nature biotechnology

SUPPLEMENTARY INFORMATION

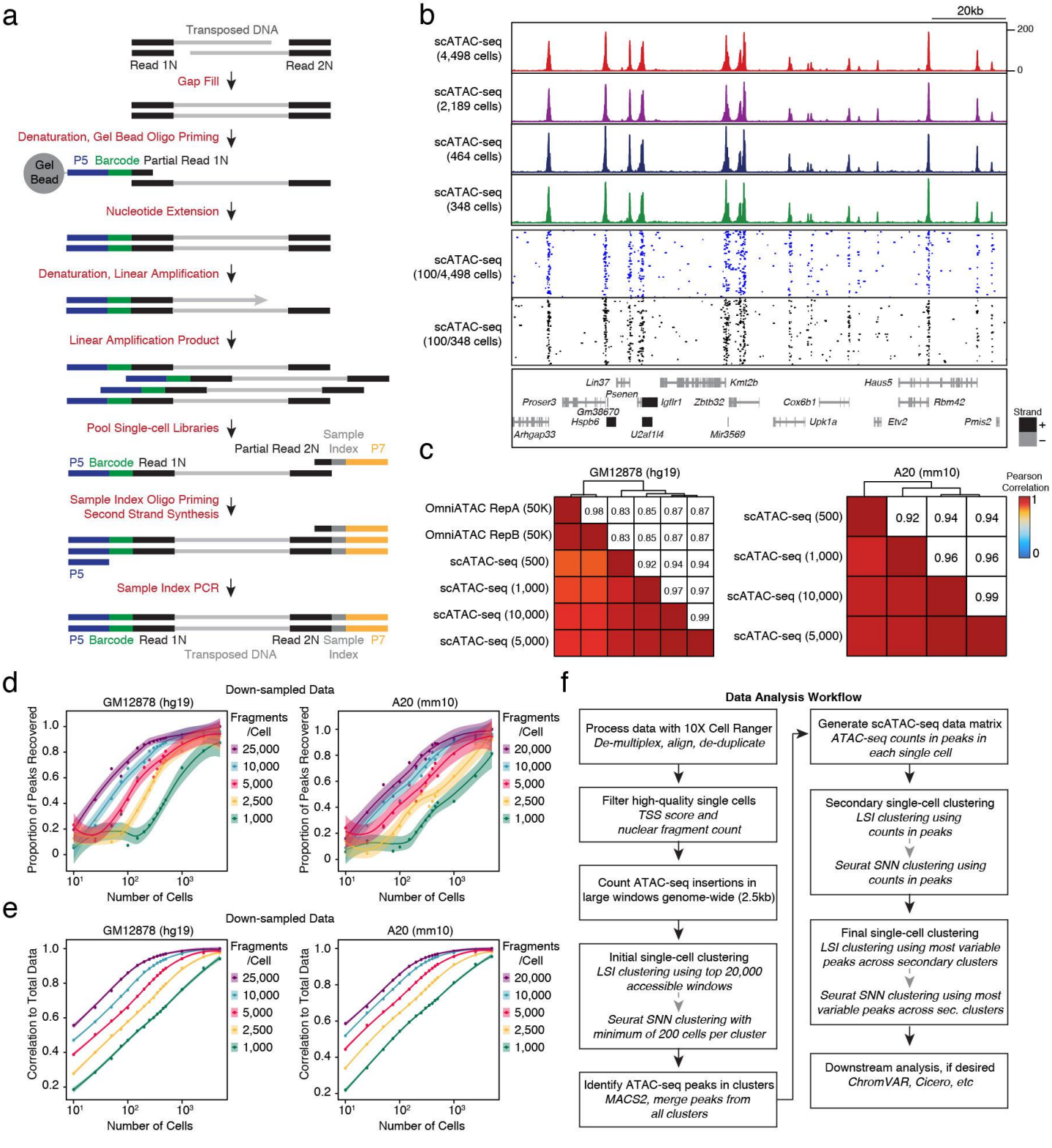
https://doi.org/10.1038/s41587-019-0206-z

In the format provided by the authors and unedited.

Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion

Ansuman T. Satpathy[1,2,11], Jeffrey M. Granja[1,3,4,11], Kathryn E. Yost[1,5,6], Yanyan Qi[1,6], Francesca Meschi[7], Geoffrey P. McDermott[7], Brett N. Olsen[7], Maxwell R. Mumbach[1,3], Sarah E. Pierce[3,5], M. Ryan Corces[1,6], Preyas Shah[7], Jason C. Bell[7], Darisha Jhutty[7], Corey M. Nemec[7], Jean Wang[7], Li Wang[7], Yifeng Yin[7], Paul G. Giresi[7], Anne Lynn S. Chang[6], Grace X. Y. Zheng[7]*, William J. Greenleaf[1,3,8,9]* and Howard Y. Chang[1,3,6,10]*

