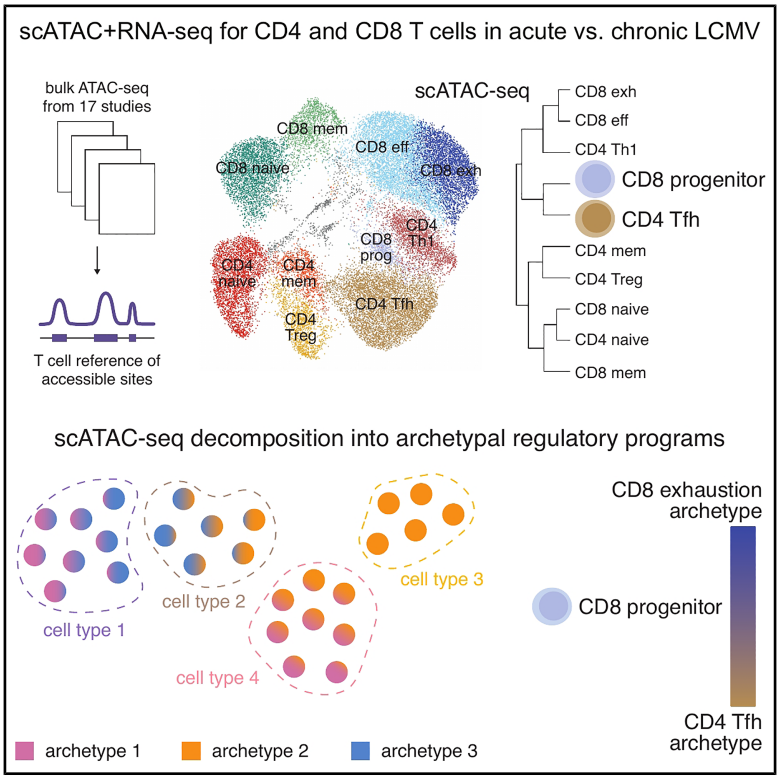
Cell Reports

Resource

# Single-cell multi-omics reveals archetypal regulatory programs shared across T cell subsets

## Graphical abstract




## Highlights

- Single-cell multiomics and ATAC-seq atlas reveal T cell regulatory archetypes
- A unified progenitor CD8 T cell state is shared between acute and chronic infection
- CD8 progenitor identity combines CD8 exhaustion and CD4 Tfh regulatory programs
- A Tfh-like differentiation gradient underlies CD8 progenitor cell heterogeneity


## Authors

Sarah K. Walker, Joris van der Veeken, Alexander Y. Rudensky, Yuri Pritykin

## Correspondence

pritykin@princeton.edu



## In brief

Walker et al. integrate single-cell multi-omics with a chromatin accessibility atlas to decode the modular regulatory architecture of T cell responses to viral infection. Using archetypal analysis, they show that progenitor CD8 T cells have a composite identity, combining regulatory programs of CD8 exhaustion and CD4 follicular helper T cells.


Walker et al., 2026, Cell Reports *45*, 118156
November 24, 2026 © 2026 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.celrep.2026.118156

CellPress

Cell Reports

CellPress
OPEN ACCESS

Resource

# Single-cell multi-omics reveals archetypal regulatory programs shared across T cell subsets

Sarah K. Walker,[1,2,6] Joris van der Veeken,[3] Alexander Y. Rudensky,[4] and Yuri Pritykin[1,5,7,*]
[1]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA
[2]Quantitative and Computational Biology Program, Princeton University, Princeton, NJ, USA
[3]Research Institute of Molecular Pathology (IMP), Vienna BioCenter (VBC), Campus-Vienna-Biocenter 1, A-1030 Vienna, Austria
[4]Howard Hughes Medical Institute, Immunology Program, and Ludwig Center, Memorial Sloan Kettering Cancer Center, New York, NY, USA
[5]Department of Computer Science, Princeton University, Princeton, NJ, USA
[6]Present address: Department of Data Science, Dana Farber Cancer Institute, Boston, MA, USA
[7]Lead contact
*Correspondence: pritykin@princeton.edu
https://doi.org/10.1016/j.celrep.2026.118156

## SUMMARY

T cells differentiate into functionally distinct subsets in response to pathogens, yet essential immune functions such as self-renewal, expansion, cytokine production, and cytotoxicity are common to multiple subsets and therefore must be programmed via shared regulatory mechanisms. To uncover these "archetypal" epigenomic programs, we integrated a single-cell multiomic (scATAC+RNA-seq) dataset of CD4 and CD8 T cells responding to acute and chronic viral infection with a comprehensive bulk ATAC-seq compendium. Our analysis reveals a modular architecture governing T cell identity, implicating distinct transcription factors in reusing regulatory programs across subsets and lineages. Notably, translationally important progenitor CD8 T cells, critical for sustained immunity and present in both acute and chronic infection, exhibit a composite regulatory state, combining a CD8 exhaustion program with a follicular helper CD4 T cell program. In sum, this resource and generalizable computational framework will aid the mechanistic dissection of adaptive immunity and immunotherapy design.

## INTRODUCTION

T cells coordinate systemic immunity and protect against infections and tumors. Over their lifespan and in response to antigen stimulation, T cells differentiate into distinct lineages and functional states, each characterized by unique epigenomic and transcriptional profiles.[1–5] However, many core immune functions, such as self-renewal, expansion, cytotoxicity, production of and response to particular cytokines, and tissue migration, are not unique to a specific T cell subset, but instead recur across subsets.[5] Therefore, these recurring functions must be driven by common underlying regulatory programs. Indeed, flow cytometry analysis has revealed that activity of many key functional molecules, surface markers, and transcription factors (TFs) spans multiple T cell populations, including across CD4 and CD8 lineages. For example, interferon-$\gamma$ (IFN-$\gamma$) production is a hallmark of both effector CD8 T cell and CD4 Th1 cell responses; adhesion molecule L-selectin (CD62L) marks both naive and central memory T cells; both follicular helper CD4 T cells (Tfh) and exhausted CD8 T cells express the inhibitory receptor programmed death 1 (PD-1) and the transcription factor TOX; and the chemokine receptor CXCR5 is shared between Tfh and progenitor exhausted CD8 T cells.[1,3–11] A central challenge in T cell biology is to dissect the shared and context-specific "archetypal" regulatory programs that govern both overlapping and specialized T cell functions. Achieving this requires comprehensive and high-resolution approaches capable of linking gene regulatory potential with transcriptional output. While flow cytometry with a limited set of markers is effective for identifying well-known populations, it limits unbiased comparisons of similarities and differences in discrete and continuous T cell phenotypes across models. Single-cell multiomic technologies that simultaneously profile chromatin accessibility and gene expression at the genome-wide level now provide an opportunity to resolve the archetypal regulatory programs systematically across heterogeneous T cell populations, following the successful use of these technologies in other biological contexts.[12–19]

The lymphocytic choriomeningitis virus (LCMV) infection model offers a powerful, gold standard experimental system to investigate regulatory programs across divergent immune responses. Profiling T cells at day 7 post-infection with the two most widely used LCMV strains, either the Armstrong strain (acute infection) or Clone 13 (chronic infection), captures a critical point of divergence between viral clearance and persistent


This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

infection leading to exhaustion.[8,9,20,21] Although these two infections are often described as eliciting largely non-overlapping T cell subsets due to differences in antigen persistence and immune outcomes, we sought to challenge this view by comprehensively exploring shared versus context-specific features of the T cell response. Including both CD4 and CD8 T cells in this analysis helps fill a significant gap in prior genomic studies of the LCMV model, which have provided foundational insights into CD8 T cells while leaving CD4 subsets and especially their relationship with CD8 responses relatively understudied.[7,22–25]

In this study, we performed a joint single-cell chromatin accessibility and transcriptomic profiling (scATAC + RNA-seq) of T cells responding to viral infection. Chromatin accessibility reflects cellular regulatory potential and can be used to infer TF activity.[17,26–31] However, unlike transcriptomics, which benefits from reliable public reference gene annotations, chromatin accessibility is more plastic and cell type-specific, and analysis requires a consensus reference set of accessible regions, which is lacking for the T cell lineage. To address this, we compiled a compendium of 285 bulk ATAC-seq profiles for sorted T cell subpopulations collected from 17 different publications and built a comprehensive reference atlas of chromatin accessibility regions across specific functional subsets of CD4 and CD8 T cells. We developed a computational framework that leverages this atlas and enables high-resolution, biologically interpretable analysis of T cell scATAC-seq and multiomic scATAC+RNA-seq data. In our multiomic data, we identified all major CD4 and CD8 T cell functional states, including naive, memory, effector, regulatory, Tfh, exhausted, and progenitor populations, each present in both LCMV Armstrong and Clone 13 responses, albeit at different frequencies. To systematically decompose the epigenomic landscape of T cells, we applied archetypal analysis to chromatin accessibility profiles, following its productive use in prior gene expression studies.[32–34] This approach identified a small number of "extreme" states, or *archetypes*, that capture the major axes of variation in the data. These archetypes recapitulated key regulatory programs, both unique to individual T cell populations and shared between populations, such as cytotoxicity, exhaustion, quiescence, and Tfh-like regulation. TF motif analysis implicated potential drivers of these programs; it was further validated using allele-specific analysis, leveraging natural genetic polymorphisms in our data, which was generated in an F1 hybrid of two mouse strains. Remarkably, unlike multiple other functional T cell states, Tcf1$^+$ progenitor CD8 T cells, a population of major translational importance,[8–10,26,35–40] did not align with a single archetype, but instead exhibited a composite epigenomic identity combining CD8 exhaustion- and CD4 Tfh-associated archetypes.

Together, our work introduces a powerful framework for high-resolution, interpretable single-cell multiomic analysis and reveals shared regulatory logic across T cell populations responding to viral infection, linking cytotoxic CD8 and helper CD4 T cell programs. By integrating computational analysis with comprehensive experimental profiling of both CD4 and CD8 T cells, we provide a valuable resource and conceptual advance for understanding the transcriptional control of adaptive immune responses. This computational and experimental framework will also be broadly applicable across experimental systems.

## RESULTS

### ATAC-seq compendium enables robust analysis of scATAC+RNA-seq data for T cells in LCMV infection

We used single-cell genomics to systematically characterize both discrete and continuous phenotypes and epigenomic states of CD4 and CD8 T cells during the early stages of the viral response. We generated single-cell multiomic datasets capturing both chromatin accessibility (scATAC-seq) and gene expression (scRNA-seq) in splenic T cells at day 7 post-infection with either LCMV Armstrong (Arm) or LCMV Clone 13 (Cl13). To enhance the resolution and reproducibility of chromatin accessibility measurements, we also generated standalone scATAC-seq data for the same cell populations (Figures 1A and S1, STAR Methods), yielding a total of 31,662 cells. These experiments were performed in F1 hybrid mice generated by crossing the laboratory C57BL/6 (B6) and wild-derived CAST/EiJ (Cast) strains, allowing both allele-agnostic analysis, which forms the primary focus of this study, and allele-specific analysis of gene expression regulation (described later in discussion).

We developed a computational framework to analyze this multiomic dataset, integrating best practices from previous methods and building a chromatin accessibility reference for T cells (STAR Methods). We systematically defined a comprehensive set of accessible regions using publicly available bulk ATAC-seq data. Building on our prior resource for CD8 T cells,[26] we assembled and uniformly reprocessed a compendium of 285 ATAC-seq profiles covering various functional states of sorted subpopulations of CD8 and CD4 T cells across 17 studies spanning diverse disease contexts, including LCMV infection and tumor models at multiple timepoints (Figure 1B; Table S1, STAR Methods). This broad coverage ensured a comprehensive framework for mapping accessible regulatory regions across the entire T cell lineage, further enhanced with peaks identified from our scATAC-seq data. The resulting set of 223,492 peaks served as the basis for subsequent analyses (Table S2).

Previous studies have shown that chromatin accessibility can better distinguish cellular phenotypes than transcriptomics.[19,29,44–47] Therefore, we used our scATAC-seq data as the primary modality for comprehensive functional annotation. Following normalization, dimensionality reduction, and integration of the standalone scATAC-seq and the scATAC-seq component of the multiome, we performed clustering to define reproducible subpopulations of CD4 and CD8 T cells (Figures 1C–1E and S2). We annotated these clusters using established marker genes, based on both their gene expression and the activity of associated chromatin accessibility peaks (Figures 1E, S2D, and S3; Tables S3–S5, STAR Methods). We identified all major CD4 and CD8 T cell functional states, including naive, memory, effector, regulatory, Tfh, exhausted, and progenitor populations. Specifically, to distinguish CD4 and CD8 T cell lineages, we used both gene expression and chromatin accessibility of the lineage-defining genes *Cd4*, *Cd8a,* and *Cd8b1*. We identified quiescent naive and memory CD4 and CD8 T cells, i.e., not actively responding to LCMV, based on the activity of *Sell*, *Lef1*, *Tcf7* (encoding protein Tcf1), *Slamf6,* and *Ifngr2*. Among CD4 T cells, we defined regulatory T cells (Treg) using *Foxp3*, follicular helper T cells (Tfh) using *Bcl6*, *Cxcr5*, *Pdcd1*, *Tox,* and *Tcf7*, and activated Th1 cells using *Ifng*.

CellPress
OPEN ACCESS

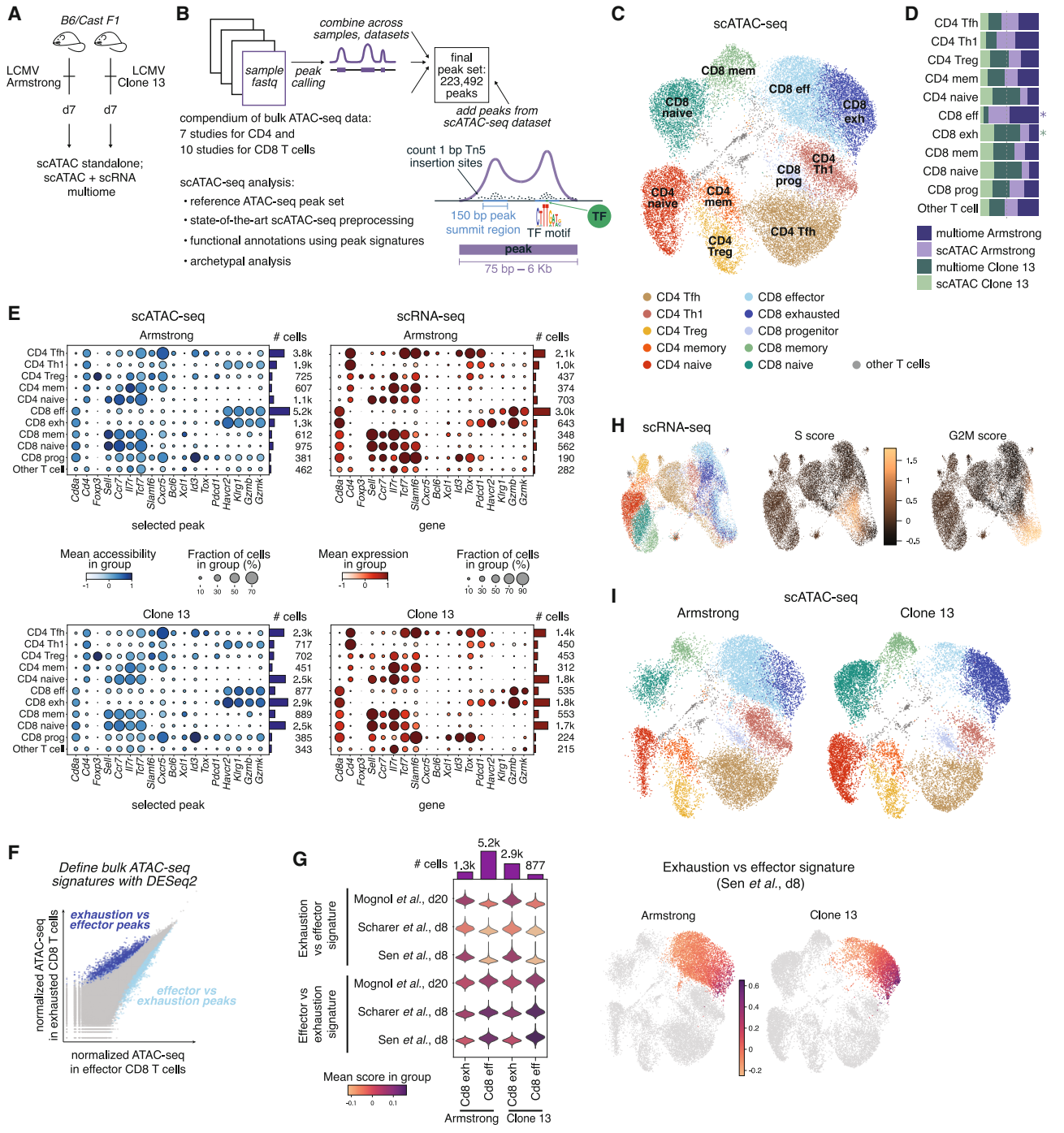



**Figure 1. Early T cell responses to acute and chronic infection share the same functional cell states**

(A) Experimental setup: splenic T cells from F1 hybrid C57BL/6J × CAST/EiJ mice infected by either Armstrong (Arm, acute infection) or Clone 13 (Cl13, chronic infection) of Lymphocytic choriomeningitis virus (LCMV) were profiled either with scATAC-seq or multiomic scATAC+scRNA-seq 7 days (d7) post-infection.

(B) Summary and schematic of the T cell scATAC-seq analysis. For the T cell ATAC-seq peak atlas, peaks were called using publicly available bulk ATAC-seq datasets for T cells from both CD4 and CD8 lineages, and combined with peaks from the scATAC-seq dataset generated in this study. Key aspects of our scATAC-seq data analysis pipeline are highlighted.

(C) Uniform manifold approximation and projection (UMAP) of scATAC-seq data with T cell functional annotations.

(D) Reproducibility of cell populations. Bar plot shows the distribution of functional cell states across the four samples (LCMV Arm and Cl13 for both standalone scATAC-seq and scATAC-seq from the multiome profiling). Asterisks, significant enrichments (hypergeometric test, odds ratio >2, $p < 1.3e{-}70$).

*(legend continued on next page)*

Among CD8 T cells, we distinguished between progenitor cells associated with *Cxcr5*, *Slamf6*, *Id3*, *Tcf7*, *Pdcd1,* and *Tox*, and a continuum of actively responding cells. We classified them into hyperactivated cells in the early state of exhaustion (hereafter referred to as exhausted), more frequent in the LCMV Cl13 response and associated with higher activity of inhibitory and exhaustion marker genes *Pdcd1*, *Tox*, *Havcr2* (encoding Tim3), *Lag3,* and *Tigit* and elevated cytotoxicity as indicated by overexpression of *Gzmb*, *Nkg7,* and *Prf1*, and the remaining effector cells, more frequent in the LCMV Arm response. To further support our annotations, we derived functional epigenomic signatures from the bulk ATAC-seq compendium (Figures 1F; Table S3, STAR Methods). To eliminate potential cross-study technical noise or batch effects, these signatures were generated by identifying differentially accessible peaks distinguishing known T cell subsets within each independent study. We then scored these peak signatures in our scATAC-seq data, assigning higher scores to cells with greater accessibility at subset-specific peaks. This analysis further supported our annotations by capturing key contrasts, e.g., between quiescent (naive/memory) and antigen-activated cells, naive and memory cells, and between effector and exhausted CD8 T cells (Figures 1G and S2F–S2H). Furthermore, all T cell subpopulations were observed in both LCMV Arm and Cl13 in both standalone scATAC-seq and the multiome scATAC-seq, demonstrating reproducibility across technologies (Figures 1D and 1E). Thus, all major functional subsets of CD4 and CD8 T cells are reproducibly present in both LCMV Arm and Cl13 responses, albeit in different frequencies.

Notably, the presence of exhausted-like CD8 T cells (a transitional, highly active state of early exhaustion) in the acute Armstrong response challenges the traditional view that exhaustion is restricted to chronic infection or cancer. However, this observation aligns with recent studies reporting transcriptional and epigenetic features of exhaustion emerging early and transiently in acute responses before viral clearance.[48,49] Our findings reinforce this emerging perspective by demonstrating that exhaustion-associated states are not exclusive to persistent antigen exposure but may be part of a broader T cell activation and differentiation landscape.

We observed that standard scATAC-seq analysis pipelines can sometimes group together cell populations that possess distinct identities, including mixing cells from the fundamentally divergent CD4 and CD8 lineages (Figures S4 and S5, STAR Methods). Furthermore, clustering based on the scRNA-seq modality frequently grouped cell populations that share dominant, transient functional states but come from distinct lineages or differentiation histories (Figures 1H and S6). While different T cell phenotypes can upregulate overlapping functional modules, such as core activation, cytokine production, or proliferation programs, this transcriptomic convergence can obscure underlying cellular identity. For example, cells sharing dominant biological programs such as quiescence (naive and memory cells from both CD4 and CD8 lineages) or rapid proliferation (cycling CD4 and CD8 T cells from across antigen-activated functional states) have highly overlapping gene expression profiles, which can result in the collapse of distinct lineages into the same scRNA-seq cluster. This challenge highlights a crucial distinction between stable cell identity and shared functional modules, aligning with the recently proposed modular framework for T cell nomenclature.[5] These observations are consistent with prior work suggesting that chromatin accessibility provides a more stable and robust signal of cell identity than transcriptomics,[19,29,44–47] and highlight the power of our computational framework integrating single-cell data with a comprehensive reference atlas.

In summary, our computational framework leveraging a comprehensive T cell ATAC-seq compendium to analyze our high-resolution multiomic dataset allowed us to accurately map functional and differentiation phenotypes in early LCMV infection. We found that all major CD4 and CD8 T cell populations are shared between the acute and chronic responses. This demonstrates that the immunological divergence between these two infection regimes is fundamentally quantitative rather than qualitative, driven by distinct shifts in cell-state frequencies rather than the emergence of response-specific cell identities.

## Shared progenitor CD8 T cell state in early LCMV Armstrong and clone 13 responses

A particularly important and intensively studied population is the progenitor CD8 T cells, essential for sustaining responses in chronic infection and cancer across both mouse models and human patients, and for enabling effective T cell reactivation in cancer immunotherapies.[8–10,26,35–40] This population is

(E) Marker gene activity across cell populations. Dot plots show normalized chromatin accessibility levels at selected peaks for marker genes (left) and normalized expression of selected marker genes (right) in different T cell functional states, for Arm (top) and Cl13 (bottom) cells. Dot size, fraction of cells with normalized accessibility or expression greater than 0; dot color, mean normalized accessibility or expression in that cell state. Bar plots on the right side show the number of cells per cell state.

(F) Definition of epigenomic signatures. Example of how epigenomic signatures are generated using `DESeq2` for bulk ATAC-seq data for two conditions from the same dataset. Peak signatures are formed from differentially accessible peaks in either direction (LFC >1 or LFC <-1, FDR $\leq$0.05). The scATAC-seq data are then scored for enrichment of these peak signatures.

(G) Scoring of exhaustion vs. effector epigenomic signatures in single cells. *Left*, violin plots showing distributions of the epigenomic signature scores in effector and exhausted CD8 T cells, separately in Arm and Cl13 samples. The signatures are for CD8 T cells at d8 in LCMV infection from multiple studies.[41–43] In each dataset, peaks significantly more accessible in exhausted than in effector cells (exhaustion vs. effector signature) and in effector than in exhausted cells (effector vs. exhaustion signature) were defined, and then scored based on accessibility of these peaks in our single-cell data. Top margin: the bar plot of the number of cells in each group. All scores were significantly different between effector and exhausted groups (FDR <1$e$−46, two-sided Mann-Whitney *U* test). *Right*, UMAP plots show the scores of the exhaustion vs. effector signature derived from Sen et al.[43] in cells annotated as either effector or exhausted CD8 T cells, separately for the Arm and Cl13 responses.

(H) scRNA-seq analysis. *Left*, UMAP of the scRNA-seq modality of the dataset showing T cell functional states annotated using the scATAC-seq analysis. *Middle* and *right*: same UMAP showing cell cycle gene signature scores (S phase, middle; G2M phase, right).

(I) Comparison of LCMV Arm and Cl13. UMAP with the scATAC-seq-based functional T cell state annotations separately for cells from Arm and Cl13 infection.

