Figure S1

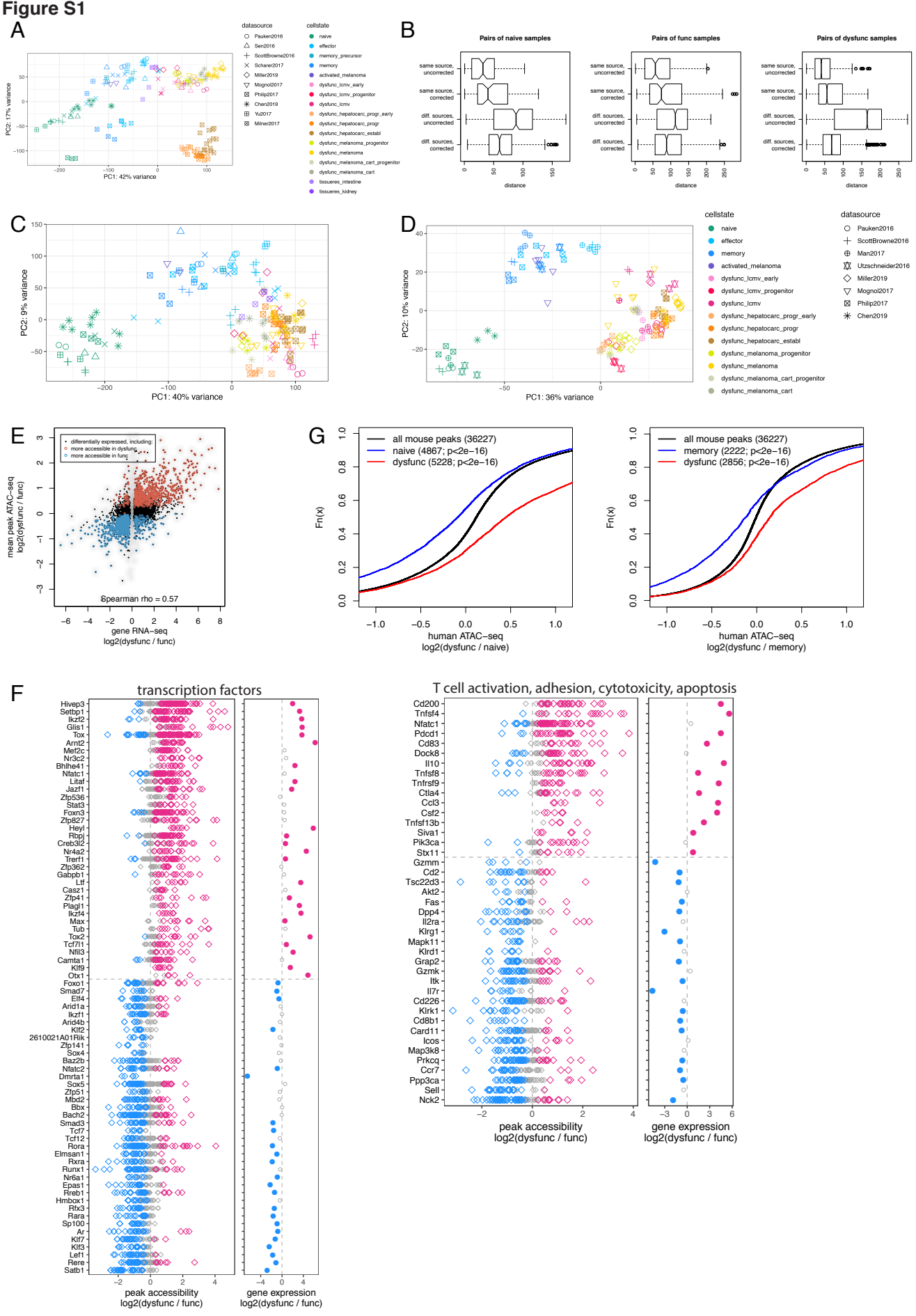

A
datasource
Pauken2016
Sen2016
ScottBrowne2016
Scharer2017
Miller2019
Mognol2017
Philip2017
Chen2019
Yu2017
Milner2017
cellstate
naive
effector
memory_precursor
memory
activated_melanoma
dysfunc_lcmv_early
dysfunc_lcmv_progenitor
dysfunc_lcmv
dysfunc_hepatocarc_progr_early
dysfunc_hepatocarc_progr
dysfunc_hepatocarc_establ
dysfunc_melanoma_progenitor
dysfunc_melanoma
dysfunc_melanoma_cart_progenitor
dysfunc_melanoma_cart
tissueres_intestine
tissueres_kidney
PC1: 42% variance
PC2: 17% variance
B
Pairs of naive samples
Pairs of func samples
Pairs of dysfunc samples
same source, uncorrected
same source, corrected
diff. sources, uncorrected
diff. sources, corrected
distance
C
PC1: 40% variance
PC2: 9% variance
D
PC1: 36% variance
PC2: 10% variance
cellstate
naive
effector
memory
activated_melanoma
dysfunc_lcmv_early
dysfunc_lcmv_progenitor
dysfunc_lcmv
dysfunc_hepatocarc_progr_early
dysfunc_hepatocarc_progr
dysfunc_hepatocarc_establ
dysfunc_melanoma_progenitor
dysfunc_melanoma
dysfunc_melanoma_cart_progenitor
dysfunc_melanoma_cart
datasource
Pauken2016
ScottBrowne2016
Man2017
Utzschneider2016
Miller2019
Mognol2017
Philip2017
Chen2019
E
differentially expressed, including:
more accessible in dysfunc
more accessible in func
Spearman rho = 0.57
mean peak ATAC−seq log2(dysfunc / func)
gene RNA−seq log2(dysfunc / func)
G
all mouse peaks (36227)
naive (4867; p<2e−16)
dysfunc (5228; p<2e−16)
Fn(x)
human ATAC−seq log2(dysfunc / naive)
all mouse peaks (36227)
memory (2222; p<2e−16)
dysfunc (2856; p<2e−16)
human ATAC−seq log2(dysfunc / memory)
F
transcription factors
Hivep3
Setbp1
Ikzf2
Glis1
Tox
Arnt2
Mef2c
Nr3c2
Bhlhe41
Nfatc1
Litaf
Jazf1
Zfp536
Stat3
Foxn3
Zfp827
Heyl
Rbpj
Creb3l2
Nr4a2
Trerf1
Zfp362
Gabpb1
Ltf
Casz1
Zfp41
Plagl1
Ikzf4
Max
Tub
Tox2
Tcf7l1
Nfil3
Camta1
Klf9
Otx1
Foxo1
Smad7
Elf4
Arid1a
Ikzf1
Arid4b
Klf2
2610021A01Rik
Zfp141
Sox4
Baz2b
Nfatc2
Dmrta1
Sox5
Zfp51
Mbd2
Bbx
Bach2
Smad3
Tcf7
Tcf12
Rora
Elmsan1
Rxra
Runx1
Nr6a1
Epas1
Rreb1
Hmbox1
Rfx3
Rara
Sp100
Ar
Klf7
Klf3
Lef1
Rere
Satb1
peak accessibility log2(dysfunc / func)
gene expression log2(dysfunc / func)
T cell activation, adhesion, cytotoxicity, apoptosis
Cd200
Tnfsf4
Nfatc1
Pdcd1
Cd83
Dock8
Il10
Tnfsf8
Tnfrsf9
Ctla4
Ccl3
Csf2
Tnfsf13b
Siva1
Pik3ca
Stx11
Gzmm
Cd2
Tsc22d3
Akt2
Fas
Dpp4
Il2ra
Klrg1
Mapk11
Klrd1
Grap2
Gzmk
Itk
Il7r
Cd226
Klrk1
Cd8b1
Card11
Icos
Map3k8
Prkcq
Ccr7
Ppp3ca
Sell
Nck2
peak accessibility log2(dysfunc / func)
gene expression log2(dysfunc / func)

Figure S1. Universal program of CD8 T cell dysfunction (related to Figure 1). A. Principal component analysis (PCA) of library-size normalized ATAC-seq read counts in 150bp windows around peak summits (functional cell state shown by color, data source by symbol) without batch-effect correction. **B.** Distributions of distances between ATAC-seq vectors of read counts in 150bp windows around peak summits. Distances were calculated in the two-dimensional PCA space built using 20000 most variable peak summit read counts, for pairs of samples in the same functional category (naïve, functional, dysfunctional), separately for pairs of samples from the same study and from different studies, before and after batch effect correction. **C.** PCA of library-size normalized batch-effect corrected ATAC-seq read counts in 150bp windows around peak summits (labels as in **A**) using 10000 most variable peak summit read counts. Here GLM-based batch effect correction used a factor encoding the functional cell state using two values, for naïve cells and all other cells. **D.** PCA of library-size normalized batch-effect corrected RNA-seq read counts for 1000 genes with most variable counts (functional cell state shown by color, data source by symbol). **E.** Scatter plot of differential expression (RNA-seq log2FC, x-axis) and differential accessibility (mean ATAC-seq log2FC over all peaks associated with a gene, y-axis) between functional and dysfunctional cells (computed as in **D**). Significantly differentially expressed genes shown as black dots. Significantly differentially accessible genes highlighted with red or blue color. **F.** "Diamond" plots of differential accessibility and differential expression between functional and dysfunctional cells for transcription factors and for genes associated with T cell activation, cytotoxicity, adhesion, and apoptosis. In each panel, left: library-size normalized batch-effect corrected ATAC-seq read count log2FC for peak summits (diamond shown in color for significantly decreased/increased, FDR < 0.05) of significantly differentially accessible genes; right: log2 fold change of RNA-seq gene expression for the same genes. **G.** Cumulative distribution function of differential accessibility (ATAC-seq log2FC) between naive and dysfunctional (left) or central memory and dysfunctional (right) CD8 T cells in human donors or cancer patients. Shown are peaks evolutionarily conserved when compared with peak summit regions in mice that were significantly more accessible in naïve or memory (blue) or dysfunctional (red) cells; background distribution is for all peaks identified in human cells that were conserved with mouse peak summit regions (black); count and p-value from Kolmogorov-Smirnov test against the background distribution.

