LETTERS
https://doi.org/10.1038/s41591-018-0008-8

nature medicine

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**[1,6,7,15]*

T cells create vast amounts of diversity in the genes that encode their T cell receptors (TCRs), which enables individual clones to recognize specific peptide–major histocompatibility complex (MHC) ligands. Here we combined sequencing of the TCR-encoding genes with assay for transposase-accessible chromatin with sequencing (ATAC-seq) analysis at the single-cell level to provide information on the TCR specificity and epigenomic state of individual T cells. By using this approach, termed transcript-indexed ATAC-seq (T-ATAC-seq), we identified epigenomic signatures in immortalized leukemic T cells, primary human T cells from healthy volunteers and primary leukemic T cells from patient samples. In peripheral blood CD4$^+$ T cells from healthy individuals, we identified cis and trans regulators of naive and memory T cell states and found substantial heterogeneity in surface-marker-defined T cell populations. In patients with a leukemic form of cutaneous T cell lymphoma, T-ATAC-seq enabled identification of leukemic and nonleukemic regulatory pathways in T cells from the same individual by allowing separation of the signals that arose from the malignant clone from the background T cell noise. Thus, T-ATAC-seq is a new tool that enables analysis of epigenomic landscapes in clonal T cells and should be valuable for studies of T cell malignancy, immunity and immunotherapy.

T lymphocytes recognize self and foreign antigens and are the central drivers of regulatory and effector immune responses. Each T cell expresses a TCR that recognizes antigens in the context of MHC molecules displayed on the surface of antigen-presenting or pathogen-infected cells. The major TCR species is composed of α- and β-subunits that are encoded by genes generated due to somatic V(D)J recombination, which results in the production of a diverse repertoire of antigen-reactive T cells, with up to a possible 10^{14} unique heterodimers in each individual[1]. As a result of antigen-specific or malignant clonal expansion, the TCR also serves as a faithful identifier of its clonal origin, as T cells expressing identical TCR-αβ pairs must almost invariably arise from a common cellular ancestor. The specific pairing of a TCR-α subunit and a TCR-β subunit from one cell is necessary to recapitulate its antigen specificity and is critical for weaponizing or disarming an immune response for immunotherapy. Therefore, the identification of TCR-αβ-encoding sequences is critical to understanding the identity of single T cells, and methods that pair TCR-αβ-encoding sequences with cell and activation states may uncover clonal gene regulatory pathways missed by ensemble measurements.

Recent advances in genome sequencing technologies have enabled single-cell gene expression and epigenetic measurements and have revealed variability in immune-cell development and responsiveness[2–5]. Our groups recently developed methods to efficiently amplify and sequence the *TRA* and *TRB* loci (which encode the TCR-α and TCR-β chains, respectively) from single T cells[6] (termed TCR-seq) and to measure epigenetic changes genome wide in single cells. The latter method, termed single-cell ATAC-seq (scATAC-seq), enables measurement of regulatory DNA elements by direct transposition of sequencing adaptors into regions of accessible chromatin[7–9]. Unlike methods to measure the transcriptome in single cells, scATAC-seq identifies cell-to-cell variation in cis regulatory elements and trans factors that drive epigenetic cell states. Moreover, analysis of single-cell epigenomic profiles can be used to reveal significant variability within cell-surface-marker-defined populations and the existence of cell states obscured by ensemble measurements[10].

Here we combined these two technologies to produce a method that allows one to study both the epigenetic landscape and the T cell specificity simultaneously at the single-cell level. This two-way analysis may facilitate discovery of antigens that drive a certain T cell fate, or conversely, cis and trans regulators that drive the expansion of a T cell clone. We refer to this method as T-ATAC-seq. The T-ATAC-seq experimental pipeline integrates scATAC-seq with targeted TCR-seq in the same single cell, followed by high-throughput sequencing and computational integration of both datasets. To demonstrate the performance and utility of T-ATAC-seq, we performed this method on 1,344 human T cells that were sorted using standard subset-specific cell surface markers, and we integrated the

---

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

analysis of regulatory landscapes with TCR identity. T-ATAC-seq in peripheral blood CD4$^+$ T cells from healthy volunteers revealed epigenomic signatures and single-cell variability of naive and memory CD4$^+$ T cells. Notably, unbiased single-cell analysis identified divergent chromatin states within cell-surface-marker-defined T cell subtypes. We extended the use of this method to clinical samples from patients with T cell leukemia. T-ATAC-seq enabled the identification of cancer-clone-specific epigenomic signatures, which were not apparent from ensemble measurements. These data demonstrate the utility of T-ATAC-seq as a new tool for single-cell epigenomic characterization of T cells for both research and clinical applications.

Results

Performance of T-ATAC-seq in human immortalized T cells. We implemented T-ATAC-seq using an automated microfluidic platform (C1; Fluidigm; Fig. 1a and Supplementary Fig. 1a). For this approach, single cells were first individually captured on the integrated fluidics circuit (IFC) in single-cell chambers and then subjected to cell lysis and DNA transposition with the prokaryotic Tn5 enzyme, which was loaded with sequencing adapters. After transposition of accessible chromatin, Tn5 was released from DNA fragments, and the *TRA* and *TRB* RNAs within each chamber were subjected to reverse transcription (RT) using primers targeting the constant-region-encoding sequences in *TRA* and *TRB*. Immediately after RT, the 5′ ends of the ATAC-seq fragments were extended, and all of the chamber contents were amplified by PCR. TCR-encoding fragments were amplified using primers targeting the sequences encoding the TCR constant and variable regions in *TRA* and *TRB*. Single-cell libraries were then collected, and the *TRA* and *TRB*, or the ATAC, amplicons were further amplified with cell-identifying barcoded primers, pooled and sequenced on a high-throughput sequencing instrument.

To assess the performance of this method, we performed T-ATAC-seq analysis in 288 single human Jurkat leukemia cells (Supplementary Fig. 1b). Combined ATAC-seq and TCR-seq (for either *TRA* or *TRB*) profiles were obtained in 93.9% of captured live cells, and 80% of live cells produced ATAC-seq and a paired *TRA* and *TRB* sequence (Fig. 1b). Next we evaluated the quantity and quality of the ATAC-seq data. Microfluidic chambers that produced low-quality data (i.e., those that corresponded to the data from empty chambers or dead-cell captures) were excluded from further analysis using cut-offs for unique nuclear fragment number and fraction of fragments in accessible chromatin sites, as previously described (Supplementary Fig. 1c,d and Methods)[8–11]. Chambers passing this filter yielded an average of 8.5×10^3 fragments that mapped to the nuclear genome, and approximately 38% of fragments were within peaks present in ensemble Jurkat ATAC-seq profiles (Fig. 1c). T-ATAC-seq data recapitulated several characteristics of ensemble ATAC-seq, including fragment-length periodicity and enrichment of fragments at transcription start sites (TSSs) (Fig. 1d and Supplementary Fig. 1c). Notably, T-ATAC-seq data quality in single cells were similar to those derived from scATAC-seq alone (Fig. 1d), demonstrating that incorporating targeted RT and PCR of the TCR-encoding transcripts did not affect the quality of the ATAC-seq data.

We next assessed the performance of T-ATAC-seq in obtaining *TRA* and *TRB* sequences from single cells. T-ATAC-seq *TRA*- and *TRB*-specific primers were designed to amplify the complementarity-determining region 3 (CDR3) in the *TRA* and *TRB* loci. *TRA* and *TRB* sequence quality was assessed by *TRA* and *TRB* sequence read number and single-cell clonal dominance, as previously described[6], and only chambers that generated high-quality *TRA* and *TRB* sequences were included in downstream analyses. On average, we obtained 2.7×10^3 reads for *TRA* and 4.2×10^3 reads for *TRB* in chambers that passed the quality-control filters (Fig. 1e and Supplementary Fig. 1d,e). In chambers that produced either ATAC-seq or TCR-seq reads, we obtained a *TRA* sequence in 89.9% of cells (249/277) and a *TRB* sequence in 79.1% of cells (219/277), which resulted in paired *TRA* and *TRB* sequences in 71% of cells (196/277) (Supplementary Fig. 1f). These efficiencies were similar to those from previous techniques that obtained TCR-encoding sequences from single cells[6,12,13]. The *TRA* and *TRB* sequences in all of the cells that passed the filter correctly identified the Jurkat TCR heterodimer as *TRBV*12-3–*TRBJ*1-2 and *TRAV*8-4–*TRAJ*3 (Fig. 1f). Finally, species mixing experiments using mouse cells (58αβ-negative T cell hybridoma cells that were transduced with sequences encoding a mouse TCR; labeled with calcein red) and human T cells (Jurkat; labeled with calcein green) confirmed that T-ATAC-seq correctly paired cells visualized on the microfluidic chip with species-specific open chromatin, and *TRA* and *TRB* sequences (Fig. 1g). Human ATAC-seq fragments were always paired with human TCRs, and mouse ATAC-seq fragments with mouse TCRs, with the exception of one doublet out of 94 cells. In summary, T-ATAC-seq efficiently and accurately pairs *TRA* and *TRB* sequence identity with chromatin accessibility in single T cells.

Single-cell epigenomic analysis using T-ATAC-seq. Single-cell epigenomic data can be assessed at the level of regulatory DNA elements or of transcription factor (TF) activity across many loci, as computed from observed versus expected fragments in TF-binding sites in each single cell, as previously described[8,11,14]. The performance of T-ATAC-seq was comparable to that of scATAC-seq in both measurements. For the first kind of measurement (level of regulatory DNA elements), aggregate T-ATAC-seq profiles from 231 single cells closely reproduced population measurements profiled by DNase I hypersensitivity sequencing (DHS-seq) and ensemble ATAC-seq generated from 10^7 or 5×10^4 cells, respectively (Fig. 2a). Single-cell profiles were strongly enriched for fragments within open chromatin sites present in ensemble profiles (Fig. 2b). For the second kind of measurement, TF motif activity in Jurkat cells that were identified using T-ATAC-seq or scATAC-seq yielded similar profiles (Supplementary Fig. 2a). Jurkat cells showed high accessibility at DNA regions that contained motifs for members of the T cell factor (TCF) and lymphoid enhancer-binding factor (LEF) family, including TCF7L2 and LEF1, and for the runt-related TF family members RUNX2 and RUNX3, as compared to single-cell profiles from H1 embryonic stem cells (ESCs), GM12878 B lymphoblastoid cells and K562 myeloid leukemia cells (Fig. 2c and Supplementary Fig. 2b,c). It is of note that TF motif enrichments (hereafter referred to as TF deviation scores) reflect the activity of all TFs with similar DNA-binding motifs, rather than any particular TF. Therefore, the high deviation scores of TCF7L2 in Jurkat cells may reflect the function of additional TCF family members, such as TCF1, which has previously been shown to function in early T cell progenitors to establish T cell fate[15]. Similarly, high RUNX2 and RUNX3 deviation scores also encompass RUNX1 activity, as seen in early T cell development[16]. Differential analysis of ATAC-seq peaks that contained binding sites for each TF identified cell-type-specific accessible sites. For example, accessible regions in Jurkat cells containing TCF7L2 motifs included promoters and enhancers for the T cell–specific genes *CD28*, *CD3D*, *CD3E* and *CD3G* (Fig. 2d). Finally, we determined how many single cells were required to reliably recapitulate ensemble ATAC-seq measurements. Notably, TF deviation scores were highly accurate even in individual cells, as compared to scores derived from ensemble ATAC-seq data (Spearman rank: $\rho=0.957$, $P<0.01$; Supplementary Fig. 2d). In contrast, accurate quantification of individual open chromatin sites required the aggregation of approximately 50 single cells in order to reflect population peak profiles (Spearman rank: $\rho=0.5$, $P<0.01$; Supplementary Fig. 2e). Therefore, our strategy to assess epigenomic signatures using T-ATAC-seq data was to first characterize cells using TF deviation

