

In the format provided by the authors and unedited.

Transcript-indexed ATAC-seq for precision immune profiling

Ansuman T. Satpathy^{1,2,14}, Naresha Saligrama^{3,14}, Jason D. Buenrostro^{4,5,14}, Yuning Wei^{1,6}, Beijing Wu⁷, Adam J. Rubin⁶, Jeffrey M. Granja^{1,7,8}, Caleb A. Lareau⁴, Rui Li^{1,6}, Yanyan Qi^{1,6}, Kevin R. Parker^{1,6}, Maxwell R. Mumbach^{1,7}, William S. Serratelli³, David G. Gennert^{1,7}, Alicia N. Schep^{1,7}, M. Ryan Corces^{1,6}, Michael S. Khodadoust⁹, Youn H. Kim⁶, Paul A. Khavari⁶, William J. Greenleaf^{7,10,11}, Mark M. Davis^{3,12,13,15*} and Howard Y. Chang^{1,6,7,15*}

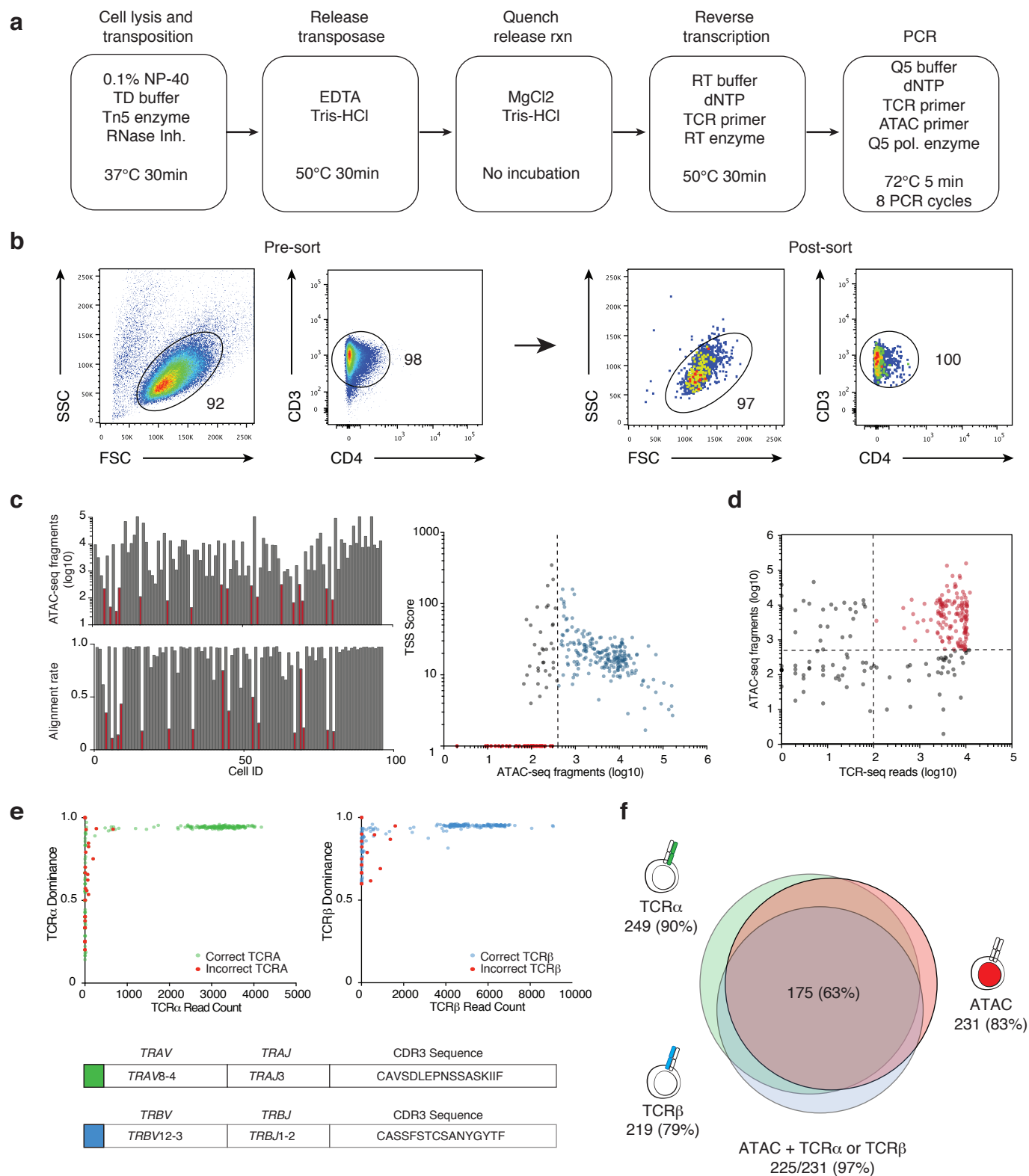
¹Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA, USA. ²Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA. ³Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA. ⁴Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard, Cambridge, MA, USA. ⁵Harvard Society of Fellows, Harvard University, Cambridge, MA, USA. ⁶Department of Dermatology, Stanford University School of Medicine, Stanford, CA, USA. ⁷Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. ⁸Biophysics Program, Stanford University School of Medicine, Stanford, CA, USA. ⁹Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. ¹⁰Department of Applied Physics, Stanford University, Stanford, CA, USA. ¹¹Chan Zuckerberg Biohub, San Francisco, CA, USA. ¹²Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA. ¹³Institute for Immunity, Transplantation and Infection, Stanford University, Stanford, CA, USA. ¹⁴These authors contributed equally: Ansuman T. Satpathy, Naresha Saligrama, Jason D. Buenrostro. ¹⁵These authors jointly supervised this work: Mark M. Davis, Howard Y. Chang. *e-mail: mmDavis@stanford.edu; howchang@stanford.edu

In the format provided by the authors and unedited.