Figure S2

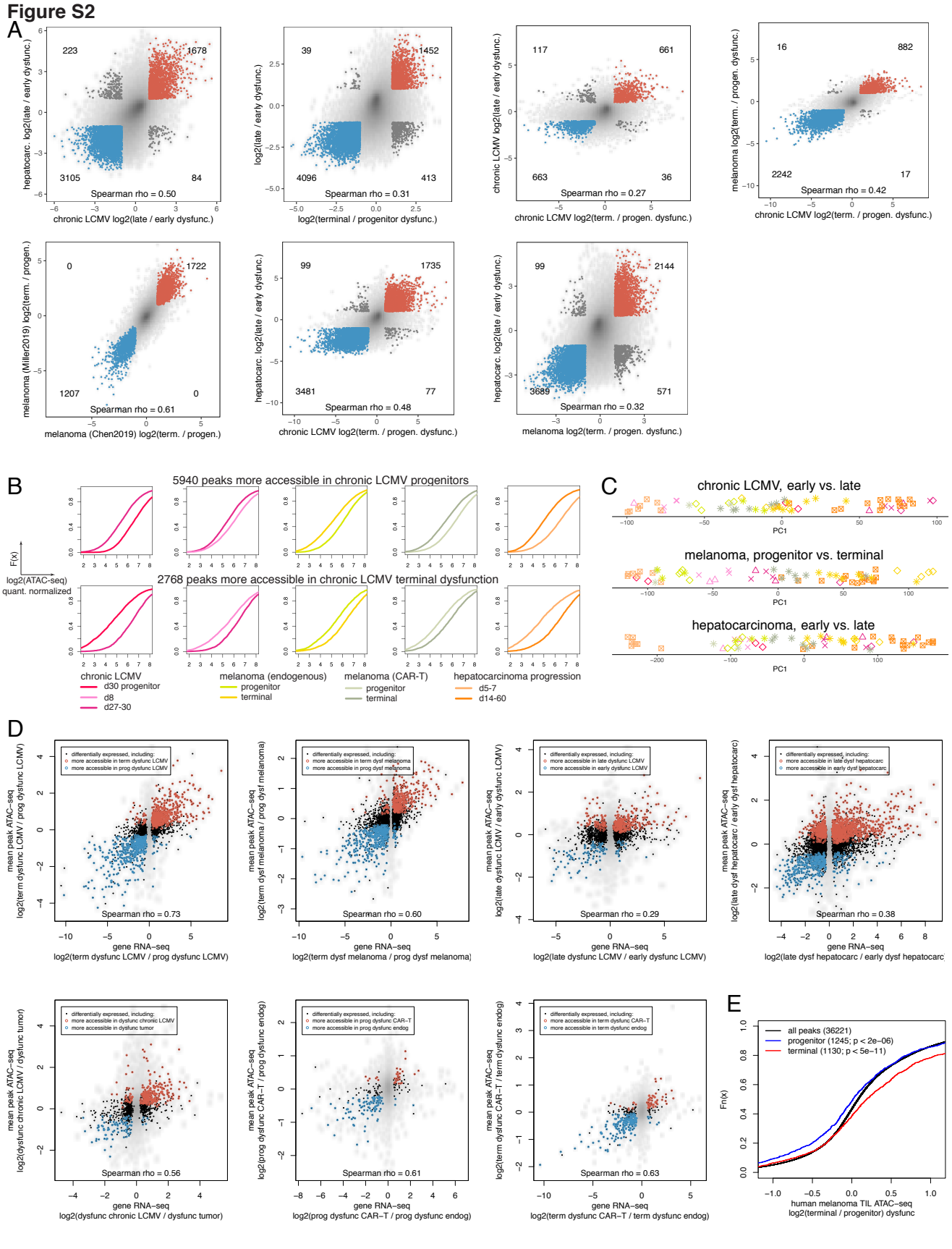

A
223
1678
3105
84
Spearman rho = 0.50
hepatocarc. log2(late / early dysfunc.)
chronic LCMV log2(late / early dysfunc.)
39
1452
4096
413
Spearman rho = 0.31
log2(late / early dysfunc.)
log2(terminal / progenitor dysfunc.)
117
661
663
36
Spearman rho = 0.27
chronic LCMV log2(late / early dysfunc.)
chronic LCMV log2(term. / progen. dysfunc.)
16
882
2242
17
Spearman rho = 0.42
melanoma log2(term. / progen. dysfunc.)
chronic LCMV log2(term. / progen. dysfunc.)
0
1722
1207
0
Spearman rho = 0.61
melanoma (Miller2019) log2(term. / progen.)
melanoma (Chen2019) log2(term. / progen.)
99
1735
3481
77
Spearman rho = 0.48
hepatocarc. log2(late / early dysfunc.)
chronic LCMV log2(term. / progen. dysfunc.)
99
2144
3689
571
Spearman rho = 0.32
hepatocarc. log2(late / early dysfunc.)
melanoma log2(term. / progen. dysfunc.)
B
5940 peaks more accessible in chronic LCMV progenitors
2768 peaks more accessible in chronic LCMV terminal dysfunction
F(x)
log2(ATAC-seq) quant. normalized
chronic LCMV
d30 progenitor
d8
d27-30
melanoma (endogenous)
progenitor
terminal
melanoma (CAR-T)
progenitor
terminal
hepatocarcinoma progression
d5-7
d14-60
C
chronic LCMV, early vs. late
melanoma, progenitor vs. terminal
hepatocarcinoma, early vs. late
PC1
D
differentially expressed, including:
more accessible in term dysfunc LCMV
more accessible in prog dysfunc LCMV
Spearman rho = 0.73
mean peak ATAC−seq log2(term dysfunc LCMV / prog dysfunc LCMV)
gene RNA−seq log2(term dysfunc LCMV / prog dysfunc LCMV)
more accessible in term dysf melanoma
more accessible in prog dysf melanoma
Spearman rho = 0.60
mean peak ATAC−seq log2(term dysf melanoma / prog dysf melanoma)
gene RNA−seq log2(term dysf melanoma / prog dysf melanoma)
more accessible in late dysfunc LCMV
more accessible in early dysfunc LCMV
Spearman rho = 0.29
mean peak ATAC−seq log2(late dysfunc LCMV / early dysfunc LCMV)
gene RNA−seq log2(late dysfunc LCMV / early dysfunc LCMV)
more accessible in late dysf hepatocarc
more accessible in early dysf hepatocarc
Spearman rho = 0.38
mean peak ATAC−seq log2(late dysf hepatocarc / early dysf hepatocarc)
gene RNA−seq log2(late dysf hepatocarc / early dysf hepatocarc)
more accessible in dysfunc chronic LCMV
more accessible in dysfunc tumor
Spearman rho = 0.56
mean peak ATAC−seq log2(dysfunc chronic LCMV / dysfunc tumor)
gene RNA−seq log2(dysfunc chronic LCMV / dysfunc tumor)
more accessible in prog dysfunc CAR−T
more accessible in prog dysfunc endog
Spearman rho = 0.61
mean peak ATAC−seq log2(prog dysfunc CAR−T / prog dysfunc endog)
gene RNA−seq log2(prog dysfunc CAR−T / prog dysfunc endog)
more accessible in term dysfunc CAR−T
more accessible in term dysfunc endog
Spearman rho = 0.63
mean peak ATAC−seq log2(term dysfunc CAR−T / term dysfunc endog)
gene RNA−seq log2(term dysfunc CAR−T / term dysfunc endog)
E
all peaks (36221)
progenitor (1245; p < 2e−06)
terminal (1130; p < 5e−11)
Fn(x)
human melanoma TIL ATAC−seq log2(terminal / progenitor) dysfunc

Figure S2. Progression from early or progenitor to terminal dysfunctional state (related to Figure 1). A. Comparisons of differential accessibility results. Each plot is a scatter plot of log2FC of library-size normalized batch-effect corrected ATAC-seq read counts in peak summit regions for two differential accessibility analyses. Color (blue, red, dark gray) is used to highlight quadrants with peak summit regions significantly differentially accessible in both pairwise comparisons. Counts in corners indicate the number of such peak summits in each quadrant. Spearman's correlation is calculated over all such peak summits. **B.** Cumulative distribution function (CDF) of log-transformed quantile-normalized library-size normalized batch-effect corrected ATAC-seq read counts from various studies in 5940 peak summits significantly more accessible in progenitor than terminally dysfunctional T cells and in 2768 peak summits significantly more accessible in terminally than progenitor dysfunctional T cells in chronic LCMV infection (Kolmogorov-Smirnov $p < 2e\text{-}16$ for all comparisons). **C.** First principal component (PC1) of PCA for library-size normalized batch-effect corrected ATAC-seq read counts in dysfunctional T cells from different studies (see Figure 1C). PCA was calculated based on significantly differentially accessible peaks in different pairwise comparisons. For clarity, vertical random jiggle is added. **D.** Scatter plots of differential expression (RNA-seq log2FC, x-axis) and differential accessibility (mean ATAC-seq log2FC over all peaks associated with a gene, y-axis) between cell states. Significantly differentially expressed genes shown as black dots. Significantly differentially accessible genes highlighted with red or blue color. **E.** Cumulative distribution function of differential accessibility (ATAC-seq log2FC) between progenitor (TIM3- CD39-) and terminally (TIM3+ CD39+) dysfunctional TILs in human tumors. Shown are peaks evolutionarily conserved when compared with peak summit regions in mouse melanoma model that were significantly more accessible in progenitor (blue) or terminally dysfunctional (red) cells; background distribution is for all peaks identified in human cells that were conserved with mouse peak summit regions (black); count and p-value from Kolmogorov-Smirnov test against the background distribution.

Figure S3

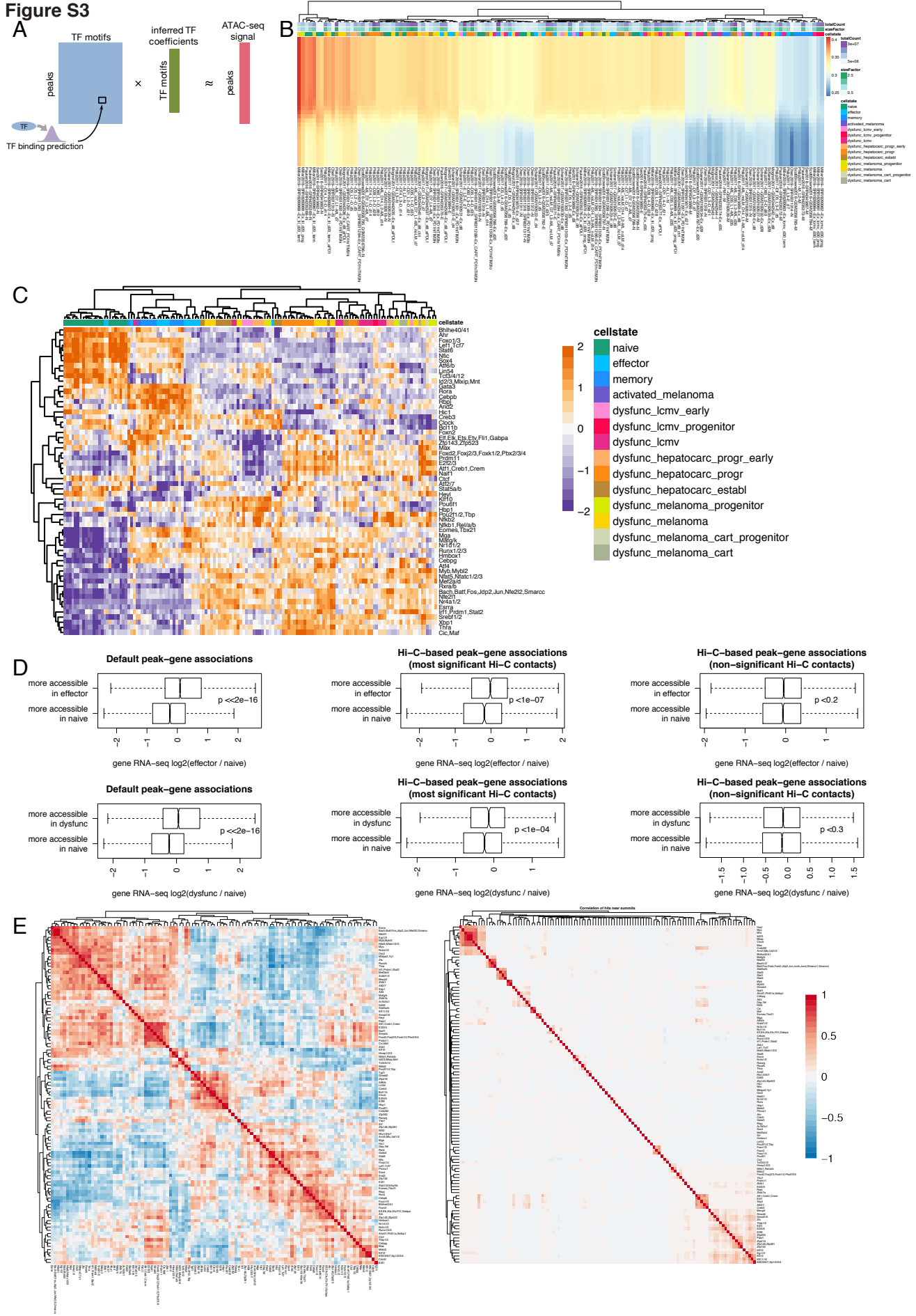
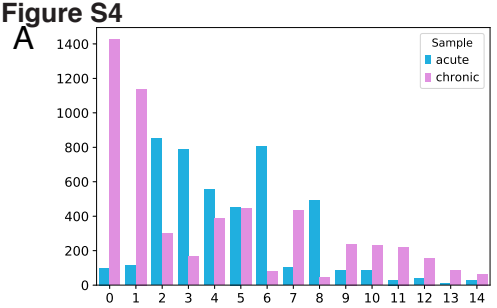
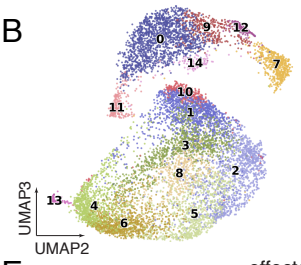
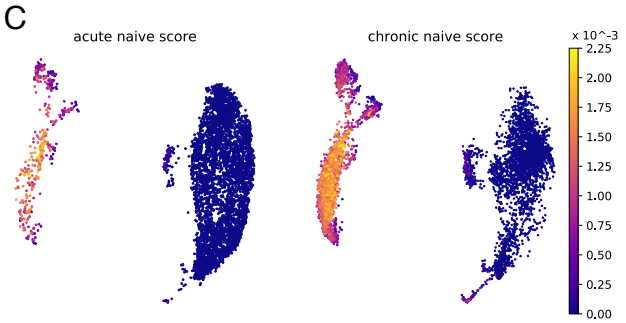