CellPress
OPEN ACCESS

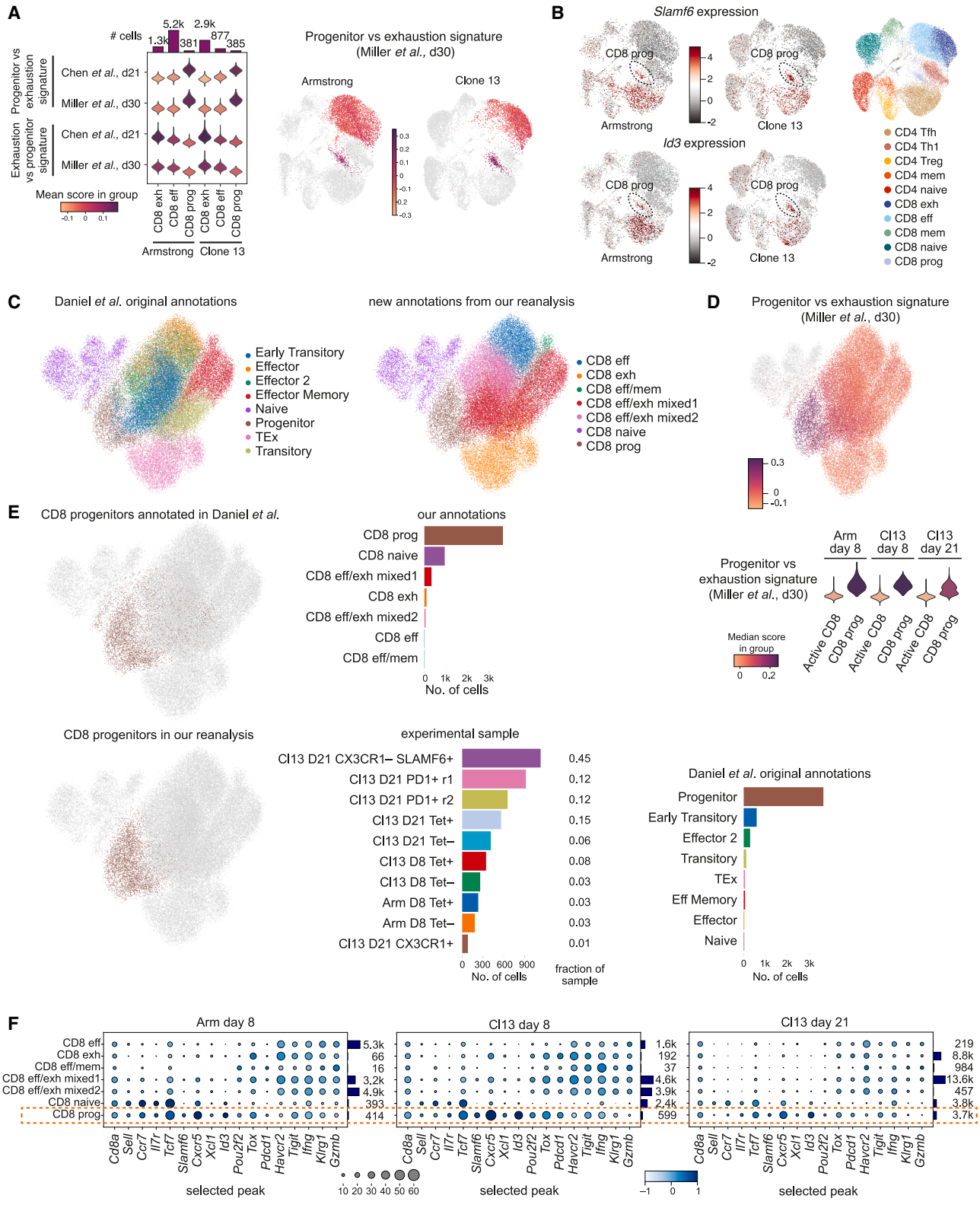



**Figure 2. Shared progenitor CD8 T cell state in early response to acute and chronic viral infection**

(A) Scoring of progenitor vs. exhaustion epigenomic signatures. *Left*, violin plots of the epigenomic signature scores compare progenitor vs. exhaustion T cell states (similar to Figure 1G). The signatures were defined using data from progenitor and terminally exhausted CD8 T cells at d21 in response to melanoma[56] or at

*(legend continued on next page)*

characterized by the marker Cxcr5 and integrates stem- or memory-like markers such as Tcf1, Slamf6, Id3, and IL-7R with exhaustion markers, including elevated levels of Tox and PD-1, within the CD8 T cell lineage. We previously reported that this progenitor CD8 T cell population is shared between acute and chronic responses, using scATAC-seq and scRNA-seq profiling in LCMV Arm and Cl13, a finding we further validated by reanalyzing multiple public scRNA-seq datasets.[26] Recent functional studies, including adoptive transfer experiments, have further confirmed that this progenitor population is shared between acute and chronic infection.[49–51] However, earlier studies, including those using high-throughput genomics, described these progenitor exhausted Cxcr5$^+$ Tcf1$^+$ CD8 T cells exclusively in chronic infection and cancer, distinguishing them from the related memory precursor effector cells (typically defined as IL-7R$^+$ Klrg1$^-$ effectors, also shown to have elevated levels of Tcf1 and Id3) described in acute infection.[20,28,29,52–55] To reconcile these historical distinctions, we explored these CD8 T cell phenotypes using multiomic single-cell data and our refined computational pipeline.

Our analysis supports the presence of this shared progenitor CD8 T cell population across both LCMV Arm and Cl13 responses (Figures 2A and 2B). This population was reproducible between standalone and multiome scATAC-seq (Figure 1D) and showed enrichment of transcriptional activity for genes associated with both progenitor exhausted and memory precursor phenotypes, such as *Cxcr5*, *Slamf6*, *Tox*, *Tcf7*, *Id3,* and *Il7r*, even when analyzed separately by infection type (Figures 1E and 2B). Moreover, cells in this population were consistently enriched for the progenitor ATAC-seq peak signatures derived from independent studies, again reproducibly across both LCMV Arm and Cl13 (Figure 2A). Taken together, these observations reinforce the existence of an epigenomically and transcriptionally similar progenitor CD8 T cell state in both acute and chronic viral infection.

Nevertheless, two prior prominent scATAC-seq analyses of CD8 T cells in LCMV Arm and Cl13, by Daniel et al.[28] and Giles et al.,[29] reported largely non-overlapping T cell subsets between the two infections, including a distinct progenitor exhausted subpopulation in LCMV Cl13 and a separate memory precursor population in LCMV Arm. We hypothesized that these apparently distinct populations might represent a single underlying state whose annotation depends on analytical parameters and the reference set of regulatory elements used. We reasoned that reanalyzing these important datasets using our comprehensive T cell compendium could provide a unified perspective and help reconcile the different interpretations.

To test this, we reanalyzed the scATAC-seq dataset from Daniel et al.[28] across multiple timepoints and conditions, including acute (Armstrong) and chronic (Clone 13) responses at day 8, alongside a late chronic timepoint at day 21 (Figures 2C–2F and S7A–S7E). By using our T cell ATAC-seq peak atlas and applying our computational pipeline to the full dataset, we identified cluster 6 as the progenitor CD8 T cell population. This cluster robustly aligned with the progenitor exhausted state, capturing 72% of the cells originally annotated as such,[28] and was overrepresented within the sorted Cx3cr1$^-$ Slamf6$^+$ cell fraction, which was experimentally enriched for progenitors in the original study (Figures 2E, S7B, and S7C). While this cluster was well-represented at the late day 21 chronic infection timepoint, it also encompassed cells from both acute and chronic conditions at day 8 (Figures 2E,2F, and S7C). Even when evaluated separately by condition, these day 8 cells demonstrated distinct progenitor marker gene activity and an enrichment for the progenitor-specific epigenomic signatures (Figures 2D and 2F), supporting our finding of a progenitor state that spans both acute and chronic contexts early in infection. Similarly, reanalysis of the day 8 scATAC-seq dataset from Giles et al.[29] using our computational pipeline identified cluster 3 as CD8 progenitors, again supported by marker gene activity and epigenomic signatures across both infections (Figure S8). This cluster contained 99.8% of cells originally annotated as pre-exhausted (from Cl13) and 92% of those labeled as memory precursors (from Arm) by the authors,[29] suggesting that the two groups represent a similar underlying progenitor state.

Furthermore, our comprehensive reanalysis of the Daniel et al.[28] dataset resolved a distinct exhausted population (clusters 3, 9, and 14) that aligned with exhaustion signatures and was enriched in the late day 21 chronic samples (Figures S7C–S7G). Notably, these clusters captured cells from both the acute Armstrong and chronic Clone 13 day 8 samples (Figures S7F and S7G). When analyzed independently, these day 8 fractions had enrichment of the exhaustion signature scores, including the early exhaustion

d30 in response to LCMV Cl13 infection.[37] All scores were significantly different between progenitor vs. effector or exhausted CD8 T cell groups (FDR <5*e*−52, two-sided Mann-Whitney *U* test). *Right*, UMAP plots showing the epigenomic signature scores of progenitor vs. terminal exhaustion from Miller et al.[37] in cells annotated as either progenitor, effector, or exhausted CD8 T cells in our data.

(B) Progenitor marker gene expression. UMAP plots show normalized scRNA-seq gene expression for the two progenitor marker genes, *Id3* and *Slamf6*, separately for Armstrong and Clone 13 cells. Progenitor CD8 T cells highlighted.

(C–F) Reanalysis of the scATAC-seq data for CD8 T cells at days 8 and 21 in response to LCMV infection from Daniel et al.[28] The dataset was preprocessed and quantified using our T cell ATAC-seq peak atlas and then analyzed using our computational pipeline. (C) UMAP plots with the cell state annotations from the publication[28] (*left*) and our annotations (*right*). (D) Scores of the progenitor vs. exhaustion epigenomic signature (as in A). *Top*, UMAP plots showing the score for all cells except naive (shown in gray). *Bottom*, violin plots comparing the score between progenitor cells and all other non-naive cells, separately for the three conditions and timepoints. (E) UMAP plots showing the progenitor cells. *Top*, cells annotated as progenitors in the publication.[28] The corresponding bar plot shows the distribution of these cells' annotations in our reanalysis. *Bottom*, cells annotated as progenitors in our reanalysis. The left barplot shows the distribution of samples of origin for these cells, showing the fraction of progenitor cells within each sample. The right bar plot shows the distribution of these cells' annotations in the original analysis. (F) Marker gene activity across cell populations. Dot plots show chromatin accessibility for selected peaks near marker genes in cell populations from our reanalysis of the data, separately for the Arm day 8 (*left*), Cl13 day 8 (*center*), and Cl13 day 21 (*right*) samples, with the progenitor cells highlighted. The dot size indicates the fraction of cells with normalized accessibility greater than 0; dot color indicates the mean normalized accessibility in that cluster. Barplots on the right side show the number of cells per cluster.

CellPress
OPEN ACCESS

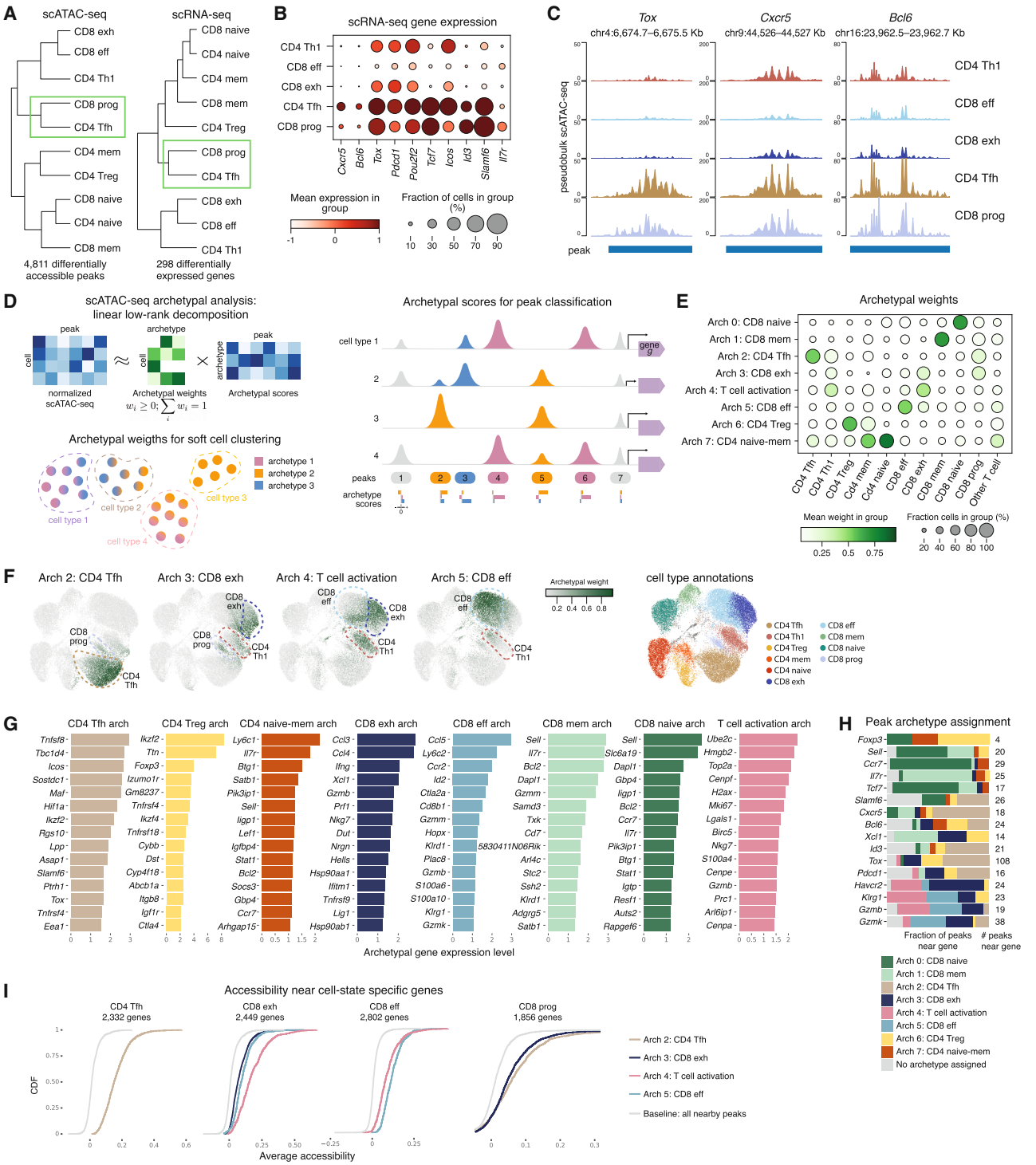



**Figure 3. Archetypal analysis of scATAC-seq data reveals the mixed regulatory program of progenitor CD8 T cells and their similarity with follicular helper CD4 T cells**

(A) Relative epigenomic and transcriptomic similarity of cell populations. Hierarchical clustering of pseudobulk scATAC-seq (*left*) and scRNA-seq (*right*) profiles for T cells in different functional states. Significantly differentially accessible peaks (LFC >1.5, FDR ≤0.05 for any cell state vs. other cells, left) and differentially expressed genes (LFC >1.5, FDR ≤0.05 for any cell state vs. other cells, right) were used for this analysis.

(B) CD8 progenitor and Tfh marker gene expression. Dot plot shows the expression of selected Tfh and CD8 progenitor marker genes for antigen-activated T cell subsets. Dot size, fraction of cells with normalized expression greater than 0; dot color, mean normalized expression in that cell state.

*(legend continued on next page)*

epigenomic signature (differentially accessible peaks) directly derived from our own day 7 dataset (Figure S7G). Because we also identified other clusters that likely contained a mixture of effector and exhausted cells, our explicit classification of these exhausted cells remains conservative. This analysis reinforces our finding that an early exhausted phenotype is reproducibly present in both acute and chronic responses.

Together, these results demonstrate that our computational framework, leveraging a comprehensive bulk ATAC-seq reference and functional epigenomic signatures, enables robust and consistent annotation of diverse T cell functional states across independent studies. Applied to both the multiomic data generated in this study and previously published datasets, this analysis supports the existence of a shared CD8 progenitor population in both acute and chronic viral responses.

## Progenitor CD8 T cells and follicular helper CD4 T cells have similar chromatin accessibility and gene expression profiles

Next, we used our single-cell multiomic dataset for systematic analysis of similarities and differences of regulatory programs across CD4 and CD8 T cell lineages in response to LCMV infection (Figures 3A, S9A, and S9B). Notably, antigen activation status appeared to be a stronger determinant of epigenomic and phenotypic similarity than lineage identity. This observation aligns with an emerging framework that views T cell identity as a modular combination of stable lineage programs and shared functional states.[5,57] For example, naive and memory CD4 and CD8 T cells clustered together, while effector and exhausted CD8 T cells clustered more closely with CD4 Th1 cells than with other CD8 T cell subsets. These results suggest that although the CD4 and CD8 lineages diverged early in thymic development,[58,59] their mature regulatory landscapes are more strongly shaped by functional status and signals from their microenvironment than by their lineage history.

Interestingly, progenitor CD8 T cells were most similar to CD4 Tfh cells in both their genome-wide chromatin accessibility and transcriptomic profiles (Figures 3A, S9A, and S9B). These two subpopulations also exhibited shared activity (in both expression and chromatin accessibility) of key marker genes of CD8 progenitors and Tfh cells, including *Cxcr5*, *Bcl6*, *Tox*, *Pdcd1*, *Pou2f2*, *Tcf7*, and *Icos*, which distinguished them from other CD4 and CD8 T cell states (Figures 3B and 3C).

On one hand, this similarity is consistent with the original characterization of progenitor exhausted CD8 T cells as a $Cxcr5^+$ subpopulation within the CD8 lineage, while Cxcr5 is a canonical marker of CD4 Tfh cells, and with reports of other important molecules (e.g., Bcl6, Tcf1) shared between CD8 progenitors and Tfh cells.[9,11,60] In contrast, these two subpopulations have distinct functional roles shaped by their different lineages and immune contexts. Thus, this striking convergence between progenitor CD8 T cell and CD4 Tfh cell states motivated us to further investigate their shared regulatory programs at finer resolution.

## Archetypal analysis of scATAC-seq data reveals a mixed regulatory program of progenitor CD8 T cells

We sought to understand how transcriptionally and epigenomically similar phenotypes can arise in distinct T cell subsets, including across lineages. In particular, we wanted to further investigate the striking similarity between CD4 Tfh and progenitor CD8 T cell states. We hypothesized that these cell states converge on similar phenotypes through shared regulatory programs, driven by common TFs, cytokines, and signaling molecules. To characterize these shared molecular programs, we needed a framework beyond discrete, non-overlapping cell type annotations.

To do so, we used archetypal analysis of the scATAC-seq data (Figure 3D). Archetypal analysis models chromatin accessibility profiles via a linear low-rank decomposition using a small number of archetypes, each representing an idealized regulatory program learned in an unsupervised manner from the data[61–63] (STAR Methods). The chromatin accessibility profile of each T cell can then be represented as a weighted combination of these archetypes. These archetypal weights provide a quantitative and interpretable framework for understanding how cell states relate to one another and how shared regulatory programs are deployed across cells. Each archetype is represented by a vector over accessibility peaks and can be interpreted post hoc based on the properties of the cells and peaks it comprises. We refer to the level of accessibility of a peak within an archetype

(C) CD8 progenitor and Tfh marker gene chromatin accessibility. Pseudobulk scATAC-seq tracks for antigen-activated T cell subsets at selected peaks (shown as horizontal blue bars at the bottom) of the genes *Tox*, *Cxcr5*, and *Bcl6*.

(D) Schematic of the archetypal analysis of scATAC-seq data. *Top left*: archetypal analysis relies on a linear low-rank matrix decomposition of the data (STAR Methods). *Bottom left*: archetypal analysis enables soft cell clustering via the archetypal weight matrix. Each cell can be assigned to multiple archetypes with some probability (referred to as an archetypal weight). Each archetype can be considered as a pseudo-cell averaging over cells with a similar peak accessibility distribution. *Right*: archetypal analysis assigns a score to each peak, reflecting its contribution to each chromatin accessibility archetype. For downstream analysis, each peak is uniquely assigned to the archetype with the highest score; peaks without a predominant archetype are not assigned to any (shown in gray). A gene may have multiple peaks coming from different archetypes, which may be associated with distinct programs of differential regulation across functional cell states and conditions.

(E–I) Results of archetypal analysis of our T cell scATAC-seq data using eight archetypes. (E) Archetypal weights across cell populations. Dot plot shows archetypal weights (probability of an archetype in a cell) for different cell states. Dot size, fraction of cells with archetypal weight greater than 0; dot color, mean archetypal weight in that cell state. (F) UMAP plots show archetypal weights for four example archetypes. Cell type annotations are shown for reference. (G) Top protein-coding genes expressed in each scATAC-seq archetype (STAR Methods). (H) Bar plot of the distributions of peak assignments to archetypes for selected genes. (I) Archetypal regulation of cell state-specific genes. Analysis of cell state-specific chromatin accessibility in peaks associated with genes with cell state-specific expression, for four selected cell states: CD4 Tfh, exhausted CD8 T cells, CD8 effectors, and CD8 progenitor cells. Genes with significant differential overexpression (LFC >0, FDR $\leq$0.05) in a cell state were selected. For these genes, cumulative distribution function (CDF) plots show the average accessibility of nearby peaks that are associated with a particular archetype. The gray lines indicate the distribution of the average accessibility for all peaks near a gene, used as a baseline. Only the archetypes that have significantly increased accessibility compared to the baseline are shown for each cell state (KS test, FDR $\leq$0.05 and difference in medians $\geq$0.03).

CellPress
OPEN ACCESS

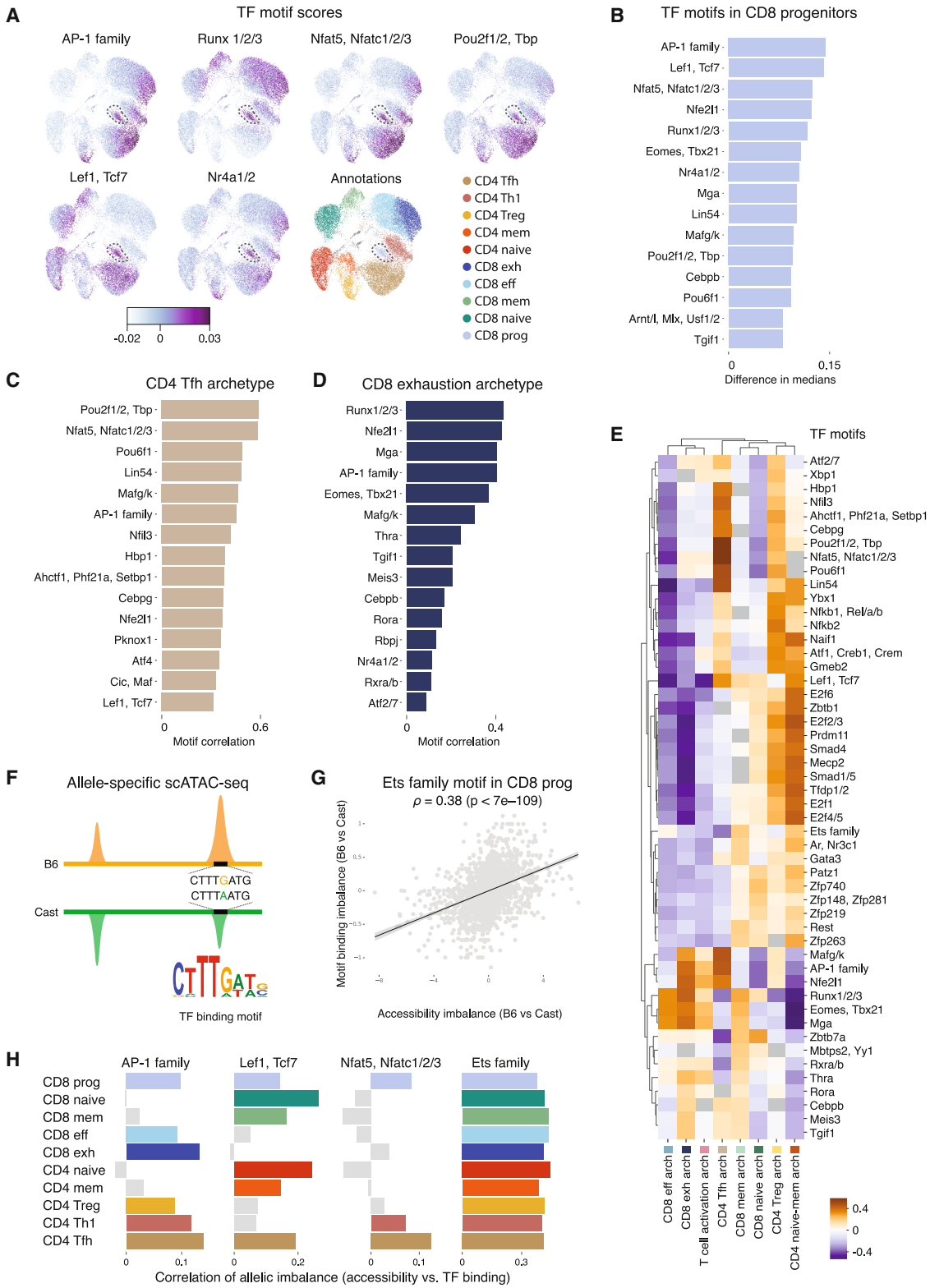



**Figure 4. Distinct transcription factor motifs are associated with archetypal regulatory programs**

(A) TF motifs across cell populations. UMAP showing single-cell motif scores for selected transcription factor (TF) motifs (STAR Methods). Cell type annotations are shown for reference.