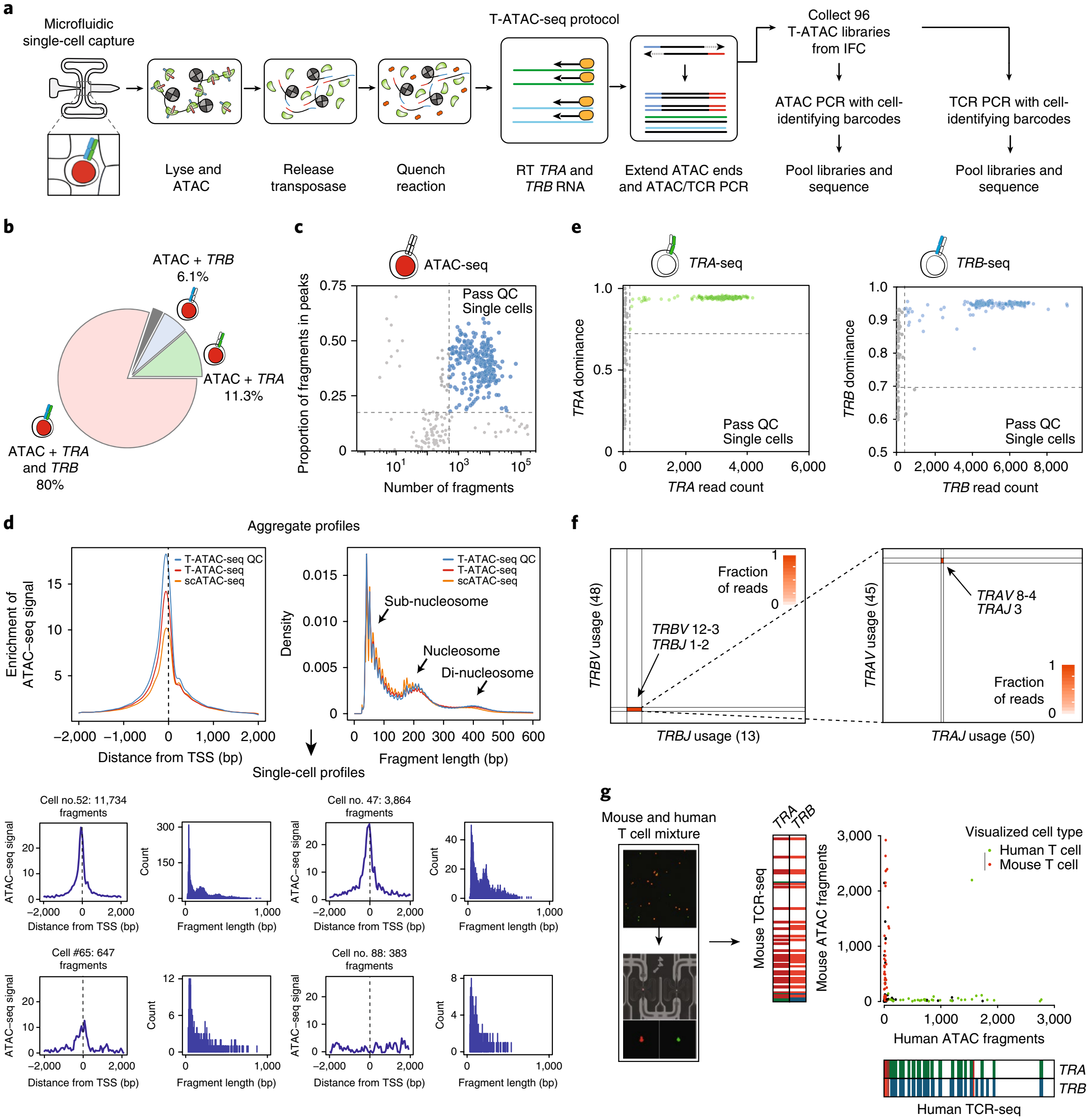

Fig. 1 | T-ATAC-seq generates open chromatin and TCR profiles in single T cells. a, Outline of the T-ATAC-seq protocol. Squares indicate individual microfluidic chambers in the IFC. T cells were individually captured and sequentially subjected to ATAC-seq (chambers 1–3), RT of *TRA* and *TRB* transcripts, and amplification of ATAC-seq and TCR-seq amplicons in nanoliter-scale reaction volumes. Single-cell libraries were then amplified with cell-identifying barcodes and analyzed by high-throughput sequencing. **b**, Pie chart indicating the overlap of TCR-seq and ATAC-seq data from single Jurkat cells ($n = 231$ single cells; $n = 3$ independent experiments) that passed quality-control filters. Shown is the proportion of cells that generated ATAC-seq profiles in which *TRA* or *TRB* sequence was also obtained. The gray bar indicates the portion of cells in which ATAC-seq data were obtained but in which *TRA* or *TRB* data were not (2.6%). **c**, T-ATAC-seq data quality-control (QC) filters. Shown is the number of unique ATAC-seq nuclear fragments in each single Jurkat cell, as compared to the percentage of fragments in ATAC-seq peaks derived from ensemble Jurkat ATAC-seq profiles ($n = 384$ single cells; $n = 4$ independent experiments). **d**, Aggregate (top) and single-cell (bottom) T-ATAC-seq profile characteristics. Shown are the enrichments of ATAC-seq Tn5 insertions around TSSs and the nucleosomal periodicity of ATAC-seq fragment lengths. Aggregate profiles obtained from all T-ATAC-seq single cells ($n = 288$ cells), T-ATAC-seq single cells that passed QC filters ($n = 231$ single cells), and scATAC-seq cells are shown ($n = 49$ cells). Fragment length indicates the genomic distance between two Tn5 insertion sites, as determined by paired-end sequencing of ATAC fragments. Density indicates the fraction of fragments with the indicated length. The cell number indicates the position of each individual cell in the IFC, and the associated fragment number indicates the number of unique nuclear fragments obtained in that cell. Count indicates the number of fragments for each fragment length. **e**, QC filters for *TRA* (left) and *TRB* (right) sequences. Shown are *TRA* or *TRB* paired-end sequencing read counts for each single cell, as compared to the TCR dominance of the top clone for each cell ($n = 288$ single cells; $n = 3$ independent experiments). TCR dominance was quantified as the fraction of reads that supported the most prevalent TCR clone by sequence identity[6]. Dashed lines represent QC filters of 100 reads and 70% dominance for Jurkat cells. **f**, Heat maps showing the *TRA* or *TRB* rearrangements identified in Jurkat cells ($n = 288$ single cells; $n = 3$ independent experiments). Each axis represents all possible genes within the indicated *TRA* or *TRB* locus. The labeled genes indicate the sequences identified using T-ATAC-seq. **g**, Mouse and human T cell mixing experiment ($n = 94$ single cells). Shown are visualized cells in the IFC (left), unique nuclear ATAC-seq fragments aligning to the mouse or human genome, and TCR-seq clones identified when compared to mouse or human references (right). In the IFC, human T cells are labeled in green, and mouse T cells are labeled in red.

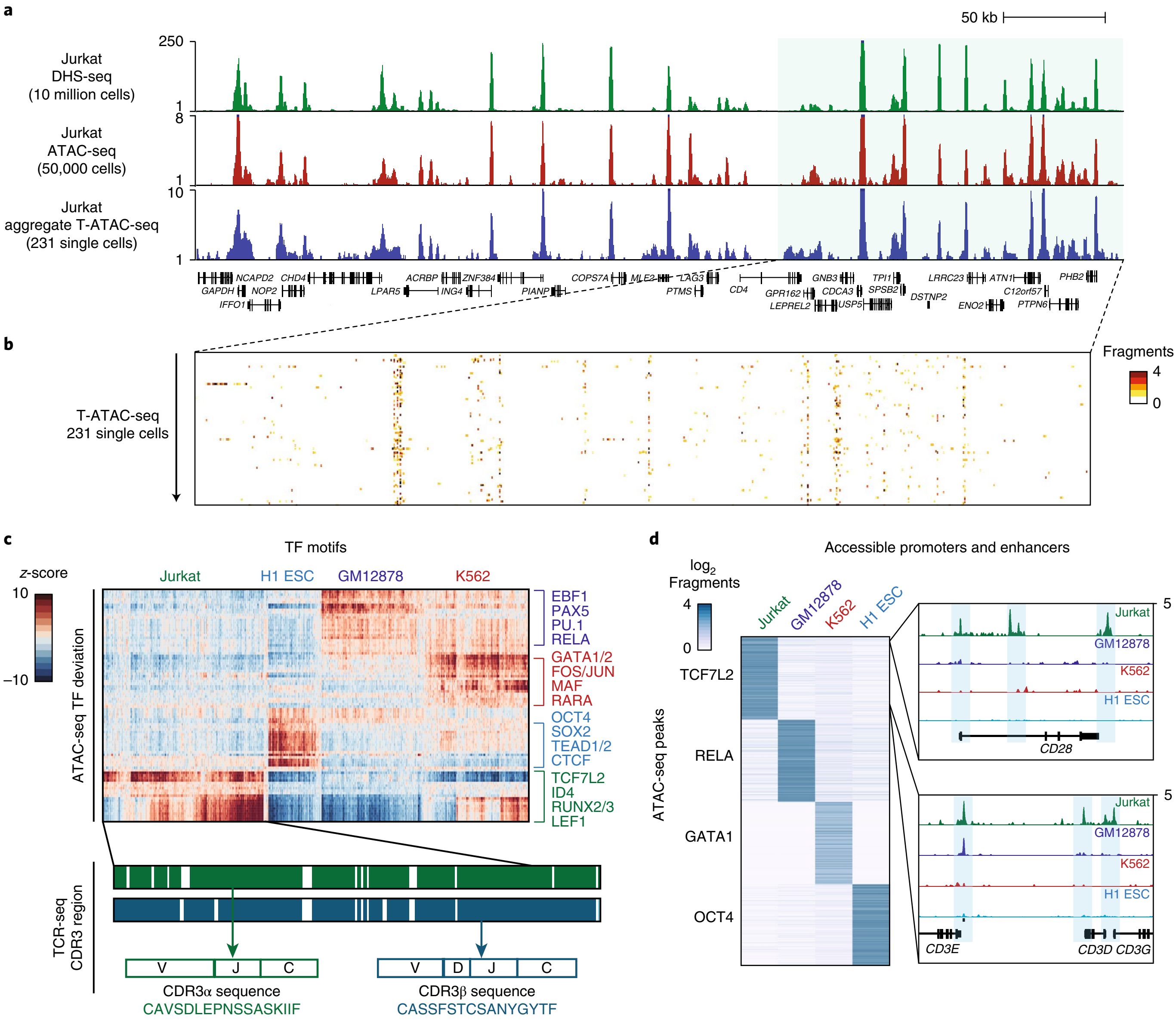

Fig. 2 | T-ATAC-seq identifies epigenomic signatures of clonal Jurkat T cells. **a**, Genome tracks showing a comparison of aggregate T-ATAC-seq profiles ($n = 231$ cells) to ensemble ATAC-seq ($n = 3$) and DHS-seq ($n = 2$) profiles. **b**, Magnified view of the indicated genome track in **a**, showing accessibility profiles for single Jurkat cells. Each pixel represents a 200-bp region. **c**, Heat map of TF deviation *z*-scores in single Jurkat cells (231 cells; $n = 3$ independent experiments) obtained using T-ATAC-seq, as compared to previously published profiles from H1 ESCs ($n = 84$ cells), GM12878 cells ($n = 159$ cells) and K562 cells ($n = 258$ cells) obtained using scATAC-seq[8]. The presence or absence of paired *TRA* or *TRB* sequences is indicated by green and blue bars. Single-letter amino acid sequences (bottom) represent the identified CDR3 region, which spans the gene sequences encoding the variable (V), diversity (D), joining (J) and constant (C) regions. **d**, Left, heat map showing ATAC-seq fragment counts in peaks (rows) containing the indicated motifs from aggregated single cells. Cell types analyzed (aggregated from single-cell profiles) are indicated above each column. Right, genome tracks for aggregated T-ATAC-seq data. Jurkat-specific peaks in the *CD28*, *CD3E*, *CD3D* and *CD3G* loci are shown.

scores and then to calculate accessibility differences at individual sites when single cells could be aggregated by their shared immunophenotype or by their *TRA* and *TRB* sequences.

T-ATAC-seq identifies single-cell regulatory signatures in primary CD4⁺ T cells. To build a comparison dataset for T-ATAC-seq profiles in primary cells and to establish T cell subset-specific chromatin landscape benchmarks, we generated ensemble ATAC-seq profiles from cell-surface-marker-defined CD4⁺ naive and memory T cell subtypes[17]. Peripheral blood CD4⁺ T cells were obtained from two healthy subjects (a total of three replicates), and T cell subsets were isolated by FACS and subjected to ATAC-seq. We profiled naive T cells ($CD4^+CD45RA^+CD25^-CD127^{hi}$), regulatory T cells ($T_{reg}$ cells; $CD4^+CD25^+CD127^{low}$), T helper (T_H) 1 cells ($CD4^+CD45RA^-CD25^-CD127^{hi}CXCR3^+CCR6^-CXCR5^-$), T_H17 cells ($CD4^+CD45RA^-CD25^-CD127^{hi}CXCR3^-CCR6^+CXCR5^-$), T_H1-17 cells ($CD4^+CD45RA^-CD25^-CD127^{hi}CXCR3^+CCR6^+CXCR5^-$) and T_H2 cells ($CD4^+CD45RA^-CD25^-CD127^{hi}CXCR3^-CCR6^-CXCR5^-$) (Supplementary Fig. 3a,b)[17]. Analysis of ensemble ATAC-seq profiles by principal component analysis (PCA) showed distinct chromatin states for each T cell subset; PC1 distinguished naive and memory T cell subtypes, PC2 distinguished T_{reg} cells from all other subtypes, and PC3 distinguished T_H1 and T_H17 subtypes (Fig. 3a). Analysis of differential ATAC-seq peaks showed that a large shift in chromatin accessibility accompanied the differentiation of naive T cells to memory T cells, with the majority of differential peaks (6,868 sites) closing in memory cells (Fig. 3b). In contrast, there were relatively fewer differences between T_H subtypes, and cell-type-specific open chromatin sites were located mainly at functional gene promoters and at distal elements (Fig. 3b–e). For example, T_{reg} cells showed increased accessibility at the promoter and upstream elements in the interleukin 2 receptor subunit alpha (*IL2RA*) locus, consistent with this gene's critical function in this cell type (Fig. 3c,d)[18]. Similarly, T_H1 and T_H1-17 cells showed increased accessibility at the interferon gamma (*IFNG*) locus, and T_H1-17 and T_H17 cells showed increased accessibility at the *IL26* and *IL22* loci, consistent with the functions of these molecules in T cell–mediated inflammation (Fig. 3e)[19,20]. Notably, all of the naive and memory T cell subtypes could be distinguished from one another when downsampled to the fragment density equivalent to that obtained by T-ATAC-seq data (1×10^3 to 1×10^4 nuclear fragments; Fig. 3f), suggesting that variability in T cell phenotypes could be determined with single-cell measurements.