A
TF motifs
peaks
TF
TF binding prediction
inferred TF coefficients
TF motifs
ATAC-seq signal
peaks
B
totalCount
sizeFactor
cellstate
naive
effector
memory
activated_melanoma
dysfunc_lcmv_early
dysfunc_lcmv_progenitor
dysfunc_lcmv
dysfunc_hepatocarc_progr_early
dysfunc_hepatocarc_progr
dysfunc_hepatocarc_establ
dysfunc_melanoma_progenitor
dysfunc_melanoma
dysfunc_melanoma_cart_progenitor
dysfunc_melanoma_cart
C
cellstate
Bhlhe40,41
Ahr
Foxo1/3
Lef1,Tcf7
Stat6
Nfic
Sox4
Atf6/b
Lin54
Tcf3/4/12
Id2/3,Mlxip,Mnt
Gata3
Rora
Cebpb
Rbpl
Ar/d2
Hic1
Creb3
Clock
Bcl11b
Foxn2
Elf,Elk,Ets,Etv,Fli1,Gabpa
Zfp143,Zfp523
Mlx
Foxd2,Foxj2/3,Foxk1/2,Pbx2/3/4
Pou2f1
Ezh2/3
Atf1,Creb1,Crem
Nair1
Ctcf
Atf2/7
Stat5a/b
Hlyl
Klf10
Pou6f1
Hbp1
Pou2f1/2,Tbp
Nfkb2
Nfkb1,Rel/a/b
Eomes,Tbx21
Mga
Mafg/k
Nr1d1/2
Runx1/2/3
Hmbox1
Cebpg
Atf4
Myb,Mybl2
Nfat5,Nfatc1/2/3
Mef2d/g
Foxa/b
Bach,Batf,Fos,Jdp2,Jun,Nfe2l2,Smarcc
Nfe2l1
Nr4a1/2
Esrra
Irf1,Prdm1,Stat2
Srebf1/2
Xbp1
Thra
Cic,Maf
naive
effector
memory
activated_melanoma
dysfunc_lcmv_early
dysfunc_lcmv_progenitor
dysfunc_lcmv
dysfunc_hepatocarc_progr_early
dysfunc_hepatocarc_progr
dysfunc_hepatocarc_establ
dysfunc_melanoma_progenitor
dysfunc_melanoma
dysfunc_melanoma_cart_progenitor
dysfunc_melanoma_cart
D
Default peak–gene associations
more accessible in effector
more accessible in naive
p <<2e−16
gene RNA−seq log2(effector / naive)
Hi–C–based peak–gene associations (most significant Hi–C contacts)
p <1e−07
Hi–C–based peak–gene associations (non–significant Hi–C contacts)
p <0.2
Default peak–gene associations
more accessible in dysfunc
more accessible in naive
p <<2e−16
gene RNA−seq log2(dysfunc / naive)
Hi–C–based peak–gene associations (most significant Hi–C contacts)
p <1e−04
Hi–C–based peak–gene associations (non–significant Hi–C contacts)
p <0.3
E
Correlation of hits near summits

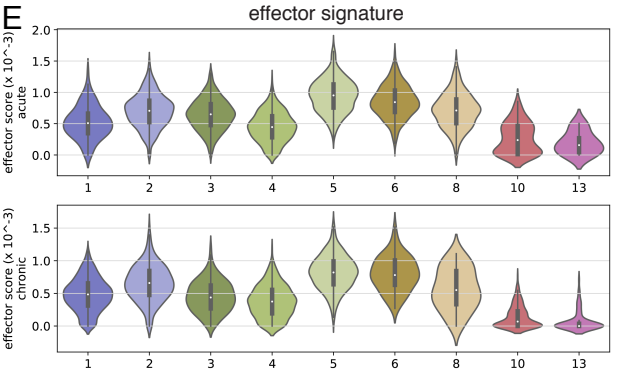
Figure S3. Predictive models of transcription factor association with chromatin accessibility (related to Figure 2). A. Schematic of the negative binomial generalized linear regression analysis to infer transcription factor (TF) associations with chromatin accessibility. **B.** Spearman correlations between library-size normalized batch-effect corrected ATAC-seq read counts and model predictions for each sample and each value of the regularization parameter. **C.** Inferred TF motif coefficients for regressing library-size normalized batch-effect corrected ATAC-seq read counts in each sample, for motifs corresponding to coefficients with the highest variance across samples, z-score normalized within each row. **D.** Top: Boxplots of log2FC of RNA-seq gene expression between naïve and effector cells for genes associated with at least one peak significantly more accessible in naïve than effector cells and associated with at least one peak significantly more accessible in effector than in naïve cells. Left: default peak-gene associations (peak associated to the closest gene in genomic coordinates, if this gene is within 50Kbp). Center: peak–gene associations defined using significant Hi-C contacts in naïve cells. Right: peak–gene associations defined using non-significant Hi-C contacts in naïve cells. Bottom: same analysis for naïve vs. dysfunctional cells. **E.** Spearman correlation of inferred TF motif coefficients across all antigen-experienced samples (left) and of TF motif match scores across the atlas of peak summit regions (right).

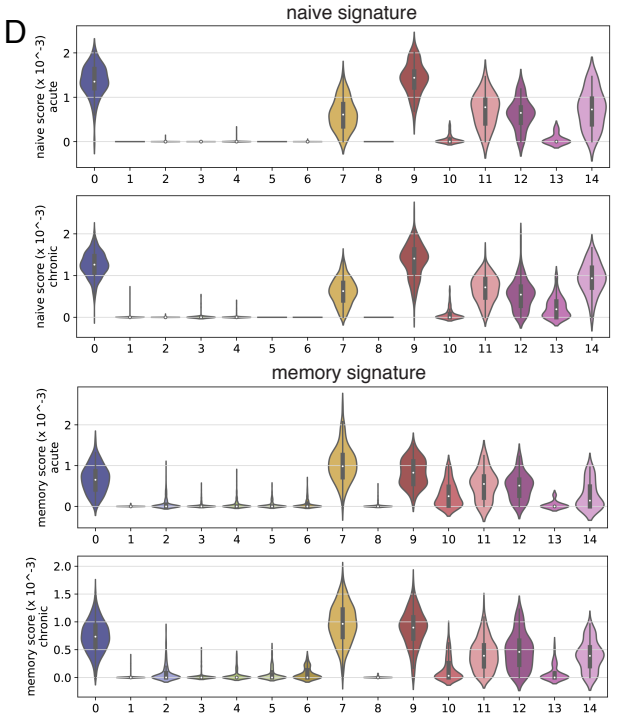
Figure S4

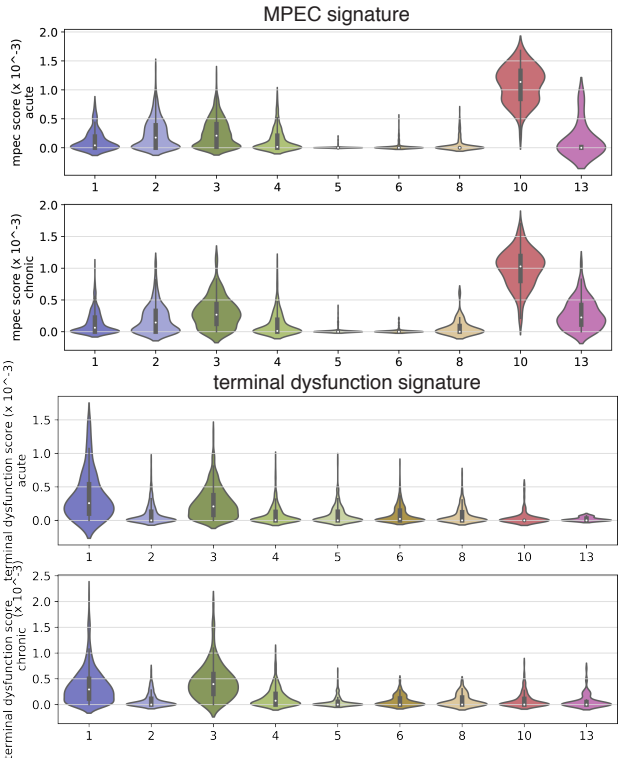

A
Sample
acute
chronic


B
UMAP3
UMAP2


C
acute naive score
chronic naive score
x 10^-3


E
effector signature
effector score (x 10^-3) acute
effector score (x 10^-3) chronic


D
naive signature
naive score (x 10^-3) acute
naive score (x 10^-3) chronic
memory signature
memory score (x 10^-3) acute
memory score (x 10^-3) chronic


MPEC signature
mpec score (x 10^-3) acute
mpec score (x 10^-3) chronic
terminal dysfunction signature
terminal dysfunction score (x 10^-3) acute
terminal dysfunction score (x 10^-3) chronic