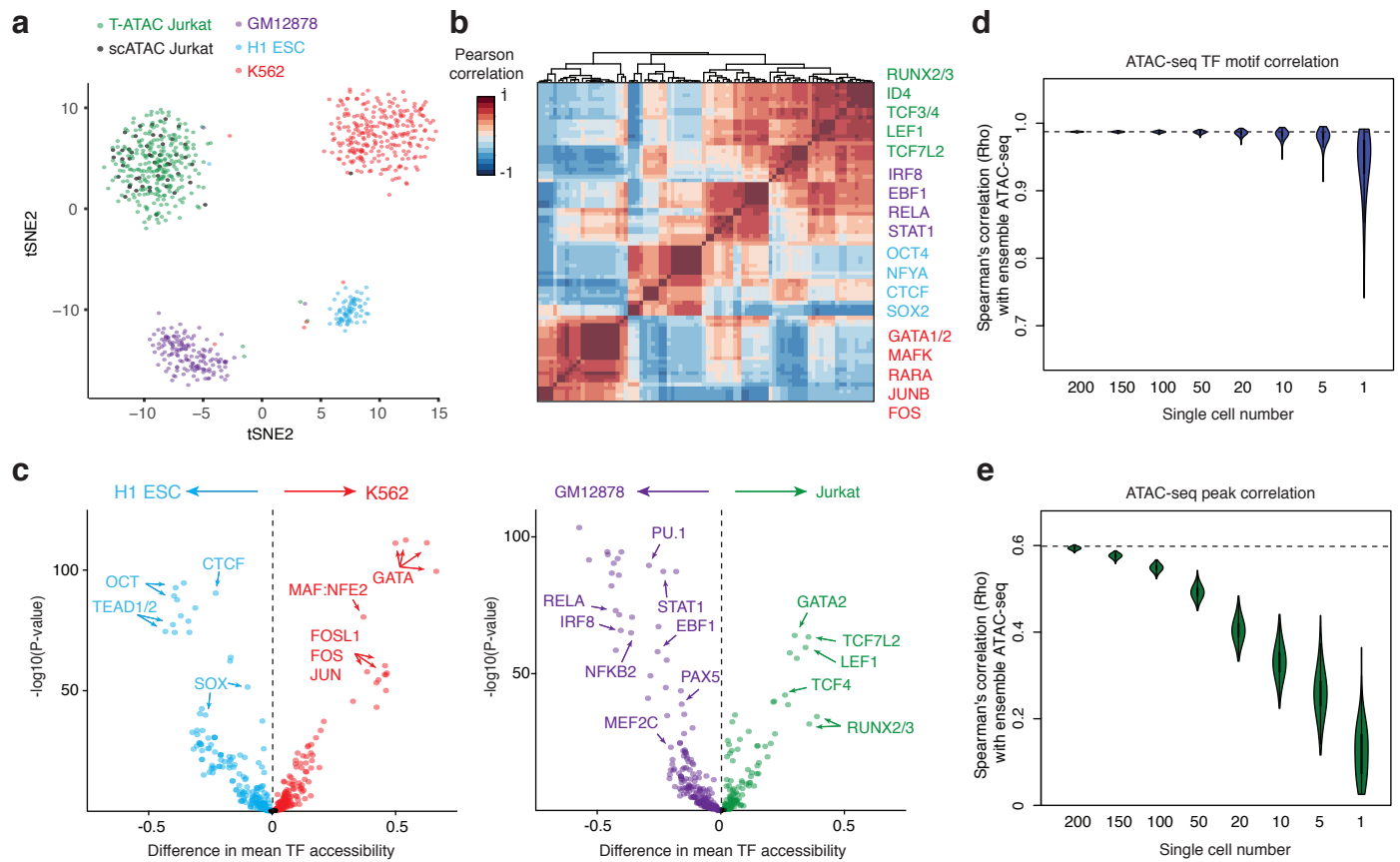
Transcript-indexed ATAC-seq for precision immune profiling

Ansuman T. **Satpathy**^{1,2,14}, Naresha **Saligrama**^{3,14}, Jason D. **Buenrostro**^{4,5,14}, Yuning **Wei**^{1,6}, Beijing **Wu**⁷, Adam J. **Rubin**⁶, Jeffrey M. **Granja**^{1,7,8}, Caleb A. **Lareau**⁴, Rui **Li**^{1,6}, Yanyan **Qi**^{1,6}, Kevin R. **Parker**^{1,6}, Maxwell R. **Mumbach**^{1,7}, William S. **Serratelli**³, David G. **Gennert**^{1,7}, Alicia N. **Schep**^{1,7}, M. Ryan **Corces**^{1,6}, Michael S. **Khodadoust**⁹, Youn H. **Kim**⁶, Paul A. **Khavari**⁶, William J. **Greenleaf**^{7,10,11}, Mark M. **Davis**^{3,12,13,15*} and Howard Y. **Chang**^{1,6,15*}