We next performed T-ATAC-seq analysis in primary human peripheral blood CD4$^+$ T cells (Fig. 3a). We sorted naive T cells (as described above), memory T cells that contained all of the T_H phenotypes (CD4$^+$CD45RA$^-$CD25$^-$CD127^{hi}) and memory T_H17 cells (CD4$^+$CD45RA$^-$CD25$^-$CD127^{hi}CXCR5$^-$CCR6$^+$) from two healthy individuals and subjected each population to T-ATAC-seq. Single-cell profiles were filtered using quality controls as described above for immortalized cells. Briefly, single primary T cells displayed high-quality ATAC-seq reads; cells passing the filter yielded an average of 2.4×10^3 fragments that mapped to the nuclear genome, and an average of 73% of fragments were within peaks derived from ensemble primary T cell ATAC-seq profiles (Supplementary Fig. 4a). T-ATAC-seq data showed enrichment of fragments at TSSs and nucleosomal periodicity of fragment lengths similar to those in the ensemble profiles (Supplementary Fig. 4a). Similarly, *TRA* and *TRB* sequencing data remained robust in captured single cells, generating on average 1.1×10^3 reads for *TRA* sequences and 4.3×10^2 reads for *TRB* sequences (Supplementary Fig. 4b,c).

We first analyzed T-ATAC-seq profiles by using a computational pipeline that integrated reference ensemble ATAC-seq data from T cells (this study) and other hematopoietic cell types[10] to phenotype individual cells (Fig. 4a). By using a previously described approach to train principal components (PCs) on ensemble ATAC-seq data and project single-cell profiles onto that PC space[10], single cells were compared against all ensemble profiles to remove contaminating non-T cells that remained after sorting (cells sorted to $>95\%$ purity). Indeed, although the majority of single-cell profiles showed the highest epigenomic correlation with ensemble T cell profiles, as compared to those of other cell types, 11/185 naive T cells, 2/134 memory T cells and 4/148 T_H17 cells showed higher similarity with other immune cell types, particularly with CD4$^+$ monocytes, and were excluded from further analysis (Supplementary Fig. 4d,e). Epigenomic profiles of the remaining T cells (450 cells) were then compared against the profiles of Jurkat cells (231 cells) and the previously published single-cell epigenomic profiles of blood monocytes (92 cells) and lymphoid-primed multipotent progenitor (LMPP) cells (89 cells)[9]. *t*-distributed stochastic neighbor embedding (t-SNE) projection[21] of single-cell epigenome profiles revealed clustering of single cells largely according to cell type, with primary T cells clustering separately from Jurkat cells, monocytes and LMPP cells (Fig. 4b). Of note, T cell profiles generated a continuous spectrum of epigenomic states, rather than distinct subpopulations of naive and memory phenotypes, suggesting substantial regulatory variability within cell-surface-marker-defined subpopulations. In particular, previous studies using high-resolution cell surface marker staining and functional analysis identified significant heterogeneity within the CD45RA$^+$ naive T cell population, including the presence of recent thymic emigrants, 'super-naive' cells, early-memory and differentiated cells, and memory stem-like cells[22–28]. Indeed, single-cell naive T cell chromatin accessibility profiles also showed a spectrum of cell states, including a small population of naive cells present in both individuals (20/174 naive cells, 11.5%) that clustered closely with memory and T_H17 cells (Fig. 4b).

We next measured TF deviation scores and variation in single cells and aggregated them by cell type. In aggregate, all of the T cells exhibited high deviations in TCF–LEF family members, as compared to monocytes, suggesting that these factors probably directed T cell lineage specification through changes in chromatin accessibility (Fig. 4c)[29]. In contrast, monocytes exhibited high activity of CCAAT/enhancer-binding protein (CEBP) family members and of the E26 transformation-specific (ETS) family TF PU.1. A comparison of naive cells and memory cells identified a large shift in epigenomic profile from high activity of TFs involved in T cell specification in naive T cells, including TCF family factors and zinc finger and BTB-domain-containing 7B (ZBTB7B), to TFs involved in T cell activation in memory cells, including the activator protein 1 (AP-1) factors FOS, JUN and basic leucine zipper ATF-like transcription factor (BATF) (Fig. 4c). Finally, comparison of memory T cells and T_H17 cells showed high activities for the STAT, GATA and IRF TFs in memory cells and for the AP-1, MAF, RUNX and RAR-related orphan receptor (ROR) TFs in T_H17 cells, consistent with the critical roles of these TFs in memory and T_H17 cells, respectively[30–38] (Fig. 4c). The cell-type-specific TFs identified in aggregated single-cell profiles were remarkably similar to those in the profiles obtained from ensemble measurements in 500× more cells. Ensemble naive T cell profiles showed similar enrichments of accessibility at TCF–LEF family members and ZBTB7B, whereas memory cells demonstrated high deviations in AP-1 factors (Supplementary Fig. 5a,b). Similarly, T_H17 cells showed high activities for ROR, AP-1 and RUNX factors, as compared to that in all of the other memory T cell types (Supplementary Fig. 5a–c). Finally, an examination of T_H1, T_H2 and T_{reg} cells identified TF signatures that aligned well with previously identified master regulators in each lineage, including TBX21 (also known as T-BET) and eomesodermin (EOMES) in T_H1 cells, GATA3 in T_H2 cells and FOXP3 in T_{reg} cells (Supplementary Fig. 5a–c).

We next integrated information from ensemble profiles and cell surface marker staining to visualize epigenomic variability in these canonical T cell populations. As observed in the t-SNE projections, CD45RA$^+$ naive T cells displayed substantial TF heterogeneity that could be divided into at least three subclusters that spanned the continuum of naive to memory cell differentiation. The majority of naive cells (132/174; 75.9%) were present in the first cluster of 'true-naive' cells, and they demonstrated high TF deviation scores for ensemble naive T cell TFs, including ZBTB7B, and low scores for ensemble memory cell TFs (Fig. 4d). A second cluster of 'early-differentiating' naive cells (22/174; 12.6%) showed lower deviation scores for naive cell TFs and higher scores for memory cell TFs, including the AP-1, IRF and STAT factors, although these were lower than those for true memory cells (Fig. 4d and Supplementary Fig. 6a,b). Finally, a small minority of naive cells existed in a differentiated state (20/174; 11.5%) with high AP-1 and RUNX activity (Fig. 4d and Supplementary Fig. 6a,b). Extensive variability was also observed in sorted memory T cells, with variation in known T_H phenotypes as expected, as well as a small fraction of cells clustering closely with naive T cells, suggesting an early differentiated memory state (Fig. 4d). The observed TF variability in T cell subtypes was greater than expected in background ATAC-seq peaks that were matched for GC bias, peak height and transposition rate, and variability was not driven by single cells with low-quality ATAC-seq data, such as low fragment numbers (Supplementary Fig. 6b,c).

Comparison of all three populations of T cells revealed two categories of factors—factors involved in general memory or naive

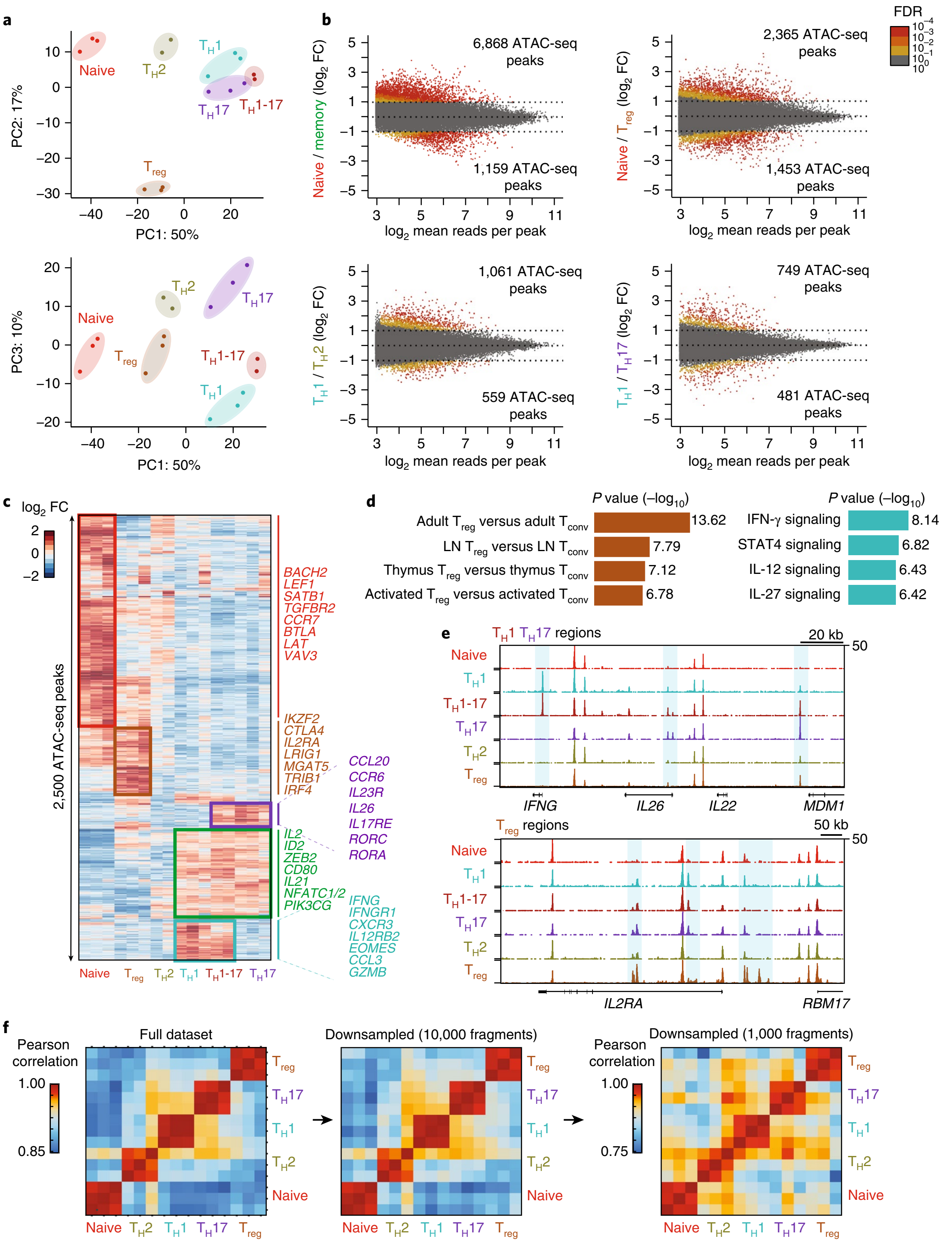

Fig. 3 | Epigenomic landscape of ensemble human CD4$^+$ T cell subtypes. a, PCA of ensemble ATAC-seq profiles from CD4$^+$ T cell subtypes using the top 2,500 variable ATAC-seq peaks, as defined by variance rank of $\log_2$(variance-stabilized read counts) ($n=3$, 3 independent experiments). Percentages indicate percentage of variance explained by each PC. **b**, Differential ATAC-seq peaks for the indicated T cell subtypes. Memory T cell signatures reflect the average accessibility in T_H1, T_H2, T_H17, and T_H1-17 cells. **c**, Heat map showing clusters for the top 2,500 varying ATAC-seq peaks. Colors indicate $\log_2$ fold change (FC) of reads in each peak compared to the mean across all T cell types. **d**, Left, MSigDB immunologic signatures of T_{reg}-specific ATAC-seq peaks as obtained from GREAT analysis. Right, MSigDB pathway signatures of T_H1-specific ATAC-seq peaks as obtained from GREAT analysis (binomial test, $n=3$ replicates per cell type, 3 independent experiments). **e**, Ensemble ATAC-seq data genome tracks for the indicated T cell subtypes. Highlighted regions show cell-type-specific ATAC-seq peaks. **f**, Pearson correlation of PC scores of ensemble ATAC-seq profiles (left) and of ensemble ATAC-seq profiles after downsampling each profile to 10,000 (middle) or 1,000 (right) fragments. Downsampling was performed by randomly selecting 10,000 or 1,000 nuclear fragments in each ensemble ATAC-seq .bam file. Heat maps demonstrate that CD4$^+$ T cell subtype profiles can be distinguished from one another by using the full dataset or profiles with a fraction of the fragments, as expected in single-cell libraries (16 ATAC-seq profiles obtained from three independent experiments).