*(legend continued on next page)*

as an archetypal score, reflecting the peak's relative contribution to that archetypal regulatory program.

We applied archetypal analysis with $k = 8$ archetypes to our scATAC-seq data for T cells from the day 7 response to LCMV (Figures 3E–3I, S9, and S10). Archetypes were annotated based on cell state annotations among cells with high archetypal weights, as well as functional annotations of genes whose expression was most strongly associated with the archetypes. Some archetypes corresponded predominantly to individual cell states, such as CD4 Treg or naive CD8 T cells, while others captured shared regulatory programs across functionally related subsets (Figures 3E and 3F). For example, we identified a T cell activation archetype shared between CD4 Th1, CD8 effector and exhausted T cells, associated with high expression of proliferation genes *Top2a*, *Mki67,* and *Birc5* and cytotoxic genes *Nkg7* and *Gzmb* (Figures 3E–3G and S10). We also distinguished between a CD8 effector archetype, predominant in effector CD8 T cells, and a separate CD8 hyperactivation or early exhaustion archetype (referred to as CD8 exhaustion), observed in exhausted CD8 T cells, but also in a subset of Th1 cells and in progenitor CD8 T cells. Notably, progenitor CD8 T cells lacked a dominant archetype and instead exhibited a mixed signature, primarily composed from the CD8 exhaustion and CD4 Tfh archetypes (Figures 3E,3F, S9C, and S9D). This was consistent with their transcriptional and epigenomic similarity to Tfh cells and their original definition as progenitor exhausted CD8 T cells. This pattern was robustly recapitulated when applying our archetypes to our reanalysis of the previously published datasets[28,29] (Figures S9G and S9H, STAR Methods). Thus, archetypal analysis effectively captured the continuum of T cell phenotypes and revealed the hybrid regulatory program of progenitor CD8 T cells.

To better understand the regulatory content of each archetype, we examined archetypal scores at individual peaks (Figures 3D and 3H). For interpretability, each peak was uniquely assigned to the archetype with the highest score for that peak (STAR Methods). We then studied the relationship between scATAC-seq peak activity and scRNA-seq gene expression across cells. We found that for genes overexpressed in specific T cell subsets, nearby peaks assigned to the corresponding archetypes had significantly elevated accessibility (Figures 3I and S9F). For example, peaks of the CD4 Tfh archetype were most accessible near genes highly expressed in CD4 Tfh cells, while peaks of CD8 effector and activation archetypes were enriched near genes upregulated in CD8 effector cells. Notably, genes overexpressed in progenitor CD8 T cells had equally strong association with both Tfh and exhausted CD8 T cell archetypes (Figure 3I). Thus, the scATAC-seq peak archetypes captured biologically meaningful regulatory programs across T cell states and revealed mixed regulation of gene activity in progenitor CD8 T cells.

We next asked whether peak archetypes also reflect gene expression regulation *within* individual T cell populations. For many key T cell marker genes, we found that different chromatin accessibility peaks within the same gene locus were assigned to distinct archetypes (Figure 3H), suggesting that these genes may be regulated by different sets of regulatory elements depending on cellular context. To explore this, we examined the correlation between individual peak accessibility and gene expression within specific T cell subsets. For example, *Tox* is a gene expressed in CD4 Tfh and exhausted CD8 T cells, and is also highly expressed in progenitor CD8 T cells[64–68] (Figure 3B). Among its 108 associated peaks, 49 were assigned to the CD4 Tfh archetype and 19 to the CD8 exhaustion archetype (Figures 3H and S11). Within CD4 Tfh cells, *Tox* peaks assigned to the Tfh archetype were most frequently significantly correlated with *Tox* expression, whereas in exhausted CD8 T cells, peaks from the CD8 T cell exhaustion archetype showed the strongest correlation (Figure S11). Thus, distinct regulatory elements, grouped by scATAC-seq archetypes, can be differentially associated with regulation of the same gene in a cell type-specific manner. This supports the view that the regulatory program of progenitor CD8 T cells integrates features of both CD4 Tfh and exhausted CD8 T cells, as reflected for example in the chromatin accessibility landscape of the *Tox* locus.

In conclusion, archetypal analysis of scATAC-seq data uncovered regulatory programs shared across T cell populations. These programs are closely linked to gene expression patterns, with the same gene regulated by distinct sets of peaks, assigned to different archetypes, in different T cell states. Archetypal analysis reveals that progenitor CD8 T cells exhibit a mixed regulatory program combining features of exhausted CD8 T cells and CD4 Tfh cells.

## TF motif analysis implicates regulatory drivers of archetypal programs and progenitor CD8 T cell identity

To better understand the transcriptional regulators shaping the landscape of continuous and discrete CD4 and CD8 T cell phenotypes during viral response, we next asked which transcription factors (TFs) are associated with the distinct chromatin

(B) CD8 progenitor motifs. Top 15 motifs significantly enriched in progenitor CD8 T cells compared to all other cells (FDR $\leq$0.05).
(C) CD4 Tfh motifs. Spearman correlations between CD4 Tfh archetypal weights and motif scores across cells. Bar plot shows the top 15 motifs with the strongest positive correlations.
(D) CD8 exhaustion motifs. Same as C, but for the CD8 T cell exhaustion archetype.
(E) Clustermap of correlations between archetypal weights and motif scores (top 10 most correlated motifs per archetype).
(F) Allele-specific scATAC-seq analysis. Schematic of the allele-specific analysis used to associate TF motifs with chromatin accessibility in ATAC-seq peaks. Using sequence differences between B6 and Cast genomes, we assigned scATAC-seq reads to parental alleles and quantified allele-specific pseudobulk chromatin accessibility per cell type. In parallel, we assessed allele-specific differences in motif matches due to genetic variants between the two genomes.
(G) Allele-specific analysis of Ets motif in CD8 progenitors. Scatterplot showing allelic imbalance of chromatin accessibility (*x*-axis) vs. allelic imbalance of Ets motif matches (*y* axis) across peaks in progenitor CD8 T cells. A significant positive Spearman correlation indicates that sequence variants with increased predicted Ets motif strength are associated with increased chromatin accessibility on the same allele, consistent with functional TF binding.
(H) Allele-specific analysis of TF motifs across cell populations. Spearman correlations between allelic imbalance of motif matches and chromatin accessibility across functional T cell states for selected motifs. Correlations that are not statistically significant (FDR >0.05) are shown in gray.

CellPress
OPEN ACCESS

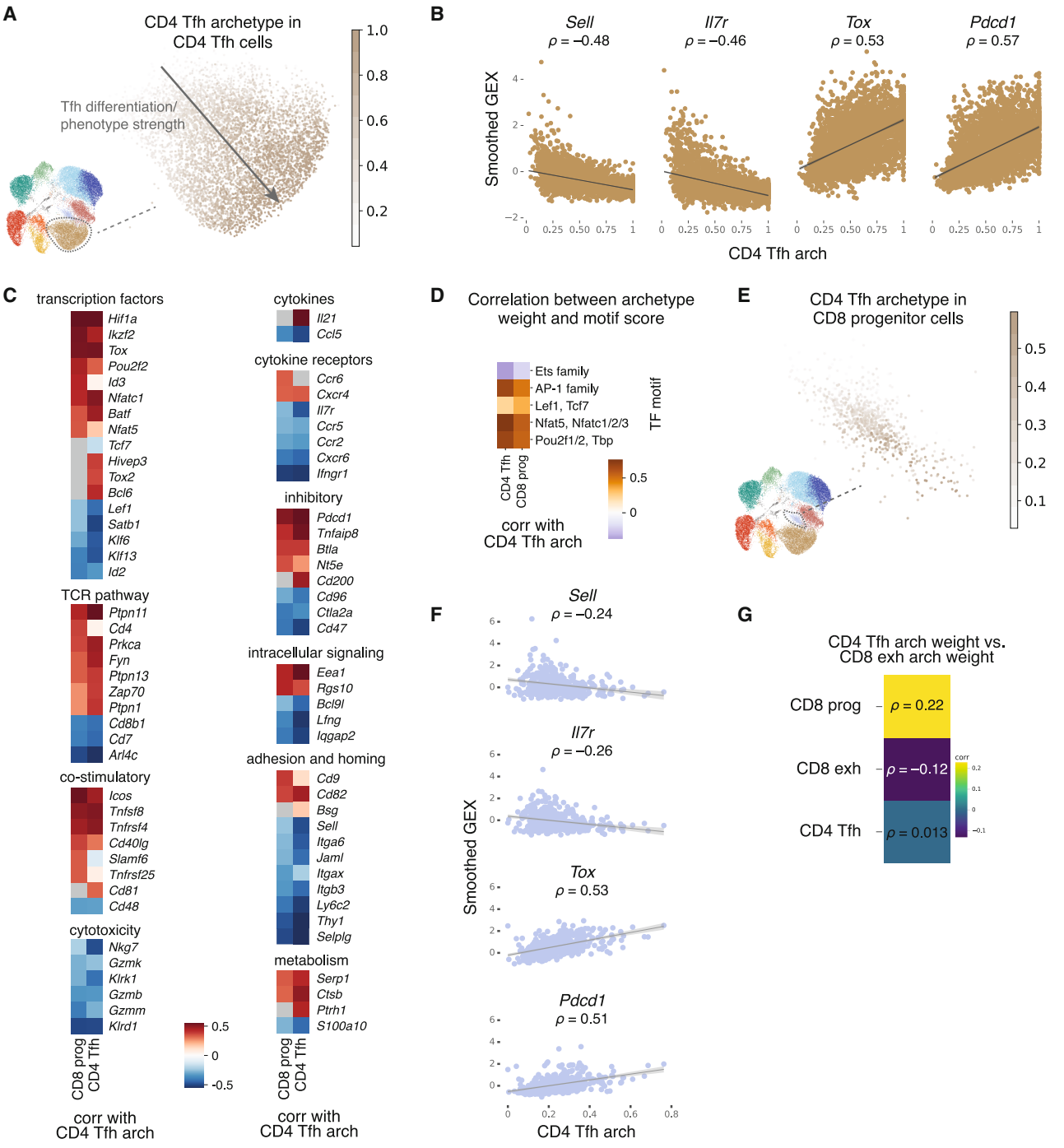



**Figure 5. Tfh archetypal weight reflects a similar gradient of differentiation in CD4 Tfh cells and progenitor CD8 T cells**

(A) Tfh archetypal gradient within Tfh cells. UMAP showing CD4 Tfh archetypal weights in the CD4 Tfh cell population (61 outlier cells not shown). We propose that this Tfh archetypal gradient reflects Tfh differentiation and/or the strength of the Tfh phenotype.

(B) Gene expression along the Tfh archetypal gradient. Scatterplots of normalized expression of selected genes in CD4 Tfh cells along the Tfh archetypal gradient. The gene expression values were averaged over neighbors in the scRNA-seq *k*-nearest-neighbor (kNN) graph ($k = 20$). A smoothed fit using ordinary least squares is shown in gray. Spearman's $\rho$ is shown on top.

(C) Correlation of Tfh archetypal gradient with gene expression. Heatmap showing Spearman correlations between CD4 Tfh archetypal weights and normalized gene expression within progenitor CD8 T cells and CD4 Tfh cells for selected genes from different functional categories. The gene expression values were averaged over neighbors in the scRNA-seq kNN graph ($k = 20$). Non-significant correlations (FDR >0.05) are shown in gray.

*(legend continued on next page)*

accessibility archetypes described above, particularly those underlying the regulatory program of progenitor CD8 T cells.

To this end, we first identified putative TF targets by scanning genomic sequences within chromatin accessibility peaks for TF motif matches. We then used these motif-based TF target signatures to score individual cells in the scATAC-seq data and compared motif scores across T cell functional states and archetypes (STAR Methods). To link TF activity with regulatory programs, we computed correlations between motif scores and archetypal weights across cells, allowing us to evaluate the relative, non-exclusive association of each TF across the entire regulatory landscape. Using this approach, we found that TF motifs enriched in progenitor CD8 T cells reflected a combination of those displaying the highest relative correlation with the CD4 Tfh and CD8 T cell exhaustion archetypes (Figures 4A–4E, S12A, and S12B). For example, progenitor cells were enriched with Nfat, Lef1/Tcf1, Pou2f1/2 (Oct1/2) motifs, which had the strongest relative association with the CD4 Tfh archetype, and with Runx and Tbet (Eomes/Tbx21) motifs, which preferentially associated with the CD8 exhaustion archetype. In addition, progenitor cells were also enriched for the AP-1 family motif; while AP-1 is a foundational hallmark of general T cell activation, our framework captured its relative coordinated association with both the Tfh and CD8 exhaustion archetypes. These results further support the idea that progenitor CD8 T cells integrate regulatory programs from both Tfh cells and exhausted CD8 T cells.

To further validate the associations between TF motifs and T cell functional phenotypes, we turned to allele-specific analysis. Because our single-cell data were generated from F1 hybrid mice (B6 × Cast), we were able to assign scATAC-seq reads to parental alleles and quantify allele-specific chromatin accessibility at regulatory regions within each T cell state. In parallel, we used known sequence variants between the two genomes to compute allele-specific motif-based predictions of TF binding (Figures 4F and 4G). We reasoned that if a TF functionally shapes the local chromatin landscape in a given cell population, a genetic variant altering its binding motif strength should cause a corresponding allelic imbalance in accessibility.[31,69] This allele-specific mapping provides a direct measurement of a motif's functional requirement for driving local epigenetic remodeling, capturing its actual impact on chromatin structure rather than the mere biochemical presence or nuclear translocation of the factor itself. Indeed, we observed a strong correlation for the Ets family motif across all T cell states (Figures 4G, 4H, and S12C), consistent with the widespread role of Ets family TFs in T cell regulation, modulated by co-factors in a cell type-specific manner.[31] Whereas the above allele-agnostic motif enrichment analysis was performed across cell states (Figures 4A–4E and S12B), this allele-specific analysis, performed *within* each cell state, provided orthogonal support for motif activity patterns. Specifically, in progenitor CD8 T cells, we observed significant allele-specific correlations with accessibility for motifs such as Lef1/Tcf7, Nfat, and AP-1, indicating their functional engagement in this population (Figures 4H and S12C). Comparing these patterns across T cell subsets revealed distinct regulatory dynamics. As expected for a core activation driver whose motif actively dictates accessibility changes upon antigen exposure, the AP-1 motif had a significant signal across all antigen-activated populations (including effector, exhausted, Treg, Th1, and Tfh cells) but not in naive or memory cells. Conversely, the Lef1/Tcf7 motif tracked with quiescence or restrained differentiation, demonstrating a significant functional impact on accessibility in naive, memory, Tfh, and progenitor cells but not in effector or exhausted CD8 T cells or Th1 cells. Remarkably, the Nfat motif demonstrated a selective profile, displaying a significant correlation with local accessibility in the CD8 progenitor, Th1, and Tfh populations. These patterns reinforce the notion that progenitor CD8 T cells integrate regulatory programs from Tfh and CD8 activation and exhaustion. These findings motivated us to further investigate the extent of underexplored epigenomic similarity between Tfh cells and progenitor CD8 T cells.

## Tfh archetype captures a similar gradient of differentiation within Tfh cells and progenitor CD8 T cells

We have shown that progenitor CD8 T cells share epigenomic and transcriptomic features with CD4 Tfh cells. To further investigate this connection, we examined the epigenomic CD4 Tfh archetype and its relationship to the progenitor CD8 T cell population.

We first examined the behavior of the CD4 Tfh archetype within the CD4 Tfh cell population itself. Within this subset, we observed a continuous gradient of the Tfh archetypal weight (Figure 5A). This gradient was positively correlated with expression of Tfh activation-associated genes such as *Tox*, *Pdcd1*, *Bcl6* and *Icos* and negatively correlated with quiescence-associated genes such as *Sell* and *Il7r* (Figures 5B and 5C). This suggested that the Tfh archetype provides a quantitative measure of Tfh cell differentiation, or the strength of the Tfh phenotype, within the CD4 Tfh population. Furthermore, we observed significant positive and negative correlations between Tfh archetypal weight and the expression of numerous functionally important T cell genes (Figures 5C and S13A). These included T cell receptor signaling genes *Zap70* and *Ptpn11*, co-stimulatory receptors *Icos*, *Tnfrsf4* and *Tnfsf8*, inhibitory molecules *Pdcd1*, *Tnfaip8* and *Nt5e*, and transcription factors *Bcl6*, *Tox*, *Hif1a*, *Ikzf2*, *Pou2f2*, *Id3*, *Nfatc1*, *Nfat5* and *Batf*. Consistently, the Tfh archetype was also correlated with the key TF motifs including

(D) Correlation of Tfh archetypal gradient with TF motifs. Heatmap showing Spearman correlations between CD4 Tfh archetypal weights and TF motif scores within progenitor CD8 T cells and CD4 Tfh cells for selected motifs.
(E) Tfh archetypal gradient within CD8 progenitors. UMAP showing the Tfh archetypal weights in the progenitor CD8 T cell population (9 outlier cells not shown).
(F) Gene expression along Tfh archetypal gradient within CD8 progenitors. Scatterplots of normalized expression of selected genes in progenitor CD8 T cells along the Tfh archetypal gradient. The gene expression values were averaged over neighbors in the scRNA-seq kNN graph ($k = 20$). A smoothed fit using ordinary least squares is shown in gray. Spearman's $\rho$ is shown on top.
(G) Comparison of Tfh and CD8 exhaustion archetypes. Heatmap comparing the Spearman correlations between Tfh and CD8 exhaustion archetypal weights calculated within three cell populations.

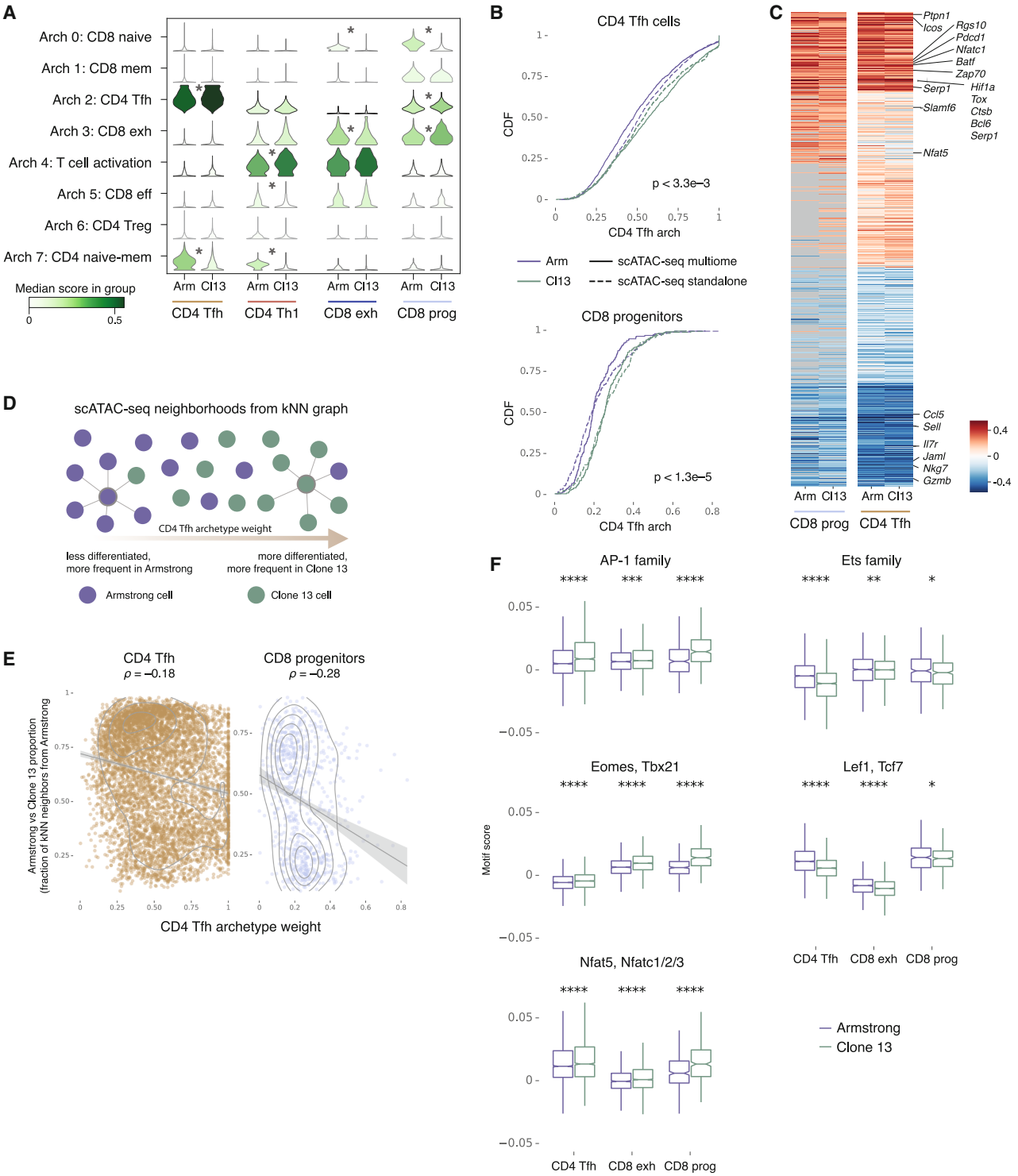



**Figure 6. Clone 13 infection accelerates progression along the Tfh differentiation gradient in CD4 Tfh and progenitor CD8 T cells**

(A) Comparison of scATAC-seq archetypes between LCMV Arm and Cl13. Violin plots show the distributions of archetypal weights for selected T cell subsets, separately in Arm and Cl13 infections. Significant differences between infections within a population are marked with asterisks (FDR $\leq$0.05, KS test; absolute median difference $\geq$0.05).

(B) Tfh archetypal gradient in Arm and Cl13. CDF curves showing the distribution of the Tfh archetypal weights for Arm and Cl13 cells in the progenitor CD8 T cell population (*top*) and the CD4 Tfh cell population (*bottom*), separately for standalone and multiome scATAC-seq data. Two-sided KS test *p*-value listed (after Benjamini-Hochberg correction) for all pairwise comparisons across all batches between Armstrong and Clone 13 cells.

*(legend continued on next page)*

Nfat, Lef1/Tcf1 and AP-1 family motifs (Figures 5D and S13F). These associations further define the functional spectrum and regulatory drivers of the Tfh archetype and support its interpretation as a measure of progressive TCR-stimulated Tfh cell activation, differentiation, and migration, enabling B cell help, while inhibiting proliferative expansion typical of effector cells.[11] These results illustrate that scATAC-seq archetypal analysis can effectively capture a continuum of differentiation within the CD4 Tfh cell population.

Given the prominent contribution of the CD4 Tfh archetype to the epigenomic phenotype of progenitor CD8 T cells (Figures 3E,3F, 4B, and 4C), next we wanted to quantitatively explore the Tfh archetype within the progenitor population. Strikingly, we found that the progenitor CD8 T cells exhibited a continuous gradient of Tfh archetypal weights, similar to that observed within CD4 Tfh cells (Figures 5C–5F, S13A, S13C, and S13F). In particular, many genes showed comparable correlations with the Tfh archetype in both the CD4 Tfh population and the CD8 progenitor population, reinforcing the epigenomic similarity between the two populations.

We also examined the contribution of the CD8 exhaustion archetype to the progenitor population. The exhaustion archetype was uncorrelated with the Tfh archetype both across the full dataset and within individual cell states, including Tfh cells, exhausted CD8 T cells and progenitor CD8 T cells (Figures 5G and S13E). Some key T cell genes, most notably the Tfh-related but also exhaustion-associated genes *Tox* and *Pdcd1*, showed similarly high expression correlations with both archetypes, particularly within progenitor cells (Figures 5C and S13E). This was consistent with the functional role of progenitors as activated cells that must balance responsiveness with restraint of proliferation. Nevertheless, most genes most correlated with the Tfh and CD8 exhaustion archetypes were distinct (Figures S13E and S13G), indicating that the two archetypes capture largely non-overlapping regulatory programs.

Together, these findings suggest that the progenitor CD8 T cell state reflects an integration of regulatory inputs from both Tfh- and exhaustion-like programs. Using scATAC-seq archetypal analysis, we reveal a Tfh-like gradient of differentiation within progenitor CD8 T cells, supporting the view that these cells, similar to CD4 Tfh cells, adopt an activated identity while avoiding full effector-like differentiation and proliferation.

## Clone 13 infection accelerates progression along the Tfh differentiation gradient in CD4 Tfh and progenitor CD8 T cells

While we have consistently observed similar T cell states and phenotypes in both LCMV Arm and Cl13 infections, the immune responses to these clones ultimately diverge over time. To investigate the basis of these differences, we examined how epigenomic and regulatory programs vary between the two infections. We focused particularly on the shared population of progenitor CD8 T cells, as they can give rise to both effector-like and terminally exhausted CD8 T cells and are therefore central to understanding how T cell response evolves in each context.