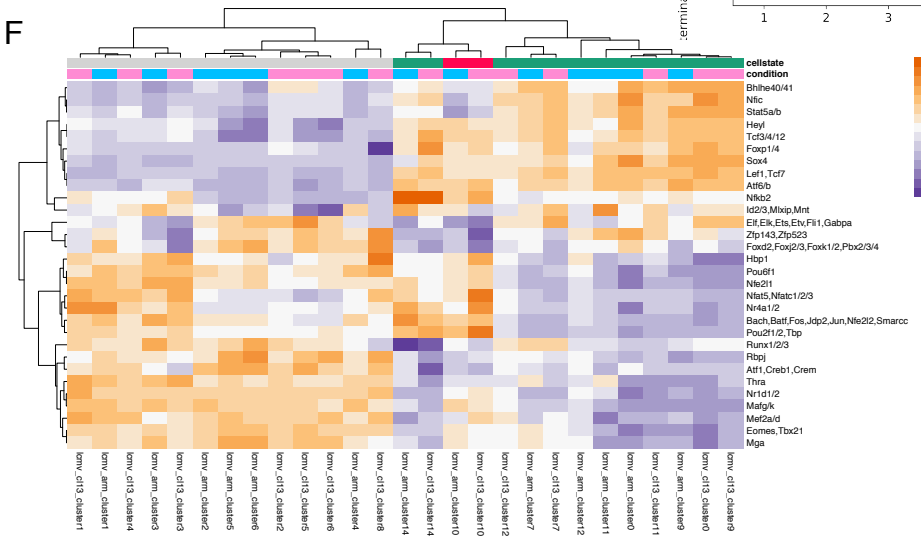

F
cellstate
condition
Bhlhe40/41
Nfic
Stat5a/b
Heyl
Tcf3/4/12
Foxp1/4
Sox4
Lef1,Tcf7
Atf6/b
Nfkb2
Id2/3,Mlxip,Mnt
Elf,Elk,Ets,Etv,Fli1,Gabpa
Zfp143,Zfp523
Foxd2,Foxj2/3,Foxk1/2,Pbx2/3/4
Hbp1
Pou6f1
Nfe2l1
Nfat5,Nfatc1/2/3
Nr4a1/2
Bach,Batf,Fos,Jdp2,Jun,Nfe2l2,Smarcc
Pou2f1/2,Tbp
Runx1/2/3
Rbpj
Atf1,Creb1,Crem
Thra
Nr1d1/2
Mafg/k
Mef2a/d
Eomes,Tbx21
Mga
cellstate
naive,memory
progenitor
activated
condition
arm
cl13

Figure S4. scATAC-seq analysis of CD8 T cells in acute and chronic infection (related to Figure 3). A. Barplot showing the number of cells in each cluster from each sample. **B.** Components 2 and 3 of the UMAP representation of TF-IDF-transformed scATAC-seq data with Louvain clusters. **C.** Heatmap for single cells in scATAC-seq data, separately for acute and chronic infection, showing the naïve cell signature derived from bulk ATAC-seq data. **D.** Violin plots for scores in scATAC-seq data clusters, separately for acute and chronic infection, of peak signatures (for naïve cells, memory cells) derived from bulk ATAC-seq data. **E.** Violin plots for scores in scATAC-seq data clusters 1-8, 10, 13, separately for acute and chronic infection, of peak signatures (for effector cells, memory precursor cells, terminally dysfunctional cells) derived from bulk ATAC-seq data. **F.** Inferred TF motif coefficients for scATAC-seq counts averaged over cells in each cluster in each of the two samples. Inferred coefficients with the highest variance are shown (z-score row normalized).

Figure S5

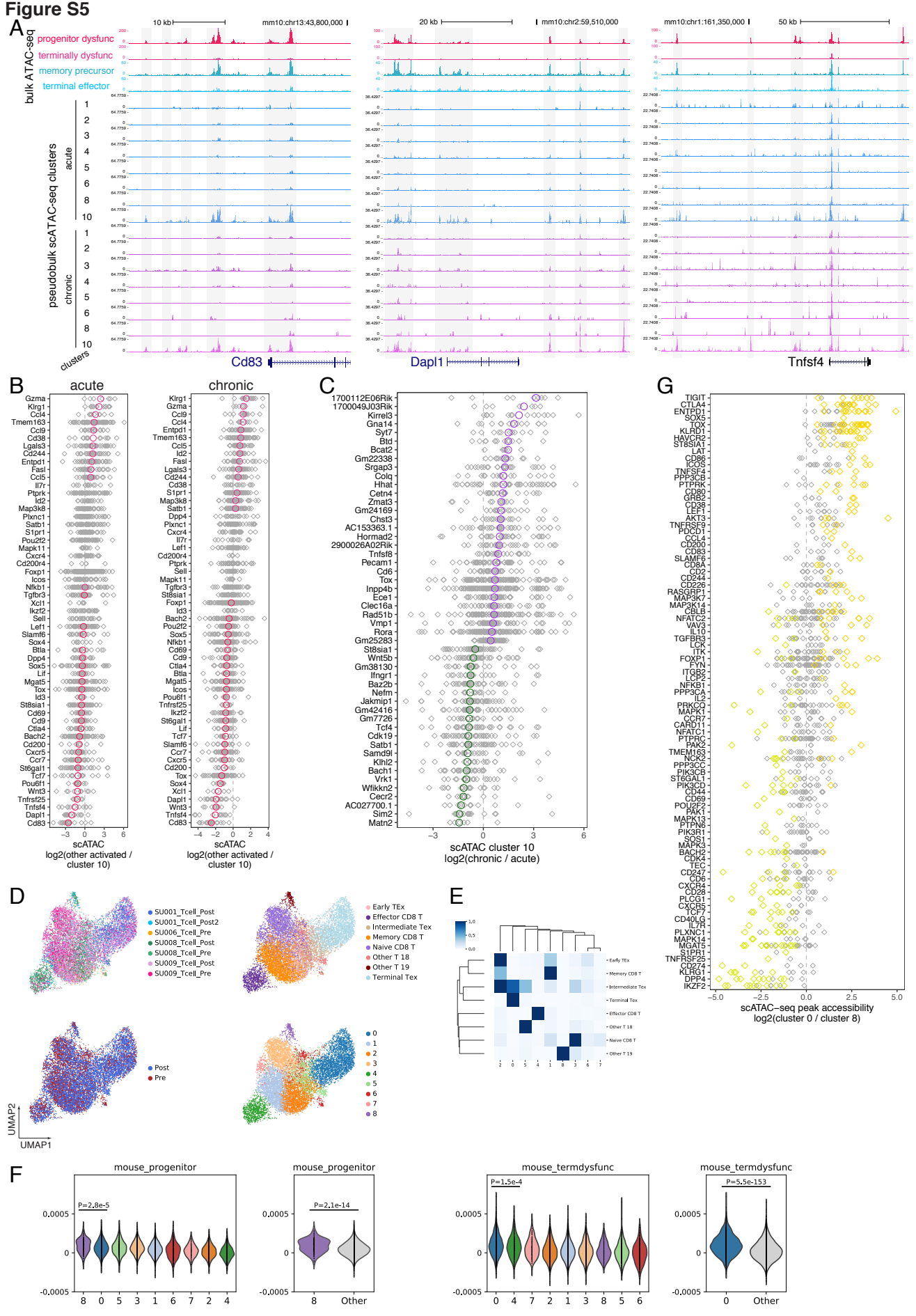

A
bulk ATAC-seq
progenitor dysfunc
terminally dysfunc
memory precursor
terminal effector
pseudobulk scATAC-seq clusters
acute
chronic
clusters
10 kb
mm10:chr13:43,800,000
20 kb
mm10:chr2:59,510,000
mm10:chr1:161,350,000
50 kb
Cd83
Dapl1
Tnfsf4
B
acute
chronic
scATAC
log2(other activated / cluster 10)
C
scATAC cluster 10
log2(chronic / acute)
G
scATAC–seq peak accessibility
log2(cluster 0 / cluster 8)
D
SU001_Tcell_Post
SU001_Tcell_Post2
SU006_Tcell_Pre
SU008_Tcell_Post
SU008_Tcell_Pre
SU009_Tcell_Post
SU009_Tcell_Pre
Early TEx
Effector CD8 T
Intermediate Tex
Memory CD8 T
Naive CD8 T
Other T 18
Other T 19
Terminal Tex
Post
Pre
UMAP2
UMAP1
E
F
mouse_progenitor
mouse_termdysfunc
P=2.8e-5
P=2.1e-14
P=1.5e-4
P=5.5e-153
Other

Figure S5. scATAC-seq analysis reveals a similar population of precursor/progenitor cells in acute and chronic infection (related to Figure 3). A. Genome browser tracks of ATAC-seq data for selected peaks. Bulk ATAC-seq for progenitor and terminally dysfunctional cells, and for terminal effector and memory precursor cells. Normalized aggregated single-cell ATAC-seq for cells in each cluster (for clusters 1-8, 10) in each of the two samples. For genes significantly more accessible in scATAC-seq cluster 10 as compared with other clusters. **B.** Differential accessibility of genes in cluster 10. Each panel shows normalized scATAC-seq pseudo-count (averaged over cells in each cluster in each sample) log2 fold change for peak summits of significantly differentially accessible genes; mean log2FC value is highlighted (with a shade of red) for genes that were overall differentially accessible. Shown are comparisons between cells in cluster 10 and cells in clusters 1-8, separately for acute and chronic infection, for genes highlighted in Figure 1E. **C.** Same as B, but for the most significantly differentially accessible genes between cells from acute and chronic infection within cluster 10. **D.** UMAP for scATAC-seq of CD8 T cells from human cancer patients (Satpathy *et al.* 2019). Top and bottom left: patient label and treatment status of each sample. Top right: clusters identified in the original publication. Bottom right: newly identified clusters (Methods). **E.** Comparison of new clusters with those from the original publication. Shown is a fraction of cells in each of the published clusters that belong to each of the new clusters 0-8. **F.** Violin plots for scores in new human scATAC-seq clusters 0-8 of peak signatures derived from bulk ATAC-seq data in mice. Shown are comparisons of all clusters (left) and of the cluster with the highest median score against the union of all other clusters (right). The results suggest a classification of cluster 8 as progenitor dysfunctional cells, and cluster 0 as terminally dysfunctional cells. **G.** Differential accessibility of genes between new human scATAC-seq clusters 0 and 8. Shown are normalized scATAC-seq pseudocount log2 fold changes for peaks of selected genes with at least one significantly differentially accessible peak (color, FDR < 0.1). Most genes show the trend similar to that in differential bulk and single-cell ATAC-seq signal between progenitor and terminally dysfunctional cells in mice, with only a few notable exceptions (e.g. KLRG1, SLAMF6, LEF1).