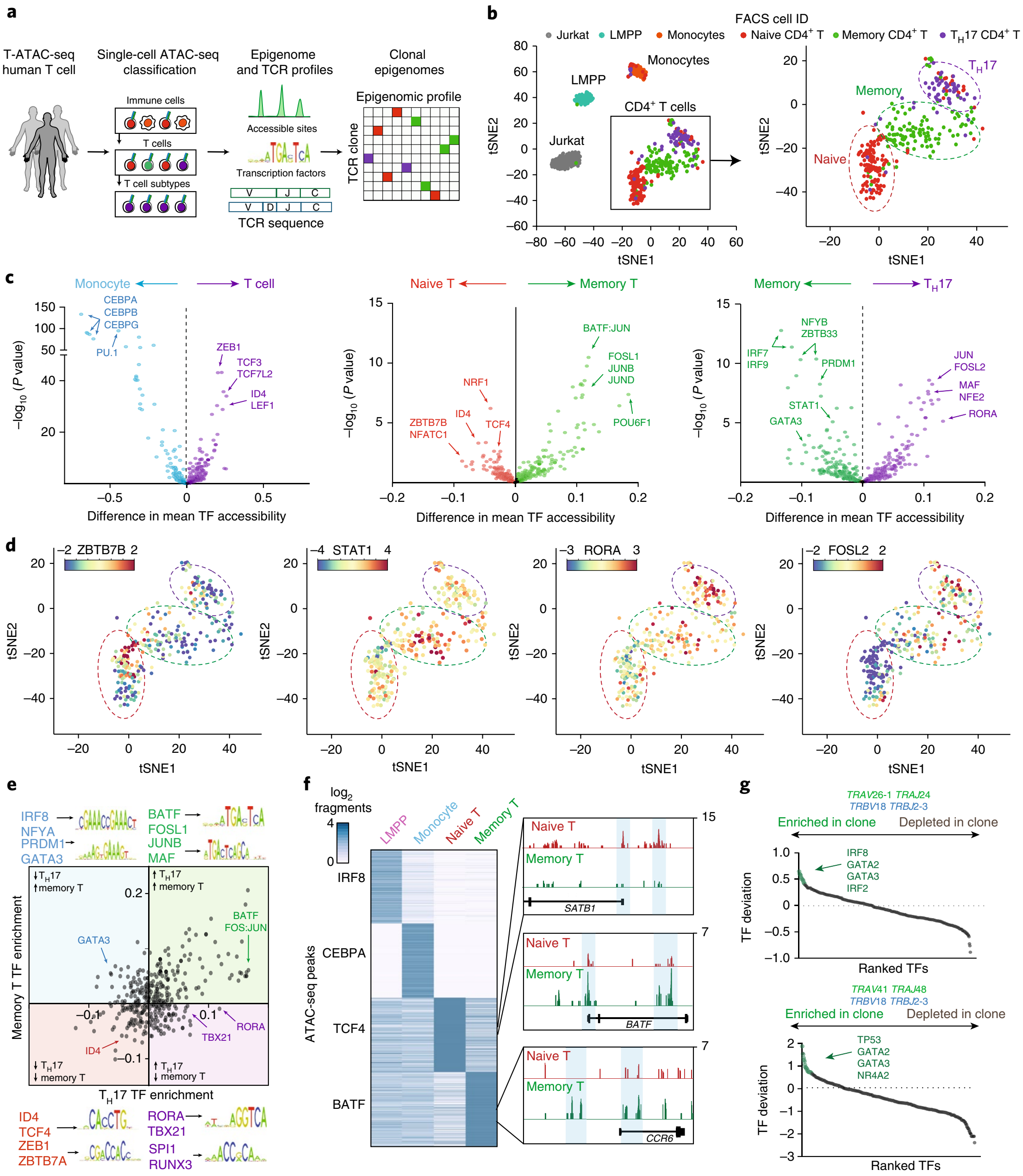

Fig. 4 | Single-cell epigenomic and TCR profiling of human CD4$^+$ T cells. a, Outline for T-ATAC-seq analysis in primary human T cells. Single cells were first sequentially classified to major blood lineages and then to T cell subsets, by similarity to ensemble reference ATAC-seq profiles. T-ATAC-seq data from classified single T cells were then analyzed for accessibility at regulatory DNA elements and TF activity using ATAC-seq data, as well as for TCR (*TRA* and *TRB*) sequence identity. Finally, integrative analysis of both data types was performed to identify epigenomic signatures in T cell clones. **b**, t-SNE projection of naive and memory T cells (T-ATAC-seq; $n=320$ cells, $n=6$ independent experiments), Jurkat T cells (T-ATAC-seq; $n=145$ cells, $n=3$ independent experiments), monocytes (scATAC-seq; $n=71$ cells) and LMPPs (scATAC-seq; $n=86$ cells). See Methods for generation of tSNE plots from high-quality single-cell libraries. **c**, TF bias-corrected deviation enrichments (chromVAR) in aggregated single-cell populations. TF enrichments were calculated as the difference in mean TF motif accessibility between two populations of single cells. Shown are enrichments for all T cells ($n = 450$ cells) as compared to monocytes ($n = 71$ cells; left), for memory T cells ($n = 132$ cells) as compared to naive T cells ($n = 154$ cells classified according to t-SNE clustering; middle) and for T$_H$17 cells ($n = 144$ cells) as compared to memory T cells (right). *P* values were calculated using a two-tailed *t*-test. **d**, t-SNE projection of single T cells colored by ZBTB7B (enriched in naive cells), STAT1 (enriched in memory cells), RORA (enriched in T$_H$17 cells) and FOSL2 (enriched in T$_H$17 cells) motif accessibility TF *z*-scores ($n=320$ cells, $n=6$ independent experiments). Scale bars indicate the range of TF *z*-scores. **e**, Mean bias-corrected deviations ranked for difference in aggregated T$_H$17 cells versus aggregated naive cells (*x* axis) and for aggregated memory (non-T$_H$17) cells versus aggregated naive cells (*y* axis; $n = 450$ cells). Shown are TF motifs for selected factors in each quadrant. For example, BATF motifs show increased accessibility in memory T cells and T$_H$17 cells. In contrast, RORA motifs show increased accessibility in T$_H$17 cells but not in memory T cells. **f**, Left, heat map showing ATAC-seq fragment counts in peaks (rows) containing the indicated motifs from aggregated single cells (sample sizes indicated in **b**). Cell types analyzed are indicated above each column. Right, aggregated single-cell genome tracks for naive T cell–specific and memory T cell–specific peaks in the *SATB1*, *BATF* and *CCR6* loci. **g**, TCR clone and CDR3-encoding sequences for two memory T cell clones (clone 1 (top) and clone 2 (bottom)) and associated TF deviation enrichments in clonal cells versus nonclonal memory T cells. Epigenomic profiles from each clone were aggregated and compared to an aggregate profile from all nonclonal memory T cells.

T cell differentiation and factors specific to T$_H$ cell subtypes (Fig. 4e). Unexpectedly, relatively few TFs were enriched in the latter category, suggesting that large-scale changes occurred during the transition from naive to memory phenotypes, which dominated the epigenomic landscape, whereas subtype-specific changes were comparatively fewer and controlled by specific factors (Fig. 4e). This principle was also supported by an unbiased analysis of TF modules, in which we correlated TF activity across single cells (Supplementary Fig. 6d). We found several TF programs that corresponded to subset-specific functions and that these TFs functioned in concert with a common memory program (Supplementary Fig. 6d). Notably, modules encompassing T$_H$1 and T$_H$2 phenotypes could be observed in this analysis, even though these populations were not specifically enriched by cell sorting, which demonstrated that this information could be derived de novo from single-cell profiles. Finally, differential analysis of ATAC-seq peaks that contained binding sites for cell-state-specific TFs identified cell-type-specific cis-regulatory elements, including SATB homeobox 1 (*SATB1*) locus elements in naive T cells and *BATF* and *CCR6* locus elements in memory T cells.

We next integrated TCR-seq results with single-cell epigenomic profiles from these healthy individuals. We identified two clonal populations within the memory population in one individual with a history of atopy, which could be identified by common expression of *TRBV*18–*TRBJ*2-3, suggesting that they may have expanded to shared antigens (Fig. 4g). Of note, neither clonotype was present in the sampled naive cells from the same individual. Analysis of epigenomic signatures in these cells revealed common high TF deviation scores for GATA factors, consistent with a T$_H$2 phenotype (Fig. 4g). In summary, these data demonstrate that T-ATAC-seq can effectively capture ensemble epigenomic measurements while simultaneously preserving single-cell regulatory and TCR information.

T-ATAC-seq reveals regulatory signatures in T cell leukemia and host immunity. We performed T-ATAC-seq analysis on clinical blood samples from patients with Sézary syndrome, which is a leukemic form of cutaneous T cell lymphoma (CTCL). Identification of cancer-cell regulatory signatures can be challenging, as only a fraction of circulating CD4$^+$ T cells are malignant, and standard immunophenotypic methods to distinguish healthy and cancer clones are imprecise and not applicable to some patients[39,40]. These observations have been the basis for the recent development of TCR clonality assays for the identification of malignant T cell expansion and minimal residual disease in clinical samples from patients with CTCL[41,42]. Therefore, we asked whether the integrated analysis of T-ATAC-seq could improve the identification of cancer-specific epigenomic signatures of malignant cells (Fig. 5a). We first isolated CD4$^+$ T cells from a patient with Sézary syndrome and subjected these cells to T-ATAC-seq (in three independent experiments). Notably, 73% of all CD4$^+$ T cells (157/215 cells) expressed a single TCRβ-encoding sequence *TRBV*7-9–*TRBJ*1-5, which represented the putative leukemic clone (Fig. 5b and Supplementary Fig. 7a). These cells showed TCR-β subunit pairing with the TCR-α subunit encoded by *TRAV*12-1–*TRAJ*26. We next aggregated all of the cells according to the leukemic or nonleukemic clonotype and compared the epigenomic profiles. Leukemic cells showed high TF deviation scores for memory T cell–specific TFs, including BATF, JUN and FOS, and GATA motifs, including the T$_H$2-specific TF GATA3 (Fig. 5c). These findings were consistent with the long-standing hypothesis, based on cytokine and cell surface marker expression, that Sézary cells represent a malignant counterpart of T$_H$2 memory T cells, which may contribute to disease persistence and pathogenesis[43,44]. t-SNE projection of single-cell T-ATAC-seq PCA scores revealed that almost all of the memory T cells in this patient were replaced by leukemic T$_H$2 cells, whereas the nonmalignant T cells were predominantly in a naive state. The nonmalignant T cell clones in the patient with CTCL exhibited strong SMAD family member 3 (SMAD3)-associated chromatin accessibility, which may reflect an immunosuppressive transforming growth factor (TGF)-β pathway (Fig. 5c,e). These findings identified a possible cause for systemic immunodeficiency associated with Sézary syndrome, as nearly all of the memory T cells had been replaced by the leukemic clone (Fig. 5d)[45]. Notably, analysis of individual cis-regulatory changes that contributed to the overall shift in TF landscape identified genes previously shown to be recurrently mutated in CTCL and other cancer types (Fig. 5e)[46,47]. These included genes involved in T cell survival and activation pathways, such as *TNFAIP3*, *PIK3CG* and *PRKCQ*. Analysis of MSigDB signatures pathways enriched in cis-elements that were more accessible in leukemic cells demonstrated that these elements were located near genes that are upregulated in T cell leukemia, as well as in other cancer types (Fig. 5f).

Finally, we asked whether the leukemia-specific signature could be identified using standard immunophenotypic FACS strategies for cancer cells. We sorted CD4$^+$ cells according to their expression of CD26 (also known as dipeptidyl peptidase 4; DPP4), a cell surface protein whose loss of expression is clinically used as a diagnostic tool to identify malignant Sézary cells (Supplementary Fig. 7b)[48]. Unexpectedly, we observed the presence of the CTCL clone in both CD26$^+$ and CD26$^-$ cell populations, demonstrating that, at least in a subset of patients, this marker does not accurately identify circulating malignant cells (Fig. 5g)[40]. Accordingly, aggregating single cells on the basis of their immunophenotype rather than their clonotype obscured cancer-specific epigenomic signatures, as memory and T$_H$2-specific TFs were not enriched in CD26$^-$ cells as compared to that in CD26$^+$ cells (Fig. 5h). T-ATAC-seq analysis of two additional patients with CTCL confirmed the superiority of TCR clonotyping over CD26 immunophenotyping to isolate leukemic clones and their epigenomic signatures (Supplementary Fig. 7c,d). Altogether, the use of T-ATAC-seq in T cell leukemia demonstrates that this method is applicable to clinical blood samples and can be used to separate clonal and nonclonal regulatory pathways in cells from the same individual.