We found that within the progenitor CD8 T cell population, the Tfh and CD8 exhaustion archetypes were significantly higher in the Cl13 response compared to Arm, whereas the CD8 naive archetype was significantly higher in Arm (Figures 6A, 6B, and S9C). Similar trends—stronger activation in Clone 13 and a more naive-like state in Armstrong—were also observed in other antigen-experienced T cell populations (Figures 6A, 6B, and S9C). This was consistent with prior reports that the overall T cell response to Cl13 is more strongly activated than to Arm and ultimately progresses toward exhaustion as a protective mechanism to limit immunopathology.[8,9]

We then examined more closely the Tfh-like differentiation trajectory within progenitor CD8 T cells that we identified (Figure 5). A similar Tfh differentiation gradient was observed in both Arm and Cl13 responses, within both CD4 Tfh and progenitor CD8 T cells, as evidenced by comparable gene expression correlations (Figures 6C and S13A). However, consistent with the overall stronger T cell activation observed in Cl13, cells from the Cl13 response progressed more rapidly along this gradient than those from Arm (Figures 6B, 6D, 6E). This finding was robust across the two scATAC-seq profiling technologies (standalone and multiome). TF motif analysis implicated potential regulators of these differential responses between Arm and Cl13. In particular, activation-associated motifs AP-1, Tbet and Nfat showed stronger signal in Cl13, while motifs associated with naive and memory-like states, such as Lef1/Tcf7 and Ets motifs, were more enriched in Arm (Figures 6F and S14E). These observations were confirmed in reanalysis of published scATAC-seq datasets (Figures S14A–S14D). Thus, we conclude that although progenitor CD8 T cells exhibit a similar Tfh-like differentiation gradient in both Arm and Cl13 responses, they advance along this gradient more rapidly in Cl13 infection. This suggests a potential for

(C) Correlation of Tfh archetype with gene expression in Arm and Cl13. Heatmap showing the correlations between the Tfh archetypal weights and normalized gene expression in progenitor CD8 T cells and CD4 Tfh cells, separately for Arm and Cl13 infections. Genes with the strongest positive or negative correlations between exhausted CD8 T cells and the CD8 exhaustion archetype, CD4 Tfh cells and Tfh archetype, and CD8 progenitors and either archetype were selected (total 1,131 genes, same as in Figure S13G).

(D) Schematic of the comparison of cells along the Tfh archetypal gradient between Arm and Cl13 infections. For each cell, the fraction of neighbors (i.e., most similar cells by scATAC-seq profile) from each infection is computed based on the scATAC-seq kNN graph constructed for all samples. We propose that cells from Clone 13 infection progress more rapidly along the Tfh differentiation trajectory as captured by the Tfh archetypal weights.

(E) Tfh archetypal gradient vs. Arm-to-Cl13 cell proportion. Scatterplot of the Tfh archetypal weights (*x* axis) and the Arm vs. Cl13 proportions (*y* axis) in CD4 Tfh cells (*left*) and progenitor CD8 T cells (*right*). The Arm vs. Cl13 proportion is calculated as the fraction of cell neighbors in the scATAC-seq kNN graph ($k = 100$) that belong to the Armstrong samples, and then smoothed over 30 nearest scATAC-seq neighbors.

(F) Comparison of TF motifs between Arm and Cl13. Distribution of selected TF motif scores for different T cell populations, colored by infection (Arm vs. Cl13). Asterisks in the plot represent statistical significance (two-sided KS test, after BH correction) of the difference between the two infection types within a population (*: FDR <0.2, **: FDR <0.05, ***: FDR <0.01, ****: FDR <0.001). Boxplots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range; outliers not shown.

further divergence of the T cell response upon persistent antigen stimulation.

## DISCUSSION

In this study, we integrated single-cell chromatin accessibility and transcriptomic profiling with a comprehensive reference of T cell regulatory regions to analyze CD4 and CD8 T cell responses to acute and chronic viral infection. Our approach enabled high-resolution, interpretable identification of all major T cell states, including naive, memory, effector, Tfh, regulatory, exhausted, and progenitor populations. We observed that each population was present in both infection types, but at different frequencies. This indicates that the immunological divergence between acute and chronic viral responses is fundamentally quantitative rather than qualitative, driven by shifts in the relative abundance of similar cell states rather than the emergence of response-specific cell identities.

We focused in particular on the clinically relevant progenitor CD8 T cell population. While the relationship between progenitor populations in acute and chronic infection has been a key area of investigation,[20,26,28,29,49–51,55] our analysis demonstrated that these cells form a highly similar population at the transcriptomic and epigenomic levels between the two infection settings. Surprisingly, we found that these CD8 progenitors exhibit striking epigenomic similarity with CD4 Tfh cells, a population with a key role in B cell help via direct physical interactions in the B cell follicles, and defined by sustained TCR signaling and restrained effector differentiation. Despite arising from the distinct lineages of CD4 and CD8 T cells, both populations share chromatin accessibility and transcriptomic profiles and co-activate transcription factors such as Tcf1, Bcl6, Nfat family factors and Tox.

To dissect this convergence, we applied archetypal analysis to the chromatin accessibility data, revealing that progenitor CD8 T cells do not represent a unique regulatory program, but rather a composite of CD8 exhaustion and Tfh archetypes. This supports a modular view of progenitor identity, arising from reuse of regulatory programs deployed in other T cell states. This "molecular hedging" likely confers a survival advantage, allowing progenitors to persist as a self-renewing reservoir by borrowing the follicular program's longevity to resist rapid terminal exhaustion. This modular state maintains a functional population that can provide effector bursts without the metabolic collapse associated with terminal states.

Concordantly, we observed that progenitor CD8 T cells combine the enrichment of TF motifs preferentially associated with Tfh state, including Lef1, Tcf1, Pou2f2, and the Nfat family, with motifs preferentially linked to CD8 exhaustion such as Runx and Tbet. While AP-1 and Nfat are canonical hallmarks of general T cell activation downstream of TCR signaling, our single-cell and allele-specific analyses capture the relative, nuanced differences in how the corresponding motifs functionally shape local chromatin accessibility across distinct contexts. Our analysis suggests that Nfat family factors, triggering Tox and PD-1 activation, may be involved in functional adaptations characteristic of both Tfh cells and CD8 progenitors, such as restrained effector differentiation immediately upon and later during persistent TCR stimulation. Activity of other factors, including AP-1 family factors, e.g., Batf, may be associated with a functional activation program shared by all antigen-experienced CD4 and CD8 T cells that is progressively lost during differentiation toward exhaustion, whereas Tbet and Runx factors may be associated with the preparation for and execution of cytotoxic effector functions. This view is consistent with and helps further refine known functionality of these factors in T cells,[20,70–77] though more specific studies of these and other factors are needed to further dissect combinatorial mechanistic regulatory programs across T cell populations.

These findings suggest possible functional and spatial links between progenitor CD8 T cells and Tfh cells. Future studies should explore whether they co-localize and interact, and thus converge to activation of similar regulatory programs despite originating from different T cell lineages. Consistently, a similar subset of CD8 T cells expressing CXCR5, exhausted genes, and memory-like genes has been discovered to localize in the B cell follicles, i.e., along with Tfh cells, but was originally defined as a less-exhausted, follicular CD8 T cell population.[78–82] Our results suggest that progenitor and follicular CD8 T cells may be the same or similar populations. Indeed, CD4 T cells and B cells influence the progenitor CD8 T cells during infection, though it remains unclear whether CD4 T cell help directly affects the formation or maintenance of the CD8 progenitors.[25,83–86] However, progenitor CD8 T cells were reported to localize in T cell zones, away from B cell follicles, at later stages of chronic infection, suggesting potential heterogeneity within the progenitor population or a dependence of spatial localization on the stage of infection progression.[60,75,87,88] Further research using multi-omics, spatial transcriptomics and proteomics, and direct cell-cell interaction measurements such as uLIPSTIC,[89] across time courses and various systems of T cell stimulation and response, could help better characterize the T cell phenotypic landscape and explore the localization and interactions of Tfh cells and progenitor and follicular CD8 cells. Perturbation experiments, including targeted depletion of one population or blockade of shared signaling pathways or transcriptional regulators, could further clarify their relationship.

In sum, our work highlights the power of multiomic and computational approaches to dissect complex T cell regulation and offers a generalizable framework for studying shared and distinct programs across immune contexts. Applying this strategy to other T cell settings, such as tumor-infiltrating lymphocytes, vaccine responses, or autoimmune diseases, could uncover novel epigenomic states and reveal common principles of immune cell plasticity. These approaches will also be broadly applicable across a range of biological contexts beyond immunology.

### Limitations of the study

While our single-cell multiomic approach provides a high-resolution view of T cell regulatory architecture, a limitation of this study is its reliance on a single snapshot at day 7 post-infection. Consequently, our data cannot fully capture temporal dynamics across the complete course of infection, missing both earlier windows critical for dissecting initial T cell priming and the establishment of the progenitor population, and later timepoints to evaluate long-term chronic antigen persistence. Additionally, our analysis identified a small, reproducible population of cells

(2.5% of total cells) exhibiting mixed lineage markers that precluded definitive assignment to major CD4 or CD8 subsets, which may represent rare transitional or innate-like populations that warrant further targeted study. Finally, our findings linking the primary regulatory programs with functional T cell states remain associative. Future spatial profiling, targeted perturbation studies across an expanded time course, and focused characterization of these non-canonical phenotypes will be required to definitively map their physical context and validate the causal roles of the implicated transcription factors.

## RESOURCE AVAILABILITY

### Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Yuri Pritykin (pritykin@princeton.edu).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

- Data were deposited in NCBI GEO under the accession number GSE300984 and are publicly available as of the date of publication. A part of the dataset (standalone scATAC-seq for LCMV Armstrong and multi-omics for antigen-activated CD8 T cells) was previously reported.[31,90] We also used existing, publicly available data, obtained from NCBI GEO at accession numbers GSE188670 and GSE199565.
- All original code was deposited in Zenodo (https://doi.org/10.5281/zenodo.22821709) and is publicly available as of the date of publication. The code used to produce the analysis presented in this study is also available on GitHub (https://github.com/pritykinlab/multimodal_tcells).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## ACKNOWLEDGMENTS

We thank all members of the Pritykin lab for helpful discussions. This work was supported by the NIH/NIAID grant DP2AI171161, AACR-Bristol-Myers Squibb Immuno-oncology Research Fellowship (19-40-15-PRIT), Ludwig Institute for Cancer Research and Weill Cancer Hub East (Y.P.); Boehringer Ingelheim, the Austrian Science Fund (FWF, 10.55776/PAT4163824 and 10.55776/PAT2786525), and the ERC grant ERC-2023-STG 101116251 (J.v.d.V); NIH grants P30 CA008748 and R01 AI034206 (A.Y.R.). A.Y.R. is an investigator with the Howard Hughes Medical Institute. We thank the Single Cell Analytics Innovation Lab at MSKCC and the head of the lab, Ronan Chaligné, for processing single-cell multi-omics samples.

## AUTHOR CONTRIBUTIONS

Conceptualization, S.K.W., J.v.d.V., and Y.P.; methodology, S.K.W., J.v.d.V., A.Y.R. and Y.P.; investigation, S.K.W. and J.v.d.V.; software, S.K.W.; writing – original draft, S.K.W. and Y.P.; writing – review and editing, S.K.W., J.v.d.V., A.Y.R., and Y.P.; funding acquisition, Y.P., J.v.d.V., and A.Y.R; supervision, A.Y.R. and Y.P.

## DECLARATION OF INTERESTS

A.Y.R. is an SAB member and has equity in Sonoma Biotherapeutics, RAPT Therapeutics, Coherus BioSciences, Santa Ana Bio, Odyssey Therapeutics, Nilo Therapeutics, and Vedanta Biosciences; he is also an SAB member of BioInvent and Amgen and a co-inventor of a CCR8+ Treg cell depletion IP licensed to Takeda, which is unrelated to the content of this publication. The remaining authors declare no competing interests.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Library preparation and sequencing
  - Read alignment
  - ATAC-seq peak atlas for T cells
  - scATAC-seq data analysis
  - ATAC-seq peak summits and TF motif analysis
  - Allele-specific analysis
  - scRNA-seq data analysis
  - Archetypal analysis
  - Analysis of other datasets
- QUANTIFICATION AND STATISTICAL ANALYSIS

## SUPPLEMENTAL INFORMATION

Supplementary data related to this article can be found online at https://doi.org/10.1016/j.celrep.2026.118156.

Received: September 25, 2025
Revised: July 22, 2026
Accepted: September 28, 2026

## REFERENCES

1. Kumar, B.V., Connors, T.J., and Farber, D.L. (2018). Human T cell development, localization, and function throughout life. Immunity *48*, 202–213. https://doi.org/10.1016/j.immuni.2018.01.007.
2. Fan, X., and Rudensky, A.Y. (2016). Hallmarks of tissue-resident lymphocytes. Cell *164*, 1198–1211. https://doi.org/10.1016/j.cell.2016.02.048.
3. Chi, H., Pepper, M., and Thomas, P.G. (2024). Principles and therapeutic applications of adaptive immunity. Cell *187*, 2052–2078. https://doi.org/10.1016/j.cell.2024.03.037.
4. Schenkel, J.M., and Pauken, K.E. (2023). Localization, tissue biology and T cell state—implications for cancer immunotherapy. Nat. Rev. Immunol. *23*, 807–823. https://doi.org/10.1038/s41577-023-00884-8.
5. Masopust, D., Awasthi, A., Bosselut, R., Brooks, D.G., Buggert, M., Chamoto, K., Cui, W., Dong, C., Farber, D.L., Gebhardt, T., et al. (2026). Guidelines for T cell nomenclature. Nat. Rev. Immunol. *26*, 298–313. https://doi.org/10.1038/s41577-025-01238-2.
6. Ruterbusch, M., Pruner, K.B., Shehata, L., and Pepper, M. (2020). In vivo CD4+ T cell differentiation and function: Revisiting the Th1/Th2 paradigm. Annu. Rev. Immunol. *38*, 705–725. https://doi.org/10.1146/annurev-immunol-103019-085803.
7. Miggelbrink, A.M., Jackson, J.D., Lorrey, S.J., Srinivasan, E.S., Waibl-Polania, J., Wilkinson, D.S., and Fecci, P.E. (2021). CD4 T-cell exhaustion: Does it exist and what are its roles in cancer? Clin. Cancer Res. *27*, 5742–5752. https://doi.org/10.1158/1078-0432.CCR-21-0206.
8. McLane, L.M., Abdel-Hakeem, M.S., and Wherry, E.J. (2019). CD8 T cell exhaustion during chronic viral infection and cancer. Annu. Rev. Immunol. *37*, 457–495. https://doi.org/10.1146/annurev-immunol-041015-055318.
9. Hashimoto, M., Kamphorst, A.O., Im, S.J., Kissick, H.T., Pillai, R.N., Ramalingam, S.S., Araki, K., and Ahmed, R. (2018). CD8 T cell exhaustion in chronic infection and cancer: Opportunities for interventions. Annu. Rev. Med. *69*, 301–318. https://doi.org/10.1146/annurev-med-012017-043208.

10. Kallies, A., Zehn, D., and Utzschneider, D.T. (2020). Precursor exhausted T cells: key to successful immunotherapy? Nat. Rev. Immunol. *20*, 128–136. https://doi.org/10.1038/s41577-019-0223-7.

11. Crotty, S. (2014). T follicular helper cell differentiation, function, and roles in disease. Immunity *41*, 529–542. https://doi.org/10.1016/j.immuni.2014.10.004.

12. Kim, Y., Greenleaf, W.J., and Bendall, S.C. (2023). Systems biology approaches to unravel lymphocyte subsets and function. Curr. Opin. Immunol. *82*, 102323. https://doi.org/10.1016/j.coi.2023.102323.

13. Efremova, M., Vento-Tormo, R., Park, J.E., Teichmann, S.A., and James, K.R. (2020). Immunology in the era of single-cell technologies. Annu. Rev. Immunol. *38*, 727–757. https://doi.org/10.1146/annurev-immunol-090419-020340.

14. Giladi, A., and Amit, I. (2018). Single-cell genomics: a stepping stone for future immunology discoveries. Cell *172*, 14–21. https://doi.org/10.1016/j.cell.2017.11.011.

15. Kiner, E., Willie, E., Vijaykumar, B., Chowdhary, K., Schmutz, H., Chandler, J., Schnell, A., Thakore, P.I., LeGros, G., Mostafavi, S., et al. (2021). Gut CD4+ T cell phenotypes are a continuum molded by microbes, not by TH archetypes. Nat. Immunol. *22*, 216–228. https://doi.org/10.1038/s41590-020-00836-7.

16. Wagner, A., Regev, A., and Yosef, N. (2016). Revealing the vectors of cellular identity with single-cell genomics. Nat. Biotechnol. *34*, 1145–1160. https://doi.org/10.1038/nbt.3711.

17. Preissl, S., Gaulton, K.J., and Ren, B. (2023). Characterizing cis-regulatory elements using single-cell epigenomics. Nat. Rev. Genet. *24*, 21–43. https://doi.org/10.1038/s41576-022-00509-1.

18. Eizenberg-Magar, I., Rimer, J., Zaretsky, I., Lara-Astiaso, D., Reich-Zeliger, S., and Friedman, N. (2017). Diverse continuum of CD4+ T-cell states is determined by hierarchical additive integration of cytokine signals. Proc. Natl. Acad. Sci. USA *114*, E6447–E6456. https://doi.org/10.1073/pnas.1615590114.

19. Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat. Genet. *48*, 1193–1203. https://doi.org/10.1038/ng.3646.

20. Chung, H.K., McDonald, B., and Kaech, S.M. (2021). The architectural design of CD8+ T cell responses in acute and chronic infection: Parallel structures with divergent fates. J. Exp. Med. *218*, e20201730. https://doi.org/10.1084/jem.20201730.

21. Badovinac, V.P., Porter, B.B., and Harty, J.T. (2002). Programmed contraction of CD8+ T cells after infection. Nat. Immunol. *3*, 619–626. https://doi.org/10.1038/ni804.

22. Walton, S., Mandaric, S., and Oxenius, A. (2013). CD4 T cell responses in latent and chronic viral infections. Front. Immunol. *4*, 105. https://doi.org/10.3389/fimmu.2013.00105.

23. Andreatta, M., Tjitropranoto, A., Sherman, Z., Kelly, M.C., Ciucci, T., and Carmona, S.J. (2022). A CD4+ T cell reference map delineates subtype-specific adaptation during acute and chronic viral infections. eLife *11*, e76339. https://doi.org/10.7554/elife.76339.

24. Ciucci, T., Vacchio, M.S., Gao, Y., Tomassoni Ardori, F., Candia, J., Mehta, M., Zhao, Y., Tran, B., Pepper, M., Tessarollo, L., et al. (2019). The emergence and functional fitness of memory CD4+ T cells require the transcription factor thpok. Immunity *50*, 91–105.e4. https://doi.org/10.1016/j.immuni.2018.12.019.

25. Zander, R., Kasmani, M.Y., Chen, Y., Topchyan, P., Shen, J., Zheng, S., Burns, R., Ingram, J., Cui, C., Joshi, N., et al. (2022). Tfh-cell-derived interleukin 21 sustains effector CD8+ T cell responses during chronic viral infection. Immunity *55*, 475–493.e5. https://doi.org/10.1016/j.immuni.2022.01.018.

26. Pritykin, Y., van der Veeken, J., Pine, A.R., Zhong, Y., Sahin, M., Mazutis, L., Pe'er, D., Rudensky, A.Y., and Leslie, C.S. (2021). A unified atlas of CD8 T cell dysfunctional states in cancer and infection. Mol. Cell *81*, 2477–2493.e10. https://doi.org/10.1016/j.molcel.2021.03.045.

27. Green, W.D., Gomez, A., Plotkin, A.L., Pratt, B.M., Merritt, E.F., Mullins, G.N., Kren, N.P., Modliszewski, J.L., Zhabotynsky, V., Woodcock, M.G., et al. (2025). Enhancer-driven gene regulatory networks reveal transcription factors governing t cell adaptation and differentiation in the tumor microenvironment. Immunity *58*, 1725–1741.e9. https://doi.org/10.1016/j.immuni.2025.04.030.

28. Daniel, B., Yost, K.E., Hsiung, S., Sandor, K., Xia, Y., Qi, Y., Hiam-Galvez, K.J., Black, M., J. Raposo, C., Shi, Q., et al. (2022). Divergent clonal differentiation trajectories of T cell exhaustion. Nat. Immunol. *23*, 1614–1627. https://doi.org/10.1038/s41590-022-01337-5.

29. Giles, J.R., Ngiow, S.F., Manne, S., Baxter, A.E., Khan, O., Wang, P., Staupe, R., Abdel-Hakeem, M.S., Huang, H., Mathew, D., et al. (2022). Shared and distinct biological circuits in effector, memory and exhausted CD8+ T cells revealed by temporal single-cell transcriptomics and epigenetics. Nat. Immunol. *23*, 1600–1613. https://doi.org/10.1038/s41590-022-01338-4.

30. Ucar, D., Márquez, E.J., Chung, C.H., Marches, R., Rossi, R.J., Uyar, A., Wu, T.C., George, J., Stitzel, M.L., Palucka, A.K., et al. (2017). The chromatin accessibility signature of human immune aging stems from CD8+ T cells. J. Exp. Med. *214*, 3123–3144. https://doi.org/10.1084/jem.20170416.

31. Zhong, Y., Walker, S.K., Pritykin, Y., Leslie, C.S., Rudensky, A.Y., and van der Veeken, J. (2022). Hierarchical regulation of the resting and activated T cell epigenome by major transcription factor families. Nat. Immunol. *23*, 122–134. https://doi.org/10.1038/s41590-021-01086-x.

32. Persad, S., Choo, Z.N., Dien, C., Sohail, N., Masilionis, I., Chaligné, R., Nawy, T., Brown, C.C., Sharma, R., Pe'er, I., et al. (2023). SEACells infers transcriptional and epigenomic cellular states from single-cell genomics data. Nat. Biotechnol. *41*, 1746–1757. https://doi.org/10.1038/s41587-023-01716-9.

33. Cook, D.P., and Vanderhyden, B.C. (2022). Transcriptional census of epithelial-mesenchymal plasticity in cancer. Sci. Adv. *8*, eabi7640. https://doi.org/10.1126/sciadv.abi7640.

34. Wei, S.C., Sharma, R., Anang, N.A.A.S., Levine, J.H., Zhao, Y., Mancuso, J.J., Setty, M., Sharma, P., Wang, J., Pe'er, D., and Allison, J.P. (2019). Negative co-stimulation constrains T cell differentiation by imposing boundaries on possible cell states. Immunity *50*, 1084–1098.e10. https://doi.org/10.1016/j.immuni.2019.03.004.

35. Sade-Feldman, M., Yizhak, K., Bjorgaard, S.L., Ray, J.P., de Boer, C.G., Jenkins, R.W., Lieb, D.J., Chen, J.H., Frederick, D.T., Barzily-Rokni, M., et al. (2018). Defining T cell states associated with response to checkpoint immunotherapy in melanoma. Cell *175*, 998–1013.

36. Blank, C.U., Haining, W.N., Held, W., Hogan, P.G., Kallies, A., Lugli, E., Lynn, R.C., Philip, M., Rao, A., Restifo, N.P., et al. (2019). Defining T cell exhaustion. Nat. Rev. Immunol. *19*, 665–674. https://doi.org/10.1038/s41577-019-0221-9.

37. Miller, B.C., Sen, D.R., Al Abosy, R., Bi, K., Virkud, Y.V., LaFleur, M.W., Yates, K.B., Lako, A., Felt, K., Naik, G.S., et al. (2019). Subsets of exhausted CD8+ T cells differentially mediate tumor control and respond to checkpoint blockade. Nat. Immunol. *20*, 326–336. https://doi.org/10.1038/s41590-019-0312-6.

38. Siddiqui, I., Schaeuble, K., Chennupati, V., Fuertes Marraco, S.A., Calderon-Copete, S., Pais Ferreira, D., Carmona, S.J., Scarpellino, L., Gfeller, D., Pradervand, S., et al. (2019). Intratumoral Tcf1+ PD-1+ CD8+ T cells with stem-like properties promote tumor control in response to vaccination and checkpoint blockade immunotherapy. Immunity *50*, 195–211.e10. https://doi.org/10.1016/j.immuni.2018.12.021.

39. Fang, Z., Ding, X., Huang, H., Jiang, H., Jiang, J., and Zheng, X. (2024). Revolutionizing tumor immunotherapy: unleashing the power of progenitor exhausted T cells. Cancer Biol. Med. *21*, 499–512. https://doi.org/10.20892/j.issn.2095-3941.2024.0105.

40. Giles, J.R., Globig, A.M., Kaech, S.M., and Wherry, E.J. (2023). CD8+ T cells in the cancer-immunity cycle. Immunity *56*, 2231–2253. https://doi.org/10.1016/j.immuni.2023.09.005.

41. Mognol, G.P., Spreafico, R., Wong, V., Scott-Browne, J.P., Togher, S., Hoffmann, A., Hogan, P.G., Rao, A., and Trifari, S. (2017). Exhaustion-associated regulatory regions in CD8+ tumor-infiltrating T cells. Proc. Natl. Acad. Sci. USA *114*, E2776–E2785. https://doi.org/10.1073/pnas.1620498114.