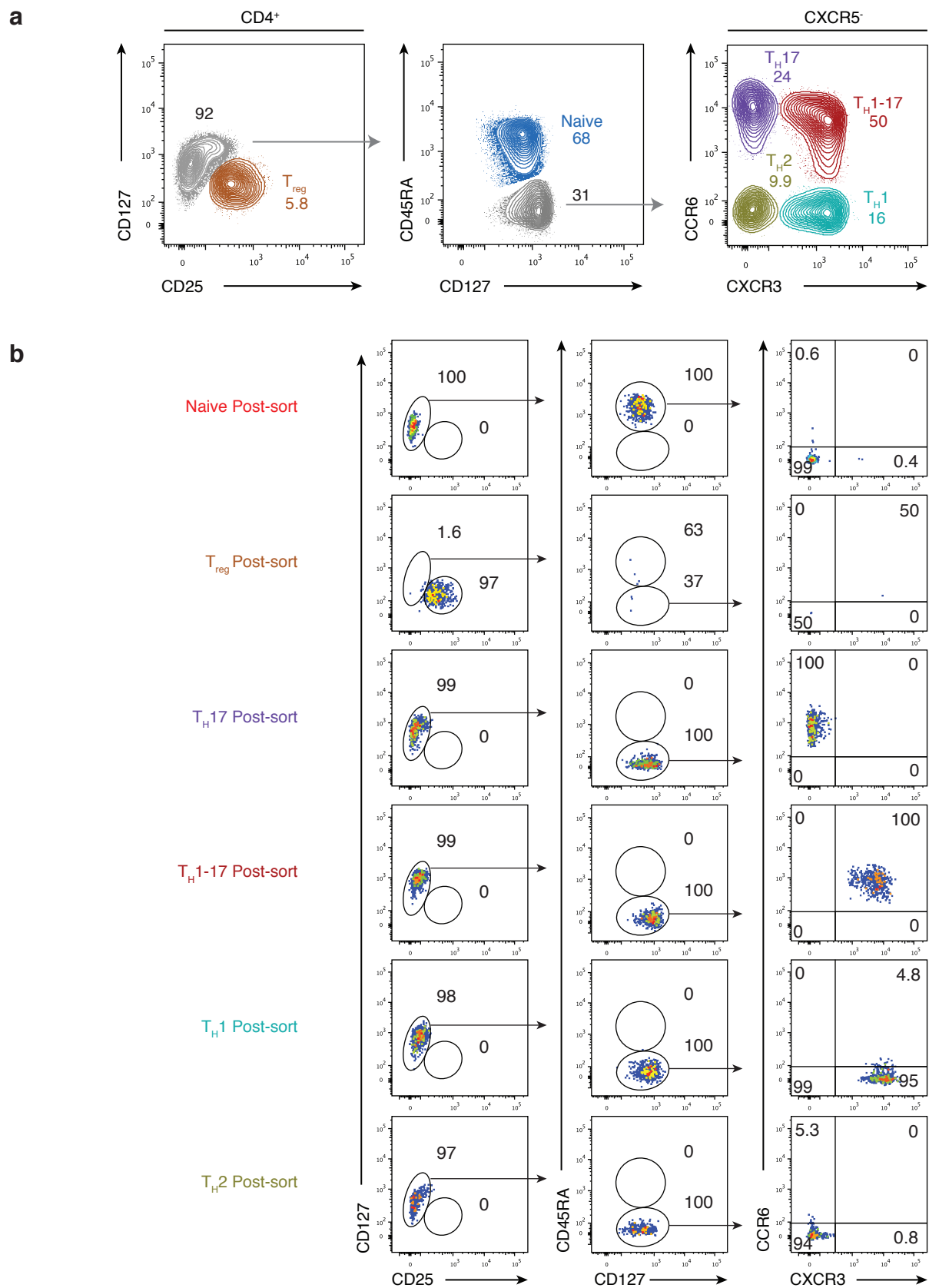
¹Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA, USA. ²Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA. ³Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA. ⁴Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard, Cambridge, MA, USA. ⁵Harvard Society of Fellows, Harvard University, Cambridge, MA, USA. ⁶Department of Dermatology, Stanford University School of Medicine, Stanford, CA, USA. ⁷Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. ⁸Biophysics Program, Stanford University School of Medicine, Stanford, CA, USA. ⁹Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. ¹⁰Department of Applied Physics, Stanford University, Stanford, CA, USA. ¹¹Chan Zuckerberg Biohub, San Francisco, CA, USA. ¹²Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA. ¹³Institute for Immunity, Transplantation and Infection, Stanford University, Stanford, CA, USA. ¹⁴These authors contributed equally: Ansuman T. Satpathy, Naresha Saligrama, Jason D. Buenrostro. contributed equally to this work. ¹⁵These authors jointly supervised this work: Mark M. Davis, Howard Y. Chang. *e-mail: mmdavis@stanford.edu; howchang@stanford.edu