Discussion

The expression of uniquely recombined TCRs on individual T cells is the central driver of immune responsiveness and connects specific antigen recognition to a particular effector function. In addition, because the diversity of possible human TCRs is estimated at ~10^{14}, single-cell TCR-seq can serve as a powerful lineage tracer, either in the context of a normal immune response or in the context of malignant transformation. Therefore, pairing TCR identity to functional phenotype represents an important strategy to investigate T cell clonal dynamics, phenotypic plasticity and tumor heterogeneity[6,12,13]. Here we reported the technical development and application of T-ATAC-seq analysis to immortalized and primary human T cells. We have found it to be robust and reproducible across T cell types and individuals and to compare favorably with previous technologies capable of assaying single-cell epigenomes. T-ATAC-seq pairs epigenomic data, identifying cis and trans determinants of cell identity, with high-fidelity RNA sequences of TCR-encoding loci to provide a platform for multi-omic investigation of T cell diversity.

We used ensemble ATAC-seq data and TF binding sites genome wide as scaffolds to map single-cell chromatin states, and we developed a step-wise approach to use single-cell chromatin accessibility to phenotype immune cells. Each single cell was sequentially classified to major blood lineages, and then to T cell subsets—a scheme that recapitulated the chromatin landscape during physiological development. Previous efforts to characterize single-cell epigenomes highlighted the presence of interpopulation and intrapopulation variability in cell lines and distinct hematopoietic cell types[8,10,14]. We demonstrated that this approach could also be

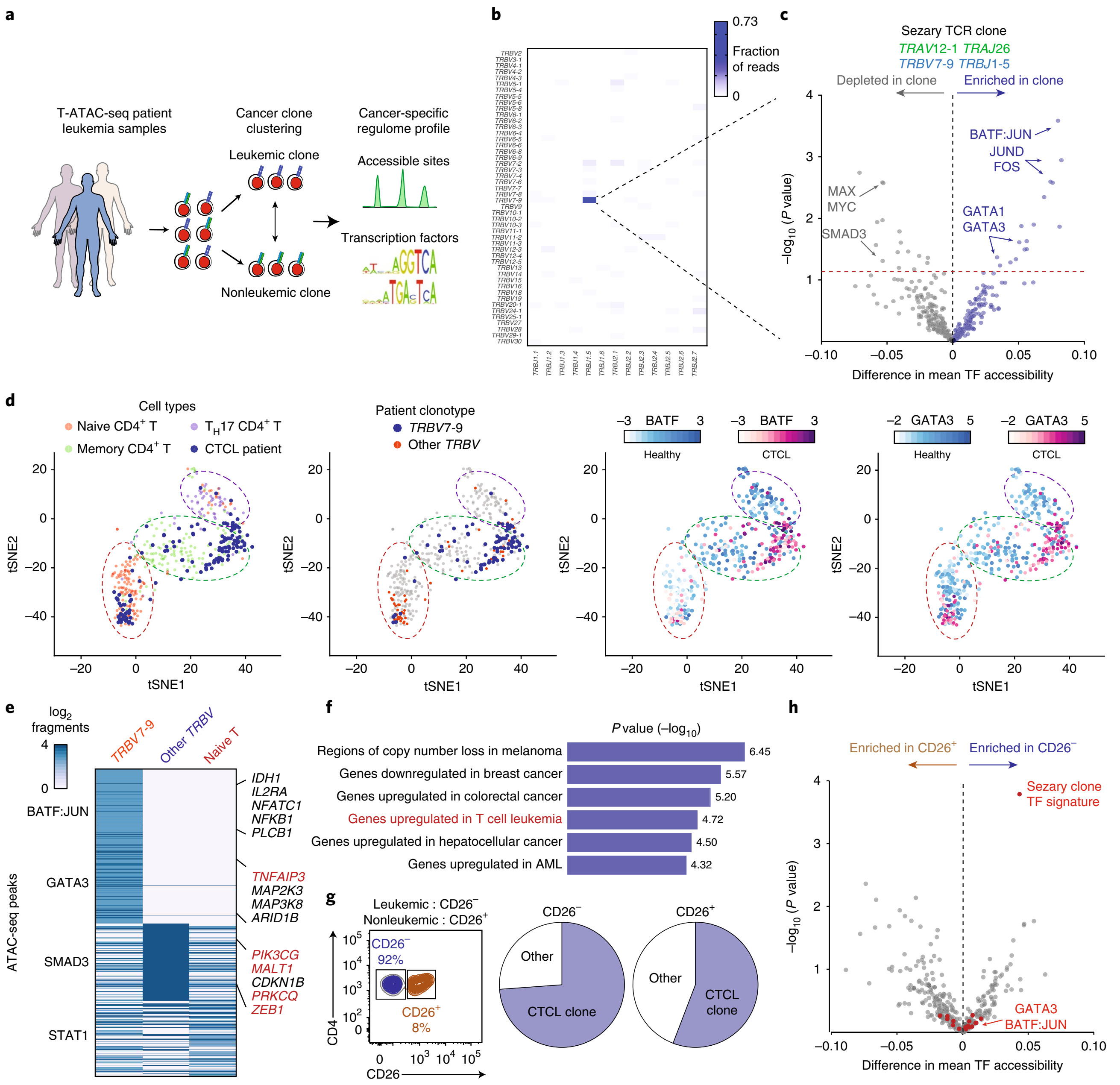

Fig. 5 | T-ATAC-seq identifies epigenomic signatures of clonal leukemic T cells in patient samples. a, Outline for T-ATAC-seq analysis in T cell samples from patients with leukemia. Single cells were first classified according to TCR sequence identity as leukemic cells or nonleukemic cells. ATAC-seq data from classified single T cells were then analyzed for accessibility at regulatory DNA elements and TF activity. **b**, Heat map showing *TRB* rearrangements in peripheral blood samples from a patient with Sézary syndrome ($n = 139$ cells, $n = 3$ independent experiments). **c**, TF bias-corrected deviation enrichments in aggregated clonal T cells as compared to all other T cells. Shown is the TCR (*TRA* and *TRB*) sequence identified in the putative leukemic T cell clone (top). TF enrichments (bottom) were calculated as the difference in mean TF motif accessibility between aggregated leukemic T cell clone profiles and nonclonal T cell profiles in the same patient. Selected TF motifs that were enriched or depleted in the T cell clone are indicated. *P* values were calculated using a two-tailed *t*-test ($n = 139$ cells, $n = 3$ independent experiments). **d**, t-SNE projection of naive and memory T cells from healthy individuals ($n = 320$ cells, $n = 6$ independent experiments) and patient cells ($n = 139$ cells, $n = 3$ independent experiments), as colored by cell ID, clonal versus nonclonal cells, BATF TF score and GATA3 TF score. Scale bars indicate range of TF *z*-scores. **e**, Heat map showing ATAC-seq fragment counts in peaks containing the indicated motifs (left). Labels indicate genes associated with differential peaks, including genes previously shown to be mutated in individuals with CTCL (red). **f**, MSigDB perturbation signatures of *TRB7-9*-specific ATAC-seq peaks, as obtained from GREAT analysis (binomial test; $n = 102$ aggregated single cells, $n = 3$ independent experiments). **g**, Sort strategy for $CD26^+$ and $CD26^-$ $CD4^+$ T cells (left), and clonal TCR profiles in each population (right; $n = 105$ cells, $n = 3$ independent experiments). The lack of CD26 expression has been previously used to distinguish leukemic cells from nonleukemic cells. **h**, TF bias-corrected deviation enrichments in aggregated $CD26^-$ cells ($n = 56$ single cells) relative to $CD26^+$ cells ($n = 49$ single cells). *P* values were calculated using a two-tailed *t*-test. TFs identified above the dashed line in **c** are highlighted in red.

informative to distinguish more subtle phenotypes in primary T cells and reveal heterogeneity in T cell populations that could appear similar by cell surface marker profiling. For example, a small fraction of naive $CD4^+$ T cells, characterized by the expression CD45RA, exhibited chromatin states more similar to those of memory T cells, showing accessibility at genomic sites bound by

AP-1 TFs. This observation is supported by previous functional studies that identified a memory T cell population with stem-like properties in the CD45RA$^+$ naive T cell gate[27]. Similarly, single cells with memory T cell or T$_H$17-defining cell surface markers displayed significant epigenomic heterogeneity, particularly in cell-type-specific TFs such as the IRF, STAT and ROR factors. These results suggest that memory T cells may exist in a phenotypic continuum, rather than in distinct quantal chromatin states[3]. Future studies with more extensive sampling of single T cells in homeostatic and inflammatory conditions could use this approach to define the continuous landscape of single T cell states and variability within cell-surface-marker-defined subtypes.

We exploited the ability of T-ATAC-seq to pair TCRs with chromatin state information to identify cancer-associated epigenomic changes in patients with T cell leukemia. The clinical diagnosis of T cell leukemia is based on several factors, including clinical presentation, histopathologic findings and the identification of a clonal T cell population. However, all of these diagnostic findings, including the expansion of T cell clones, are often present in benign inflammatory skin conditions, and it remains a significant challenge to distinguish small populations of malignant cells from benign, but oligoclonal, T cell proliferations[42,49]. Using T-ATAC-seq, we were able to define epigenomic signatures of clonal cancer cells that were missed by ensemble or standard FACS-based separation methods, demonstrating the promise of this approach. This result has potentially significant clinical applications, as recent studies have described distinct epigenomic classifications of CTCL that are associated with differential responses to currently used clinical therapies that target the epigenome, such as histone deacetylase inhibitors[50,51]. Future studies on larger patient cohorts are needed to establish whether integration of epigenomic information with T cell clonality can (i) improve diagnostic precision as compared to the standard clinical techniques currently in use and (ii) predict or monitor successful clinical responses to therapies that target the epigenome.

More broadly, T-ATAC-seq represents an important technical advance toward achieving an atlas of human cell types and states[52], in that it is able to generate genome-wide chromatin accessibility maps, while simultaneously preserving and measuring RNA sequence. T-ATAC-seq may be particularly well-suited for the examination of TF activity and specific enhancer elements that underlie cell states, as compared to existing methods that pair whole-transcriptome profiles with TCR sequences in single cells. Although we used an unbiased approach and sequenced all of the captured cells, which is applicable to settings of significant clonal expansion such as CTCL, the use of T-ATAC-seq to interrogate rare clonal populations may be technically challenging at the current throughput of 96 cells per microfluidic chip. One strategy to address this challenge may be to selectively sequence single-cell epigenomes after identifying TCRs of interest (or vice versa); however, further technical improvements focused on increasing throughput of T-ATAC-seq will be critical for the analysis of rare T cell clones. Given the inherent challenges in obtaining large amounts of RNA from T cells, as compared to other cell types, we believe that this strategy should be easily adaptable to other cell types for which RNA is more abundant. In particular, T-ATAC-seq could be adapted to determine RNA sequences of other cell-identity-specific transcripts, such as those encoding B cell receptors, olfactory receptors, long noncoding RNAs (lncRNAs) and cytokines, or perhaps with additional technical development, even to measure whole transcriptomes. Finally, the sequential reaction conditions used to assay chromatin and RNA sequences from single cells can be easily scaled-up to obtain both types of information from ensemble samples for which material is limited, such as rare cell types or clinical samples.

We envision that T-ATAC-seq will be complementary to approaches for unbiased identification of TCR ligands, enabling integration of T cell epigenomic state, TCR sequence and TCR ligands[53,54]. The application of this strategy to human diseases such as cancer and autoimmune disease, particularly in the context of immunotherapy, could be invaluable in generating comprehensive profiles of beneficial and harmful T cell responses, the regulatory networks underlying either response and the antigens that drive these networks.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41591-018-0008-8.

