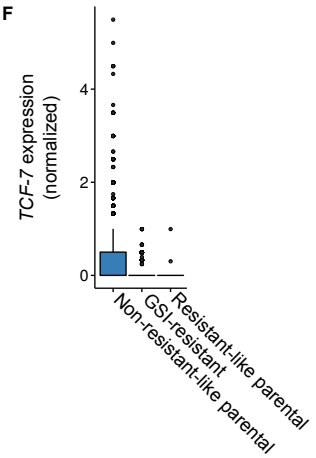


Figure S7: MYC and TCF-1 potentially bind differentially accessible elements of resistant-like parental cells. (**A**) Metascape analysis with DisGeNET database (left) and Ontology gene sets (right) showing that transcription factors with enriched binding motifs at the differential accessible elements of resistant-like parental cells are associated with pathways involved in T cell development and malignancies (highlighted in red, see STAR Methods). (**B**) *MYC* expression levels in GSI-resistant, resistant-like parental, and non-resistant-like parental cells identified in the TooManyCells tree ($n = 7{,}371$ cells). (**C**) Transcription factors from HOMER with both significant upregulation and more accessible binding motif recognition sites in resistant-like parental compared to non-resistant-like parental cells. (**D**) TooManyPeaks tree of parental and GSI-resistant DND-41 cells as in Figure 3A showing the accessibility of the Notch-dependent and Notch-independent *MYC* enhancers ($n = 7{,}989$ cells). (**E**) Transcription factors from JASPAR (Sandelin et al., 2004) with both significantly lower expression and accessible binding motif recognition sites at the *LINC00977*-proximal enhancer E2 of the resistant-like compared to non-resistant-like parental cells. (**f**) *TCF-7* expression levels in GSI-resistant, resistant-like parental, and non-resistant-like parental cells identified in the TooManyCells tree (Schwartz et al., 2020) ($n = 7{,}371$ cells). Related to Figure 3.