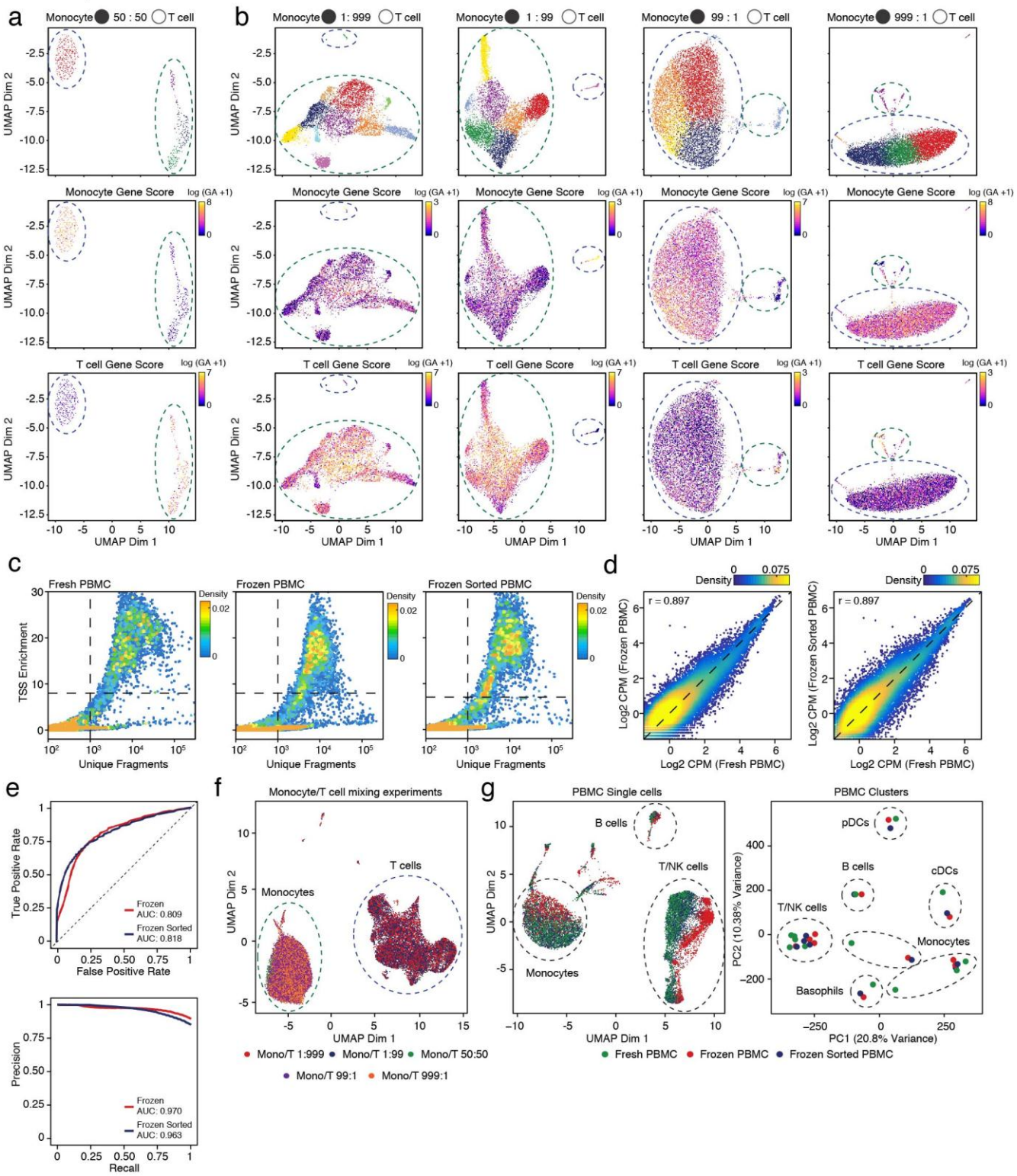
[1]Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA, USA. [2]Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA. [3]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. [4]Biophysics Program, Stanford University School of Medicine, Stanford, CA, USA. [5]Cancer Biology Program, Stanford University School of Medicine, Stanford, CA, USA. [6]Department of Dermatology, Stanford University School of Medicine, Redwood City, CA, USA. [7]10x Genomics, Inc., Pleasanton, CA, USA. [8]Department of Applied Physics, Stanford University, Stanford, CA, USA. [9]Chan Zuckerberg Biohub, San Francisco, CA, USA. [10]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA. [11]These authors contributed equally: Ansuman T. Satpathy, Jeffrey M. Granja.
*e-mail: grace@10xgenomics.com; wjg@stanford.edu; howchang@stanford.edu



Supplementary Figure 1

Droplet-based scATAC-seq workflow and quality control measurements.