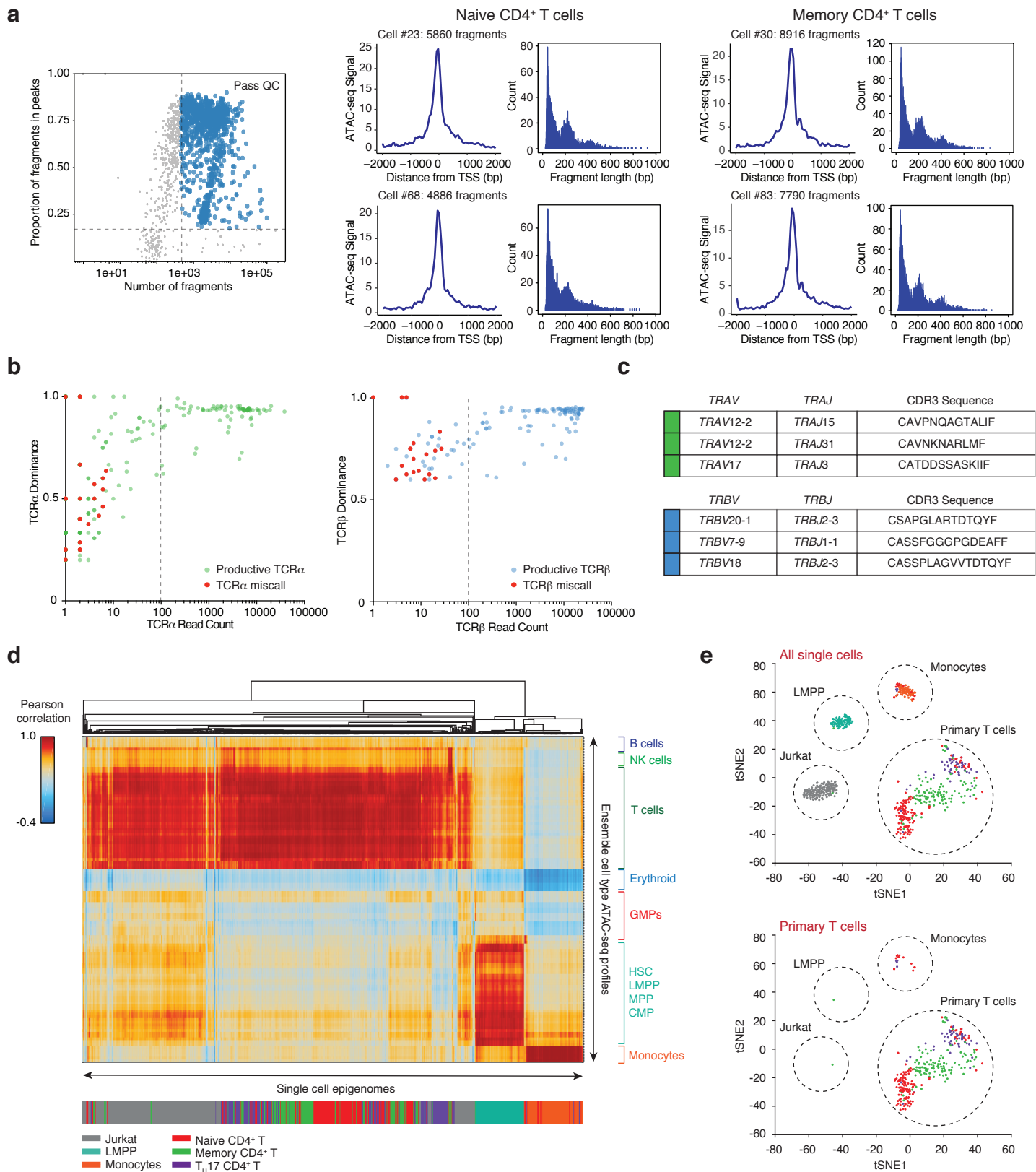
Supplementary Figure 1. T-ATAC-seq protocol and quality control measurements. (a) Detailed T-ATAC-seq protocol outlining biochemical reactions occurring in each microfluidic chamber in the IFC. Microfluidic chambers are indicated as boxes. (b) FACS analysis of Jurkat cells pre- (left) and post-sort (right). Cells were sorted for single live cells prior to loading in the IFC. Numbers represent the percentage of cells within the indicated gate. (c) ATAC-seq quality measurements for single Jurkat cells in a single experiment (96 cells). Plots show unique nuclear ATAC-seq fragments (left, top) and read alignment rate (left, bottom) for each single cell. Single cells from three independent T-ATAC-seq experiments on Jurkat cells are compared for ATAC-seq fragment number vs TSS enrichment rate (right). Dashed line indicates quality filter of 500 unique nuclear fragments per cell. (d) Comparison of ATAC-seq fragments and TCR-seq paired-end reads in single Jurkat cells from three individual T-ATAC-seq experiments. Dashed lines indicate quality filters of 500 unique nuclear fragments for ATAC-seq data and 100 reads for TCR-seq data. (e) TCR-seq quality control measurements in Jurkat cells. Shown are TCR α or TCR β reads in single cells and associated dominance of the major TCR clone (top). TCRs that passed filter correctly obtained Jurkat TCR sequences (green and blue dots), while TCRs not meeting quality filters identified incorrect TCRs (red dots). CDR3 sequences and gene usage of the Jurkat TCR are shown in the bottom panels. (f) Overlap of single-cell TCR-seq and ATAC-seq data in all cells in which either TCR or ATAC sequence was obtained.