Figure S6

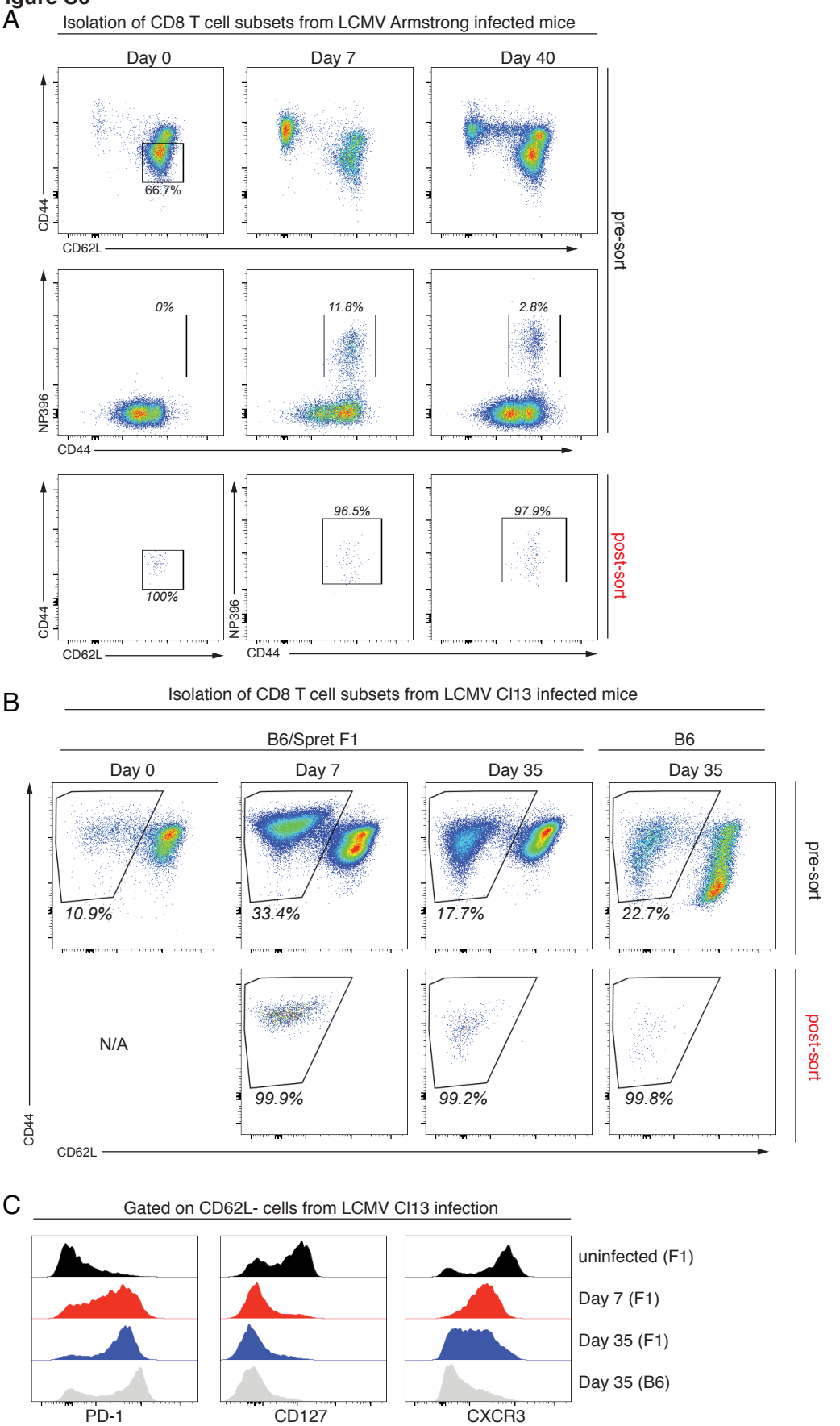

A
Isolation of CD8 T cell subsets from LCMV Armstrong infected mice
Day 0
Day 7
Day 40
66.7%
CD44
CD62L
pre-sort
0%
11.8%
2.8%
NP396
CD44
100%
96.5%
97.9%
post-sort
B
Isolation of CD8 T cell subsets from LCMV Cl13 infected mice
B6/Spret F1
B6
Day 0
Day 7
Day 35
Day 35
10.9%
33.4%
17.7%
22.7%
pre-sort
N/A
99.9%
99.2%
99.8%
post-sort
CD44
CD62L
C
Gated on CD62L- cells from LCMV Cl13 infection
uninfected (F1)
Day 7 (F1)
Day 35 (F1)
Day 35 (B6)
PD-1
CD127
CXCR3

Figure S6. Sorting for CD8 T cell subpopulations for scRNA-seq profiling (related to Figure 4). A-B. Sorting strategy for isolating CD8 T cells responding to acute and chronic LCMV infection. **C.** Protein expression in cells isolated from chronic infection.

Figure S7

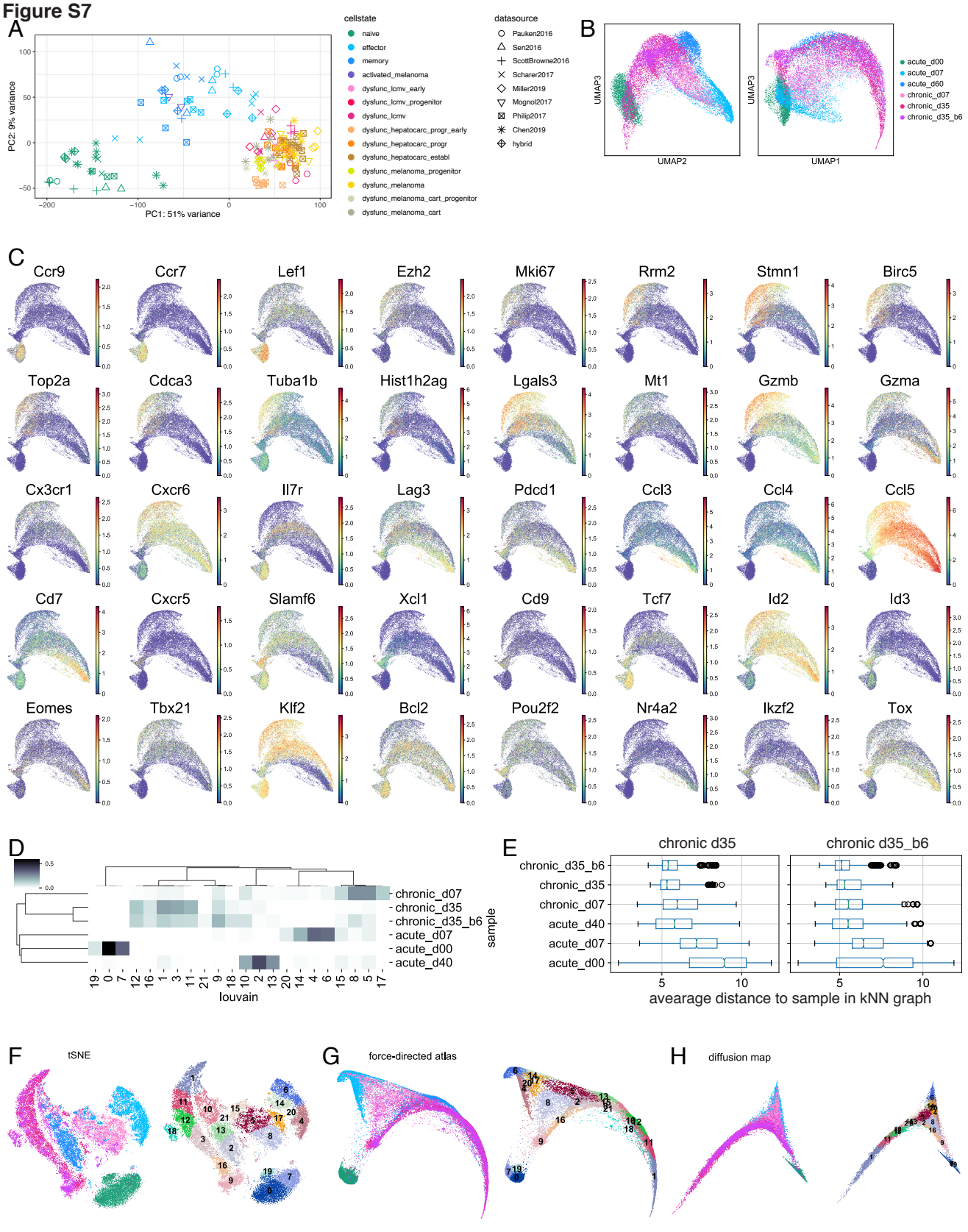

A
PC2: 9% variance
PC1: 51% variance
cellstate
naive
effector
memory
activated_melanoma
dysfunc_lcmv_early
dysfunc_lcmv_progenitor
dysfunc_lcmv
dysfunc_hepatocarc_progr_early
dysfunc_hepatocarc_progr
dysfunc_hepatocarc_establ
dysfunc_melanoma_progenitor
dysfunc_melanoma
dysfunc_melanoma_cart_progenitor
dysfunc_melanoma_cart
datasource
Pauken2016
Sen2016
ScottBrowne2016
Scharer2017
Miller2019
Mognol2017
Philip2017
Chen2019
hybrid
B
UMAP3
UMAP2
UMAP1
acute_d00
acute_d07
acute_d60
chronic_d07
chronic_d35
chronic_d35_b6
C
Ccr9
Ccr7
Lef1
Ezh2
Mki67
Rrm2
Stmn1
Birc5
Top2a
Cdca3
Tuba1b
Hist1h2ag
Lgals3
Mt1
Gzmb
Gzma
Cx3cr1
Cxcr6
Il7r
Lag3
Pdcd1
Ccl3
Ccl4
Ccl5
Cd7
Cxcr5
Slamf6
Xcl1
Cd9
Tcf7
Id2
Id3
Eomes
Tbx21
Klf2
Bcl2
Pou2f2
Nr4a2
Ikzf2
Tox
D
chronic_d07
chronic_d35
chronic_d35_b6
acute_d07
acute_d00
acute_d40
sample
louvain
E
chronic d35
chronic d35_b6
chronic_d35_b6
chronic_d35
chronic_d07
acute_d40
acute_d07
acute_d00
avearage distance to sample in kNN graph
F
tSNE
G
force-directed atlas
H
diffusion map

Figure S7. Dimensionality reduction and visualization of scRNA-seq data (related to Figure 4). A. Principal component analysis (PCA) of library-size normalized batch-effect corrected ATAC-seq read counts in 150bp windows around peak summits (functional cell state shown by color, data source by symbol), including samples from naïve, effector, and memory cells from hybrid F1(B6xSpret) mice. This analysis suggests that CD8 T cell functional states are similar between hybrid and B6 mice. **B.** Projections of three-dimensional UMAP representation of library-size normalized scRNA-seq data. **C.** Log-transformed gene expression (library-size normalized scRNA-seq UMI counts) on UMAP. **D.** Cluster composition of samples. Heatmap showing for each cluster what fraction of cells that cluster occupies in each sample. **E.** Boxplot of the kNN graph distances from cells in samples “chronic_d35” (left) and “chronic_d35_b6” (right) to cells in other samples. For each cell C, average distance in the kNN graph from C to cells in each sample was calculated, and distribution of these values across cells C was plotted. **F-H.** Results of dimensionality reduction methods applied to normalized scRNA-seq data.