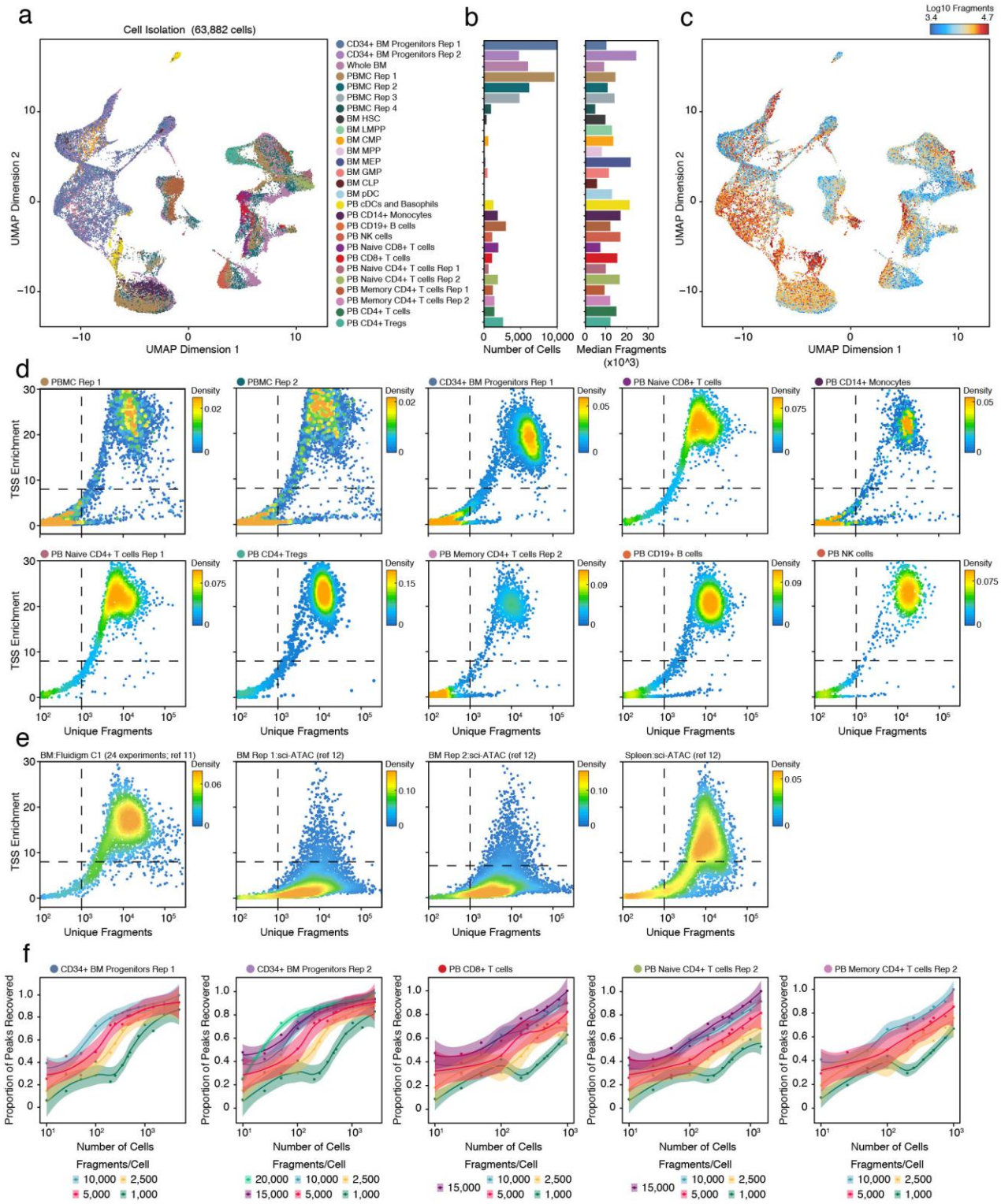
(**a**) Protocol steps for scATAC-seq in droplets. (**b**) Genome tracks showing the comparison of aggregate scATAC-seq profiles from A20 B lymphocytes (top panel). scATAC-seq profiles were obtained from four independent experiments, as indicated. The bottom panel shows accessibility profiles of 100 random single A20 cells from two cell mixing experiments. Each pixel represents a 100bp region. (**c**) Left plot: Pearson correlation heatmaps of log-normalized reads in bulk GM12878 Omni-ATAC-seq peaks in aggregate scATAC-seq profiles generated from varying numbers of single cells, or from published Omni-ATAC profiles[5] (n=100,000 ATAC-seq peaks). Right plot: Pearson correlation heatmaps of log-normalized reads in aggregate scATAC-seq profiles from A20 cells (n=100,000 ATAC-seq peaks, identified in an aggregate profile from all cells). Numbers in parentheses indicate the cell loading concentration. (**d**) Peak recovery analysis with subsampled cells and unique fragments as determined by x-axis and colors, respectively. scATAC-seq cells were subsampled to the indicated unique fragments, and the proportion of peaks recovered from the aggregate profile was calculated as a function of number of cells analyzed. GM12878 cells generated a median of 29,451 unique nuclear fragments per cell (top level of down-sampling was 25,000) while A20 cells generated a median of 20,809 unique nuclear fragments per cell (top level of down-sampling was 20,000). The center line represents the Loess fit, and shaded regions indicate 95% confidence interval (n=16 sub-sampled profiles at each read depth). (**e**) Pearson correlation analysis with subsampled cells and unique fragments as determined by x-axis and colors, respectively. scATAC-seq profiles were subsampled to the indicated unique fragments, and Pearson correlation to the aggregate profile was calculated as a function of number of cells analyzed (n=16 sub-sampled profiles at each read depth). **(f)** Analysis workflow for scATAC-seq data in this study.



Supplementary Figure 2

scATAC-seq performance in frozen cells and synthetic cell mixtures.