Received: 8 June 2017; Accepted: 7 February 2018;

References

1. Davis, M. M. & Bjorkman, P. J. T cell antigen receptor genes and T cell recognition. *Nature* **334**, 395–402 (1988).
2. Shalek, A. K. et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363–369 (2014).
3. Gaublomme, J. T. et al. Single-cell genomics unveils critical regulators of T$_H$17 cell pathogenicity. *Cell* **163**, 1400–1412 (2015).
4. Paul, F. et al. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell* **163**, 1663–1677 (2015).
5. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
6. Han, A., Glanville, J., Hansmann, L. & Davis, M. M. Linking T cell receptor sequence to functional phenotype at the single-cell level. *Nat. Biotechnol.* **32**, 684–692 (2014).
7. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
8. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
9. Corces, M. R. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).
10. Buenrostro, J. D. et al. Single-cell epigenomics maps the continuous regulatory landscape of human hematopoietic differentiation. *bioRxiv* http://dx.doi.org/10.1101/109843 (2017).
11. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
12. Stubbington, M. J. T. et al. T cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* **13**, 329–332 (2016).
13. Afik, S. et al. Targeted reconstruction of T cell receptor sequence from single-cell RNA-seq links CDR3 length to T cell differentiation state. *Nucleic Acids Res.* **45**, e148 (2017).
14. Cusanovich, D. A. et al. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
15. Weber, B. N. et al. A critical role for TCF-1 in T lineage specification and differentiation. *Nature* **476**, 63–68 (2011).
16. Collins, A., Littman, D. R. & Taniuchi, I. RUNX proteins in transcription factor networks that regulate T cell lineage choice. *Nat. Rev. Immunol.* **9**, 106–115 (2009).
17. Morita, R. et al. Human blood CXCR5$^+$CD4$^+$ T cells are counterparts of T follicular cells and contain specific subsets that differentially support antibody secretion. *Immunity* **34**, 108–121 (2011).
18. Fontenot, J. D., Rasmussen, J. P., Gavin, M. A. & Rudensky, A. Y. A function for interleukin 2 in Foxp3-expressing regulatory T cells. *Nat. Immunol.* **6**, 1142–1151 (2005).
19. Ouyang, W., Kolls, J. K. & Zheng, Y. The biological functions of T helper 17 cell effector cytokines in inflammation. *Immunity* **28**, 454–467 (2008).
20. Meller, S. et al. T$_H$17 cells promote microbial killing and innate immune sensing of DNA via interleukin 26. *Nat. Immunol.* **16**, 970–979 (2015).
21. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
22. Kimmig, S. et al. Two subsets of naive T helper cells with distinct T cell receptor excision circle content in human adult peripheral blood. *J. Exp. Med.* **195**, 789–794 (2002).
23. Boursalian, T. E., Golob, J., Soper, D. M., Cooper, C. J. & Fink, P. J. Continued maturation of thymic emigrants in the periphery. *Nat. Immunol.* **5**, 418–425 (2004).
24. Harari, A., Vallelian, F. & Pantaleo, G. Phenotypic heterogeneity of antigen-specific CD4 T cells under different conditions of antigen persistence and antigen load. *Eur. J. Immunol.* **34**, 3525–3533 (2004).

25. Zhao, C. & Davies, J. D. A peripheral CD4$^+$ T cell precursor for naive, memory and regulatory T cells. *J. Exp. Med.* **207**, 2883–2894 (2010).
26. Song, K. et al. Characterization of subsets of CD4$^+$ memory T cells reveals early branched pathways of T cell differentiation in humans. *Proc. Natl Acad. Sci. USA* **102**, 7916–7921 (2005).
27. Gattinoni, L. et al. A human memory T cell subset with stem-cell-like properties. *Nat. Med.* **17**, 1290–1297 (2011).
28. Weiskopf, D. et al. Dengue virus infection elicits highly polarized CX3CR1$^+$ cytotoxic CD4$^+$ T cells associated with protective immunity. *Proc. Natl Acad. Sci. USA* **112**, E4256–E4263 (2015).
29. Yui, M. A. & Rothenberg, E. V. Developmental gene networks: a triathlon on the course to T cell identity. *Nat. Rev. Immunol.* **14**, 529–545 (2014).
30. Zheng, W. & Flavell, R. A. The transcription factor GATA-3 is necessary and sufficient for T$_H$2 cytokine gene expression in CD4 T cells. *Cell* **89**, 587–596 (1997).
31. Lohoff, M. & Mak, T. W. Roles of interferon-regulatory factors in T helper cell differentiation. *Nat. Rev. Immunol.* **5**, 125–135 (2005).
32. Ivanov, I. I. et al. The orphan nuclear receptor ROR-γt directs the differentiation program of proinflammatory IL-17$^+$ T helper cells. *Cell* **126**, 1121–1133 (2006).
33. Yang, X. O. et al. T helper 17 lineage differentiation is programmed by orphan nuclear receptors ROR-α and ROR-γ. *Immunity* **28**, 29–39 (2008).
34. Bauquet, A. T. et al. The co-stimulatory molecule ICOS regulates the expression of c-Maf and IL-21 in the development of follicular T helper cells and T$_H$17 cells. *Nat. Immunol.* **10**, 167–175 (2009).
35. Schraml, B. U. et al. The AP-1 transcription factor Batf controls T$_H$17 differentiation. *Nature* **460**, 405–409 (2009).
36. O'Shea, J. J., Lahesmaa, R., Vahedi, G., Laurence, A. & Kanno, Y. Genomic views of STAT function in CD4$^+$ T helper cell differentiation. *Nat. Rev. Immunol.* **11**, 239–250 (2011).
37. Rutz, S. et al. Transcription factor c-Maf mediates the TGF-β-dependent suppression of IL-22 production in T$_H$17 cells. *Nat. Immunol.* **12**, 1238–1245 (2011).
38. Ciofani, M. et al. A validated regulatory network for T$_H$17 cell specification. *Cell* **151**, 289–303 (2012).
39. Bigler, R. D., Boselli, C. M., Foley, B. & Vonderheid, E. C. Failure of anti–T cell receptor Vβ antibodies to consistently identify a malignant T cell clone in Sézary syndrome. *Am. J. Pathol.* **149**, 1477–1483 (1996).
40. Kelemen, K., Guitart, J., Kuzel, T. M., Goolsby, C. L. & Peterson, L. C. The usefulness of CD26 in flow cytometric analysis of peripheral blood in Sézary syndrome. *Am. J. Clin. Pathol.* **129**, 146–156 (2008).
41. Weng, W. K. et al. Minimal residual disease monitoring with high-throughput sequencing of T cell receptors in cutaneous T cell lymphoma. *Sci. Transl. Med.* **5**, 214ra171 (2013).
42. Sufficool, K. E. et al. T cell clonality assessment by next-generation sequencing improves detection sensitivity in mycosis fungoides. *J. Am. Acad. Dermatol.* **73**, 228–236 (2015).
43. Rook, A. H., Vowels, B. R., Jaworsky, C., Singh, A. & Lessin, S. R. The immunopathogenesis of cutaneous T cell lymphoma. Abnormal cytokine production by Sézary T cells. *Arch. Dermatol.* **129**, 486–489 (1993).
44. Vowels, B. R. et al. T$_H$2 cytokine mRNA expression in skin in cutaneous T cell lymphoma. *J. Invest. Dermatol.* **103**, 669–673 (1994).
45. Krejsgaard, T., Odum, N., Geisler, C., Wasik, M. A. & Woetmann, A. Regulatory T cells and immunodeficiency in mycosis fungoides and Sézary syndrome. *Leukemia* **26**, 424–432 (2012).
46. Ungewickell, A. et al. Genomic analysis of mycosis fungoides and Sézary syndrome identifies recurrent alterations in TNFR2. *Nat. Genet.* **47**, 1056–1060 (2015).
47. Choi, J. et al. Genomic landscape of cutaneous T cell lymphoma. *Nat. Genet.* **47**, 1011–1019 (2015).
48. Bernengo, M. G. et al. Prognostic factors in Sézary syndrome: a multivariate analysis of clinical, hematological and immunological features. *Ann. Oncol.* **9**, 857–863 (1998).
49. Kirsch, I. R. et al. TCR sequencing facilitates diagnosis and identifies mature T cells as the cell of origin in CTCL. *Sci. Transl. Med.* **7**, 308ra158 (2015).
50. Bolden, J. E., Peart, M. J. & Johnstone, R. W. Anticancer activities of histone deacetylase inhibitors. *Nat. Rev. Drug Discov.* **5**, 769–784 (2006).
51. Qu, K. et al. Chromatin accessibility landscape of cutaneous T cell lymphoma and dynamic response to HDAC inhibitors. *Cancer Cell* **32**, 27–41.e4 (2017).
52. Regev, A. et al. The Human Cell Atlas. *Elife* **6**, e27041 (2017).
53. Birnbaum, M. E. et al. Deconstructing the peptide–MHC specificity of T cell recognition. *Cell* **157**, 1073–1087 (2014).
54. Newell, E. W. & Davis, M. M. Beyond model antigens: high-dimensional methods for the analysis of antigen-specific T cells. *Nat. Biotechnol.* **32**, 149–157 (2014).

Acknowledgements

We thank members of the Chang, Davis and Greenleaf laboratories, including Y. Shen and K. Qu, for helpful discussions. We thank X. Ji, D. Wagh and J. Coller at the Stanford Functional Genomics Facility. This work was supported by the Parker Institute for Cancer Immunotherapy (A.T.S., H.Y.C. and M.M.D.), the US National Institutes of Health (NIH) grants P50HG007735 (H.Y.C. and W.J.G.), 5U19AI057229 (M.M.D.), and U19AI057266 (W.J.G), and the Scleroderma Research Foundation (H.Y.C.). A.T.S. was supported by a Parker Bridge Scholar Award from the Parker Institute for Cancer Immunotherapy and a Cancer Research Institute Irvington Fellowship from the Cancer Research Institute. N.S. was supported by the National Multiple Sclerosis Society Postdoctoral Fellowship. J.D.B. acknowledges the Broad Institute Fellows and Harvard Society of Fellows programs for funding. M.R.C. was supported by a grant from the Leukemia and Lymphoma Society Career Development Program. W.J.G. is a Chan Zuckerberg Biohub investigator. M.M.D. is an investigator of the Howard Hughes Medical Institute. Sequencing was performed by the Stanford Functional Genomics Facility (which is supported by NIH grant S10OD018220).

Author contributions

A.T.S., N.S., M.M.D. and H.Y.C. conceived the project; A.T.S, N.S. and J.D.B. performed experiments and analyzed data; B.W. and Y.Q. performed T-ATAC-seq experiments; R.L., J.M.G., M.R.M. and D.G.G. performed ensemble ATAC-seq experiments and analyzed data; W.S.S. performed TCR-seq experiments; Y.W., A.J.R., K.R.P., C.A.L., A.N.S. and M.R.C. analyzed data; M.S.K. and Y.H.K. obtained clinical specimens; H.Y.C, M.M.D, W.J.G. and P.A.K. guided experiments and data analysis; and A.T.S, M.M.D. and H.Y.C. wrote the manuscript with input from all of the authors.

Competing interests

H.Y.C. and W.J.G. are founders of Epinomics and members of its scientific advisory board. H.Y.C. is a founder of Accent Therapeutics and a member of its scientific advisory board. H.Y.C. is a member of the scientific advisory board of Spring Discovery.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41591-018-0008-8.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.M.D. or H.Y.C.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Human subjects. This study was approved by, and performed in compliance with, the ethical regulations of the Stanford University Administrative Panels on Human Subjects in Medical Research. Written informed consent was obtained from all participants.

Cell culture and T cell isolation. Jurkat cells were obtained from the ATCC (clone E6-1) and were cultured in RPMI-1640 medium (Thermo Fisher Scientific) with 10% FBS and penicillin–streptomycin. For single-cell experiments with Jurkat cells, the cells were sorted into a single-cell suspension before capture on the C1 IFC microfluidic chips (Fluidigm). Mouse 58αβ-negative hybridoma cells were retrovirally transduced with a paired TCR-αβ-encoding sequence, and these cells were used in the mouse and human cell mixing experiments[55,56]. CD4$^+$ T cells from healthy volunteers or patients with Sézary syndrome were enriched from peripheral blood using the RosetteSep Human CD4$^+$ T Cell Enrichment Cocktail (StemCell Technology). For single-cell experiments, CD4$^+$ T cells were sorted as naive T cells (CD4$^+$CD25$^-$CD45RA$^+$), memory T cells (CD4$^+$CD25$^-$CD45RA$^-$) or T$_H$17 cells (CD4$^+$CD25$^-$CD45RA$^-$CCR6$^+$CXCR5$^-$). 200,000 cells from two healthy volunteers were sorted into RPMI-1640 medium supplemented with 10% FBS, washed and loaded onto the C1 IFC microfluidic chips, as described below. For ensemble ATAC-seq experiments, CD4$^+$ T cells were sorted as naive T cells (CD4$^+$CD25$^-$CD45RA$^+$), T$_{reg}$ cells (CD4$^+$CD25$^+$IL7R^{lo}), T$_H$1 cells (CD4$^+$CD25$^-$IL7R^{hi}CD45RA$^-$CXCR3$^+$CCR6$^-$), T$_H$2 cells (CD4$^+$CD25$^-$IL7R^{hi}CD45RA$^-$CXCR3$^-$CCR6$^-$), T$_H$17 cells (CD4$^+$CD25$^-$IL7R^{hi}CD45RA$^-$CXCR3$^-$CCR6$^+$) and T$_H$1-17 cells (CD4$^+$CD25$^-$IL7R^{hi}CD45RA$^-$CXCR3$^+$CCR6$^+$) (Supplementary Fig. 5). 55,000 cells from two healthy volunteers (three replicates total) were sorted into RPMI-1640 medium supplemented with 10% FBS, washed with PBS and immediately transposed as described below. Post-sort purities of > 95% were confirmed by flow cytometry for all of the samples.