42. Scharer, C.D., Bally, A.P.R., Gandham, B., and Boss, J.M. (2017). Cutting edge: Chromatin accessibility programs CD8 T cell memory. J. Immunol. *198*, 2238–2243. https://doi.org/10.4049/jimmunol.1602086.

43. Sen, D.R., Kaminski, J., Barnitz, R.A., Kurachi, M., Gerdemann, U., Yates, K.B., Tsao, H.W., Godec, J., LaFleur, M.W., Brown, F.D., et al. (2016). The epigenetic landscape of T cell exhaustion. Science *354*, 1165–1169. https://doi.org/10.1126/science.aae0491.

44. Ma, S., Zhang, B., LaFave, L.M., Earl, A.S., Chiang, Z., Hu, Y., Ding, J., Brack, A., Kartha, V.K., Tay, T., et al. (2020). Chromatin potential identified by shared single-cell profiling of rna and chromatin. Cell *183*, 1103–1116.e20. https://doi.org/10.1016/j.cell.2020.09.056.

45. González, A.J., Setty, M., and Leslie, C.S. (2015). Early enhancer establishment and regulatory locus complexity shape transcriptional programs in hematopoietic differentiation. Nat. Genet. *47*, 1249–1259. https://doi.org/10.1038/ng.3402.

46. Pauken, K.E., Sammons, M.A., Odorizzi, P.M., Manne, S., Godec, J., Khan, O., Drake, A.M., Chen, Z., Sen, D.R., Kurachi, M., et al. (2016). Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. Science *354*, 1160–1165. https://doi.org/10.1126/science.aaf2807.

47. Satpathy, A.T., Granja, J.M., Yost, K.E., Qi, Y., Meschi, F., McDermott, G.P., Olsen, B.N., Mumbach, M.R., Pierce, S.E., Corces, M.R., et al. (2019). Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat. Biotechnol. *37*, 925–936. https://doi.org/10.1038/s41587-019-0206-z.

48. Fagerberg, E., Attanasio, J., Dien, C., Singh, J., Kessler, E.A., Abdullah, L., Shen, J., Hunt, B.G., Connolly, K.A., De Brouwer, E., et al. (2025). KLF2 maintains lineage fidelity and suppresses CD8 T cell exhaustion during acute LCMV infection. Science *387*, eadn2337. https://doi.org/10.1126/science.adn2337.

49. Chu, T., Wu, M., Hoellbacher, B., de Almeida, G.P., Wurmser, C., Berner, J., Donhauser, L.V., Gerullis, A.K., Lin, S., Cepeda-Mayorga, J.D., et al. (2025). Precursors of exhausted T cells are pre-emptively formed in acute infection. Nature *640*, 782–792. https://doi.org/10.1038/s41586-024-08451-4.

50. McManus, D.T., Valanparambil, R.M., Medina, C.B., Scharer, C.D., McGuire, D.J., Sobierajska, E., Hu, Y., Chang, D.Y., Wieland, A., Lee, J., et al. (2025). An early precursor CD8+ T cell that adapts to acute or chronic viral infection. Nature *640*, 772–781. https://doi.org/10.1038/s41586-024-08562-y.

51. Gago da Graça, C., Sheikh, A.A., Newman, D.M., Wen, L., Li, S., Shen, J., Zhang, Y., Gabriel, S.S., Chisanga, D., Seow, J., et al. (2025). Stem-like memory and precursors of exhausted T cells share a common progenitor defined by ID3 expression. Sci. Immunol. *10*, eadn1945. https://doi.org/10.1126/sciimmunol.adn1945.

52. Joshi, N.S., Cui, W., Chandele, A., Lee, H.K., Urso, D.R., Hagman, J., Gapin, L., and Kaech, S.M. (2007). Inflammation directs memory precursor and short-lived effector CD8+ T cell fates via the graded expression of T-bet transcription factor. Immunity *27*, 281–295. https://doi.org/10.1016/j.immuni.2007.07.010.

53. Kaech, S.M., and Cui, W. (2012). Transcriptional control of effector and memory CD8+ T cell differentiation. Nat. Rev. Immunol. *12*, 749–761. https://doi.org/10.1038/nri3307.

54. Yang, C.Y., Best, J.A., Knell, J., Yang, E., Sheridan, A.D., Jesionek, A.K., Li, H.S., Rivera, R.R., Lind, K.C., D'cruz, L.M., et al. (2011). The transcriptional regulators Id2 and Id3 control the formation of distinct memory CD8+ T cell subsets. Nat. Immunol. *12*, 1221–1229. https://doi.org/10.1038/ni.2158.

55. Utzschneider, D.T., Gabriel, S.S., Chisanga, D., Gloury, R., Gubser, P.M., Vasanthakumar, A., Shi, W., and Kallies, A. (2020). Early precursor T cells establish and propagate T cell exhaustion in chronic infection. Nat. Immunol. *21*, 1256–1266. https://doi.org/10.1038/s41590-020-0760-z.

56. Chen, J., López-Moyado, I.F., Seo, H., Lio, C.W.J., Hempleman, L.J., Sekiya, T., Yoshimura, A., Scott-Browne, J.P., and Rao, A. (2019). NR4A transcription factors limit CAR T cell function in solid tumours. Nature *567*, 530–534. https://doi.org/10.1038/s41586-019-0985-x.

57. Szabo, P.A., Levitin, H.M., Miron, M., Snyder, M.E., Senda, T., Yuan, J., Cheng, Y.L., Bush, E.C., Dogra, P., Thapa, P., et al. (2019). Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. Nat. Commun. *10*, 4706. https://doi.org/10.1038/s41467-019-12464-3.

58. Singer, A., Adoro, S., and Park, J.H. (2008). Lineage fate and intense debate: myths, models and mechanisms of CD4-versus CD8-lineage choice. Nat. Rev. Immunol. *8*, 788–801. https://doi.org/10.1038/nri2416.

59. Steier, Z., Aylard, D.A., McIntyre, L.L., Baldwin, I., Kim, E.J.Y., Lutes, L.K., Ergen, C., Huang, T.S., Robey, E.A., Yosef, N., et al. (2023). Single-cell multiomic analysis of thymocyte development reveals drivers of CD4+ T cell and CD8+ T cell lineage commitment. Nat. Immunol. *24*, 1579–1590. https://doi.org/10.1038/s41590-023-01584-0.

60. Im, S.J., Hashimoto, M., Gerner, M.Y., Lee, J., Kissick, H.T., Burger, M.C., Shan, Q., Hale, J.S., Lee, J., Nasti, T.H., et al. (2016). Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. Nature *537*, 417–421. https://doi.org/10.1038/nature19330.

61. Morup, M., and Hansen, L.K. (2012). Archetypal analysis for machine learning and data mining. Neurocomputing *80*, 54–63.

62. Cutler, A., and Breiman, L. (1994). Archetypal analysis. Technometrics *36*, 338–347. https://doi.org/10.1080/00401706.1994.10485840.

63. Damle, A., and Sun, Y. (2017). A geometric approach to archetypal analysis and nonnegative matrix factorization. Technometrics *59*, 361–370. https://doi.org/10.1080/00401706.2016.1247017.

64. Niu, H., and Wang, H. (2023). TOX regulates T lymphocytes differentiation and its function in tumor. Front. Immunol. *14*, 990419. https://doi.org/10.3389/fimmu.2023.990419.

65. Khan, O., Giles, J.R., McDonald, S., Manne, S., Ngiow, S.F., Patel, K.P., Werner, M.T., Huang, A.C., Alexander, K.A., Wu, J.E., et al. (2019). TOX transcriptionally and epigenetically programs CD8+ T cell exhaustion. Nature *571*, 211–218. https://doi.org/10.1038/s41586-019-1325-x.

66. Alfei, F., Kanev, K., Hofmann, M., Wu, M., Ghoneim, H.E., Roelli, P., Utzschneider, D.T., von Hoesslin, M., Cullen, J.G., Fan, Y., et al. (2019). TOX reinforces the phenotype and longevity of exhausted T cells in chronic viral infection. Nature *571*, 265–269. https://doi.org/10.1038/s41586-019-1326-9.

67. Yao, C., Sun, H.W., Lacey, N.E., Ji, Y., Moseman, E.A., Shih, H.Y., Heuston, E.F., Kirby, M., Anderson, S., Cheng, J., et al. (2019). Single-cell rna-seq reveals TOX as a key regulator of CD8+ T cell persistence in chronic infection. Nat. Immunol. *20*, 890–901. https://doi.org/10.1038/s41590-019-0403-4.

68. Scott, A.C., Dündar, F., Zumbo, P., Chandran, S.S., Klebanoff, C.A., Shakiba, M., Trivedi, P., Menocal, L., Appleby, H., Camara, S., et al. (2019). TOX is a critical regulator of tumour-specific T cell differentiation. Nature *571*, 270–274. https://doi.org/10.1038/s41586-019-1324-y.

69. van der Veeken, J., Glasner, A., Zhong, Y., Hu, W., Wang, Z.M., Bou-Puerto, R., Charbonnier, L.M., Chatila, T.A., Leslie, C.S., and Rudensky, A.Y. (2020). The transcription factor Foxp3 shapes regulatory T cell identity by tuning the activity of trans-acting intermediaries. Immunity *53*, 971–984.e5. https://doi.org/10.1016/j.immuni.2020.10.010.

70. Martinez, G.J., Hu, J.K., Pereira, R.M., Crampton, J.S., Togher, S., Bild, N., Crotty, S., and Rao, A. (2016). Cutting edge: NFAT transcription factors promote the generation of follicular helper T cells in response to

acute viral infection. J. Immunol. *196*, 2015–2019. https://doi.org/10.4049/jimmunol.1501841.

71. Martinez, G.J., Pereira, R.M., Äijö, T., Kim, E.Y., Marangoni, F., Pipkin, M.E., Togher, S., Heissmeyer, V., Zhang, Y.C., Crotty, S., et al. (2015). The transcription factor NFAT promotes exhaustion of activated CD8+ T cells. Immunity *42*, 265–278. https://doi.org/10.1016/j.immuni.2015.01.006.

72. Zhang, X., Zhang, C., Qiao, M., Cheng, C., Tang, N., Lu, S., Sun, W., Xu, B., Cao, Y., Wei, X., et al. (2022). Depletion of BATF in CAR-T cells enhances antitumor activity by inducing resistance against exhaustion and formation of central memory cells. Cancer Cell *40*, 1407–1422.e7. https://doi.org/10.1016/j.ccell.2022.09.013.

73. Tsao, H.W., Kaminski, J., Kurachi, M., Barnitz, R.A., DiIorio, M.A., LaFleur, M.W., Ise, W., Kurosaki, T., Wherry, E.J., Haining, W.N., et al. (2022). Batf-mediated epigenetic control of effector CD8+ T cell differentiation. Sci. Immunol. *7*, eabi4919. https://doi.org/10.1126/sciimmunol.abi4919.

74. Seo, H., Chen, J., González-Avalos, E., Samaniego-Castruita, D., Das, A., Wang, Y.H., López-Moyado, I.F., Georges, R.O., Zhang, W., Onodera, A., et al. (2019). TOX and TOX2 Transcription Factors Cooperate with NR4A Transcription Factors to Impose CD8+ T Cell Exhaustion, *116* (Proceedings of the National Academy of Sciences), pp. 12410–12415.

75. Beltra, J.C., Manne, S., Abdel-Hakeem, M.S., Kurachi, M., Giles, J.R., Chen, Z., Casella, V., Ngiow, S.F., Khan, O., Huang, Y.J., et al. (2020). Developmental relationships of four exhausted CD8+ T cell subsets reveals underlying transcriptional and epigenetic landscape control mechanisms. Immunity *52*, 825–841.e8. https://doi.org/10.1016/j.immuni.2020.04.014.

76. Zehn, D., Thimme, R., Lugli, E., de Almeida, G.P., and Oxenius, A. (2022). Stem-like precursors are the fount to sustain persistent CD8+ T cell responses. Nat. Immunol. *23*, 836–847. https://doi.org/10.1038/s41590-022-01219-w.

77. Seth, A., Yokokura, Y., Choi, J.Y., Shyer, J.A., Vidyarthi, A., and Craft, J. (2023). AP-1–independent NFAT signaling maintains follicular T cell function in infection and autoimmunity. J. Exp. Med. *220*, e20211110. https://doi.org/10.1084/jem.20211110.

78. Lv, Y., Ricard, L., Gaugler, B., Huang, H., and Ye, Y. (2022). Biology and clinical relevance of follicular cytotoxic T cells. Front. Immunol. *13*, 1036616. https://doi.org/10.3389/fimmu.2022.1036616.

79. He, R., Hou, S., Liu, C., Zhang, A., Bai, Q., Han, M., Yang, Y., Wei, G., Shen, T., Yang, X., et al. (2016). Follicular CXCR5-expressing CD8+ T cells curtail chronic viral infection. Nature *537*, 412–428. https://doi.org/10.1038/nature19317.

80. Leong, Y.A., Chen, Y., Ong, H.S., Wu, D., Man, K., Deleage, C., Minnich, M., Meckiff, B.J., Wei, Y., Hou, Z., et al. (2016). CXCR5+ follicular cytotoxic T cells control viral infection in B cell follicles. Nat. Immunol. *17*, 1187–1196. https://doi.org/10.1038/ni.3543.

81. Chen, Y., Yu, M., Zheng, Y., Fu, G., Xin, G., Zhu, W., Luo, L., Burns, R., Li, Q.Z., Dent, A.L., et al. (2019). CXCR5+ PD-1+ follicular helper CD8 T cells control B cell tolerance. Nat. Commun. *10*, 4415. https://doi.org/10.1038/s41467-019-12446-5.

82. Yu, D., and Ye, L. (2018). A portrait of CXCR5+ follicular cytotoxic CD8+ T cells. Trends Immunol. *39*, 965–979. https://doi.org/10.1016/j.it.2018.10.002.

83. Chen, Z., Ji, Z., Ngiow, S.F., Manne, S., Cai, Z., Huang, A.C., Johnson, J., Staupe, R.P., Bengsch, B., Xu, C., et al. (2019). TCF-1-centered transcriptional network drives an effector versus exhausted CD8 T cell-fate decision. Immunity *51*, 840–855.e5. https://doi.org/10.1016/j.immuni.2019.09.013.

84. Kanev, K., Wu, M., Drews, A., Roelli, P., Wurmser, C., von Hösslin, M., and Zehn, D. (2019). Proliferation-competent Tcf1+ CD8 T Cells in Dysfunctional Populations Are CD4 T Cell Help Independent, *116* (Proceedings of the National Academy of Sciences), pp. 20070–20076.

85. van der Heide, V., Laghlali, G., Davenport, B., Cubitt, B., Roudko, V., Choo, D., Jhun, K., Humblin, E., Vaidya, A., Angeliadis, K., et al. (2025). Prolonged but finite antigen presentation promotes reversible defects of "helpless" memory CD8+ T cells. Immunity *58*, 1742–1761.e14. https://doi.org/10.1016/j.immuni.2025.05.025.

86. Manes, C., Moreno, M.G., Cimons, J., D'Antonio, M.A., Brunetti, T.M., Harbell, M.G., Selva, S., Mizenko, C., Borko, T.L., Lasda, E.L., et al. (2025). B cells shape naive CD8+ T cell programming. J. Clin. Investig. *135*, 12.

87. Burger, M.L., Cruz, A.M., Crossland, G.E., Gaglia, G., Ritch, C.C., Blatt, S.E., Bhutkar, A., Canner, D., Kienka, T., Tavana, S.Z., Barandiaran, A.L., et al. (2021). Antigen dominance hierarchies shape TCF1+ progenitor CD8 T cell phenotypes in tumors. Cell *184*, 4996–5014.e26. https://doi.org/10.1016/j.cell.2021.08.020.

88. Schenkel, J.M., Herbst, R.H., Canner, D., Li, A., Hillman, M., Shanahan, S.L., Gibbons, G., Smith, O.C., Kim, J.Y., Westcott, P., Hwang, W.L., et al. (2021). Conventional type I dendritic cells maintain a reservoir of proliferative tumor-antigen specific TCF-1+ CD8+ T cells in tumor-draining lymph nodes. Immunity *54*, 2338–2353.e6. https://doi.org/10.1016/j.immuni.2021.08.026.

89. Nakandakari-Higa, S., Walker, S., Canesso, M.C.C., van der Heide, V., Chudnovskiy, A., Kim, D.Y., Jacobsen, J.T., Parsa, R., Bilanovic, J., Parigi, S.M., et al. (2024). Universal recording of immune cell interactions in vivo. Nature *627*, 399–406. https://doi.org/10.1038/s41586-024-07134-4.

90. Avdeeva, M., Walker, S.K., van der Veeken, J., Rudensky, A.Y., and Pritykin, Y. (2026). ArchVelo: archetypal velocity modeling for single-cell multi-omic trajectories. Nat. Commun. *17*, 7228. https://doi.org/10.1038/s41467-026-74000-4.

91. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., and Li, W. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.

92. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

93. Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. *5*, 1752–1779. https://doi.org/10.1214/11-aoas466.

94. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

95. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., and Davies, R.M. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

96. Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat. Genet. *53*, 403–411. https://doi.org/10.1038/s41588-021-00790-6.

97. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

98. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.

99. Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics *26*, 2204–2207. https://doi.org/10.1093/bioinformatics/btq351.

100. Lianoglou, S., Garg, V., Yang, J.L., Leslie, C.S., and Mayr, C. (2013). Ubiquitously transcribed genes use alternative polyadenylation to

achieve tissue-specific expression. Genes Dev. *27*, 2380–2396. https://doi.org/10.1101/gad.229328.113.

101. Schep, A. (2022). motifmatchr: Fast motif matching in R. R package version *1*. https://doi.org/10.18129/B9.bioc.motifmatchr.
102. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.
103. Kibirige, H., Lamp, G., Katins, J., gdowding, austin, Finkernagel, F., matthias, k, Funnell, T., Arnfred, J., Blanchard, D., Kishimoto, P.N., et al. (2022). has2k1/plotnine: Version v0.10.1. Zenodo. https://doi.org/10.5281/zenodo.7124917.
104. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.r., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.
105. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS Comput. Biol. *9*, e1003118. https://doi.org/10.1371/journal.pcbi.1003118.
106. Martens, L.D., Fischer, D.S., Yépez, V.A., Theis, F.J., and Gagneur, J. (2024). Modeling fragment counts improves single-cell ATAC-seq analysis. Nat. Methods *21*, 28–31. https://doi.org/10.1038/s41592-023-02112-6.
107. Lause, J., Berens, P., and Kobak, D. (2021). Analytic Pearson residuals for normalization of single-cell RNA-seq UMI data. Genome Biol. *22*, 258. https://doi.org/10.1186/s13059-021-02451-7.
108. Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: Computational identification of cell doublets in single-cell transcriptomic data. Cell Syst. *8*, 281–291.e9. https://doi.org/10.1016/j.cels.2018.11.005.
109. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| anti-TCRβ eFluor 610 (Clone H57-597) | Thermo Fisher | **Catalog #** 61–5961-82 |
| anti-NK1.1 eFluor450 (PK136) | Thermo Fisher | **Catalog #** 48–5941-82 |
| anti-CD4 PerCP-Cy5.5 (RM4-5) | Thermo Fisher | **Catalog #** 46–0042-82 |
| anti-CD8a AF700 | Thermo Fisher | **Catalog #** 56–0081-82 |
| Bacterial and virus strains | | |
| LCMV Armstrong | In house | N/A |
| LCMV Clone 13 | In house | N/A |
| Chemicals, peptides, and recombinant proteins | | |
| Ghost Dye Live/Dead stain | Tonbo Biosciences | 13–0870-T100 |
| Critical commercial assays | | |
| Chromium Single Cell ATAC Reagent Kits | 10x Genomics | CG000168 Rev A |
| Chromium Next GEM Single Cell Multiome ATAC + Gene Expression | 10x Genomics | CG000338 Rev A |
| Dynabeads FlowComp mouse CD8 kit | Thermo Scientific | **Catalog number** 11462D |
| Dynabeads FlowComp mouse CD4 kit | Thermo Scientific | **Catalog number** 11461D |
| Deposited data | | |
| Raw and analyzed data | This paper | GEO: GSE300984 |
| Mouse reference genome, *mm10* | Genome Reference Consortium | https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001635.20/ |
| Raw and analyzed data | Giles et al.[29] | GSE199565 |
| Raw and analyzed data | Daniel et al.[28] | GSE188670 |
| Experimental models: Organisms/strains | | |
| Foxp3$^{GFP-DTR}$ mice | The Jackson Laboratory | Strain #:016958 |
| CAST/EiJ | The Jackson Laboratory | Strain #:000928 |
| Software and algorithms | | |
| cellranger-atac | 10x Genomics | v1.2.0 |
| cellranger-arc | 10x Genomics | v1.0.1 |
| MACS2 | Zhang et al.[91] | 2.2.7.1 |
| Bowtie2 | Langmead and Salzberg et al.[92] | 2.4.1 |
| IDR | Li et al.[93] | v2.0.3 |
| Bedtools | Quinlan and Hall et al.[94] | 2.29.1 |
| Samtools | Danecek et al.[95] | 1.10 |
| ArchR | Granja et al.[96] | 1.0.2 |
| DESeq2 | Love et al.[97] | 1.32.0 |
| scanpy | Wolf et al.[98] | v1.9.1 |
| bedGraphToBigWig | Kent et al.[99] | 2.10 |
| biosignals | Lianoglou et al.[100] | 0.1.0 |
| motifmatchr | Schep et al.[101] | 1.14.0 |
| py_pcha | Mørup and Hansen et al.[61] | v0.1.3 |
| scipy | Virtanen et al.[102] | v1.8.0 |
| plotnine | Kibirige et al.[103] | 0.10.1 |
| Other | | |
| Resource website | This paper | https://github.com/prityklinlab/multimodal_tcells |

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Experimental procedures for mouse work, cell isolation and single-cell genomics were as previously described.[31] Briefly, animals were housed at the Memorial Sloan Kettering Cancer Center (MSKCC) animal facility under specific pathogen-free conditions on a 12-h light:dark cycle under ambient conditions with free access to food and water. Mice used in the present study had no previous history of experimentation or exposure to drugs. Male Cast/Eij mice were purchased from Jackson Laboratory and bred to female *Foxp3*$^{GFP-DTR/GFP-DTR}$ mice on a C57BL/6 (B6) genetic background to generate F1 offspring. Adult mice (aged >6 weeks) were used for experiments. All animal experiments in this study were approved by the Sloan Kettering Institute (SKI) Institutional Animal Care and Use Committee under protocol 08–10-023.

## METHOD DETAILS

### Library preparation and sequencing

*Foxp3*$^{GFP-DTR}$ B6/Cast F1 males were infected with LCMV Armstrong ($2 \times 10^5$ pfu I.P.) or LCMV Clone 13 ($2 \times 10^6$ pfu I.V./retroorbital). Spleens were harvested at day 7 post-infection, a timepoint selected to capture the endogenous T cell response at its peak activation and to analyze the early molecular divergence between the acute and chronic regimes before terminal differentiation occurs. The difference in pooling ($n = 3$ for Armstrong, $n = 4$ for Clone 13) was a practical decision based on the number of available mice. T cells were enriched using anti-CD4 and anti-CD8 dynabead flowcomp kits (mixing CD4 and CD8 antibody). Following enrichment, cells were stained with the following panel of antibodies and BV510 ghost dye: NK1.1: BV421; CD62L: BV605; CD4: PerCP-Cy5.5; CD44: APC; TCRβ: PE-Tx-red. A 50–50% mix of $CD4^+$ and $CD8^+$ T cells (2 million cells total) was sorted into complete RPMI with 10% FCS, washed once, resuspended at 1M cells/mL and used as input material for single cell analysis.

For the standalone scATAC-seq samples, libraries were prepared according to the 10x scATAC-seq protocol from 10x Genomics Chromium (Single Cell ATAC Reagent Kits User Guide (catalog no. CG000168, Rev A)). Briefly, cells were centrifuged (300g, 5min, 4°C) and permeabilized with 100μL of chilled lysis buffer (10mM Tris-HCl, pH 7.4, 10mM NaCl, 3mM MgCl2, 0.1% Tween 20, 0.1% IGEPAL CA-630, 0.01% digitonin and 1% bovine serum albumin). The sample was incubated on ice for 3–5min and resuspended with 1mL of chilled wash buffer (10mM Tris-HCl, pH 7.4, 10mM NaCl, 3mM MgCl2, 0.1% Tween 20 and 1% bovine serum albumin). After centrifugation (500g, 5min, 4°C), the pellet was resuspended in 100 μL of chilled Nuclei Buffer (catalog no. 2000153, 10x Genomics). Nuclei were counted using a hemocytometer, and finally the nucleus concentration was adjusted to 3,000 nuclei per μL. We used 15,360 nuclei as input for tagmentation. Nuclei were diluted to 5 μL with 1× Nuclei Buffer (10x Genomics) and mixed with ATAC buffer (10x Genomics) and ATAC enzyme (10x Genomics) for tagmentation (60 min, 37°C). ScATAC-seq libraries were generated using the Chromium Chip E Single Cell ATAC kit (10x Genomics, catalog no. 1000086) and indices (Chromium i7 Multiplex Kit N, Set A, 10x Genomics, catalog no. 1000084) following the manufacturer's instructions. Final libraries were quantified using a Qubit fluorimeter (Life Technologies) and the nucleosomal pattern was verified using a BioAnalyzer (Agilent). Libraries were sequenced on a NovaSeq6000 (Illumina) with the following read lengths: 50 + 8+16 + 50 (Read 1+Index 1+Index 2+Read 2).