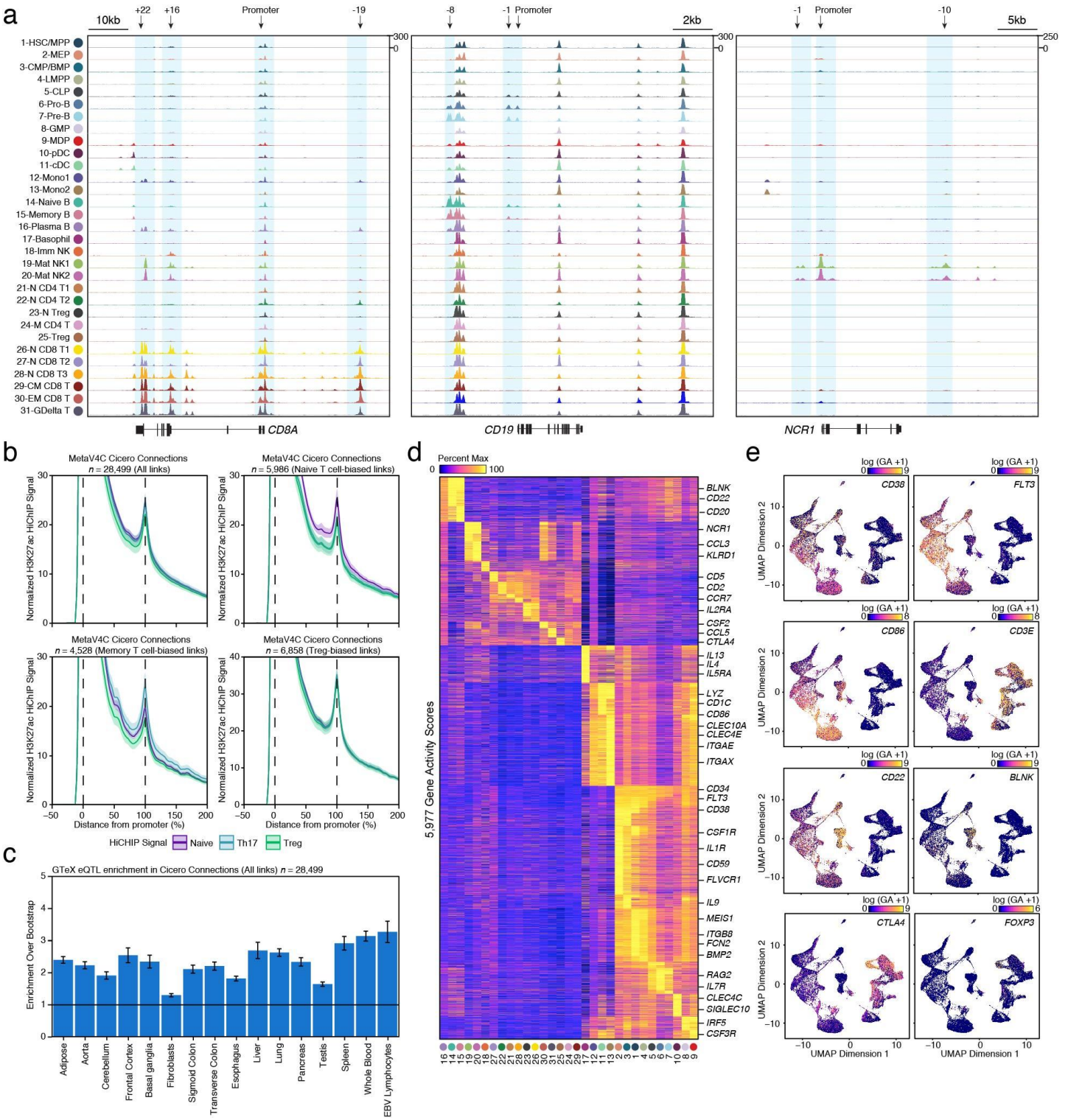
(**a**) Synthetic immune cell mixture quality control experiments. Sorted human monocytes or T cells were mixed at the indicated ratio and analyzed with scATAC-seq. Plots show the UMAP of scATAC-seq profiles (top), and gene scores for monocyte- or T cell-associated cardinal genes (see Methods) in each single cell (middle and bottom). Dashed circles indicate monocyte and T cell identity of single cells as determined by ATAC-seq profiles. Colors indicate cluster identity defined *de novo*. (**b**) Sorted human monocytes or T cells were mixed at the indicated ratios and analyzed with scATAC-seq and analyzed as described in (a). (**c**) Comparison of data quality from fresh and frozen PBMCs, and frozen PBMCs sorted for live cells. Representative ATAC-seq data quality control filters by sample source. Shown are the number of unique ATAC-seq nuclear fragments in each single cell (each dot) compared to TSS enrichment of all fragments in that cell. Dashed lines represent the filters for high-quality single-cell data (1,000 unique nuclear fragments and TSS score greater than or equal to 8). (**d**) One-to-one plots of log-normalized reads in aggregated scATAC-seq in profiles generated from the indicated cell source (fresh, frozen, or frozen sorted PBMCs). Peaks were defined in fresh samples. Numbers indicate Pearson correlation value. (**e**) ROC (top) and Precision-vs-Recall (bottom) curves showing recovery of fresh PBMC peaks with frozen or frozen sorted cells. True positive peaks were defined as those identified in fresh PBMC scATAC-seq profiles. (**f**) Integrated UMAP of all scATAC-seq profiles from monocyte/T cell mixing experiments in (a-b). This indicates that strong clustering batch effects are not seen between experiments. (**g**) UMAP and PCA analysis of of scATAC-seq profiles (left) and clusters (right) identified in fresh, frozen, or frozen-sorted PBMCs using fresh PBMC peaks.



Supplementary Figure 3

Sample descriptions and quality control of scATAC-seq hematopoiesis data.

(**a**) UMAP projection of 63,882 scATAC-seq profiles of bone marrow and peripheral blood immune cell types. Dots represent individual cells, and colors indicate the experimental source of each cluster, as labeled on the right of the plot (see Methods). (**b**) Bar plots indicate the number of scATAC-seq profiles obtained from each experimental source of cells (left), and the median number of unique nuclear fragments in single cells (right). (**c**) UMAP projection of 63,882 scATAC-seq profiles of bone marrow and peripheral blood immune cell types. Colors represent the log10 number of unique nuclear fragments per single cell. (**d**) Representative scATAC-seq data quality control filters by sample source. Shown are the number of unique ATAC-seq nuclear fragments in each single cell (each dot) compared to TSS enrichment of all fragments in that cell. Dashed lines represent the filters for high-quality single-cell data (1,000 unique nuclear fragments and TSS score greater than or equal to 8). (**e**) Single-cell ATAC-seq data quality control filters in profiles generated using the C1 microfluidic system[11] (Fluidigm; left) or sci-ATAC-seq[12] (middle and right panels). **(f)** Peak recovery analysis with subsampled cells and unique fragments as determined by x-axis and colors, respectively. scATAC-seq cells were subsampled to the indicated unique fragments, and the proportion of peaks recovered from the aggregate profile was calculated as a function of number of cells analyzed. The center line represents the Loess fit, and shaded regions indicate 95% confidence interval (n=10 sub-sampled profiles at each read depth).