Antibodies. The following antibodies were used in this study: PERCP–Cy5.5-conjugated anti-human-CD45RA (clone HI100, lot no. B213966, cat. no. 304107, BioLegend), anti-human-CD127 conjugated to Brilliant Violet 510 (clone A019D5, lot no. B197159, cat. no. 351331, BioLegend), allophycocyanin (APC)–Cy7-conjugated anti-human-CD4 (clone OKT4, lot no. B207751, cat. no. 317417, BioLegend), phycoerythrin (PE)-conjugated anti-human-CCR6 (clone G034E3, lot no. B203239, cat. no. 353409, BioLegend), fluorescein isothiocyanate (FITC)-conjugated anti-human-CD25 (clone BC96, lot no. B168869, cat. no. 302603, BioLegend), anti-human-CXCR3 conjugated to Brilliant Violet 421 (clone G025H7, lot no. B206003, cat. no. 353715, BioLegend), Alexa-Fluor-647-conjugated anti-human-CXCR5 (clone RF8B2, lot no. 5302868, cat. no. 558113, BD Pharmingen), PE-conjugated anti-human-CD26 (clone 2A6, lot no. 4301881, cat. no. 12-0269-42, Thermo Fisher) and anti-human-CD3E conjugated to Pacific Blue (clone UCHT1, lot no. 4341657, cat. no. 558117, BD Biosciences). All of the antibodies were validated by the manufacturer in human peripheral blood samples, used at a 1:200 dilution and compared to isotype and no-staining control samples.

Ensemble ATAC-seq. *Cell isolation and transposase reaction.* Cells were isolated and subjected to ATAC-seq as previously described[7]. Briefly, 55,000 cells were pelleted after sorting and washed once with 100 µl PBS. Cell pellets were then resuspended in 50 µl lysis buffer (10 mM Tris-HCl, pH 7.4, 3 mM MgCl$_2$, 10 mM NaCl, 0.1% NP-40 (Igepal CA-630)) and immediately centrifuged at 500 *g* for 10 min at 4 °C. The nuclei-containing pellets were resuspended in 50 µl transposition buffer (25 µl 2 × TD buffer, 22.5 µl dH$_2$0, 2.5 µl Illumina Tn5 transposase) and incubated at 37 °C for 30 min. Transposed DNA was purified with MinElute PCR Purification Kit (Qiagen) and eluted in 10 µl EB buffer.

Primary data processing and peak calling. ATAC-seq libraries were prepared as previously described, barcoded and sequenced on an Illumina Nextseq instrument at the Stanford Functional Genomics Facility[7]. Adaptor sequence trimming, mapping to the human hg19 reference genome using Bowtie2 and PCR duplicate removal using Picard Tools were performed. All samples were merged for peak calling using MACS2. The number of aligned reads, Tn5 offset corrected, mapped to the union peak set for each sample was quantified using intersectBed in BedTools. Peak counts were normalized using the 'CQN' package in R. Peak intensity was defined as the variance-stabilized log$_2$ counts using the 'DESeq2' package in R. After these steps, an $N \times M$ data matrix was obtained, where N indicates the number of merged peaks, M indicates the number of samples, and the value $D_{i,j}$ indicates the number of reads that fall within peak i ($i = 1$ to N) in sample j ($j = 1$ to M). Pearson correlation was calculated based on the log$_2$-normalized counts of all the peaks. Unsupervised correlation of the Pearson correlation matrix was performed using Cluster 3.0 and visualized in Java Treeview.

Transcript-indexed single-cell ATAC-seq (T-ATAC-seq). *Step 1. Cell isolation and loading onto the IFC.* We adapted the C1 Single-Cell Auto Prep System with its Open App program (Fluidigm) to perform T-ATAC-seq. Single T cells were captured using the C1 IFC microfluidic chips (small; 5–10 µm), and custom-built T-ATAC-seq scripts were generated using the C1 Script Builder Software (scripts available from Fluidigm and upon request). Jurkat cells or peripheral blood T cells were first isolated by FACS-based sorting and then washed three times in C1 DNA Seq Cell Wash Buffer (Fluidigm). Cells were resuspended in DNA Seq Cell Wash Buffer at a concentration of 300 cells/µl and mixed with C1 Cell Suspension Reagent at a ratio of 3:2. 15 µl of this cell mix was loaded onto the IFC. After cell loading, the captured cells were visualized by imaging on a Leica CTR 6000 microscope.

Step 2. Microfluidic reactions on the IFC: reagents and conditions. On the C1, cells were subjected sequentially to lysis and transposition, transposase release, MgCl$_2$ quenching, reverse transcription and PCR, as described (Fig. 1a and Supplementary Fig. 1a), using the custom T-ATAC-seq script 'T-ATAC-seq: Sample Prep (1,861 ×, 1,862 ×, 1,863 ×)'. For lysis and transposition (in chamber 1), 30 µl of Tn5 transposition mix was prepared (22.5 µl 2 × TD buffer, 2.25 µl transposase (Nextera DNA Sample Prep Kit, Illumina), 2.25 µl C1 Loading Reagent without salt (Fluidigm), 0.45 µl 10% NP40, 0.30 µl water, and 2.25 µl Superase In RNase inhibitor (20 U/µl; Thermo Fisher Scientific)). For transposase release (in chamber 2), 20 µl of Tn5 release buffer mix was prepared (2 µl 500 mM EDTA, 1 µl C1 Loading Reagent without salt and 17 µl 10 mM Tris-HCl buffer, pH 8). For MgCl$_2$ quenching (in chamber 3), 20 µl of MgCl$_2$ quenching buffer mix was prepared (18 µl 50 mM MgCl$_2$, 1 µl C1 Loading Reagent without salt and 1 µl 10 mM Tris-HCl buffer, pH 8). For reverse transcription (in chamber 4), 30 µl of RT mix was prepared (15.55 µl water, 3.7 µl 10 × Sensiscript RT buffer (Qiagen), 3.7 µl 5 mM dNTPs, 1.5 µl C1 Loading Reagent without salt (Fluidigm), 1.85 µl Sensiscript (Qiagen) and 3.7 µl 6 µM TCR primer mix (described below)). Finally, for PCR of *TRA*, and *TRB* and ATAC fragments (in chamber 5), 30 µl of PCR mix was prepared (8.62 µl water, 13.4 µl 5 × Q5 polymerase buffer (NEB), 1.2 µl 5 mM dNTPs, 1.5 µl C1 Loading Reagent without salt, 0.67 µl Q5 polymerase (2 U/µl; NEB), 0.8 µl 25 µM non-indexed custom Nextera ATAC-seq PCR primer 1, 0.8 µl 25 µM non-indexed custom Nextera ATAC-seq primer 2 and 3 µl 6 µM TCR primer mix). The primer sequences for the non-indexed custom Nextera ATAC-seq primers are listed in Supplementary Table 1 of a prior study[8].

7 µl lysis and transposition mix, 7 µl transposase release buffer, 7 µl MgCl$_2$ quenching buffer, 24 µl RT mix and 24 µl PCR mix were added to the IFC inlets. On the IFC, Tn5 lysis and transposition reaction was carried out for 30 min at 37 °C. Next, transposase release was carried out for 30 min at 50 °C. MgCl$_2$ quenching buffer was immediately added, and chamber contents were immediately incubated with RT mix for 30 min at 50 °C. Finally, gap-filling and eight cycles of PCR were performed using the following conditions: 72 °C for 5 min and then thermocycling at 94 °C for 30 s, 62 °C for 60 s and 72 °C for 60 s. The amplified transposed DNA was harvested in a total of 13.5 µl C1 Harvest Reagent. Following completion of the on-chip protocol (~4–5 h), chamber contents were transferred to 96-well PCR plates, mixed and divided for further amplification of ATAC-seq fragments (5 µl) or TCR-seq fragments (6–7 µl).

Step 3. Amplification of TCR-seq libraries. The *TRA* and *TRB* sequences (collectively referred to as TCR sequences) from single cells were obtained by a series of three PCR reactions (phases) as previously described, with slight modifications for implementation on the IFC[6,57]. The design principles and validation of all TCR primers were described previously[6], and primer sequences are listed in Supplementary Table 1 in that study[6]. To integrate TCR sequence amplification into the T-ATAC-seq protocol, the RT and first-phase PCR was performed in chambers 4 and 5 of the IFC using the conditions described above. The phase 1 TCR primer mix included multiple Vα- and Vβ-region-specific primers and Cα- and Cβ-region-specific; each V-region-specific primer was at a concentration of 0.06 µM, and each C-region-specific primer was at a concentration of 0.3 µM. RT was performed using the Cα- and Cβ-region-specific primers, and the cDNA was then subjected to eight cycles of PCR using both Vα- and Vβ-region-specific primers and Cα- and Cβ-region-specific primers (simultaneously, as ATAC fragments were also being amplified in the same chamber using distinct primers, as described above).

For off-chip phase 1 PCR, following completion of the on-chip protocol, 6–7 µl of the harvested libraries were further amplified using TCR primers. First, an additional eight cycles of PCR was performed using the following cycling conditions: 95 °C for 15 min and thermocycling at 94 °C for 30 s, 62 °C for 1 min and 72 °C for 1 min; 72 °C 10 min; and a hold at 4 °C.

For off-chip phase 2 PCR, a 1-µl aliquot of this final phase 1 product was used as a template for a 12-µl phase 2 PCR reaction. The following cycling conditions were used for a 25-cycle phase 2 PCR: 95 °C for 15 min and thermocycling at 94 °C for 30 s, 64 °C for 1 min and 72 °C for 1 min; 72 °C for 5 min; and a hold at 4 °C. For the phase 2 reaction, multiple internally nested Vα-, Vβ-, Cα- and Cβ-specific primers were used (V primers 0.6 µM, C primers 0.3 µM). The phase 2 primers targeting the V-region contained a common 23-b sequence at the 5′ end to enable further amplification (during the phase 3 reaction) with a common 23-b primer.

For off-chip phase 3 PCR, 1 µl of the final phase 2 PCR product was used as a template for a 14-µl phase 3 PCR reaction, which incorporated barcodes and enabled sequencing on the Illumina MiSeq platform. For the phase 3 PCR reaction, amplification was performed using a 5′ barcoding primer (0.05 µM) containing the common 23-b sequence and a 3′ barcoding primer (0.05 µM) containing sequence of a third internally nested Cα-specific and/or Cβ-specific primer, and Illumina paired-end primers (0.5 µM each). The following cycling conditions were used for a 25-cycle phase 3 PCR: 95 °C for 15 min and thermocycling at 94 °C for 30 s, 66 °C for 30 s and 72 °C for 1 min; 72 °C for 5 min; and a hold at 4 °C. The final phase 3 barcoding PCR reactions for the *TRA* and *TRB* sequences were done separately. For the phase 3 reaction, 0.5 µM of the 3′ Cα-specific barcoding primer and the 3′ Cβ-specific barcoding primer were used. In addition to the common 23-b sequence at the 3′ end (which enabled amplification of products from the second reaction) and a common 23-b sequence at the 5′ end (which enabled amplification with Illumina paired-end primers), each 5′ barcoding primer contained a unique 5-b barcode that specified the plate and a unique 5-b barcode that specified the row within the plate. In addition to the internally nested TCR C-region-specific sequence and a common 23-b sequence at the 3′ end (which enabled amplification with Illumina paired-end primers), each 3′ barcoding primer contained a unique 5-nt barcode that specified the column within the plate.

For library purification and sequencing, after the phase 3 PCR reaction, each PCR product should have had a unique set of barcodes incorporated that specified the plate, row and column, as well as Illumina paired-end sequences that enabled sequencing on the Illumina MiSeq platform. The PCR products were combined at equal proportions by volume and run on a 1.2% agarose gel; a band ~350–380 bp in size was excised and gel-purified using a Qiaquick gel extraction kit (Qiagen). This purified product was then sequenced.

Step 4. Amplification of ATAC-seq libraries. 5 µl of harvested libraries were amplified in a 50-µl PCR reaction for an additional 17 cycles with 1.25 µM Nextera dual-index PCR primers[8] in 1 × NEBnext High-Fidelity PCR Master Mix, using the following PCR conditions: 72 °C for 5 min; 98 °C for 30 s; and thermocycling at 98 °C for 10 s, 72 °C for 30 s and 72 °C for 1 min. The PCR products were pooled and purified on a single MinElute PCR purification column (Qiagen). Libraries were quantified using qPCR prior to sequencing.

Data processing of single-cell TCR-seq libraries. TCR sequencing data were analyzed as previously described[6,57]. Briefly, raw sequencing data were demultiplexed using a custom computational pipeline, and primer dimers were removed. All paired-end reads were assembled by finding a consensus of at least 100 b in the middle of each read. A consensus sequence was obtained for each TCR gene. Because multiple TCR genes might have been present in a given well, we established sequence-identity cutoffs according to sequence-identity distributions in each experiment (generally > 80% sequence identity within a given well). The sequence-identity cutoff ensured that all of the sequences derived from the same transcript would be properly assigned, even given a PCR error rate of 1/9,000 bases and a sequencing error rate of up to 0.4%. TCR V, D and J segments were assigned by VDJFasta. For downstream analysis, an additional read cut-off of 100 reads was used for each identified TCR sequence. For confirmation of identified *TRB* sequences, select patient samples were also sequenced by immunoSEQ (Adaptive Biotechnologies), according to the Survey protocol.