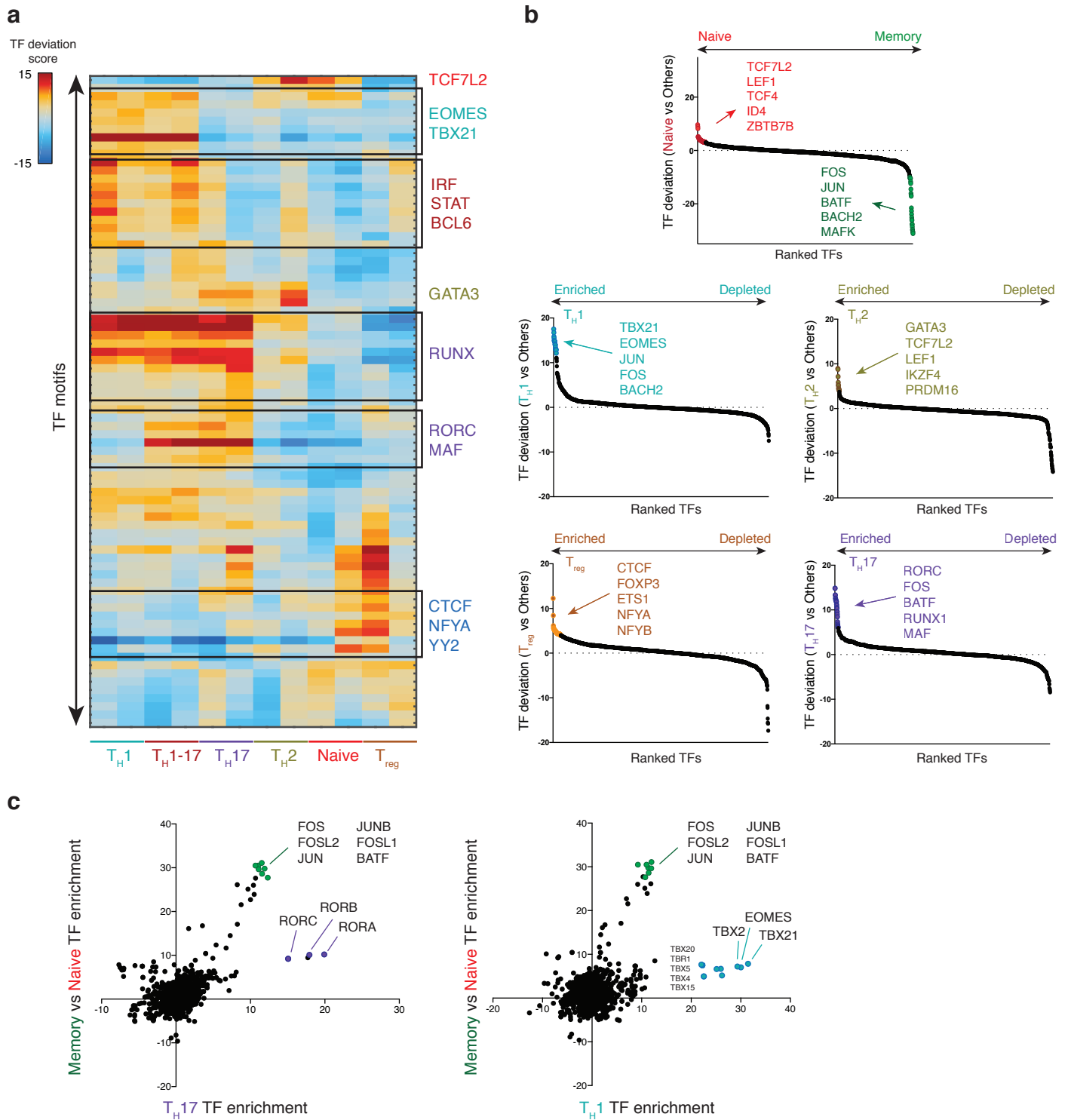
Supplementary Figure 2. T-ATAC-seq analysis of Jurkat cells. (a) t-SNE projection of T-ATAC-seq data from single Jurkat cells (231 cells, 3 independent experiments), scATAC-seq data from single Jurkat cells (49 cells, 1 independent experiment), and scATAC-seq data from previously published single GM12878 (159 cells), H1 ESC (84 cells), and K562 (258 cells)⁸. (b) Pearson correlation of TF deviation z-scores in single cells described in a. (c) TF bias-corrected deviation enrichments in aggregated single-cell populations described in a. TF bias-corrected deviation enrichments in K562 cells compared to H1 ESCs (left) and in Jurkat cells compared to GM12878 cells (right). TF enrichments are calculated as the difference in mean TF motif accessibility between two populations of single cells. P-values were calculated using a two-tailed t-test. (d) Spearman's correlation of TF z-score in ensemble ATAC-seq data in Jurkat cells compared to aggregated single cells. The number of aggregated single cells in each comparison is noted on the x-axis. Cells were obtained from 3 independent experiments. Thick black boxes within the violin plots extend from the 25th to the 75th percentile, and boundaries of the violin are max and min values. (e) Spearman's correlation of fragment counts in ATAC-seq peaks in ensemble ATAC-seq data in Jurkat cells compared to fragment counts in aggregated single cells. The number of aggregated single cells in each comparison is noted on the x-axis. Cells were obtained from 3 independent experiments. Thick black boxes within the violin plots extend from the 25th to the 75th percentile, and boundaries of the violin are max and min values.