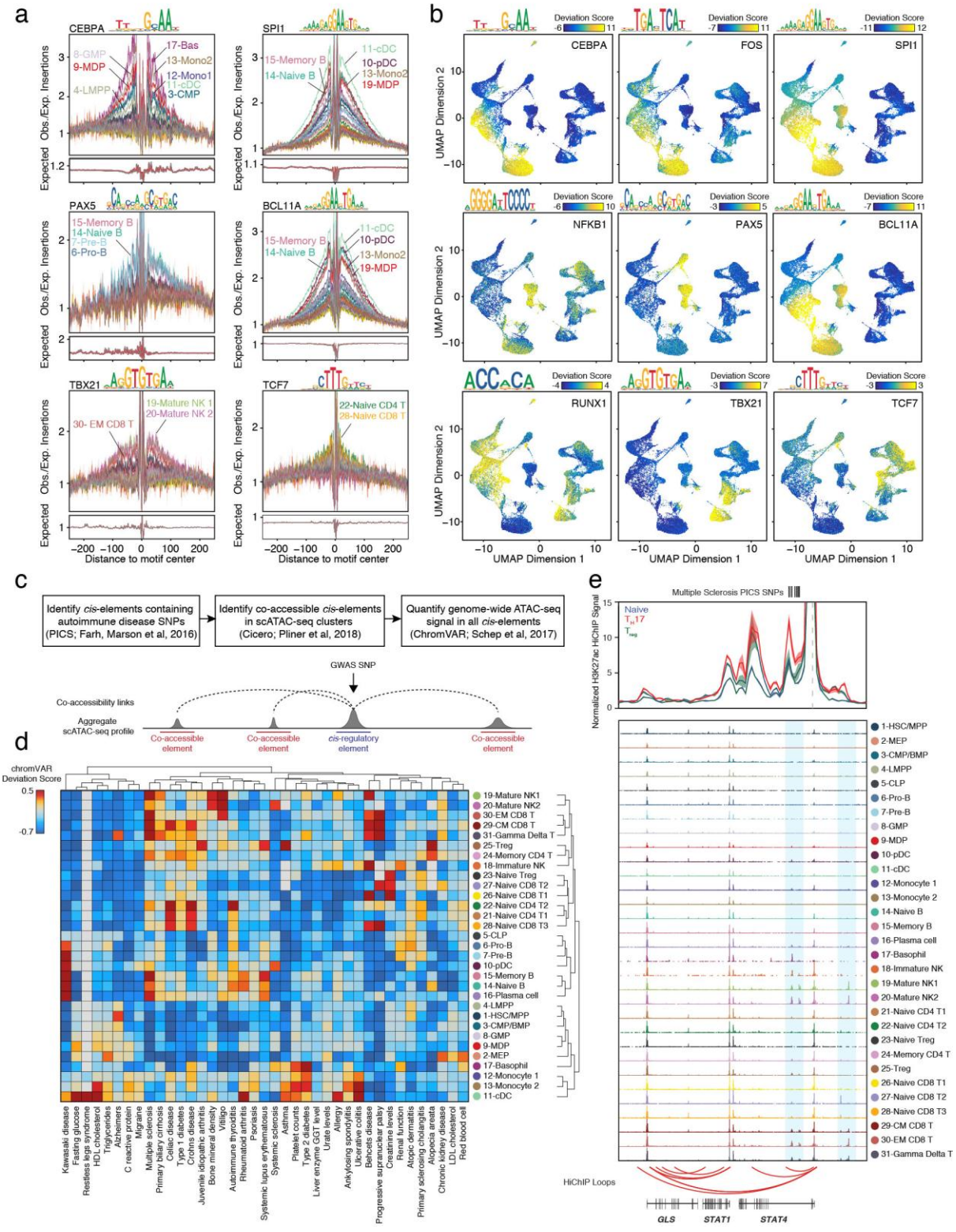


Supplementary Figure 4

***Cis*-regulatory elements in hematopoiesis and co-accessibility validation.**

(**a**) Genome tracks of aggregate scATAC-seq data, clustered as indicated in Figure 2b. Arrows indicate the position and distance (in kb)

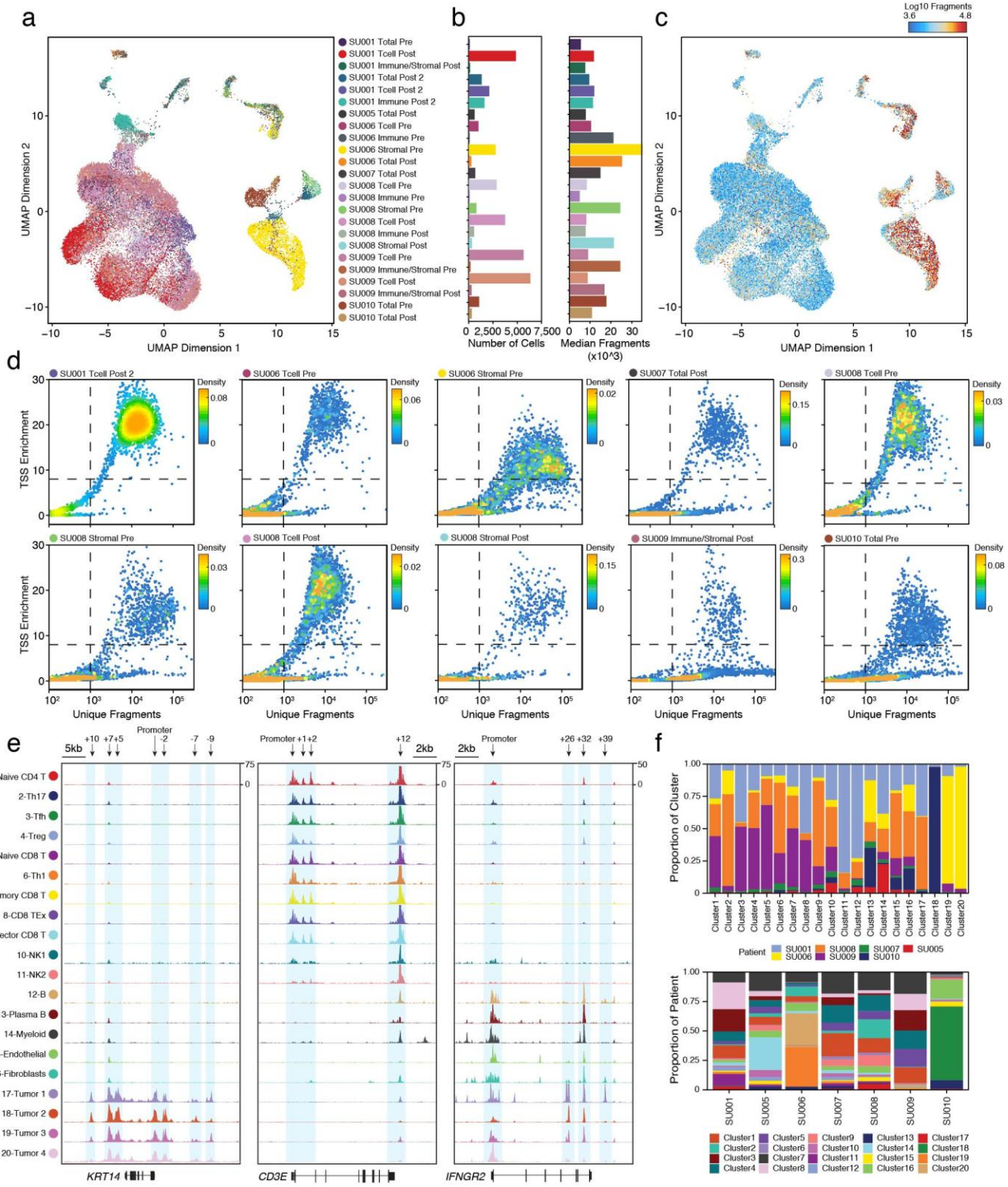
of intragenic or distal enhancers in each gene locus. (**b**) MetaV4C plot of H3K27ac HiChIP data demonstrating HiChIP signal at Cicero-identified co-accessible *cis*-elements (linked to promoter elements). Each plot shows the aggregate HiChIP signal (from 3 T cell types, n=2 biologically independent HiChIP profiles per cell type) between linked *cis*-elements identified in scATAC-seq data. Each link is scaled so that the 0% position indicates the promoter site and the 100% position indicates the linked *cis*-element site. The peak indicates an enrichment of HiChIP signal at the linked peaks compared to surrounding genomic regions. Biased links are identified by differential peak analysis in the indicated scATAC-seq clusters. The center line represents the Loess fit, and shaded regions indicate 95% confidence interval. (**c**) Support for Cicero-identified co-accessible *cis*-elements by GTEX eQTL data. Shown is the mean enrichment of eQTL signal (determined in the indicated tissue type, bars indicate standard deviation from n=250 simulations) in co-accessible sites linked to promoter elements described in GTEX vs 250 permutations of ATAC-seq peak to genes. Greater enrichment is observed in immune tissues because scATAC-seq data profiled the relevant cell types. (**d**) Heatmap of log-normalized gene scores for the indicated genes. (**e**) UMAP projection colored by log normalized gene scores demonstrating the accessibility of *cis*-regulatory elements linked (using Cicero) to the indicated gene.