Figure S8

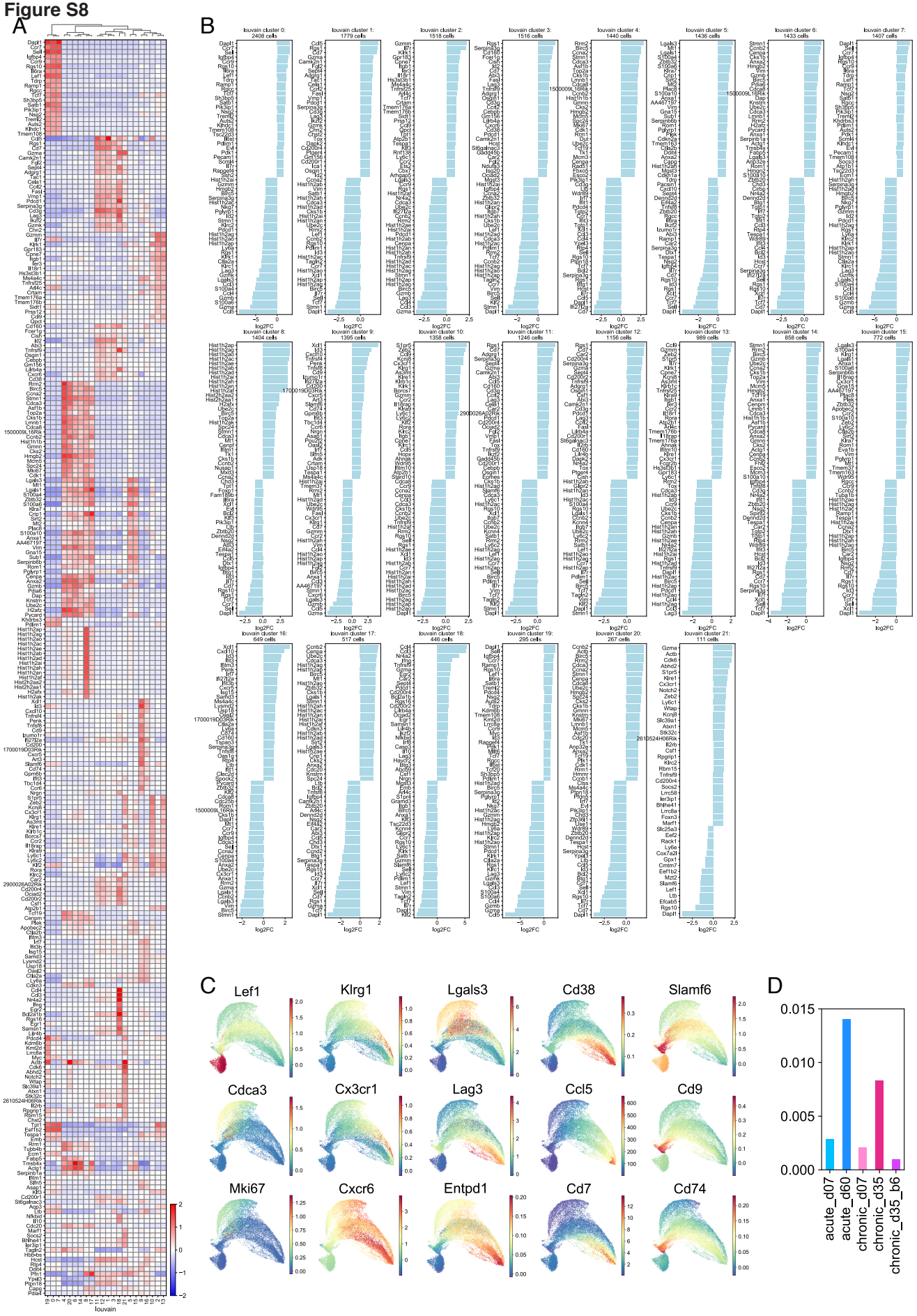

A
B
louvain cluster 0: 2408 cells
louvain cluster 1: 1779 cells
louvain cluster 2: 1518 cells
louvain cluster 3: 1516 cells
louvain cluster 4: 1440 cells
louvain cluster 5: 1436 cells
louvain cluster 6: 1433 cells
louvain cluster 7: 1407 cells
louvain cluster 8: 1404 cells
louvain cluster 9: 1395 cells
louvain cluster 10: 1358 cells
louvain cluster 11: 1246 cells
louvain cluster 12: 1156 cells
louvain cluster 13: 989 cells
louvain cluster 14: 858 cells
louvain cluster 15: 772 cells
louvain cluster 16: 649 cells
louvain cluster 17: 517 cells
louvain cluster 18: 446 cells
louvain cluster 19: 295 cells
louvain cluster 20: 267 cells
louvain cluster 21: 111 cells
log2FC
louvain
C
Lef1
Klrg1
Lgals3
Cd38
Slamf6
Cdca3
Cx3cr1
Lag3
Ccl5
Cd9
Mki67
Cxcr6
Entpd1
Cd7
Cd74
D
acute_d07
acute_d60
chronic_d07
chronic_d35
chronic_d35_b6

Figure S8. Differential expression analysis of scRNA-seq data (related to Figure 4). A. Average scRNA-seq gene expression (library-size normalized UMI counts, z-score row normalized) across clusters for 261 differentially expressed genes obtained as the union of three overlapping gene sets: genes significantly overexpressed in a cluster as compared with a union of all other clusters (top 20 genes per cluster, log2FC > 0.7, adjusted p-value < 0.001, total 224 genes), genes significantly overexpressed in a non-naïve-cell cluster as compared with a union of all other non-naïve-cell clusters (top 20 genes per cluster, log2FC > 1.0, adjusted p-value < 0.001, total 188 genes), genes significantly differentially expressed between a pair of non-naïve-cell clusters (top 5 genes, log2FC > 1.5, adjusted p-value < 0.001, total 98 genes). **B.** Barplots of scRNA-seq gene expression log2 fold change values for top 30 genes significantly differentially over-expressed and under-expressed in each cluster. **C.** MAGIC-imputed gene expression for selected genes. **D.** Fraction of cells classified as naïve (clusters 0, 7, 19) in individual samples.

Figure S9

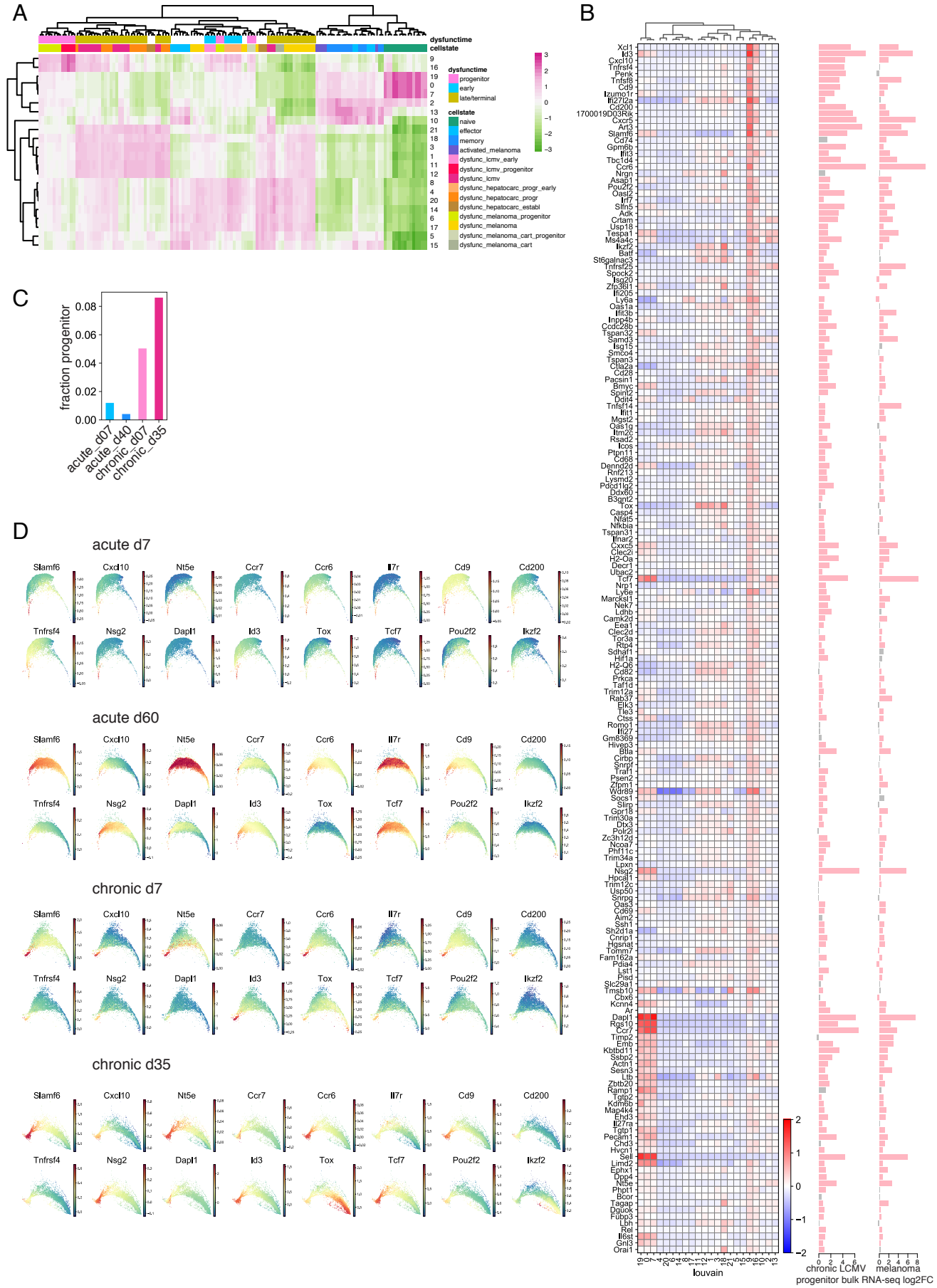

A
dysfunctime
cellstate
dysfunctime
progenitor
early
late/terminal
cellstate
naive
effector
memory
activated_melanoma
dysfunc_lcmv_early
dysfunc_lcmv_progenitor
dysfunc_lcmv
dysfunc_hepatocarc_progr_early
dysfunc_hepatocarc_progr
dysfunc_hepatocarc_establ
dysfunc_melanoma_progenitor
dysfunc_melanoma
dysfunc_melanoma_cart_progenitor
dysfunc_melanoma_cart
B
louvain
chronic LCMV
melanoma
progenitor bulk RNA-seq log2FC
C
fraction progenitor
acute_d07
acute_d40
chronic_d07
chronic_d35
D
acute d7
acute d60
chronic d7
chronic d35
Slamf6
Cxcl10
Nt5e
Ccr7
Ccr6
Il7r
Cd9
Cd200
Tnfrsf4
Nsg2
Dapl1
Id3
Tox
Tcf7
Pou2f2
Ikzf2

Figure S9. scRNA-seq cluster 9 consists of precursor/progenitor cells in acute and chronic infection (related to Figure 4). **A.** ssGSEA association of library size-normalized batch effect-corrected bulk RNA-seq data with library size-normalized scRNA-seq data averaged over cells in each cluster (z-score column normalized). Clusters 0, 7, 19 were most strongly associated with naïve cells, clusters 2, 10, 13 with memory cells, clusters 1, 3, 11, 12, 18, 21 with late and terminally dysfunctional cells, clusters 4, 5, 6, 8, 14, 15, 17, 20 with effector and early and progenitor dysfunctional cells, clusters 9 and 16 with progenitor cells and with early dysfunctional cells. **B.** Average scRNA-seq gene expression (library-size normalized UMI counts) in each cluster for genes significantly overexpressed in cluster 9 as compared with all other cells and as compared with all other cells except naïve cells (clusters 0, 7, 19). **C.** Fraction of cells that belong to cluster 9 in individual samples. **D.** MAGIC-imputed gene expression in individual samples.