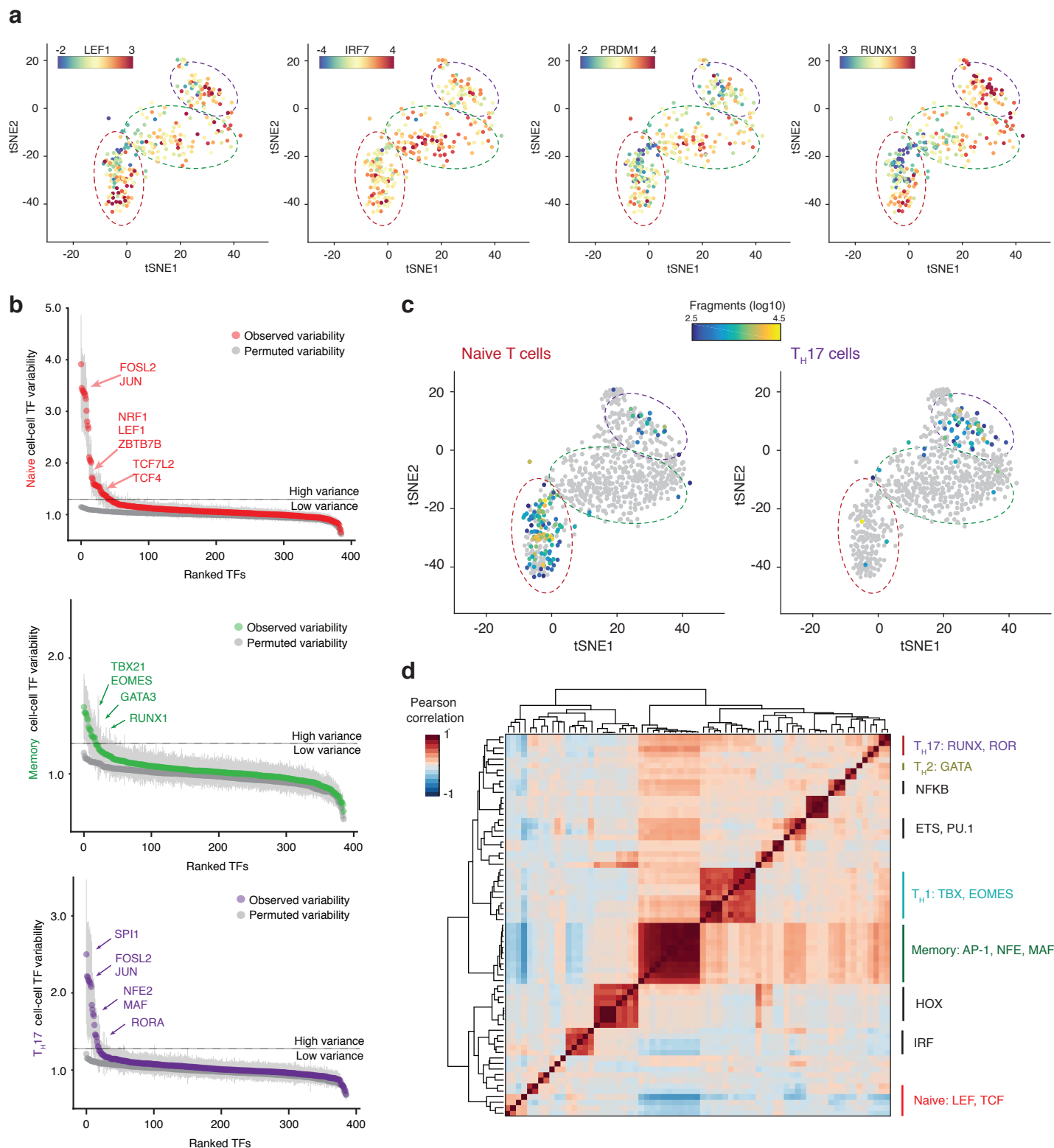
Supplementary Figure 3. FACS analysis of CD4⁺ T cell subtypes. (a) Peripheral blood cells were stained for expression of the indicated markers. Two-color histograms are shown for live cells pre-gated as indicated above the diagram. Numbers represent the percentage of cells within the indicated gate. Data are representative of three independent experiments. (b) Post-sort purities for CD4⁺ T cell subtypes. Numbers represent the percentage of cells within the indicated gate. All data in this Figure are representative of 3 independent experiments.