Supplementary Figure 5

TF motif accessibility in hematopoiesis and cell type-specific GWAS enrichment.

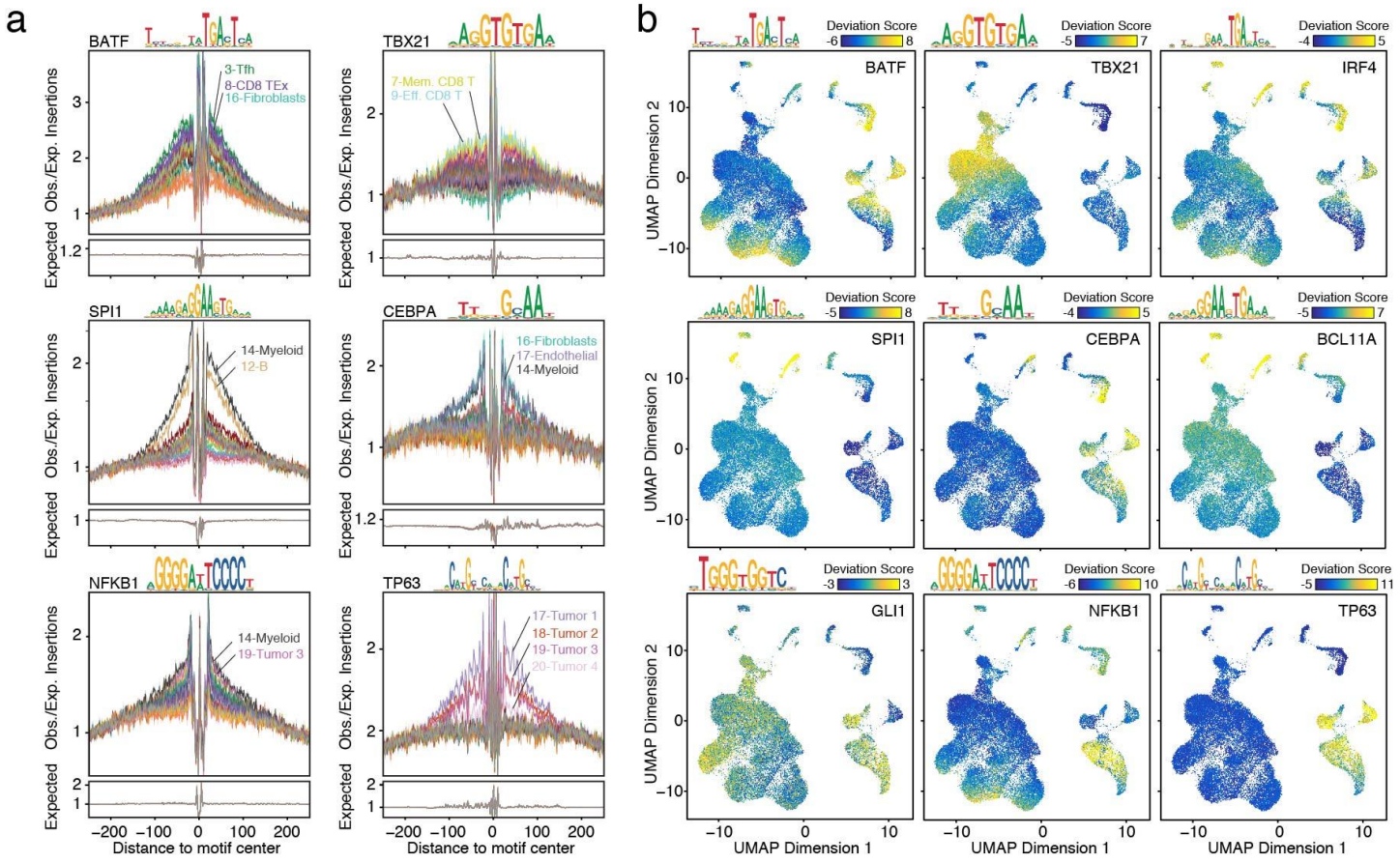
(**a**) Example TF footprints and motifs in the indicated scATAC-seq clusters identified in Fig 2b. The Tn5 insertion bias track is shown below. (**b**) UMAP projection of scATAC-seq profiles colored by chromVAR TF motif bias-corrected deviations for the indicated factors. (**c**) Analysis workflow for GWAS enrichment scores using Cicero co-accessibility. (**d**) Heatmap showing GWAS deviation scores for PICS SNPs associated with the indicated diseases. PICS SNPs were identified previously[21]. (**e**) Example of increased ATAC-seq signal in a GWAS-containing cis-element in NK and T cell scATAC-seq clusters. The HiChIP plot (top) demonstrates increased H3K27ac HiChIP signal between the *STAT4* promoter and the highlighted *cis*-elements. The center line represents the Loess fit, and shaded regions indicate 95% confidence interval (n=2 biologically independent HiChIP profiles per cell type).