Figure S10

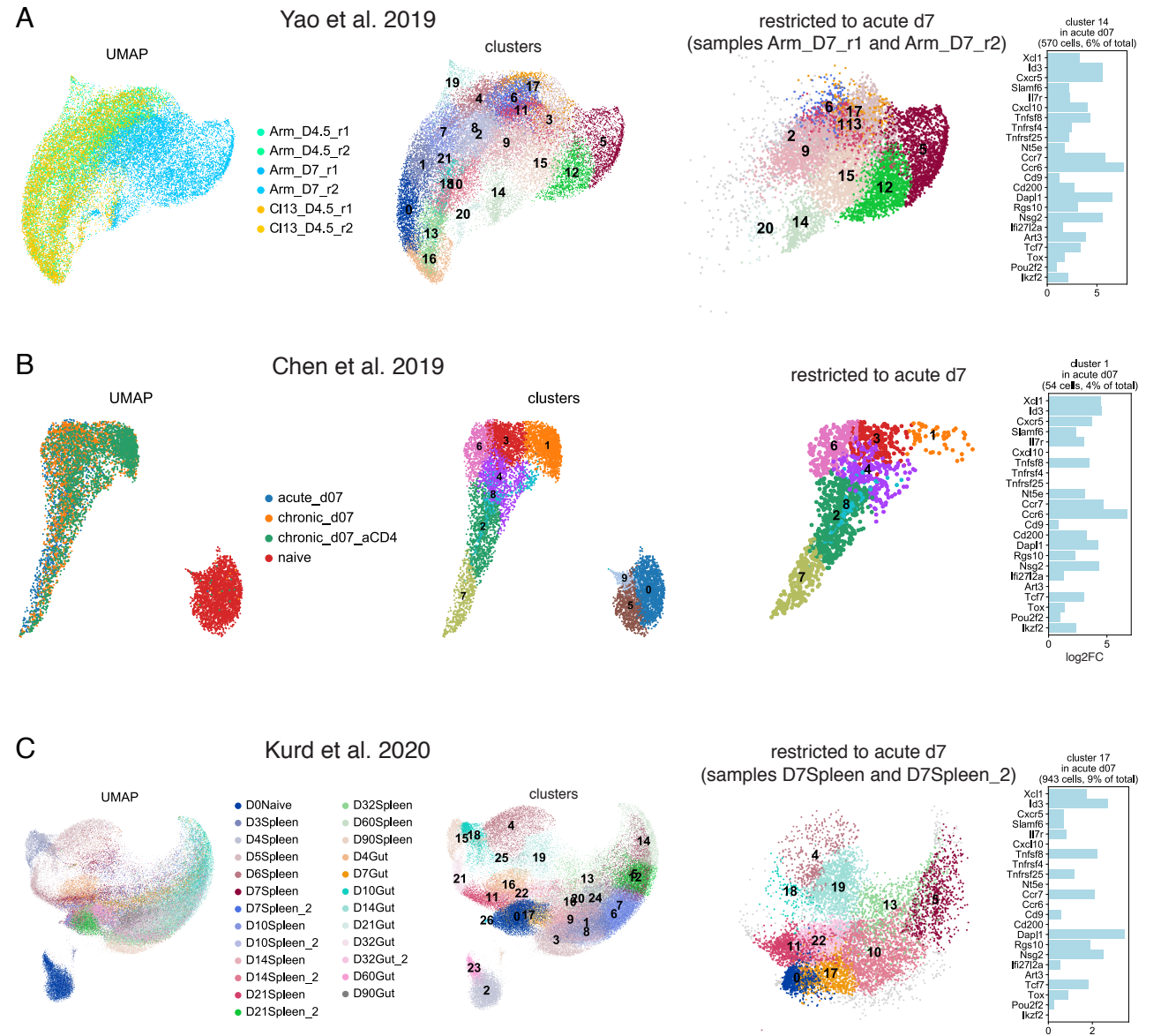

A
Yao et al. 2019
UMAP
clusters
Arm_D4.5_r1
Arm_D4.5_r2
Arm_D7_r1
Arm_D7_r2
Cl13_D4.5_r1
Cl13_D4.5_r2
restricted to acute d7
(samples Arm_D7_r1 and Arm_D7_r2)
cluster 14
in acute d07
(570 cells, 6% of total)
Xcl1
Id3
Cxcr5
Slamf6
Il7r
Cxcl10
Tnfsf8
Tnfrsf4
Tnfrsf25
Nt5e
Ccr7
Ccr6
Cd9
Cd200
Dapl1
Rgs10
Nsg2
Ifi27l2a
Art3
Tcf7
Tox
Pou2f2
Ikzf2
B
Chen et al. 2019
UMAP
clusters
acute_d07
chronic_d07
chronic_d07_aCD4
naive
restricted to acute d7
cluster 1
in acute d07
(54 cells, 4% of total)
log2FC
C
Kurd et al. 2020
UMAP
clusters
D0Naive
D3Spleen
D4Spleen
D5Spleen
D6Spleen
D7Spleen
D7Spleen_2
D10Spleen
D10Spleen_2
D14Spleen
D14Spleen_2
D21Spleen
D21Spleen_2
D32Spleen
D60Spleen
D90Spleen
D4Gut
D7Gut
D10Gut
D14Gut
D21Gut
D32Gut
D32Gut_2
D60Gut
D90Gut
restricted to acute d7
(samples D7Spleen and D7Spleen_2)
cluster 17
in acute d07
(943 cells, 9% of total)

Figure S10. Analysis of published scRNA-seq data confirms presence of progenitor-like subset at d7 in acute LCVM infection (related to Figure 4). Enrichment of progenitor markers in a cluster of cells profiled at d7 in acute LCMV infection in three studies (Yao *et al.* 2019, Kurd *et al.* 2020, Chen *et al.* 2020) (**A-C**). Shown are significant ($q < 0.01$) log2 fold change values of expression in a selected cluster as compared with cells from all other clusters, as identified by diffxpy (barplot, rightmost column). Clusters occupying at least 1% of the total data set are labeled on UMAP (third column). UMAP and clustering of the whole dataset are shown (leftmost two columns).

Figure S11

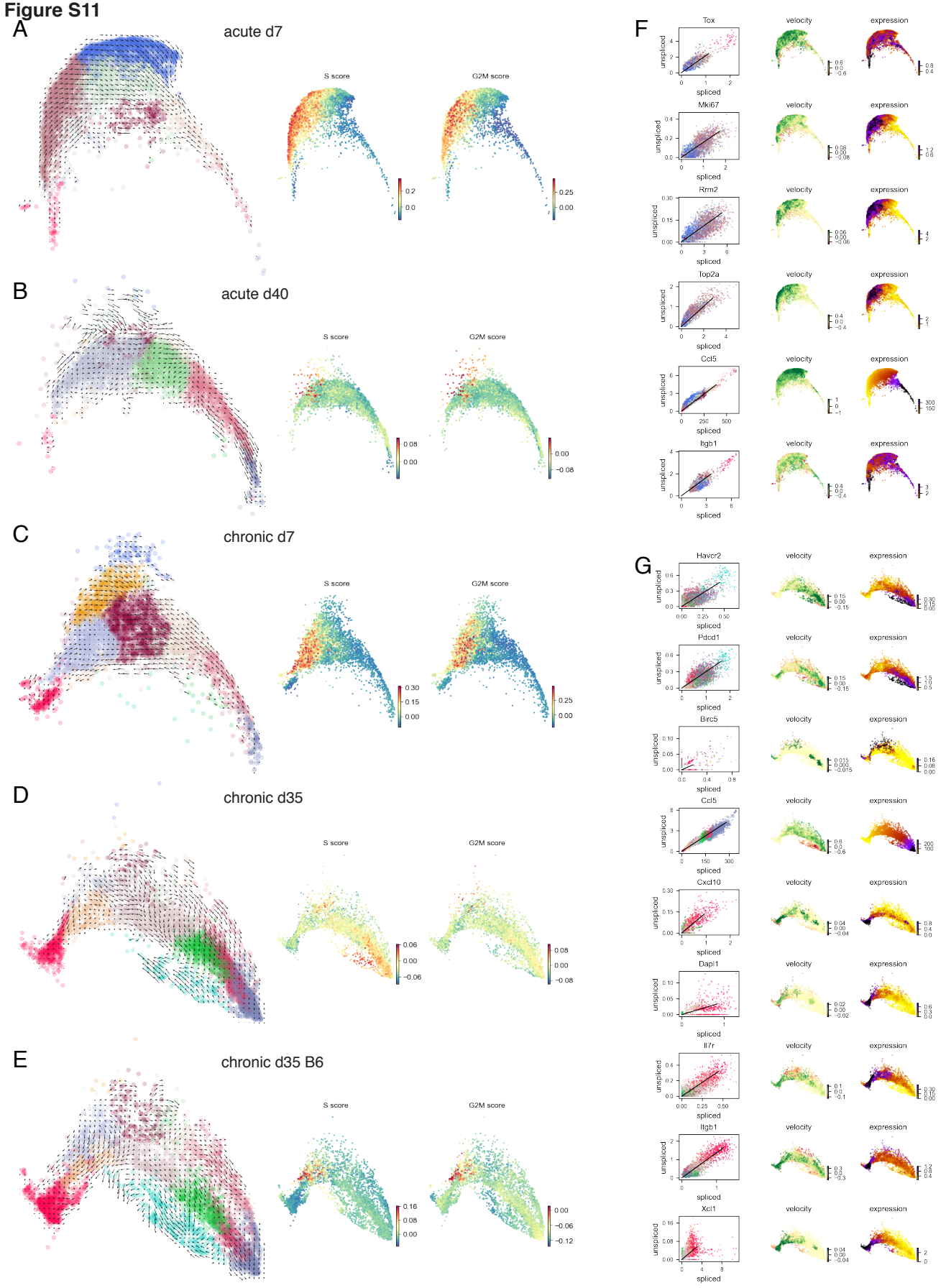

A
acute d7
S score
G2M score
B
acute d40
S score
G2M score
C
chronic d7
S score
G2M score
D
chronic d35
S score
G2M score
E
chronic d35 B6
S score
G2M score
F
Tox
Mki67
Rrm2
Top2a
Ccl5
Itgb1
velocity
expression
unspliced
spliced
G
Havcr2
Pdcd1
Birc5
Ccl5
Cxcl10
Dapl1
Il7r
Itgb1
Xcl1
velocity
expression
unspliced
spliced

Figure S11. RNA velocity analysis (related to Figure 4). A-E. Vector maps of RNA velocities (left) and cell cycle gene signature scores (right) are shown for scRNA-seq samples “acute_d07” (**A**), “acute_d40” (**B**), “chronic_d07” (**C**), “chronic_d35” (**D**), and “chronic_d35_b6” (**E**). **F-G.** Phase maps of unspliced and spliced read counts (left), RNA velocities (middle), and gene expression values (averaged over nearest neighbors in the kNN graph, right) for selected genes in cells from samples “acute_d07” (**F**) and “chronic_d35” (**G**).