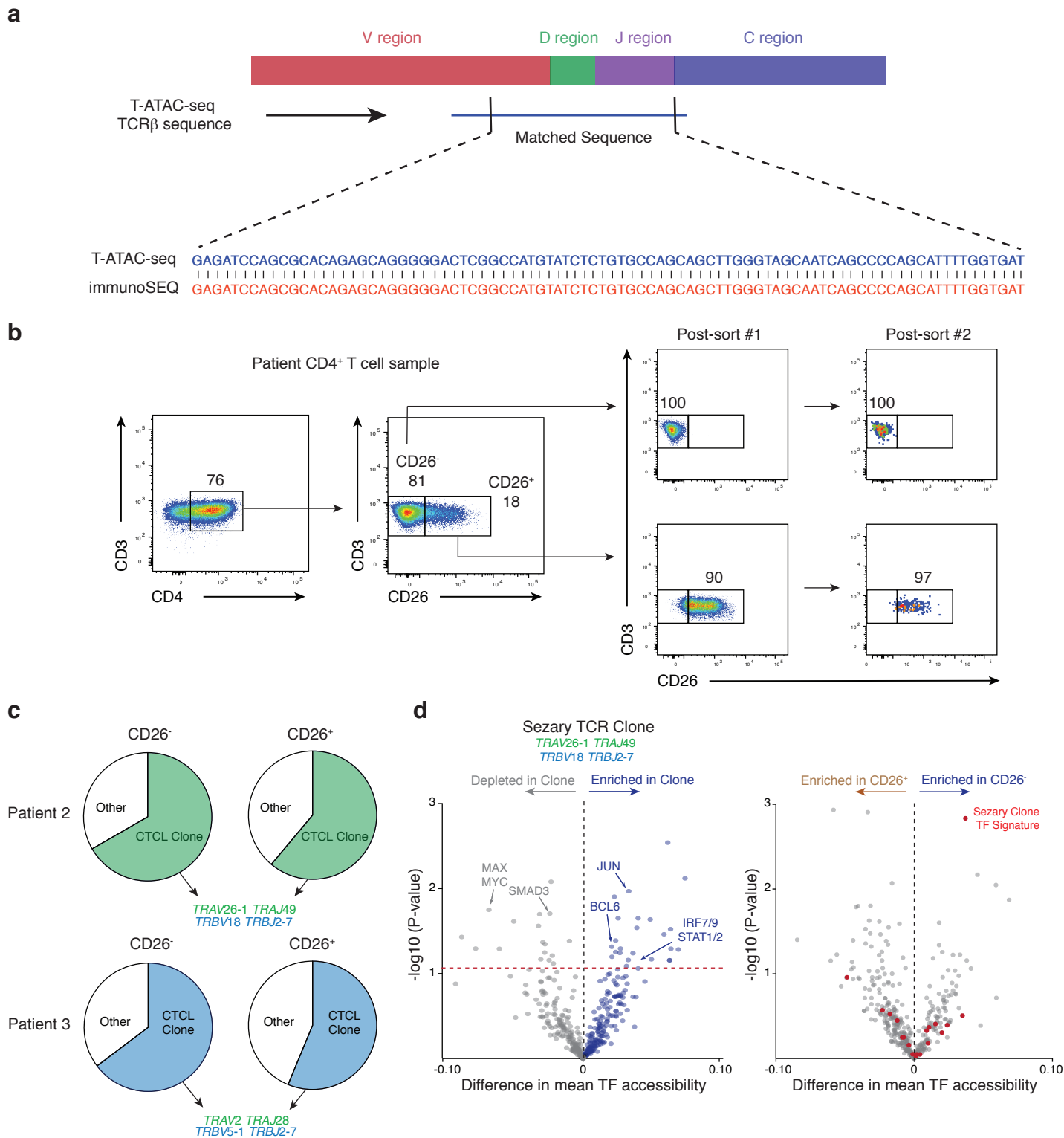
Supplementary Figure 4. T-ATAC-seq data quality measurements in primary human T cells. (a) T-ATAC-seq data quality control filters. Shown are the number of unique ATAC-seq nuclear fragments in each single primary T cell compared to the percentage of fragments in ATAC-seq peaks derived from ensemble T cell ATAC-seq data (left), and individual single-cell profiles for four cells (right). Single cells show enrichment at transcription start sites (TSS) and nucleosomal periodicity of ATAC-seq fragment lengths. Fragment length indicates the genomic distance between two Tn5 insertion sites, as determined by paired-end sequencing of ATAC fragments. (b) Quality measurements of TCR-seq profiles from T-ATAC-seq in single cells. Shown are TCR α and TCR β reads in each single cell compared to TCR read dominance in each cell. Red dots indicate incorrect TCR calls (either TCR γ or TCR δ sequences). (c) Example sequences for productive TCR data are shown. (d) Pearson correlation of PC scores of single cells (x-axis; 879 single cells) and ensemble cells (y-axis; 93 ensemble cell types). Ensemble cell types are derived from data generated in this study (T cell subtypes, 3 independent experiments) and from Corces et al. (2016)⁹. (e) t-SNE projection of PCA scores for all cells (top) or primary human T cells (bottom). Single T cells that cluster with monocytes/LMPs likely represent FACS contaminants and are removed from downstream analysis.



Supplementary Figure 5. TF motif profiles in ensemble ATAC-seq data from CD4⁺ T cell subtypes. (a) Heat map of TF deviation z-scores for ensemble T cell ATAC-seq profiles. (b) Ranked TF motif z-scores for each T cell subtype compared to the average TF z-score calculated in all other subtypes. (c) Comparison of TF deviation z-score enrichment (mean difference) in T_H17 cells vs all other T cell subtypes (left; x-axis) and in T_H1 cells vs all other T cell subtypes (right; x-axis). The enrichment of TFs in each subtype is compared to their enrichment in all T cell memory subtypes compared to naive cells (y-axis). Labels highlight TFs that are specifically enriched in T_H17 or T_H1 cells (purple and blue dots) or TFs that are enriched in all memory cells (green dots). Data shown in b and c were obtained from 3 independent experiments.