For the multiomic scATAC-seq + scRNA-seq samples, libraries were prepared according to the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression User Guide (CG000338 Rev A). Briefly, cells were centrifuged (300g, 5 min, 4°C) and permeabilized with 100 μL chilled lysis buffer (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM MgCl2, 0.1% Tween 20, 0.1% IGEPAL-CA630, 0.01% digitonin, 1% BSA). The samples were incubated on ice for 3–5 min and resuspended with 1 mL chilled wash buffer (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM MgCl2, 0.1% Tween 20 and 1% BSA and RNase Inhibitor 1 U/μL; Lucigen NxGen, F83923-1). After centrifugation (500g, 5 min, 4°C), the pellets were resuspended in 100 μL chilled Nuclei buffer (2000153, 10x Genomics). Nuclei were counted using a haemocytometer, and finally the nucleus concentration was adjusted to 3,000 nuclei per μL. 15,400 permeabilized nuclei were diluted to 5 μL in isotonic Tagmentation Buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 3 mM MgCl2, RNase Inhibitor 1 U/μL; Lucigen NxGen, F83923-1). Nuclei were mixed with 10 μL of a transposition master mix consisting of 7 μL ATAC Buffer B (10x Genomics, 2000193) and 3 μL ATAC Enzyme B (Tn5 transposase; 10x Genomics, 2000265) per reaction. Transposition was performed at 37°C for 60 min, followed by a brief hold at 4°C. Multiomic libraries were generated using the Chromium NextGEM Chip J (10x Genomics, 2000264).

### Read alignment

For read alignment and initial generation of count matrices, cellranger-atac and cellranger-arc count were run twice for each sample, once with the standard 10x mm10 reference genome and a second time with a custom Cast pseudogenome reference to enable allele-specific analysis, as previously described.[31] This Cast pseudogenome was prepared by introducing variants into the B6 reference genome. For the scRNA-seq modality, a final allele-specific BAM file for each sample was generated by combining the B6 and Cast GEX BAM file output by the two runs of CellRanger. In this combined BAM file, each read was labeled either with the genome that had the higher mapping alignment score or as ambiguous if the mapping scores were similar. Similarly for the scATAC-seq data, an allele-specific BAM file was generated for each sample by combining the B6 and Cast ATAC BAM file output by the two runs of CellRanger, labeling reads as B6, Cast, or ambiguous based on alignment scores. A final BED file of single nucleotide insertion sites (shifted by +4/-5) was generated from each BAM file.

## ATAC-seq peak atlas for T cells

### *Bulk ATAC-seq peaks*

We built our T cell peak atlas by extending a previously published compendium of CD8 T cell peaks.[26] We extended the compendium with a collection of CD4 and CD8 T cell peaks using seven previously published studies (Table S1). For these studies, FASTQ files were downloaded from SRA. Bowtie2[92] was used to align the FASTQ files for each sample to mm10 genome using the flags `--no-unal`, `--no-mixed`, and `--no-discordant`. Samtools[95] was used to turn the SAM files produced by Bowtie2 into sorted BAM files. The sample BAM files were then turned into BED files using BEDtools[94] `bamtobed` function, and then shifted by +4/-5. For samples with technical replicates, the shifted BED files were concatenated together and sorted. MACS2[91] was then run on the shifted BED files for each sample, using the flags `--nomodel`, `--shift 0`, `--extsize 76`, `-p 0.1`, `-B`, `--SPMR`, and `--keep-dup 'auto'`. In addition, MACS2 was run on all of the shifted BED files for all samples in a study, using the same flags. To identify peaks that were reproducible across biological replicates, IDR v2.0.3[93] was then run using all biological replicates of a sample from a study (denoted with the `--samples` flag) versus the background peak list run on all samples in the study (denoted with the `--peak-list` flag). The final IDR peaks for each study were then sorted and merged together using bedtools.

### *Single-cell ATAC-seq peaks*

An initial analysis of the scATAC-seq component of our multiomic dataset was performed using ArchR[96] on the standard mm10 CellRanger output in order to filter out low-quality cells and generate a single-cell peak set. An ArchR project was generated using ArrowFiles with all default parameters. A tile matrix was generated using `tileSize` of 500 and `binarize` set to False. Cells were filtered out based on low number of `nFrags`, with different thresholds used for each of the four samples (1200 for multiome Armstrong, 1500 for multiome Clone 13, 5000 for scATAC Armstrong, and 6000 for scATAC Clone 13). Doublets were called using the `addDoubletScores` function with `k = 15`. Cells with a `DoubletEnrichment` score greater than or equal to 3 were filtered out, removing 3,819 cells. Iterative LSI was performed with the `addIterativeLSI` function, called with `iterations = 8`, `clusterParams = list(resolution = c(1.5), sampleCells = 10000, n.start = 10, maxClusters = 25)`, `totalFeatures = 1e08`, `varFeatures = 5e06`, `force = TRUE`, and `firstSelection = ''Var''`.

Batch correction was performed using the `addHarmony` function[104] in ArchR on the IterativeLSI dimension-reduced matrix, for the four samples (scATAC Armstrong, scATAC Clone 13, multiome Armstrong, multiome Clone 13). A UMAP was generated on the batch-corrected Harmony reduced dimensions, using the `addUMAP` function. A gene score matrix was generated with the `addGeneScoreMatrix` function and all default parameters, and clusters were generated using the `addClusters` function with `resolution = 1.2` and `reducedDims = ''Harmony''`. Clusters were assigned to different T cell states based on marker gene scores from the GeneScoreMatrix, and one cluster was found to be B cells.

Peaks on the single-cell data were then called using MACS2 for all T cells using a BED file of single base pair insertion sites with the flags `--nomodel`, `--shift −75`, `--extsize 150`, `-q 0.001`, `-B`, `--SPMR`, and `--keep-dup 'all'`.

### *Merging bulk and single-cell ATAC-seq peaks*

The high quality bulk ATAC-seq peaks for CD4 T cells from IDR analysis were read in to R as GRanges[105] objects and combined using the `GenomicRangesreduce` function, to generate a simplified set. These were then filtered to peaks with a size greater than 76. This peak set, a set of previously published peaks from CD8 T cell data,[26] and the scATAC-seq peaks were combined using the `GenomicRangesreduce` function, and again filtered to peaks with a size greater than 76. This generated a set of 223,820 peaks, which was then filtered to only those on the standard chromosomes, resulting in a final set of 223,492 peaks (Table S2). Nearest genes to peaks were determined with a custom script using the bedtools[94] `nearest` function.

### *Peak signatures*

We learned bulk ATAC-seq peak signatures by using `DESeq2`[97] to compare chromatin accessibility between all samples from pairs of conditions within a study. For each study, peaks were filtered to those where at least 20% of the samples had counts greater than or equal to 10 reads. Then, differentially accessible peaks were calculated using `DESeq2` between the two conditions. Significantly differentially accessible peaks between the two conditions of interest in either direction (log2 fold change >1, adjusted $p$-value $\leq$0.05) were stored as epigenomic peak signatures (Table S3). In the scATAC-seq data, these peak signatures were scored using `score_genes` function in scanpy.[98]

## scATAC-seq data analysis

Peak counts were generated separately for each of the four samples (standalone scATAC-seq Armstrong, multiome scATAC-seq Armstrong, standalone scATAC-seq Clone 13, and multiome scATAC-seq Clone 13) in an allele-specific way to count reads aligned to the B6 genome separately from reads aligned to the Cast genome, using a custom script from the single basepair BED file of Tn5 insertions. The allele-specific counts were then summed together for each sample to generate total counts (across both alleles). The `scanpy`[98] package was used for all downstream analyses, plots were generated using `plotnine`,[103] and statistical tests were performed using scipy.[102] Cells were filtered out if they had low total counts across all genes, or low number of peaks with non-zero count per cell. Peaks were filtered so that they were present in at least 0.5% of cells in any one of the four samples, resulting in a dataset of 200,771 peaks and 32,100 cells. Fragment counts were generated by taking the ceiling of peak counts/2, to remove the bias of even-numbered counts.[106] Pearson residual normalized values[107] were then generated separately for each of the four samples, using theta parameter for overdispersion set to 0.3 for the two scATAC-seq standalone samples and 1 for the two scATAC-seq multiome samples (theta was chosen such that the distributions for normalized peak means was centered on 0 and

normalized peak variances were centered around 1). Two additional peaks were filtered out due to non-finite values after Pearson residual normalization, resulting in a dataset of 200,769 peaks. Then PCA was run with 100 PCs using the Pearson residual values, and due to a batch effect between multiome vs. standalone scATAC-seq, Harmony[104] within `scanpy` was run to integrate the multiome and standalone components of the data. A kNN graph of single-cell chromatin accessibility similarities was built on the Harmony-integrated PCA matrix using 30 PCs, 30 nearest neighbors and the cosine similarity metric. Preliminary Leiden clustering was performed with a resolution of 0.7, and UMAP was generated with the default parameters. One cluster at this preliminary analysis stage (cluster 11, 212 cells, 0.7% of all cells) was removed due to high chromatin accessibility and transcript expression at genes encoding well-defined B cell markers such as *Cd79a*, *Cd79b*, *Cd19*, *Ms4a1*, suggesting this cluster was likely B cell contamination; this filtering resulted in a dataset of 31,888 cells. After that, a cell similarity graph was rebuilt using 30 nearest neighbors, 30 PCs, cosine metric on the Harmony-integrated PCA matrix, Leiden clustering was rerun with resolution of 1, and UMAP was recalculated with default parameters. An additional cluster (cluster 15, 0.7% of all cells) was filtered out due to high doublet scores from running the `scrublet`[108] function in `scanpy`. This resulted in a final dataset of 31,662 cells and 200,769 peaks.

Cells were assigned to annotations based on accessibility at marker genes and using the epigenomic signature scores from our bulk ATAC-seq compendium.

### ATAC-seq peak summits and TF motif analysis

For each of the four scATAC-seq samples, the 1 bp-resolution Tn5 insertion BED file was intersected with the peak regions using bedtools `intersect`. These were then sorted, and a bedGraph file was generated from the BED using a custom script. The bedGraph file was then turned into a bigWig using the `bedGraphToBigWig`[99] method. The bigWig files were read into R as RleLists, the signal was accumulated over the four samples, and then subset to the peak regions. Using the `biosignals`[100] package, 150 bp windows of maximum signal within the peaks were uncovered. These summit regions were then used to predict TF binding to both the B6 and Cast genomes using the package `motifmatchr`[101] and a *p*-value threshold of 1e −3. For allele-specific analysis, imbalanced TF motif match was determined as the difference in predicted binding between B6 and Cast, with an absolute (non-zero) difference less than 4 (i.e., the sites with the absolute difference >4 were excluded from the analysis as outliers). For allele-agnostic TF motif analysis, the maximum predicted binding value to either the B6 or Cast genome was used for a TF motif in a summit. To explore these TF motifs in our scATAC-seq data, we first looked for enrichment of TF motifs in differentially accessible peaks for each cell type (pval $\leq$0.05, LFC >0). We used the hypergeometric test to determine enrichment of summits within differentially accessible peaks with predicted TF motif matches. In addition, cells were scored for TF motifs by using scanpy `score_genes()` function with the set of peaks that contain at least one summit with a TF motif match.

### Allele-specific analysis

Allele-specific counts were generated separately for the four samples over 475,357 summit windows within peaks. Counts were either assigned to the B6 genome, the Cast genome, or were considered ambiguous and ignored in downstream analyses. The `scanpy` package was used for all downstream analyses, and an AnnData object was generated by concatenating all four samples and then filtered to cells in our final dataset. Counts were then normalized by first generating a scaling factor for each cell, summing read counts for B6 and Cast counts. The scaling factor for each cell was then divided by the mean scaling factor over all cells. Finally the B6 and Cast count matrices were separately divided by the cell-based scaling factor. Allele-specific imbalance values (log2 fold changes) were then generated using these normalized counts to compare accessibility for B6 reads vs. Cast reads within each cell state. P-values were generated using a paired Wilcoxon test, and then corrected using Benjamini-Hochberg multiple hypothesis correction, and considered significantly imbalanced if FDR $\leq$0.05. We used Spearman correlation to compare allelic imbalance between predicted TF motif matches and chromatin accessibility. Correlations in each cell state were calculated only for motifs with at least 1000 imbalanced summits.

### scRNA-seq data analysis

The cell by gene count matrices were generated using a custom script, using only exonic reads. All analyses were performed using `scanpy`. QC steps were performed separately for the Armstrong samples and Clone 13 samples. Cells were first filtered out if having high mitochondrial UMI counts (using standard CellRanger filtered matrix output). Then cells were also filtered out if containing a low number of genes with non-zero count (fewer than 200), and genes were filtered out if present in fewer than 3 cells in either sample, resulting in 15,397 genes for further analysis. Cells were filtered out if having large total counts (25,000 for the clone 13 sample and 35,000 for the Armstrong sample). Pearson residual normalization was performed on each sample separately with `theta = 10` for both. The two samples were then concatenated together, resulting in a final dataset of 19,186 cells. Standard single-cell RNA-seq analysis steps were run: PCA, kNN neighborhood graph construction with 100 neighbors and 40 PCs and cosine similarity metric, Leiden clustering with resolution of 1.5, and UMAP generated. One cluster was filtered out as likely B cell contamination (high *Cd79a* expression), resulting in 19,107 cells. Cells were then filtered out if not present in the scATAC-seq data component, and cell type annotations from the scATAC-seq analysis were used to annotate cells. This resulted in a final dataset of 18,972 cells. Cell cycle gene signature scores were generated using scanpy `score_genes_cell_cycle` function and the G2M and S genes from `cc.genes` lists in the package Seurat.[109]

### Archetypal analysis

Archetypal analysis of the scATAC-seq data was performed using the `py_pcha` package[61] using the transposed Pearson residual normalized count matrix. We used $k = 8,12,15$, ultimately selecting $k = 8$ for further analysis. To establish the biological and functional definitions of each archetype, we used a post hoc multi-modal approach combining cell-state enrichment with marker gene association. Crucially, the archetypal decomposition itself is a completely unsupervised process that does not rely on prior biological labels. Functional annotations were assigned to the resulting archetypes post-decomposition solely to facilitate biological interpretation. Each archetype was labeled based on the predominant cell state of the individual cells exhibiting the highest archetypal weights, representing the cells for which that specific archetype was the most defining regulatory feature. While some archetypes were highly subset-specific (e.g., Archetype 0 tracking with naive CD8 T cells), others captured shared programs across lineages (e.g., Archetype 4 tracking with activation features across multiple T cell subsets), which were further validated by checking the expression of canonical marker genes strongly associated with those weights.

To characterize archetype-specific peaks, the matrix of archetype scores ($H \in \mathbb{R}^{k \times p}$, where $p$ is the number of peaks) was used to uniquely assign peaks to the archetype with the highest value (if greater than the threshold value of 0.1). To transfer and apply archetypes to other datasets (such as the Daniel et al.[28] and Giles et al.[29] datasets), the Scanpy `score_genes()` function was used to score the peaks assigned to each archetype (Table S6). These scores were then linearly scaled to the 0–1 range and normalized to sum to 1 for each cell.

To generate gene expression vectors for each archetype, we calculated weighted averages of cell expression vectors. In archetypal analysis, we decompose our original cell by peak ($n \times p$) data matrix into two matrices.[63] More explicitly, $X \sim WH$, where $X$ is our original cell by peak ($\mathbb{R}^{n \times p}$) normalized count matrix, $W \in \mathbb{R}^{n \times k}$ is the archetypal weights (providing the probability each cell is assigned to each archetype), and $H \in \mathbb{R}^{k \times p}$ is the archetypal score matrix, representing accessibility profiles for each archetype. The archetype scores matrix $H$ is generated directly from the original data matrix: $H = BX$, because each archetype profile is a linear composition of the cells in the dataset. We used the weights from the $B \in \mathbb{R}^{k \times n}$ matrix (representing how strong an effect each cell has on an archetype) to calculate the weighted average of expression profiles for each archetype, only for cells in the multiome.

### Analysis of other datasets

**Daniel** et al.[28] Fragment files were downloaded from SRA for day 8 and day 21 samples. BED files with single base pair Tn5 insertion sites were generated from the fragment files using a custom script, and then intersected with our peak atlas BED file using bedtools `intersect`. Counts were generated over all peaks in our peak atlas using a custom script from the Tn5 insertion BED file. The QC analysis was done for each sample separately, with cells being filtered based on the total read counts, and then peaks filtered based on presence in sufficient number of cells across all samples. One sample, the day 21 CX3CR1$^{-}$ sample, had substantially lower counts compared to all other samples and therefore was excluded from subsequent analysis. Fragment counts were generated from raw counts using the ceiling of peak counts/2 to remove the bias toward even counts, and then normalized with Pearson residuals, as described above. The samples were then concatenated, resulting in a dataset of 59,146 cells and 111,405 peaks. PCA was run, kNN neighborhood graph was generated using 40 neighbors and 50 PCs with cosine similarity metric, and UMAP and Leiden clustering with resolution of 0.3 were run. Scoring of progenitor vs. exhaustion peak signatures was calculated using `scanpy` default `score_genes()` function.

**Giles** et al.[29] Fragment files were downloaded from SRA for three samples: Armstrong infection 8 days post infection, Clone 13 infection 8 days post infection, and naive cells. BED files with single base pair Tn5 insertion sites were generated from the fragment files using a custom script, and then intersected with our peak atlas BED file using bedtools intersect. Counts were generated over all peaks in our peak atlas using a custom script. Basic QC steps (filtering cells due to low total counts and peaks due to low counts across cells) were run in scanpy, separately for each of the three samples. Peak counts were first fragment-adjusted (ceiling of peak counts/2) to remove the bias toward even counts and then normalized with Pearson residuals, as described above. Samples were then concatenated, resulting in a matrix of 36,959 cells and 144,833 peaks. PCA was run, kNN graph was generated with 30 neighbors, 40 PCs, and cosine similarity metric, UMAP was run with default parameters, and Leiden clustering was run with a resolution value of 0.4. Scoring of progenitor vs. exhaustion peak signatures was calculated using scanpy default `score_genes` function.

## QUANTIFICATION AND STATISTICAL ANALYSIS

All details of data quantification, statistical and computational analyses used in this study can be found in the main and supplementary figures, the corresponding figure legends, results and method details section of STAR Methods.

**Cell Reports, Volume *45***

## Supplemental information

# Single-cell multi-omics reveals archetypal regulatory programs shared across T cell subsets

**Sarah K. Walker, Joris van der Veeken, Alexander Y. Rudensky, and Yuri Pritykin**

# Document S1: Supplementary Figures S1–S14.

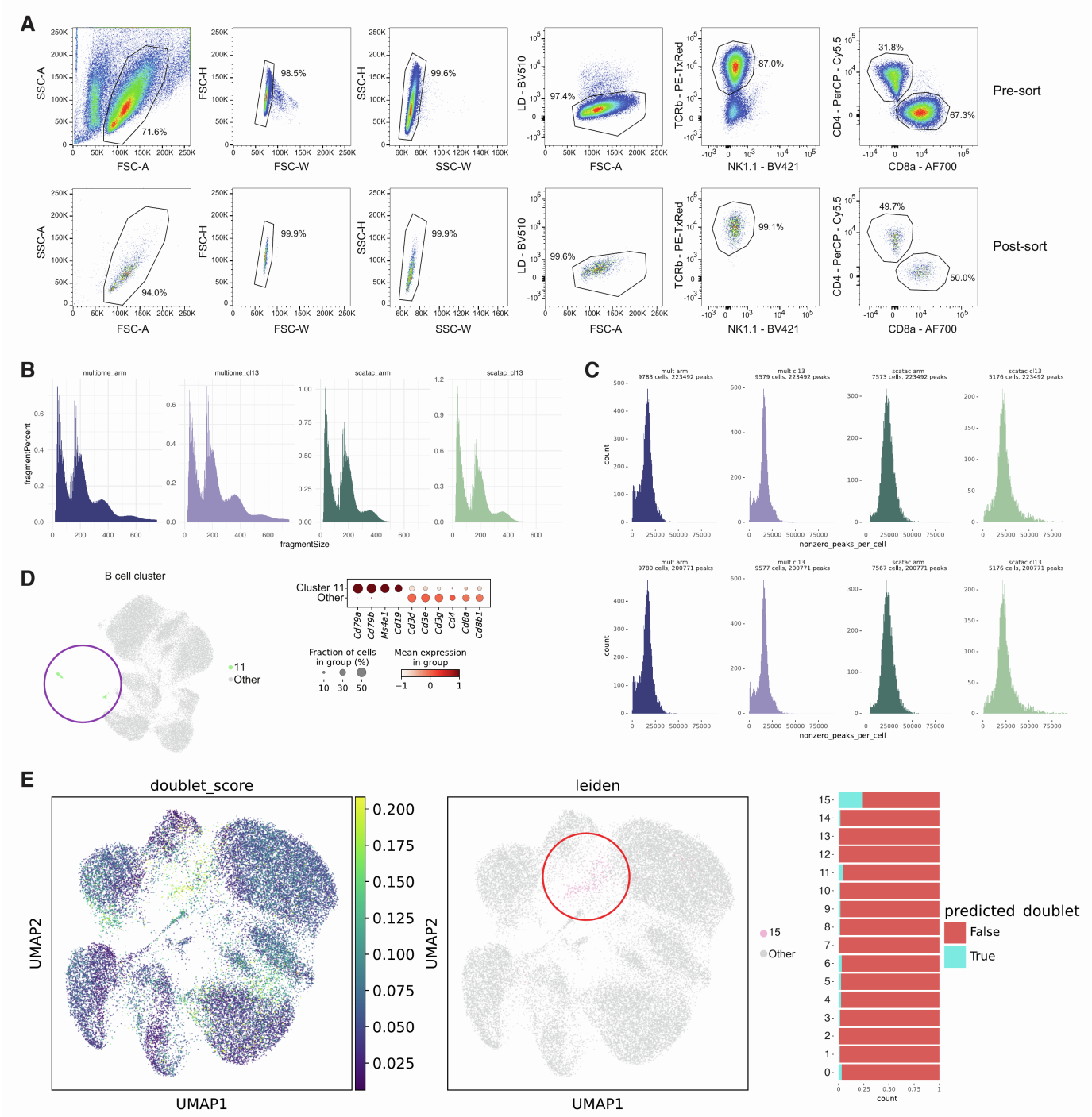



**Figure S1. Sorting of T cells and initial quality control of the scATAC-seq data (related to Figure 1).**
**(A)** Flow cytometry sorting strategy: splenic CD4 and CD8 T cells from mice infected with LCMV were isolated for multiomic profiling (see **STAR Methods**).

**(B)** Quality control (QC) plot showing the distribution of fragment sizes for each of the four samples. Expected modes corresponding to fragments spanning one, two or three nucleosomes, as well as characteristic 10 bp periodicity, confirm high data quality.
**(C)** QC plots showing the number of peaks with non-zero scATAC-seq signal (1 bp Tn5 insertion count) per cell from before (*top*) and after (*bottom*) filtering out low quality cells and peaks, across all four samples.
**(D)** *Left*, UMAP plot from preliminary analysis highlighting preliminary cluster 11 (212 cells, 0.7% of total), which consisted of B cell contamination. *Right,* dotplot showing expression of B cell and T cell marker genes in preliminary cluster 11 vs. all other cells. Dot size, fraction of cells with normalized expression greater than 0; dot color, mean normalized expression. This B cell contamination cluster was removed from the data prior to reclustering for subsequent analysis.
**(E)** *Left*, UMAP plots showing doublet scores calculated with Scrublet and Leiden cluster 15. *Right*, barplot showing the fraction of predicted doublets (as identified by Scrublet) within each Leiden cluster. Cluster 15 contained the highest fraction of predicted doublets and was therefore removed from subsequent analysis.