Data processing of single-cell ATAC-seq libraries. All single-cell ATAC-seq libraries were sequenced using paired-end, dual-index sequencing. ATAC-seq data were pre-processed as previously described[8]. Briefly, adaptor sequences were trimmed, sequences were mapped to the hg19 reference genome (or mm9 for mixing experiments) using Bowtie2 and the parameter –X2000, and PCR duplicates were removed. Reads that mapped to mitochondria and unmapped contigs were also removed and not considered in further analyses. Filtered single-cell libraries were required to contain > 15% of unique fragments in called peaks from ensemble profiles (described below) and a library size of > 500 unique nuclear fragments for most of the downstream analysis. For t-SNE projections, a further filtering step was performed to include only high-quality libraries that contained > 40% of unique fragments in called peaks and a library size of > 500 unique nuclear fragments. For example, conclusions regarding primary T cell subsets were derived from 450 single T cells that passed the 15% fragments in the peaks cut-off. t-SNE projections showed 320 high-quality cells that passed the 40% fragments in peaks cut-off (using the 455,057 peaks described below) to ensure that all conclusions based on clustering results were also true for high-quality single-cell libraries.

We validated that the ATAC-seq libraries did not contain contaminating fragments from TCR libraries in the T-ATAC-seq protocol. First, the phase 1 TCR primer mix used on the IFC (described above) was designed to exclude ATAC-seq Nextera-primer-binding sites. Therefore, TCR-encoding fragments present in the ATAC-seq library would not amplify in library preparation steps or be sequenced. Second, we did not observe TCR library fragments in filtered and aligned ATAC-seq reads. Third, ATAC-seq data derived from T-ATAC-seq in Jurkat cells displayed similar accessibility and TF motif measurements as ATAC-seq data derived from scATAC-seq in Jurkat cells.

Principal component analysis (PCA) and t-SNE clustering. We performed PCA projections of ensemble ATAC-seq and single-cell T-ATAC-seq profiles as previously described[10,11]. For ensemble ATAC-seq T cell profiles, after removing unmapped contigs, 97,395 peaks were used for further downstream analysis, and PCA analysis was performed on the 2,500 peaks that exhibited the highest variance across T cell subtypes ($\log_2$ variance-stabilized). For single-cell T-ATAC-seq analysis of primary T cells, we called peaks on a reference set of ensemble ATAC-seq profiles encompassing a wide array of hematopoietic cell types that included previously published hematopoietic progenitors and end-stage cell types[9,10], as well as CD4$^+$ T subtypes generated in this study (Supplementary Figs. 4 and 5d). After removing peaks that aligned to annotated promoters, chromosome X, chromosome Y and unmapped contigs, 455,057 peaks were used for the PCA projection analysis. To normalize ensemble ATAC-seq profiles, we identified 18,858 low-variance promoters across all ensemble samples and normalized each sample by the mean fragment counts within the low-variance promoters. We subsequently performed PCA on the normalized values aggregated by similar ensemble cell types, resulting in 24 PCs. To score single cells for each component, we used the weighted coefficients for each peak and PC (determined using PCA-SVD of the ensemble data above) and calculated the product of the weighted PC coefficients by the centered count values for each cell; taking the sum of this value resulted in a matrix of cells by PCs. We then normalized each cell across the PC-scored values using the sum-of-squares. The matrix of cells by PCs, normalized by the sum-of-squares, was used as an input to a MATLAB implementation of t-SNE (https://lvdmaaten.github.io/tsne/). Data were visualized with scHemeR[10].

Transcription factor deviation and variability scores using ChromVAR. Single-cell ATAC-seq data processing and calculation of TF deviation were performed using chromVAR[11]. Human TF motifs were obtained from the JASPAR database[58] and included many T cell–specific motifs derived from high-throughput 'systematic evolution of ligands by exponential enrichment' (SELEX) and chromatin immunoprecipitation with sequencing (ChIP-seq) experiments[59]. All analysis was repeated using a curated list of human TF motifs from the cisBP database, without substantial differences[11,60]. JASPAR motif results are presented in all of the figures, except for Supplementary Fig. 5. Briefly, for each TF, 'raw accessibility deviations' were computed by subtracting the expected number of ATAC-seq fragments in peaks for a given motif (from the population average) from the observed number of ATAC-seq fragments in peaks for each single cell. For accessibility deviation calculations in primary T cells, we used either 455,057 hematopoietic peaks (as defined above) or a subset of 87,360 peaks called from ensemble T cell subsets, monocyte and LMPP cell data, with similar results. For accessibility deviation calculations in Jurkat cells and other cell lines, we used 114,654 peaks called from ensemble DHS-seq profiles from Jurkat, K562, GM12878, and H1 ESC (ENCODE). Next, the accessibility deviation value for each cell was subtracted by the mean deviation calculated for sets of ATAC-seq peaks with similar accessibility and GC content (background peak set) to obtain a bias-corrected deviation value, and additionally divided by the s.d. of the deviation calculated for the background peak sets to obtain a z-score. For TF differences between single cells or aggregate single-cell populations, either bias-corrected deviations or z-scores were used to identify cell-specific motifs, as indicated in the figure legends. Volcano plots were generated by calculating the mean difference in the bias-corrected TF deviation score between two aggregate single-cell populations. Significance was tested by using a two-tailed Student's t-test. The variability of a TF motif across single cells was determined by computing the s.d. of the z-scores across the cells[8,11]. The expected value of this metric was 1 if the motif was no more variable than the background peak sets for that motif.

Modification of T-ATAC-seq for additional RNA targets. For method development and RT primer troubleshooting, the T-ATAC-seq protocol could be performed on 1,000 cells in microcentrifuge tubes, with each reaction performed in 1,000 × volume. Following lysis, transposition and transposase release, RNA could be reverse-transcribed and subjected to PCR amplification to check RNA quality and quantity for a chosen primer set.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary.

Code availability. All custom code used in this work is available upon request.

Data availability. All ensemble and single-cell sequencing data are available through the Gene Expression Omnibus (GEO) under accession GSE107817. Two replicates of the ensemble ATAC-seq data for naive, T_H17 and T_{reg} cells were previously published and are available under GEO accession GSE101498[61]. In addition, we have generated an open-access interactive web browser, which enables single-cell TCR-seq and ATAC-seq TF deviation exploration (Supplementary Fig. 8; http://tcr.buenrostrolab.com). This browser includes all single-cell data presented in the study, links to ensemble ATAC-seq profile browsers, and processed T-ATAC-seq data matrices.

A WashU browser session with ensemble T cell subtype ATAC-seq data is available at http://epigenomegateway.wustl.edu/browser/?genome=hg19&session=N7ew2XJpWK&statusId=293545209/. A WashU browser session with ensemble and aggregated single-cell Jurkat ATAC-seq data is available at http://epigenomegateway.wustl.edu/browser/?genome=hg19&session=j5Qd1YqeVz&statusId=1775458173.

References

55. Letourneur, F. & Malissen, B. Derivation of a T cell hybridoma variant deprived of functional T cell receptor α- and β-chain transcripts reveals a nonfunctional α-mRNA of BW5147 origin. *Eur. J. Immunol.* **19**, 2269–2274 (1989).
56. Huse, M. et al. Spatial and temporal dynamics of T cell receptor signaling with a photoactivatable agonist. *Immunity* **27**, 76–88 (2007).
57. Glanville, J. et al. Identifying specificity groups in the T cell receptor repertoire. *Nature* **547**, 94–98 (2017).
58. Mathelier, A. et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **44.D1**, D110–D115 (2016).
59. Jolma, A. et al. DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339 (2013).
60. Weirauch, M. T. et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
61. Mumbach, M. R. et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* **49**, 1602–1612 (2017).

Corresponding author(s): Mark M. Davis, Howard Y. Chang

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.	Sample sizes were chosen to provide sufficient confidence to validate methodological conclusions of the applicability of T-ATAC-seq. In general, sample sizes for primary cells were calculated in order to capture at least 20-50 single cells per aggregate cell state. The rationale for obtaining 20-50 cells per state is based on the ability of aggregated single-cell ATAC-seq data to accurately replicate ensemble profiles. This is described in detail in Supplementary Figure 2d-e. Single cells needed to obtain this number were estimated from flow cytometry of healthy T cell samples or clinical phenotyping of Sezary samples. Sample sizes for cell line data were calculated based on generating matched single-cell data for comparison to previously published cell line scATAC-seq data.

2. Data exclusions

Describe any data exclusions.	No inclusion or exclusion criteria were used for human studies. No data were excluded from the manuscript.

3. Replication

Describe whether the experimental findings were reliably reproduced.	All results presented in manuscript were reliably reproduced.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.	No randomization of human participants was used. The experiments were designed to demonstrate the applicability of T-ATAC-seq to human T cell samples and not to determine treatment or clinical outcome.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.	No blinding was used. The experiments were designed to demonstrate the applicability of T-ATAC-seq to human T cell samples and not the effect of treatments or perturbations on phenotypes.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	GraphPad Prism (Version 7) was used for statistical tests on ensemble population data. For single-cell ATAC-seq analysis, we used chromVAR, which is described in the methods, the original publication, and available on GitHub. Further downstream analysis of TF matrices obtained from chromVAR was performed using MATLAB (Version 8), Cluster 3.0, Java Treeview (Version 3.0), and RStudio (Version 1.0.136). Visualization of scATAC-seq data was performed with scHemeR (described in the original publication) and viewed on tcr.buenrostrolab.com.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	There are no restrictions on data availability.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	We used anti-human CD45RA-PERCPCy5.5 (Clone HI100, Lot# B213966, Cat# 304107, Biolegend), anti-human CD127-Brilliant Violet 510 (Clone A019D5, Lot# B197159, Cat# 351331, Biolegend), anti-human CD4-APC-Cy7 (Clone OKT4, Lot# B207751, Cat# 317417, Biolegend), anti-human CCR6-PE (Clone G034E3, Lot# B203239, Cat# 353409, Biolegend), anti-human CD25-FITC (Clone BC96, Lot# B168869, Cat# 302603, Biolegend), anti-human CXCR3-Brilliant Violet 421 (Clone G025H7, Lot# B206003, Cat# 353715, Biolegend), anti-human CXCR5-AlexaFluor647 (Clone RF8B2, Lot# 5302868, Cat# 558113, BD Pharmingen), anti-human CD26-PE (Clone 2A6, Lot# 4301881, Cat# 12-0269-42, Thermo Fisher), and anti-human CD3E-Pacific Blue (Clone UCHT1, Lot# 4341657, Cat# 558117, BD Biosciences). All antibodies were validated by the manufacturer in human peripheral blood samples, used at a 1:200 dilution, and compared to isotype and no staining control samples.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	Jurkat cells were obtained from ATCC (Clone E6-1).
b. Describe the method of cell line authentication used.	Jurkat cells were used immediately used for experiments after acquisition from the commercial source. Jurkat cells were further authenticated by FACS for CD3 and CD4 prior to use in experiments.
c. Report whether the cell lines were tested for mycoplasma contamination.	All cell lines tested negative for mycoplasma contamination prior to use in experiments.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	None of the cell lines used in this study are listed in this database.

▸ Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	N/A

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	This study was approved by the Stanford University Administrative Panels on Human Subjects in Medical Research, and written informed consent was obtained from all participants. Healthy human subjects were male, ages 30-50. Leukemic patients were female, ages 40-70.

Corresponding author(s): Mark M. Davis, Howard Y. Chang

☐ Initial submission ☐ Revised version ☒ Final submission

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.	CD4+ T cells from healthy volunteers or Sezary syndrome patients were enriched from peripheral blood using the RosetteSep Human CD4+ T Cell Enrichment Cocktail (StemCell Technology). Jurkat cells were obtained from ATCC (Clone E6-1) and cultured in RPMI- 1640 Medium with 10% FBS and Penicillin/Streptomycin.
6. Identify the instrument used for data collection.	BD FACSAria II
7. Describe the software used to collect and analyze the flow cytometry data.	Flowjo v10
8. Describe the abundance of the relevant cell populations within post-sort fractions.	>95%. Examples of post-sort purities for each population are provided in the supplementary information.
9. Describe the gating strategy used.	For single-cell experiments, CD4+ T helper cells were sorted as naive T cells (CD4+CD25-CD45RA+), memory T cells (CD4+CD25-CD45RA-), or TH17 cells (CD4+CD25-CD45RA-CCR6+CXCR5-). For ensemble ATAC-seq experiments, CD4+ T helper cells were sorted as naive T cells (CD4+CD25-CD45RA+), Treg (CD4+CD25+IL7Rlo), TH1 (CD4 +CD25-,IL7Rhi,CD45RA-,CXCR3+,CCR6-), TH2 (CD4 +CD25-,IL7Rhi,CD45RA-,CXCR3-,CCR6-), TH17 (CD4 +CD25-,IL7Rhi,CD45RA-,CXCR3-,CCR6+), and TH1-17 (CD4 +CD25-,IL7Rhi,CD45RA-,CXCR3+,CCR6+). For single-cell Sezary cell experiments, cells were sorted as CD4+CD26+ or CD4+CD26- populations. Example gating strategies for each population are provided in Supplementary Figures 3 and 7.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