Supplementary Figure 6

Sample descriptions and quality control of scATAC-seq profiles of the BCC TME.

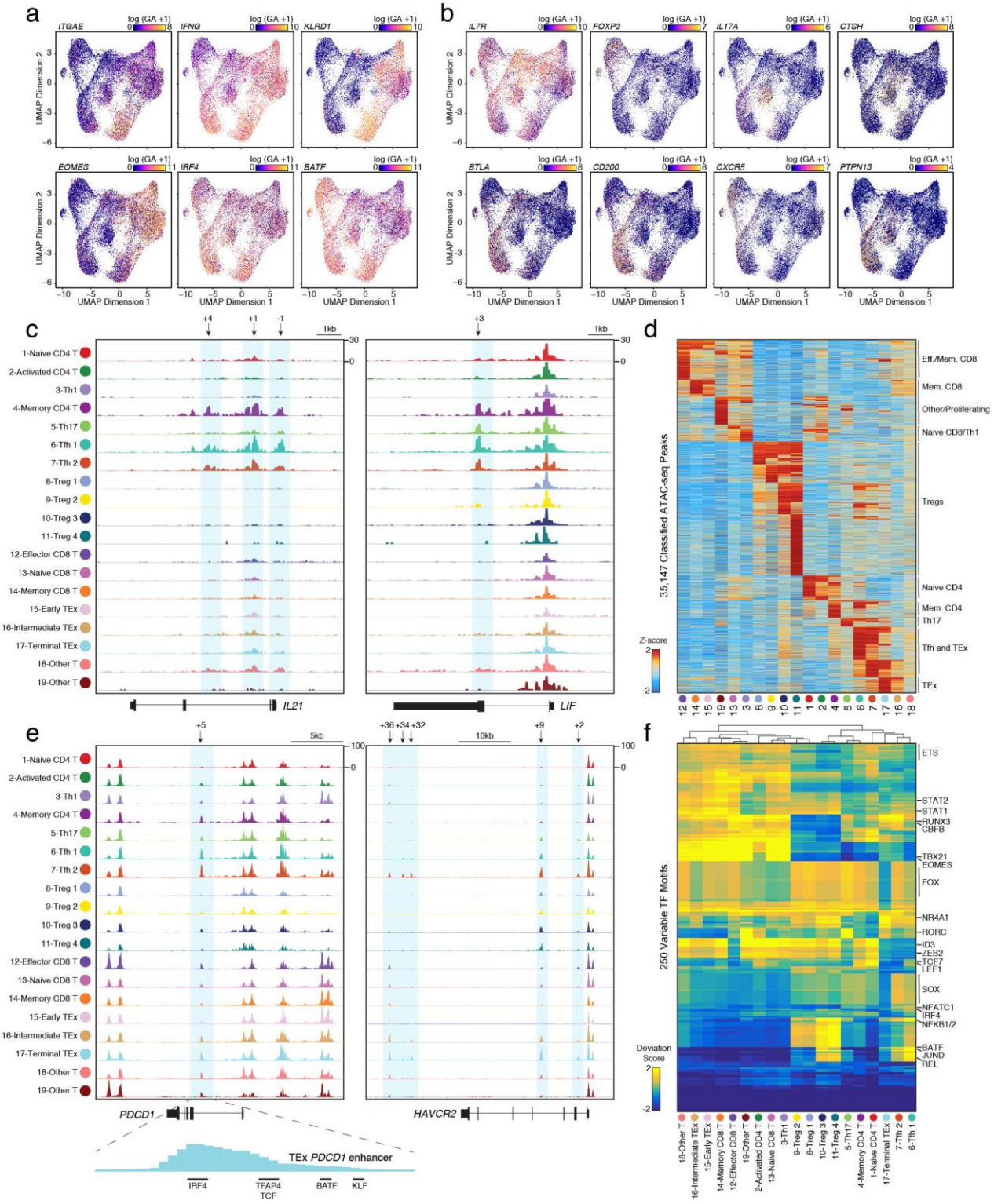
(**a**) UMAP projection of 37,818 scATAC-seq profiles of BCC TME cell types. Dots represent individual cells, and colors indicate the experimental source of each cluster, as labeled on the right of the plot (see Methods). 'Total' samples were sorted as all live cells in a single BCC biopsy. 'T cell' samples were sorted as $CD45^{+}CD3^{+}$ cells in the biopsy. 'Immune' samples were sorted as $CD45^{+}CD3^{-}$ cells in the biopsy. 'Stromal' samples were sorted as $CD45^{-}CD3^{-}$ cells in the biopsy. (**b**) Bar plots indicate the number of scATAC-seq profiles obtained from each experimental source of cells (left), and the median number of unique nuclear fragments in single cells (right). (**c**) UMAP projection of 37,818 scATAC-seq TME profiles. Colors represent the log10 number of unique nuclear reads per single cell. (**d**) Representative ATAC-seq data quality control filters by sample source. Shown are the number of unique ATAC-seq nuclear fragments in each single cell (each dot) compared to TSS enrichment of all fragments in that cell. Dashed lines represent the filters for high-quality single-cell data (1,000 unique nuclear fragments and TSS score greater than or equal to 8). (**e**) Genome tracks of aggregate scATAC-seq data, clustered as indicated in Figure 4b. Arrows indicate the position and distance (in kb) of intragenic or distal enhancers in each gene locus. (**f**) Bar plots indicating the relative proportion of cells from each patient detected in each cluster (top) and the relative proportion of cells from each cluster detected in each patient (bottom).



Supplementary Figure 7

TF motif accessibility in the BCC TME.