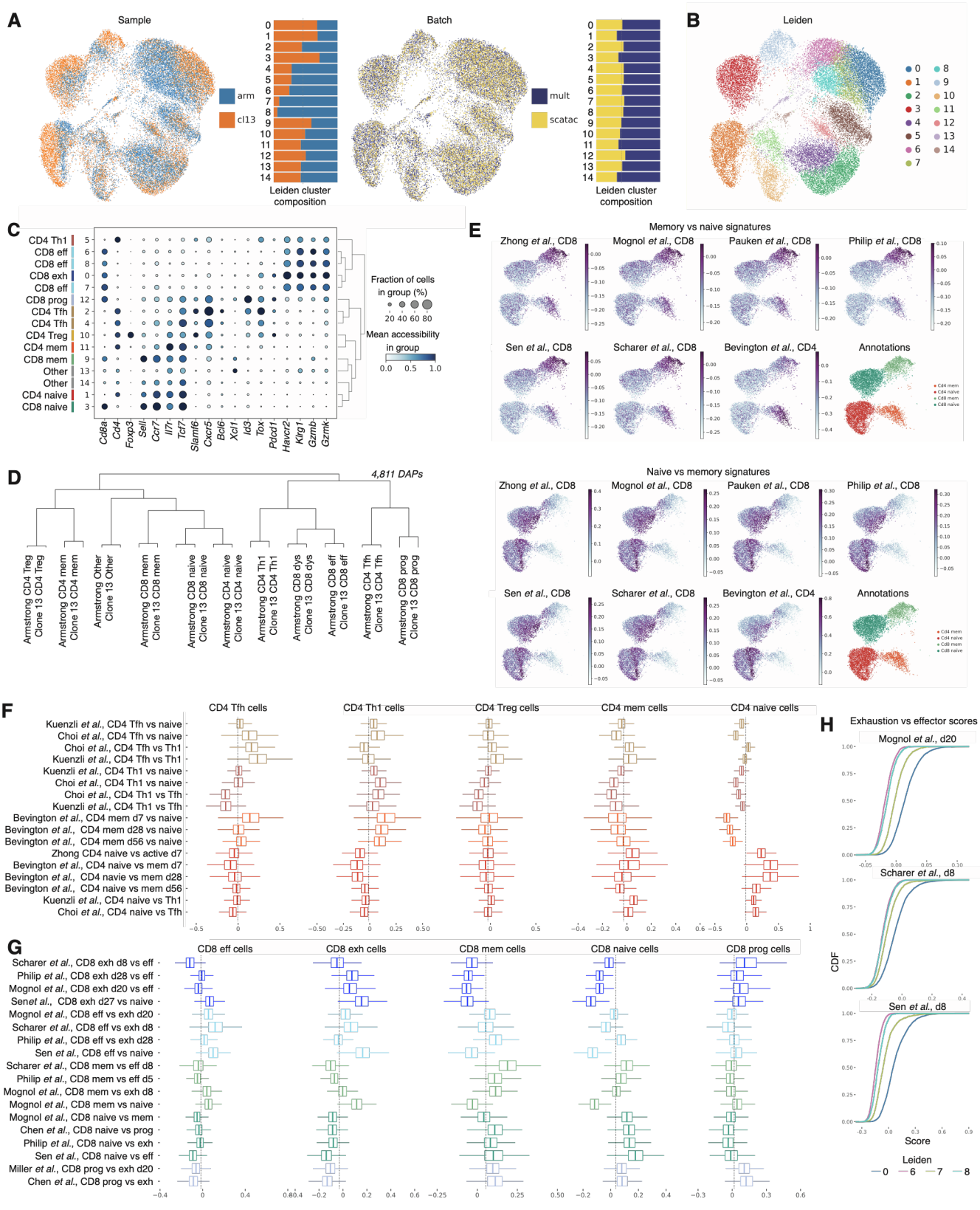

A
Sample
arm
cl13
Leiden cluster composition
Batch
mult
scatac
Leiden cluster composition
B
Leiden
C
CD4 Th1
CD8 eff
CD8 eff
CD8 exh
CD8 eff
CD8 prog
CD4 Tfh
CD4 Tfh
CD4 Treg
CD4 mem
CD8 mem
Other
Other
CD4 naive
CD8 naive
Cd8a
Cd4
Foxp3
Sell
Ccr7
Il7r
Tcf7
Slamf6
Cxcr5
Bcl6
Xcl1
Id3
Tox
Pdcd1
Havcr2
Klrg1
Gzmb
Gzmk
Fraction of cells in group (%)
Mean accessibility in group
D
4,811 DAPs
Armstrong CD4 Treg
Clone 13 CD4 Treg
Armstrong CD4 mem
Clone 13 CD4 mem
Armstrong Other
Clone 13 Other
Armstrong CD8 mem
Clone 13 CD8 mem
Armstrong CD8 naive
Clone 13 CD8 naive
Armstrong CD4 naive
Clone 13 CD4 naive
Armstrong CD4 Th1
Clone 13 CD4 Th1
Armstrong CD8 dys
Clone 13 CD8 dys
Armstrong CD8 eff
Clone 13 CD8 eff
Armstrong CD4 Tfh
Clone 13 CD4 Tfh
Armstrong CD8 prog
Clone 13 CD8 prog
E
Memory vs naive signatures
Zhong et al., CD8
Mognol et al., CD8
Pauken et al., CD8
Philip et al., CD8
Sen et al., CD8
Scharer et al., CD8
Bevington et al., CD4
Annotations
Naive vs memory signatures
F
CD4 Tfh cells
CD4 Th1 cells
CD4 Treg cells
CD4 mem cells
CD4 naive cells
Kuenzli et al., CD4 Tfh vs naive
Choi et al., CD4 Tfh vs naive
Choi et al., CD4 Tfh vs Th1
Kuenzli et al., CD4 Tfh vs Th1
Kuenzli et al., CD4 Th1 vs naive
Choi et al., CD4 Th1 vs naive
Choi et al., CD4 Th1 vs Tfh
Kuenzli et al., CD4 Th1 vs Tfh
Bevington et al., CD4 mem d7 vs naive
Bevington et al., CD4 mem d28 vs naive
Bevington et al., CD4 mem d56 vs naive
Zhong CD4 naive vs active d7
Bevington et al., CD4 naive vs mem d7
Bevington et al., CD4 navie vs mem d28
Bevington et al., CD4 naive vs mem d56
Kuenzli et al., CD4 naive vs Th1
Choi et al., CD4 naive vs Tfh
G
CD8 eff cells
CD8 exh cells
CD8 mem cells
CD8 naive cells
CD8 prog cells
Scharer et al., CD8 exh d8 vs eff
Philip et al., CD8 exh d28 vs eff
Mognol et al., CD8 exh d20 vs eff
Senet al., CD8 exh d27 vs naive
Mognol et al., CD8 eff vs exh d20
Scharer et al., CD8 eff vs exh d8
Philip et al., CD8 eff vs exh d28
Sen et al., CD8 eff vs naive
Scharer et al., CD8 mem vs eff d8
Philip et al., CD8 mem vs eff d5
Mognol et al., CD8 mem vs exh d8
Mognol et al., CD8 mem vs naive
Mognol et al., CD8 naive vs mem
Chen et al., CD8 naive vs prog
Philip et al., CD8 naive vs exh
Sen et al., CD8 naive vs eff
Miller et al., CD8 prog vs exh d20
Chen et al., CD8 prog vs exh
H
Exhaustion vs effector scores
Mognol et al., d20
Scharer et al., d8
Sen et al., d8
CDF
Score
Leiden
0
6
7
8

**Figure S2. Analysis and functional annotation of the scATAC-seq data (related to Figure 1).**
**(A)** UMAP plot of the scATAC-seq component of the data, colored by (*left*) sample (infection type) and (*right*) experimental batch (scATAC-seq from multiome and standalone scATAC-seq experiment). Accompanying bar plots show composition of Leiden clusters by (*left*) infection type and (*right*) batch (scATAC-seq from multiome vs. standalone scATAC-seq). This analysis illustrates presence of cells from each cluster in both infection type conditions and approximately equal presence in both batches, suggesting reproducibility and little batch effect for cluster definitions.
**(B)** UMAP colored by Leiden clusters.
**(C)** Dotplot showing normalized scATAC-seq accessibility of a selected peak for each marker gene across scATAC-seq Leiden clusters, illustrating functional cell state annotations of the clusters. Cluster assignments to functional cell states are shown on the left. A dendrogram on the right shows how Leiden clusters relate to each other. Dot size, fraction of cells with normalized accessibility greater than 0; dot color, mean normalized accessibility in a cluster.
**(D)** Dendrogram of hierarchical clustering over 4,811 differentially accessible peaks (peaks enriched in at least one cluster, with log2 fold change >1.5 and $p \leq 0.05$), for T cell functional states, separated by infection type. This analysis illustrates similarity of cells with the same functional annotation from the two distinct infection responses, and relative similarity of different functional states.
**(E)** UMAP of the four clusters annotated as naive and memory populations, colored by scores from memory vs. naive (*top*) and naive vs. memory (*bottom*) epigenomic signatures derived from different bulk ATAC-seq dataset (see **STAR Methods**). Shown also is a UMAP with final functional cell state annotations. This analysis illustrates naive and memory functional cell state annotations.
**(F)** Boxplots showing distributions of scores from CD4 T cell epigenomic signatures derived from bulk ATAC-seq data, grouped by category, for CD4 T cell states.
**(G)** Boxplots showing distributions of scores from CD8 T cell epigenomic signatures derived from bulk ATAC-seq data, grouped by category, for CD8 T cell states.
**(H)** Cumulative distribution function (CDF) plots showing the distribution of scores from three exhausted vs. effector epigenomic signatures derived from bulk ATAC-seq data, separately for each of the four clusters annotated as either effector or exhausted CD8 T cells. Distribution of scores for cluster 0 is significantly higher than that for all other clusters (FDR < 2.4e−112), justifying annotation of cluster 0 as exhausted CD8 T cells.
Boxplots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; outliers not shown.

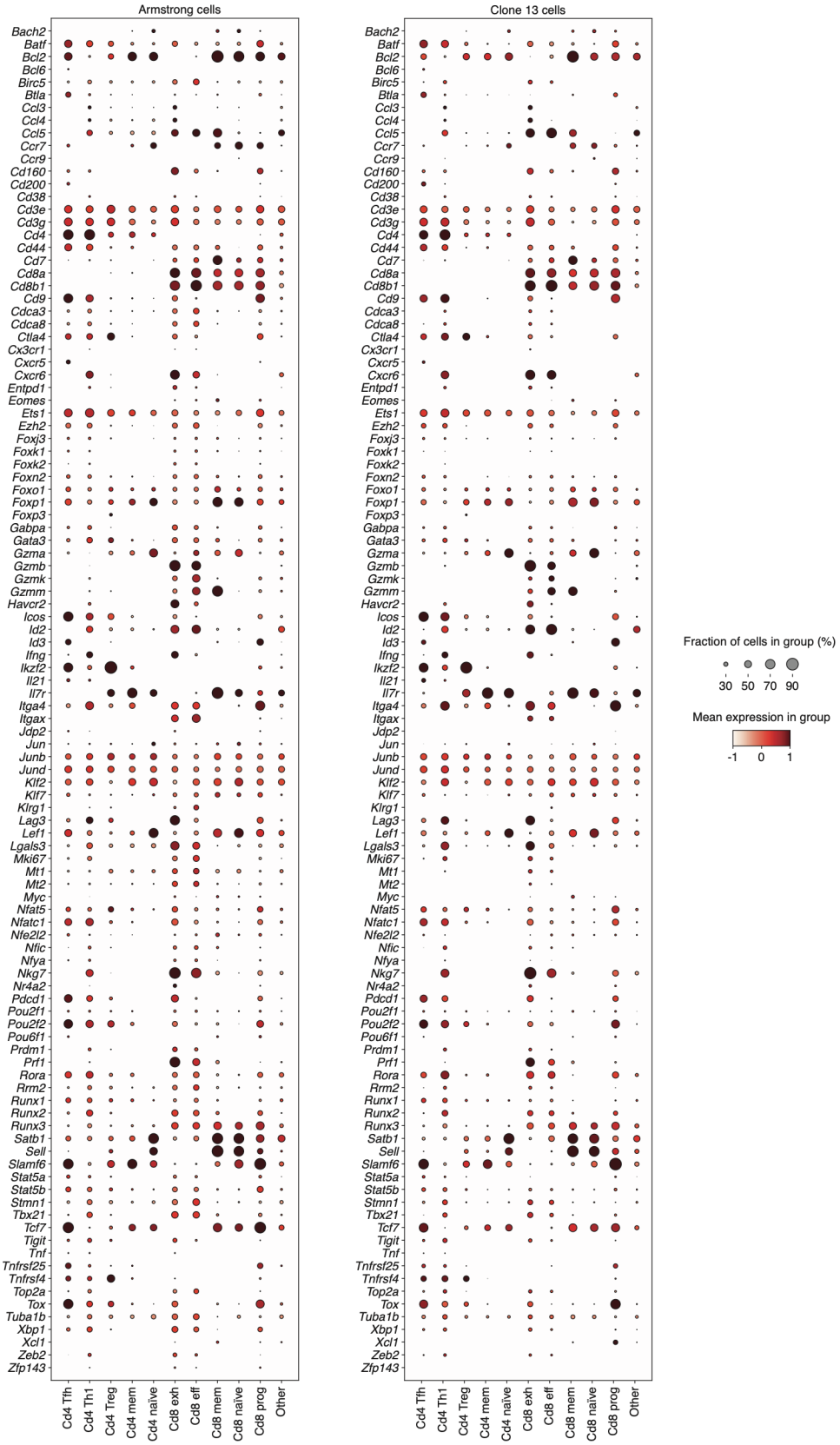

Armstrong cells
Clone 13 cells
Bach2
Batf
Bcl2
Bcl6
Birc5
Btla
Ccl3
Ccl4
Ccl5
Ccr7
Ccr9
Cd160
Cd200
Cd38
Cd3e
Cd3g
Cd4
Cd44
Cd7
Cd8a
Cd8b1
Cd9
Cdca3
Cdca8
Ctla4
Cx3cr1
Cxcr5
Cxcr6
Entpd1
Eomes
Ets1
Ezh2
Foxj3
Foxk1
Foxk2
Foxn2
Foxo1
Foxp1
Foxp3
Gabpa
Gata3
Gzma
Gzmb
Gzmk
Gzmm
Havcr2
Icos
Id2
Id3
Ifng
Ikzf2
Il21
Il7r
Itga4
Itgax
Jdp2
Jun
Junb
Jund
Klf2
Klf7
Klrg1
Lag3
Lef1
Lgals3
Mki67
Mt1
Mt2
Myc
Nfat5
Nfatc1
Nfe2l2
Nfic
Nfya
Nkg7
Nr4a2
Pdcd1
Pou2f1
Pou2f2
Pou6f1
Prdm1
Prf1
Rora
Rrm2
Runx1
Runx2
Runx3
Satb1
Sell
Slamf6
Stat5a
Stat5b
Stmn1
Tbx21
Tcf7
Tigit
Tnf
Tnfrsf25
Tnfrsf4
Top2a
Tox
Tuba1b
Xbp1
Xcl1
Zeb2
Zfp143
Cd4 Tfh
Cd4 Th1
Cd4 Treg
Cd4 mem
Cd4 naive
Cd8 exh
Cd8 eff
Cd8 mem
Cd8 naive
Cd8 prog
Other
Fraction of cells in group (%)
30 50 70 90
Mean expression in group
-1 0 1

**Figure S3. Expression of selected genes across functional cell states (related to Figure 1).** Dotplot showing normalized scRNA-seq expression of genes with important functions in T cells, across functional cell states derived from scATAC-seq analysis, separated by infection type (left, Arm; right, Cl13). Dot size, fraction of cells with normalized expression greater than 0; dot color, mean normalized expression.

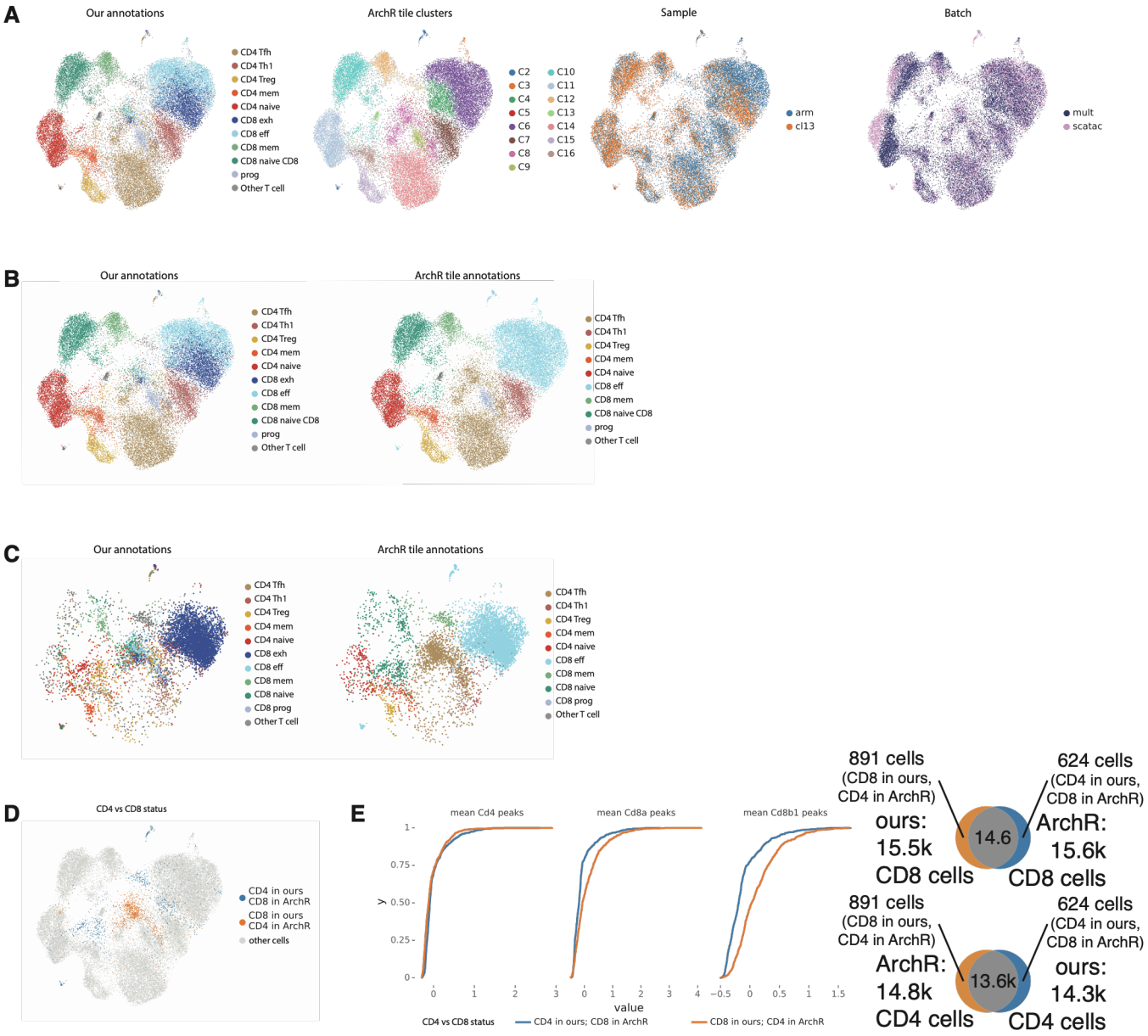



**Figure S4. Comparison of our scATAC-seq analysis with ArchR tile-based analysis (related to Figure 1).**

**(A)** UMAP generated using the default ArchR pipeline with a TileMatrix for downstream analyses. UMAP plots are colored by (left to right): our final cell state annotations, ArchR tile-based clusters, sample (infection type), and experimental batch (multiome scATAC-seq vs. standalone scATAC-seq). Cluster C1 consisted primarily of B cell contamination and was removed.

**(B)** ArchR UMAP colored by our cell state annotations and by annotations derived from ArchR tile-based clustering. ArchR clusters were annotated based on predominant overlapping annotations from our analysis (which were based on marker gene activity and literature-derived epigenomic signatures);

following a standard approach in single-cell genomics analysis, all cells in each ArchR cluster received the same annotation.
**(C)** ArchR UMAP plots showing cells with conflicting annotations between our analysis and the ArchR tile-based analysis. The discrepancy was observed across nearly all clusters. In particular, following the logic above, both ArchR clusters C4 and C6, overlapping with exhausted cells defined in our analysis, predominantly contained effector cells defined in our analysis and therefore were annotated as effector CD8 T cells.
**(D, E)** Comparisons of cells that are annotated as CD4 vs. CD8 T cells in our vs. ArchR analysis. In our analysis, 15.5K cells were annotated as CD8 T cells vs. 15.6K CD8 T cells in ArchR analysis; 14.8K cells were annotated as CD4 T cells in our analysis vs. 14.3K CD4 T cells in ArchR analysis. We considered two groups of discrepant cells (1,515 cells total): 891 cells annotated as CD8 T cells in our analysis but CD4 T cells by ArchR (shown in orange), and 624 cells annotated as CD4 T cells in our analysis but CD8 T cells by ArchR (shown in blue).
**(D)** UMAP plot showing the discrepant annotations.
**(E)** CDF plots showing distributions of chromatin accessibility for the two groups of cells with discrepant annotations, for peaks of the lineage-defining genes *Cd4*, *Cd8a*, and *Cd8b1*. Venn diagrams illustrate the definitions of the two groups of cells with discrepant annotations. Accessibility values were averaged across peaks near each gene for each cell. All pairwise differences between blue and orange distributions are significant (FDR < 5.4e−19, two-sided KS test): “orange” cells are more likely to be CD8 T cells and “blue” cells are more likely to be CD4 T cells, suggesting that our annotations are more accurate than those from ArchR analysis.

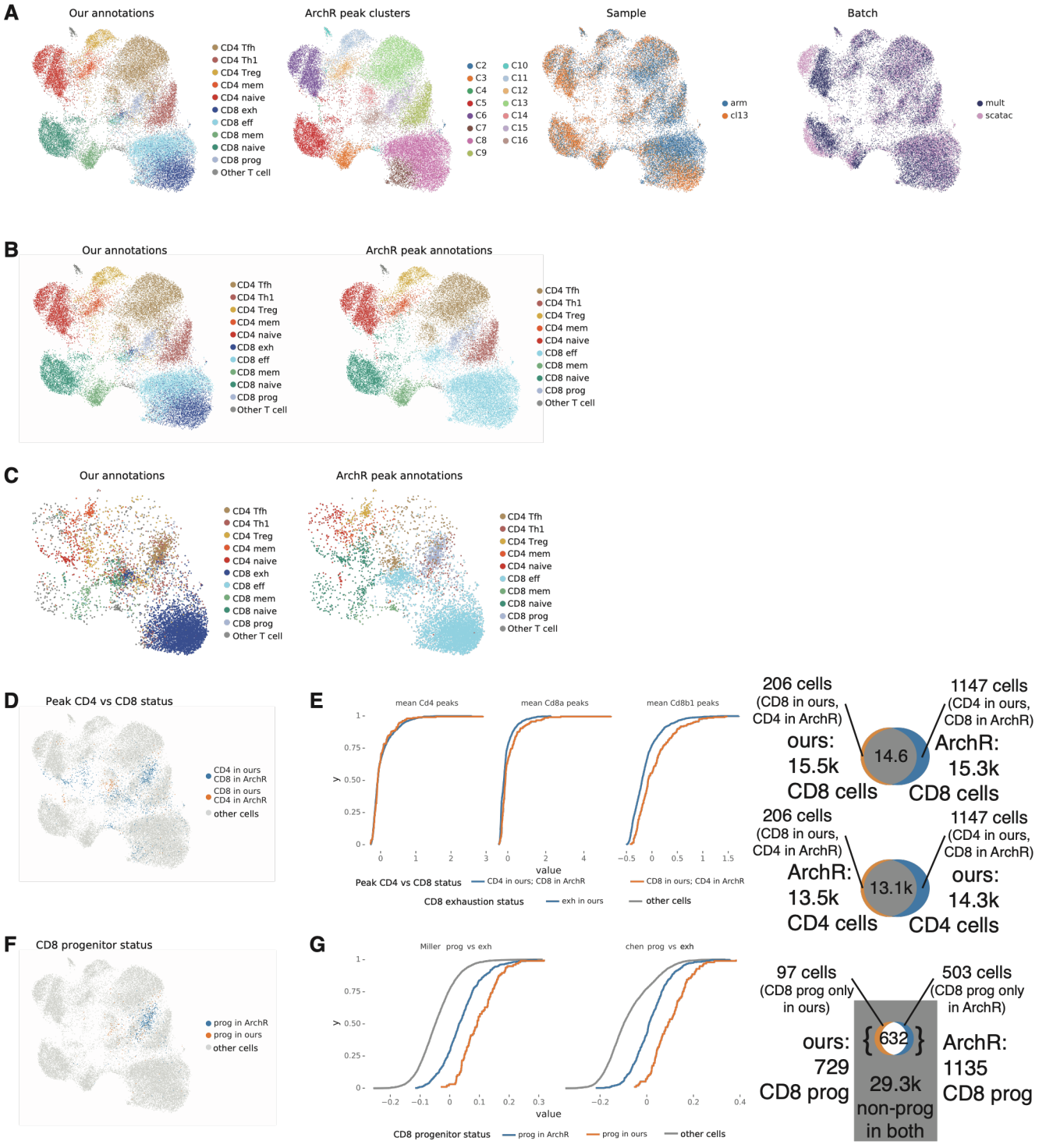



**Figure S5. Comparison of our scATAC-seq analysis with ArchR peak-based analysis (related to Figure 1).**

Same comparative analysis as in **Fig. S4** but for the peak-based instead of tile-based ArchR analysis.