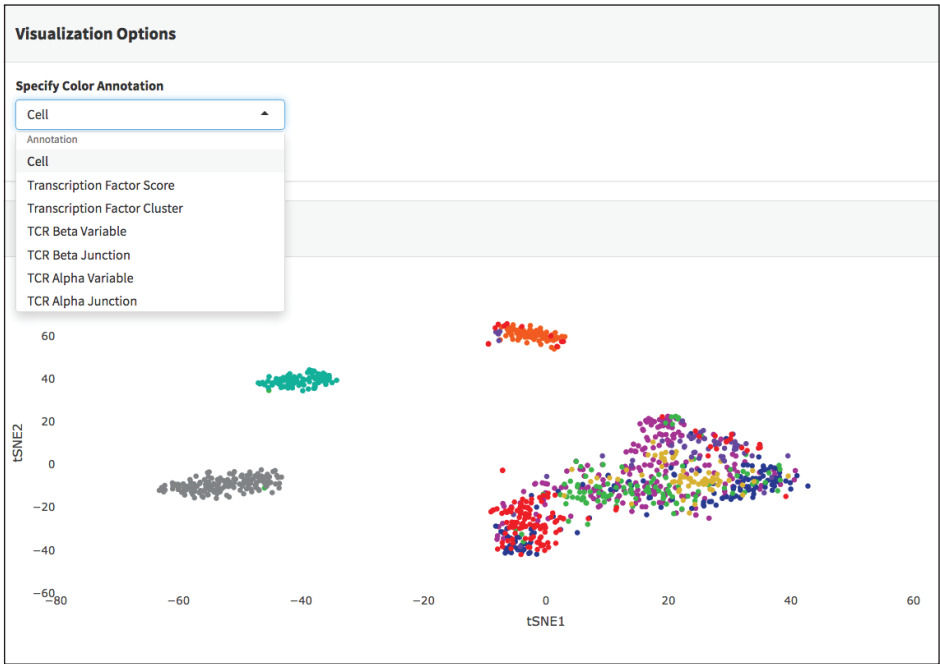
Supplementary Figure 6. Single-cell epigenomic variability in naive and memory CD4⁺ T cells. (a) t-SNE projection of sorted single naive, memory, and T_H17 T cells (320 cells, 6 independent experiments). Cells are colored by LEF1, IRF7, PRDM1, and RUNX1 motif accessibility TF scores. Scale bars indicate range of TF z-scores. (b) Cell-to-cell variability of TF motif accessibility in single T cells. Shown is the observed TF variability in sorted T cell populations (colored) and error estimates (gray). Variability measured from a permuted background is shown in gray dots for comparison (see Methods for details of background calculations). Selected high variance TFs are indicated by arrows. (c) t-SNE projection of single T cells. Naive T cells (left; 145 cells) or T_H17 cells (right; 60 cells) are colored by the number of unique nuclear ATAC-seq fragments obtained in that cell. Gray cells are additional populations of T cells (including Memory T and CTCL cells). These plots demonstrate that distinct clusters of naive T cells are not a byproduct of differences in ATAC-seq fragments obtained per cell. Highlighted cells were obtained from 2 independent experiments for each subtype. (d) Pearson correlation of TF deviation z-scores for single naive, memory, and T_H17 cells (450 single cells, 6 independent experiments). TF modules associated with canonical T helper cell phenotypes are indicated on the right.



Supplementary Figure 7. T cell clonality in CD26⁺ and CD26⁻ CTCL cells. (a) Confirmation of TCRβ sequence obtained in clonal CTCL cells using T-ATAC-seq with immunoSEQ profiling from the same sample (Adaptive Biotechnologies; related to Fig. 5b). (b) Representative FACS strategy and post-sort purities for CD26⁺ and CD26⁻ CTCL cells. Peripheral CD4⁺ blood cells were stained for expression of the indicated markers. Numbers represent the percentage of cells within the indicated gate. Cells were double-sorted to ensure high purity of the desired populations. Data are representative of 3 independent experiments. (c) Expanded TCR clones are present in CD26⁺ and CD26⁻ CD4⁺ T cell populations in CTCL patients #2 and #3. (d) TF bias-corrected deviation enrichments in aggregated clonal T cells from patient #2 compared to all other T cells (left). TF deviations enrichments in aggregated clonal cells are not enriched in CD26⁻ cells compared to CD26⁺ cells (right). Clone-specific TFs are highlighted in red. 112 single cells were obtained from 2 independent experiments, and p-values were calculated using a two-tailed t-test.

a

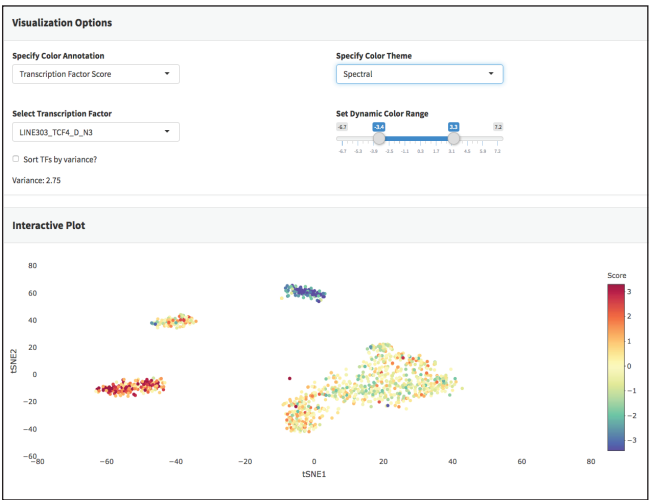
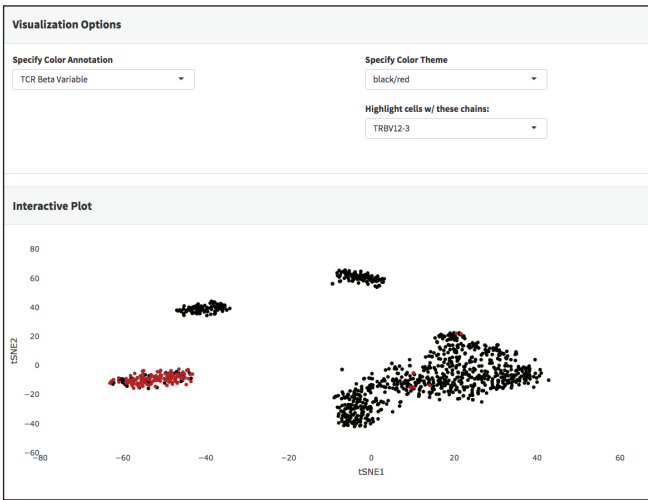
Visualization of single-cell clustering



b

TCR α and β sequences

TF motif deviations



Supplementary Figure 8. Open access T-ATAC-seq data browser. (a) Browser screenshot from tcr.buenrostrolab.com showing drop-down menu options to navigate single-cell TF deviation scores and TCR sequences. (b) Example screenshots showing single cells with TRBV12-3 identity (left) and colored by TCF4 TF deviation z-score (right).