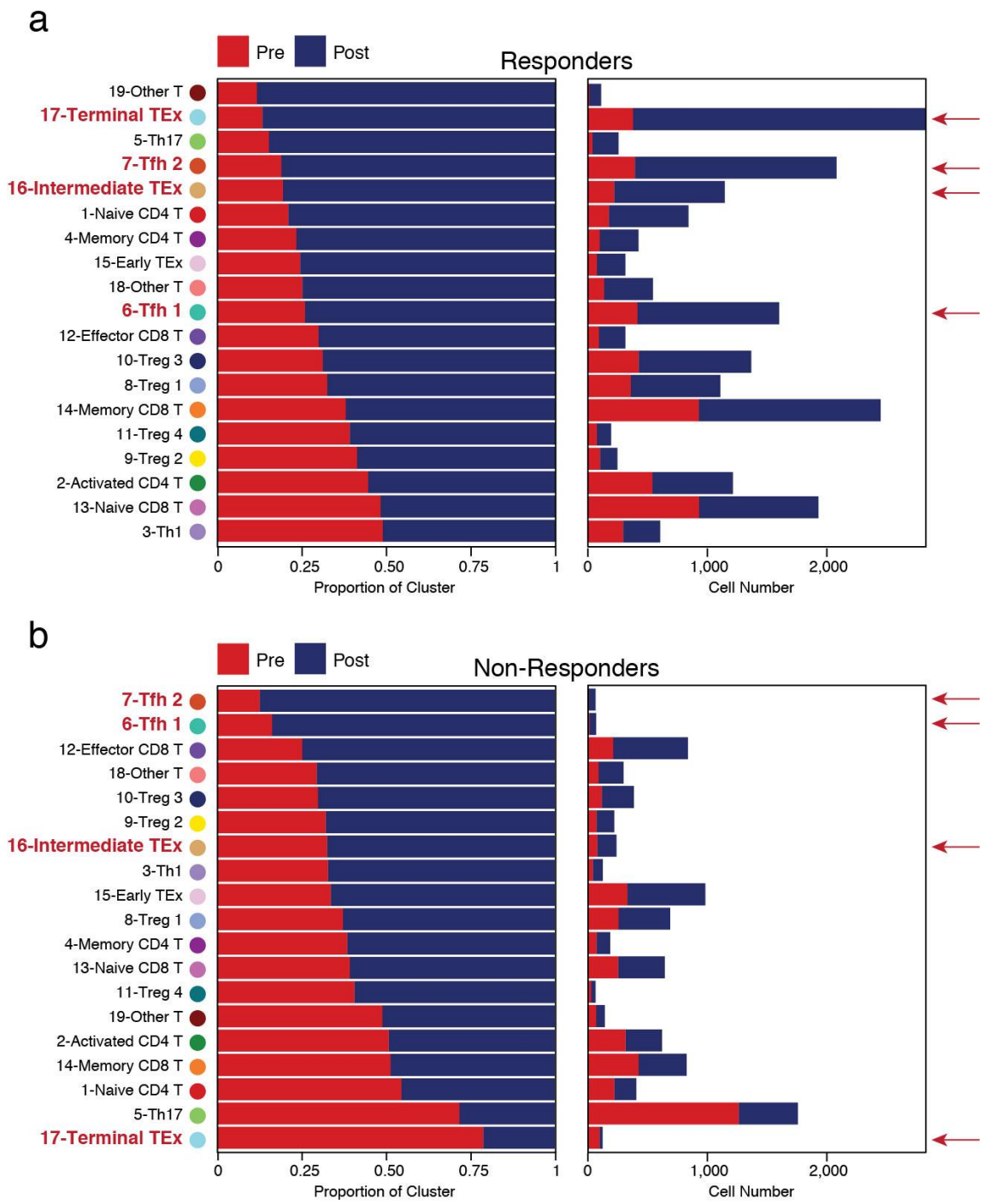
(**a**) Example TF footprints and motifs in the indicated scATAC-seq clusters identified in Fig 4b. The Tn5 insertion bias track is shown below. (**e**) UMAP projection of scATAC-seq profiles colored by chromVAR TF motif bias-corrected deviations for the indicated factors.



Supplementary Figure 8

Regulatory landscapes of tumor-infiltrating T cell subsets.

(**a**) UMAP projection of intratumoral T cell scATAC-seq data colored by log normalized gene scores, demonstrating the accessibility of *cis*-regulatory elements linked (using Cicero) to the indicated $CD8^+$ T cell signature genes. (**b**) UMAP projection of intratumoral T cell scATAC-seq data colored by log normalized gene scores, demonstrating the accessibility of *cis*-regulatory elements linked (using Cicero) to the indicated $CD4^+$ T cell signature genes. (**c**) Genome tracks of Tfh signature genes in aggregate scATAC-seq data, clustered as indicated in Figure 5a. Arrows indicate the position and distance (in kb) of intragenic or distal enhancers in each gene locus. (**d**) Heatmap of Z-scores of 35,147 *cis*-regulatory elements in scATAC-seq clusters derived from (b). Labels indicate cell type-specific accessibility of regulatory elements. (**e**) Genome tracks of $CD8^+$ TEx signature genes in aggregate scATAC-seq data, demonstrating the overlap of TEx and Tfh regulatory elements. Arrows indicate the position and distance (in kb) of intragenic or distal enhancers in each gene locus. Selected TF binding motifs present in the +5kb enhancer of *PDCD1* are shown (bottom). Lines indicate the binding motif location. (**f**) Heatmap representation of ATAC-seq chromVAR bias-corrected deviations in the 250 most variable TFs across all intratumoral T cell scATAC-seq clusters, as identified in Figure 5a. Cluster identities are indicated at the bottom of the plot.



Supplementary Figure 9

T cell cluster dynamics pre- and post-therapy according to clinical outcome.

(**a**) Shown is the pre- and post-therapy frequency (left) and cell number (right) for each T cell cluster identified in Figure 5. Each plot is generated from all responder patient samples aggregated together. (**b**) Pre- and post-therapy frequency (left) and cell number (right) for each T cell cluster in non-responder patients.