Figure S12

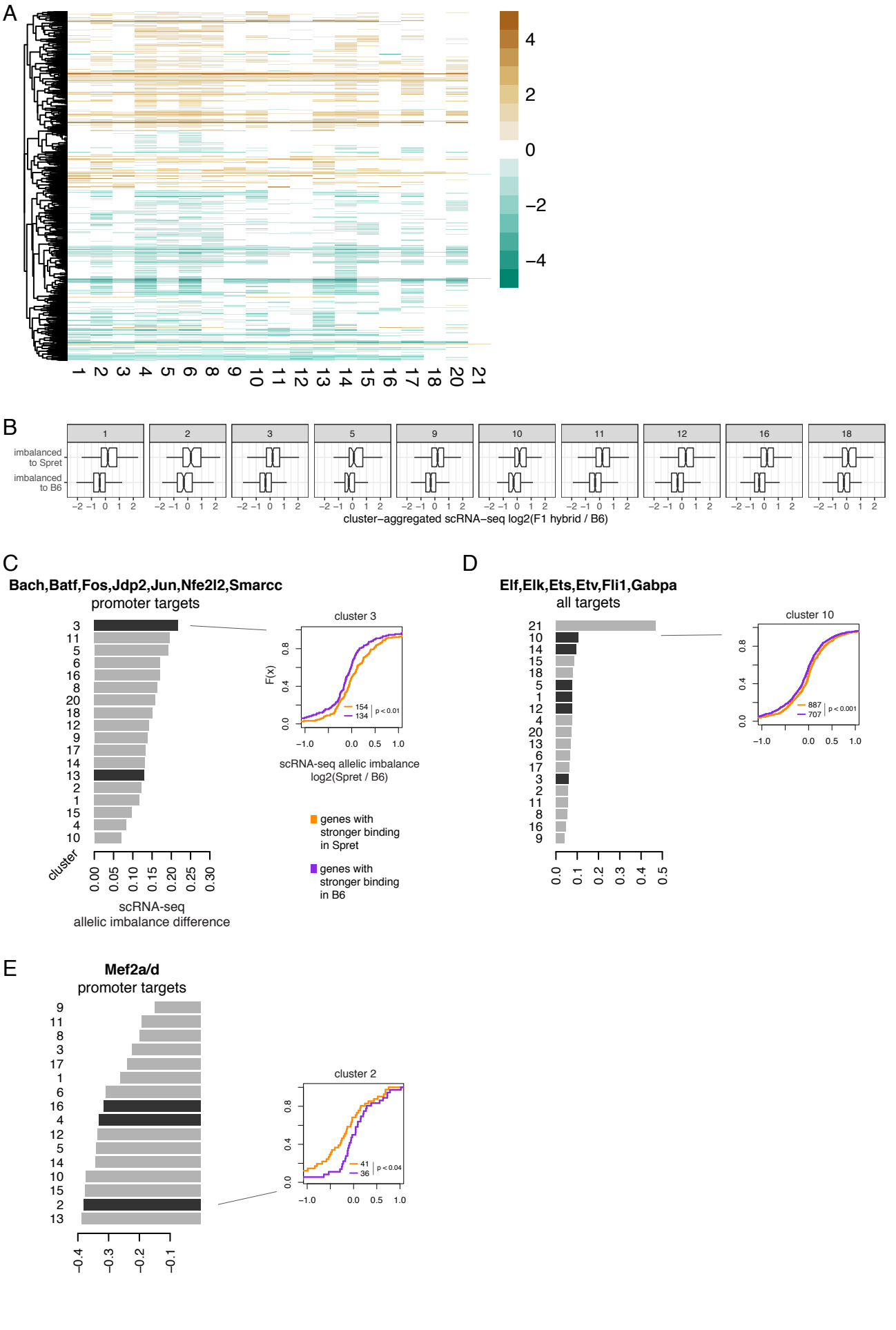

A
4
2
0
−2
−4
1 2 3 4 5 6 8 9 10 11 12 13 14 15 16 17 18 20 21
B
1 2 3 5 9 10 11 12 16 18
imbalanced to Spret
imbalanced to B6
−2 −1 0 1 2
cluster−aggregated scRNA−seq log2(F1 hybrid / B6)
C
Bach,Batf,Fos,Jdp2,Jun,Nfe2l2,Smarcc
promoter targets
3 11 5 6 16 8 20 18 12 9 17 14 13 2 1 15 4 10
cluster
0.00 0.05 0.10 0.15 0.20 0.25 0.30
scRNA-seq
allelic imbalance difference
cluster 3
F(x)
0.0 0.4 0.8
154
134
p < 0.01
−1.0 0.0 0.5 1.0
scRNA-seq allelic imbalance
log2(Spret / B6)
genes with stronger binding in Spret
genes with stronger binding in B6
D
Elf,Elk,Ets,Etv,Fli1,Gabpa
all targets
21 10 14 15 18 5 1 12 4 20 13 6 17 3 2 11 8 16 9
0.0 0.1 0.2 0.3 0.4 0.5
cluster 10
0.0 0.4 0.8
887
707
p < 0.001
−1.0 0.0 0.5 1.0
E
Mef2a/d
promoter targets
9 11 8 3 17 1 6 16 4 12 5 14 10 15 2 13
−0.4 −0.3 −0.2 −0.1
cluster 2
0.0 0.4 0.8
41
36
p < 0.04
−1.0 0.0 0.5 1.0

Figure S12. Allele-specific analysis of TF binding and scRNA-seq data (related to Figure 5). A. Allelic imbalance of gene expression in clusters of scRNA-seq data (color, $p < 0.05$, Mann-Whitney U). **B.** Boxplots of differential expression (log2 fold change) between cells within the same scRNA-seq cluster profiled at d35 in chronic infection in B6 and in B6/Spret F1 mice, for genes with significant allelic imbalance towards B6 or Spret in B6/Spret F1 mice. Cluster number indicated on top of each boxplot. Significant difference between log2FC distributions ($p < 2e{-}4$ for clusters 2, 5, 10, 18, $p < 2e{-}16$ for all other clusters, Mann-Whitney U test) in all cases. **C-E.** Allelic specificity expression analysis of scRNA-seq data. CDF plots: allelic imbalance between B6 and Spret of library-size normalized scRNA-seq counts aggregated over cells in a selected cluster, for genes predicted to be bound more strongly in B6 or Spret using sequence motif analysis in 150bp windows around summits of their promoter peaks. Barplots: summary of the above analysis for each TF motif over all clusters; black bar indicates significant association (Kolmogorov-Smirnov $p < 0.05$).

Figure S13

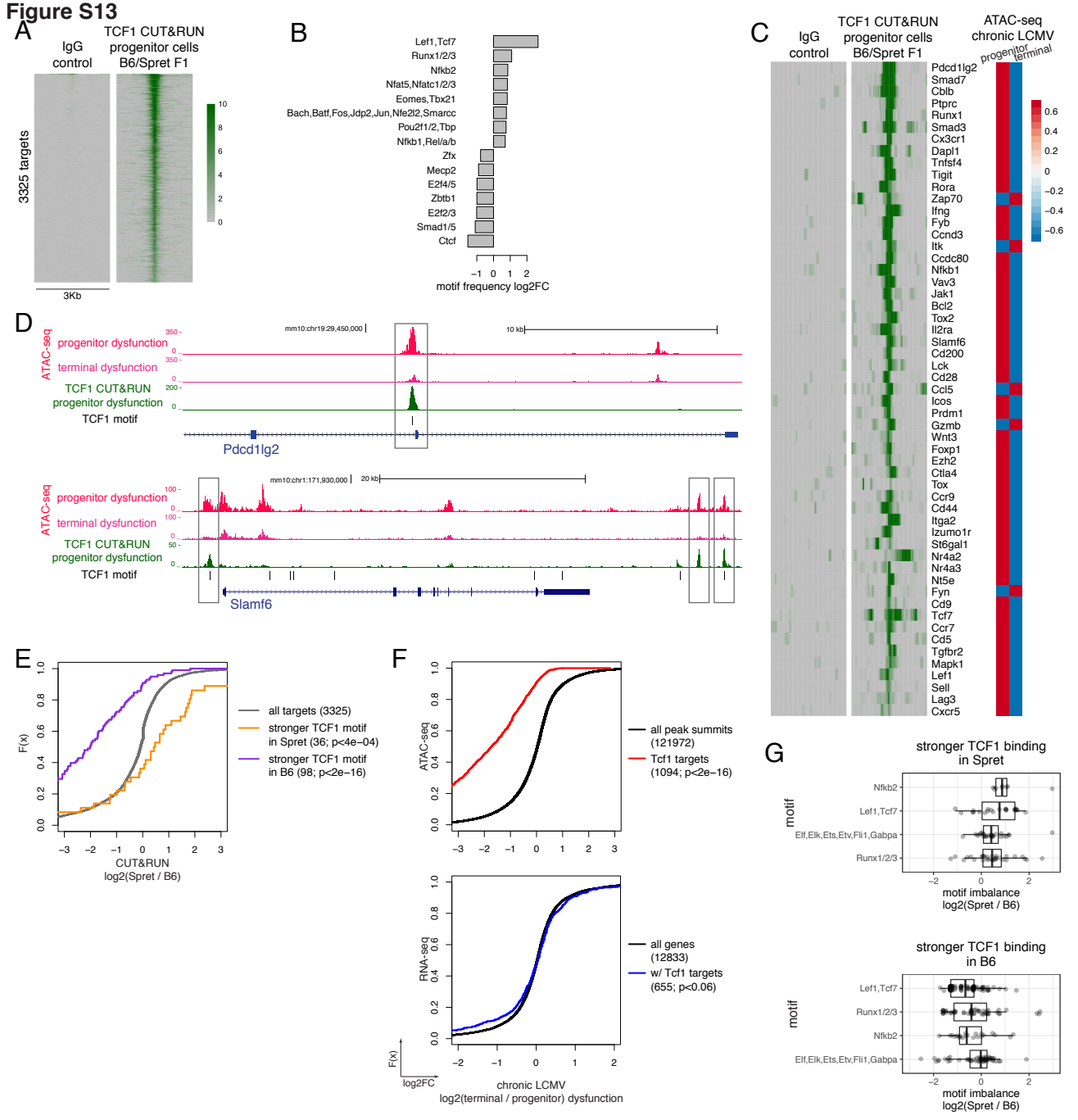

A
IgG control
TCF1 CUT&RUN progenitor cells B6/Spret F1
3325 targets
3Kb
B
Lef1,Tcf7
Runx1/2/3
Nfkb2
Nfat5,Nfatc1/2/3
Eomes,Tbx21
Bach,Batf,Fos,Jdp2,Jun,Nfe2l2,Smarcc
Pou2f1/2,Tbp
Nfkb1,Rel/a/b
Zfx
Mecp2
E2f4/5
Zbtb1
E2f2/3
Smad1/5
Ctcf
motif frequency log2FC
C
IgG control
TCF1 CUT&RUN progenitor cells B6/Spret F1
ATAC-seq chronic LCMV
progenitor
terminal
Pdcd1lg2
Smad7
Cblb
Ptprc
Runx1
Smad3
Cx3cr1
Dapl1
Tnfsf4
Tigit
Rora
Zap70
Ifng
Fyb
Ccnd3
Itk
Ccdc80
Nfkb1
Vav3
Jak1
Bcl2
Tox2
Il2ra
Slamf6
Cd200
Lck
Cd28
Ccl5
Icos
Prdm1
Gzmb
Wnt3
Foxp1
Ezh2
Ctla4
Tox
Ccr9
Cd44
Itga2
Izumo1r
St6gal1
Nr4a2
Nr4a3
Nt5e
Fyn
Cd9
Tcf7
Ccr7
Cd5
Tgfbr2
Mapk1
Lef1
Sell
Lag3
Cxcr5
D
ATAC-seq
progenitor dysfunction
terminal dysfunction
TCF1 CUT&RUN progenitor dysfunction
TCF1 motif
mm10:chr19:29,450,000
10 kb
Pdcd1lg2
mm10:chr1:171,930,000
20 kb
Slamf6
E
F(x)
CUT&RUN log2(Spret / B6)
all targets (3325)
stronger TCF1 motif in Spret (36; p<4e−04)
stronger TCF1 motif in B6 (98; p<2e−16)
F
ATAC-seq
all peak summits (121972)
Tcf1 targets (1094; p<2e−16)
RNA-seq
all genes (12833)
w/ Tcf1 targets (655; p<0.06)
F(x)
log2FC
chronic LCMV log2(terminal / progenitor) dysfunction
G
stronger TCF1 binding in Spret
motif
Nfkb2
Lef1,Tcf7
Elf,Elk,Ets,Etv,Fli1,Gabpa
Runx1/2/3
motif imbalance log2(Spret / B6)
stronger TCF1 binding in B6
motif imbalance log2(Spret / B6)

Figure S13. CUT&RUN in progenitor dysfunctional CD8 T cells from B6/Spret F1 hybrid mice maps allele-specific TCF1 binding (related to Figure 5). A. Normalized TCF1 CUT&RUN signal in TCF1-bound ATAC-seq peak summits (library-size normalized read counts in 50bp bins in 3000bp window around peak summit, ordered by total signal). **B.** Log2 fold change enrichment of motif frequency in TCF1 CUT&RUN peaks relative to all ATAC-seq peak summits. **C.** Examples of TCF1 binding sites. **D.** Genome browser tracks of bulk ATAC-seq for progenitor and terminally dysfunctional cells and TCF1 CUT&RUN in progenitor dysfunctional cells for selected loci. **E.** CDF curves for allelic imbalance of CUT&RUN signal between B6 and Spret for all TCF1 targets (black) and for those predicted to be bound more strongly in B6 (purple) or Spret (orange) using sequence motif analysis. **F.** Top: CDF plots of library size-normalized batch effect-corrected ATAC-seq signal for TCF1-bound sites vs. all ATAC-seq peak summits. Bottom: CDF plots of library-size normalized batch-effect corrected RNA-seq gene expression for genes with TCF1 binding vs. all expressed genes. Shown is the comparison between progenitor and terminally dysfunctional cells in chronic LCMV infection. **G.** Imbalance of predicted TF binding using sequence motif analysis in peaks with stronger TCF1 binding in B6 or in Spret allele as identified by CUT&RUN.