**(A)** UMAP generated using the default ArchR pipeline using an ArchR-generated PeakMatrix with MACS2 peak calling. UMAP plots are colored by (left to right): our final cell state annotations, ArchR peak-based clusters, sample (infection type), and experimental batch (multiome scATAC-seq vs. standalone scATAC-seq). Cluster C1 consisted primarily of B cell contamination and was removed.
**(B)** UMAP from ArchR peak-based analysis colored by our cell state annotations and by annotations derived from ArchR peak-based clustering. ArchR clusters were annotated based on predominant overlapping annotations from our analysis (which were based on marker gene activity and literature-derived epigenomic signatures); following a standard approach in single-cell genomics analysis, all cells in each ArchR cluster received the same annotation.
**(C)** UMAP from ArchR peak-based analysis showing cells with conflicting annotations between our analysis and the ArchR peak-based analysis. The discrepancy was observed across nearly all clusters. In particular, following the logic above, both ArchR peak-based clusters C7 and C8, overlapping with exhausted cells defined in our analysis, predominantly contained effector cells defined in our analysis and therefore were annotated as effector CD8 T cells; no ArchR clusters were annotated as exhausted.
**(D, E)** Comparisons of cells that are annotated as CD4 vs. CD8 T cells in our vs. ArchR peak-based analysis (1,353 cells). Analogous to **Fig. S4D,E**. Pairwise differences between the blue and orange distributions are significant for *Cd8a* and *Cd8b1* (FDR < 1.7e−10, two-sided KS test): “orange” cells are more likely to be CD8 than CD4 T cells, suggesting that our annotations are more accurate than those from ArchR peak-based analysis.
**(F, G)** Comparisons of cells that are annotated as progenitor CD8 T cells in our analysis or ArchR peak-based analysis. In our analysis, 729 cells were annotated as CD8 progenitors vs. 1,135 CD8 progenitors in ArchR peak-based analysis. We considered two groups of cells with discrepant annotations (608 cells total): 97 cells annotated as CD8 progenitors in our but not ArchR peak-based analysis (shown in orange), and 503 cells annotated as CD8 progenitors in ArchR peak-based but not in our analysis (shown in blue). Other 29.3K cells were not annotated as CD8 progenitors in both analyses (shown in gray).
**(F)** UMAP plots showing the discrepant annotations.
**(G)** CDF plots showing distributions of progenitor vs. exhausted epigenomic signature scores (derived from our bulk ATAC-seq compendium) for cells with discrepant annotations. Venn diagrams illustrate the definitions of the two groups of cells with discrepant annotations. The orange values are significantly larger than blue values in both comparisons (FDR < 2.4e−12, two-sided KS test), suggesting that our annotations are more accurate.

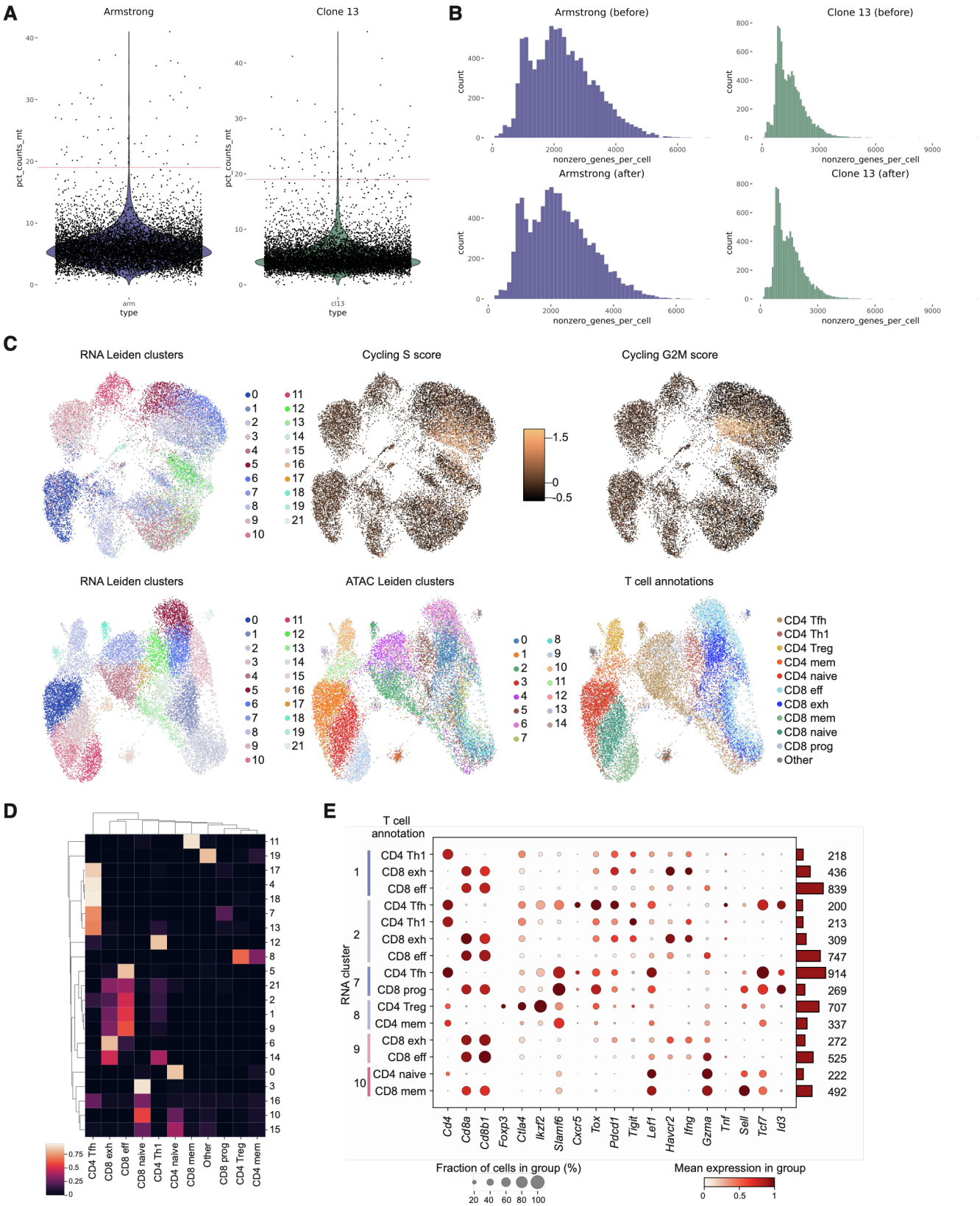

A
Armstrong
Clone 13
pct_counts_mt
arm
cl13
type
B
Armstrong (before)
Clone 13 (before)
Armstrong (after)
Clone 13 (after)
count
nonzero_genes_per_cell
C
RNA Leiden clusters
Cycling S score
Cycling G2M score
ATAC Leiden clusters
T cell annotations
CD4 Tfh
CD4 Th1
CD4 Treg
CD4 mem
CD4 naive
CD8 eff
CD8 exh
CD8 mem
CD8 naive
CD8 prog
Other
D
E
T cell annotation
RNA cluster
Cd4
Cd8a
Cd8b1
Foxp3
Ctla4
Ikzf2
Slamf6
Cxcr5
Tox
Pdcd1
Tigit
Lef1
Havcr2
Ifng
Gzma
Tnf
Sell
Tcf7
Id3
Fraction of cells in group (%)
Mean expression in group

**Figure S6. Analysis of the scRNA-seq data (related to Figure 1).**
**(A)** Violin plots showing the distribution of the percentage of reads aligned to mitochondria per cell for the Armstrong (*left*) and Clone 13 (*right*) sample. Cells exceeding the 19% mitochondrial read threshold (red line) were removed for downstream analysis.
**(B)** QC plots showing the distribution of the number of genes with non-zero expression per cell before (*top*) and after (*bottom*) filtering out low quality cells, separately for the Armstrong (*left*) and the Clone 13 (*right*) sample.
**(C)** Top row: scATAC-seq UMAP plots colored by scRNA-seq Leiden clusters (*left*) and cell cycle gene signature scores (S phase, *center*, and G2M phase, *right*) in the scRNA-seq data. Bottom row: scRNA-seq UMAP plots colored by scRNA-seq Leiden clusters (*left*), scATAC-seq Leiden clusters (*center*) and our final functional T cell state annotations based on the scATAC-seq analysis (*right*).
**(D)** Clustered heatmap showing the correspondence between scRNA-seq-based Leiden clusters and our final cell state annotations (based on the scATAC-seq analysis). Each square represents the fraction of cells from a given scRNA-seq cluster assigned to each functional cell state annotation; rows sum to 1.
**(E)** Dotplot showing expression of selected marker genes across scRNA-seq-based Leiden clusters, further stratified by their assigned scATAC-seq-derived functional annotations. Bar plot on the right margin shows the number of cells in each group; only groups of > 200 cells are shown. Dot size, fraction of cells with normalized expression greater than 0; dot color, mean normalized expression. We conclude that scRNA-seq clustering cannot resolve between distinct functional annotations, including between cells from CD4 and CD8 lineages (confirmed by expression of lineage-defining markers *Cd4*, *Cd8a*, *Cd8b1*). Indeed, if annotations were defined based on scRNA-seq clusters (and following a standard approach in single-cell genomics analysis, all cells in each scRNA-seq cluster would receive the same annotation), this would result in combining functionally distinct cells into the same annotation. For example, scRNA-seq cluster 1 contains both CD4 Th1 cells and CD8 effector and exhausted cells; scRNA-seq cluster 2 contains both CD4 Th1 and Tfh cells and CD8 effector and exhausted cells; scRNA-seq cluster 10 contains both naive CD4 T cells and memory CD8 T cells, etc., which would receive the same annotation within the cluster, resulting in misannotation.

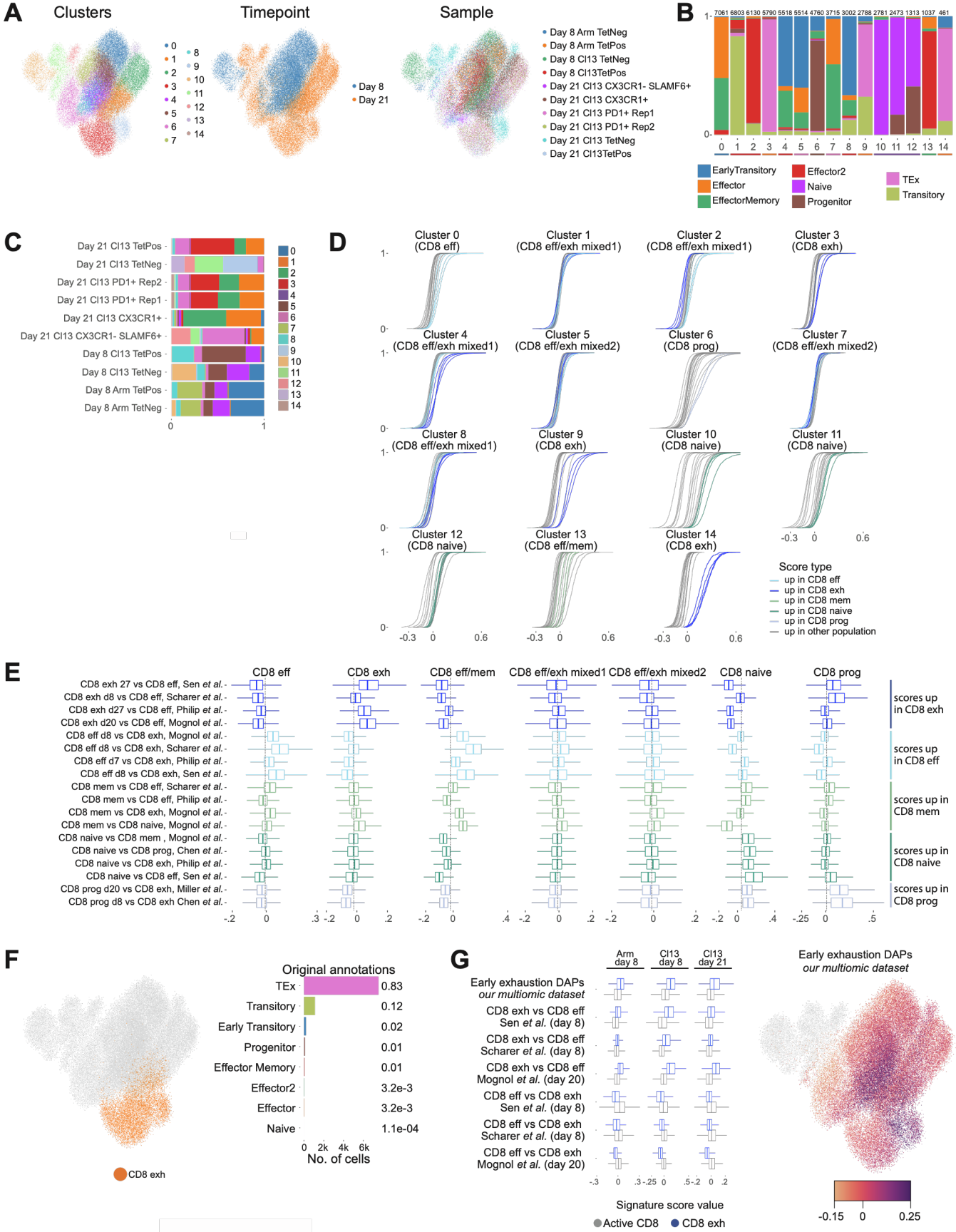

A
Clusters
Timepoint
Sample
Day 8
Day 21
Day 8 Arm TetNeg
Day 8 Arm TetPos
Day 8 Cl13 TetNeg
Day 8 Cl13TetPos
Day 21 Cl13 CX3CR1- SLAMF6+
Day 21 Cl13 CX3CR1+
Day 21 Cl13 PD1+ Rep1
Day 21 Cl13 PD1+ Rep2
Day 21 Cl13 TetNeg
Day 21 Cl13TetPos
B
EarlyTransitory
Effector
EffectorMemory
Effector2
Naive
Progenitor
TEx
Transitory
C
Day 21 Cl13 TetPos
Day 21 Cl13 TetNeg
Day 21 Cl13 PD1+ Rep2
Day 21 Cl13 PD1+ Rep1
Day 21 Cl13 CX3CR1+
Day 21 Cl13 CX3CR1- SLAMF6+
Day 8 Cl13 TetPos
Day 8 Cl13 TetNeg
Day 8 Arm TetPos
Day 8 Arm TetNeg
D
Cluster 0 (CD8 eff)
Cluster 1 (CD8 eff/exh mixed1)
Cluster 2 (CD8 eff/exh mixed1)
Cluster 3 (CD8 exh)
Cluster 4 (CD8 eff/exh mixed1)
Cluster 5 (CD8 eff/exh mixed2)
Cluster 6 (CD8 prog)
Cluster 7 (CD8 eff/exh mixed2)
Cluster 8 (CD8 eff/exh mixed1)
Cluster 9 (CD8 exh)
Cluster 10 (CD8 naive)
Cluster 11 (CD8 naive)
Cluster 12 (CD8 naive)
Cluster 13 (CD8 eff/mem)
Cluster 14 (CD8 exh)
Score type
up in CD8 eff
up in CD8 exh
up in CD8 mem
up in CD8 naive
up in CD8 prog
up in other population
E
CD8 eff
CD8 exh
CD8 eff/mem
CD8 eff/exh mixed1
CD8 eff/exh mixed2
CD8 naive
CD8 prog
CD8 exh 27 vs CD8 eff, Sen et al.
CD8 exh d8 vs CD8 eff, Scharer et al.
CD8 exh d27 vs CD8 eff, Philip et al.
CD8 exh d20 vs CD8 eff, Mognol et al.
CD8 eff d8 vs CD8 exh, Mognol et al.
CD8 eff d8 vs CD8 exh, Scharer et al.
CD8 eff d7 vs CD8 exh, Philip et al.
CD8 eff d8 vs CD8 exh, Sen et al.
CD8 mem vs CD8 eff, Scharer et al.
CD8 mem vs CD8 eff, Philip et al.
CD8 mem vs CD8 exh, Mognol et al.
CD8 mem vs CD8 naive, Mognol et al.
CD8 naive vs CD8 mem , Mognol et al.
CD8 naive vs CD8 prog, Chen et al.
CD8 naive vs CD8 exh, Philip et al.
CD8 naive vs CD8 eff, Sen et al.
CD8 prog d20 vs CD8 exh, Miller et al.
CD8 prog d8 vs CD8 exh Chen et al.
scores up in CD8 exh
scores up in CD8 eff
scores up in CD8 mem
scores up in CD8 naive
scores up in CD8 prog
F
CD8 exh
Original annotations
TEx 0.83
Transitory 0.12
Early Transitory 0.02
Progenitor 0.01
Effector Memory 0.01
Effector2 3.2e-3
Effector 3.2e-3
Naive 1.1e-04
No. of cells
G
Arm day 8
Cl13 day 8
Cl13 day 21
Early exhaustion DAPs our multiomic dataset
CD8 exh vs CD8 eff Sen et al. (day 8)
CD8 exh vs CD8 eff Scharer et al. (day 8)
CD8 exh vs CD8 eff Mognol et al. (day 20)
CD8 eff vs CD8 exh Sen et al. (day 8)
CD8 eff vs CD8 exh Scharer et al. (day 8)
CD8 eff vs CD8 exh Mognol et al. (day 20)
Signature score value
Active CD8
CD8 exh
Early exhaustion DAPs
our multiomic dataset

**Figure S7. Additional analysis of the CD8 T cell scATAC-seq data from Daniel *et al.* (related to Figure 2).**

**(A)** UMAP of the Daniel *et al*. dataset (same as in **Fig. 2C**), colored by Leiden clusters defined in our reanalysis (left), timepoint post infection (center), and data sample (right).

**(B)** Barplot showing the distribution of annotations from the original publication across our newly defined clusters. Clusters are grouped according to our new annotations: cluster 0, CD8 effectors; clusters 1, 2, 4 and 8, CD8 eff/exh mixed1 (mixed effector and exhausted CD8 T cells, subset 1); clusters 3, 9 and 14, CD8 exhausted; clusters 5 and 7, CD8 eff/exh mixed2 (mixed effector and exhausted CD8 T cells, subset 2); cluster 6, CD8 progenitors; clusters 10, 11 and 12, CD8 naive; cluster 13, CD8 effector/memory. The justification of these annotations is presented in panels C–G.

**(C)** Bar plot showing the distribution of our Leiden clusters across each data sample.

**(D)** CDF curves showing the distribution of epigenomic signature scores across Leiden clusters. Each epigenomic signature is defined as the set of peaks with significantly higher accessibility in one condition vs. another, as derived from reanalysis of published ATAC-seq data (as in **Fig. 1F,G, Fig. 2A,D, Fig. S2E–H**). Signatures are grouped by cell state, e.g. all signatures for effector vs. another cell state, for exhausted vs. another cell state, etc. (see panel E for the full list). For each cluster, scores of the relevant signatures are shown in the corresponding color, other ones in gray. This analysis justifies functional annotation of each cluster: for example, cluster 0 has enrichment of the effector signatures (i.e. signatures of effector cells vs. another cell state), clusters 3, 9 and 14 have enrichment of the exhaustion signatures (i.e. signatures of exhausted cells vs. another cell state), etc.

**(E)** Boxplots showing distributions of epigenomic signature scores, grouped by category (as in panel D), for cells from different functional annotations defined in our reanalysis (see panels A–D). Boxplots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; outliers not shown. Despite multiple populations with mixed annotations, this analysis further justifies our key functional annotations: effector cells have higher effector signature scores, exhausted cells have higher exhaustion signature scores, and progenitor cells have higher progenitor signature scores.

**(F–G)** Further justification of the exhaustion annotation across both infection types and timepoints.

**(F)** UMAP plot highlighting the cells identified as exhausted in our reanalysis (clusters 3, 9 and 14). The corresponding barplot shows the distribution of these cells' annotations in the original analysis.

**(G)** *Left*, boxplots showing distributions of epigenomic signature scores for cells identified as exhausted in our reanalysis vs. all other antigen-activated cells (i.e. all other cells except those labeled as naive in the original annotations), separately for cells from Arm day 8, Cl13 day 8 and Cl13 day 21 samples. Shown are the exhaustion vs. effector signatures and effector vs. exhaustion signatures (as in panels D,E). In addition, we derived the epigenomic signature (differentially accessible peaks, DAP, defined as LFC > 1, padj ≤ 0.05) of early exhaustion from our multiomic data (exhausted vs. other, LCMV day 7, as in **Fig. 1C**). Boxplots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; outliers not shown.

*Right*, UMAP showing the score of the epigenomic signature of early exhaustion derived from our data, for all antigen-activated cells (i.e. excluding naive).

Cells that we annotated as exhausted indeed have higher exhaustion vs. effector signature scores and lower effector vs. exhaustion signature scores across all conditions, including at day 8 in response to LCMV Armstrong.

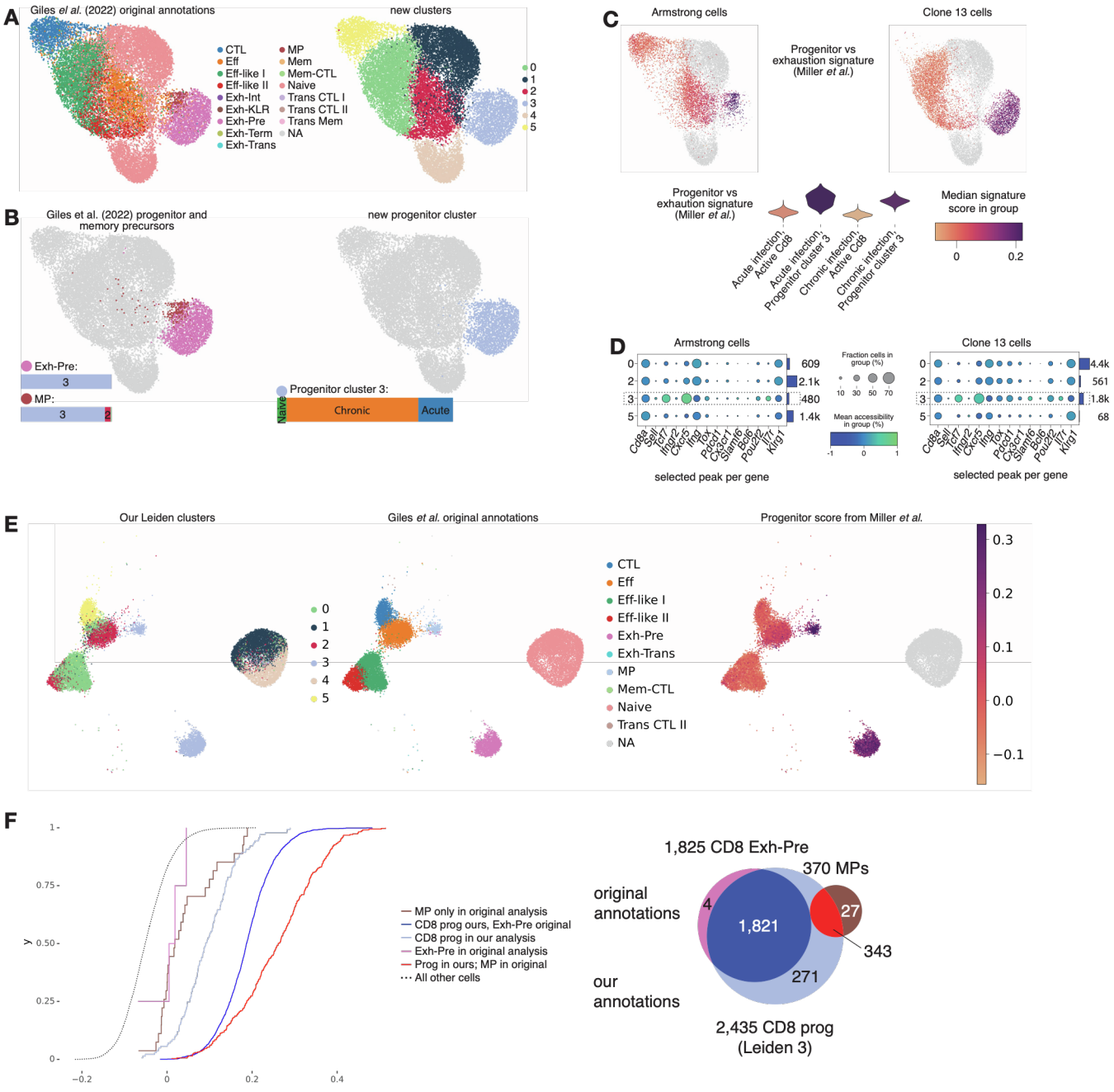



**Figure S8. Reanalysis of the CD8 T cell scATAC-seq data (LCMV, day 8 post infection) from Giles *et al.* (related to Figure 2).**

**(A)** UMAP plots with the cell state annotations from the original publication (Giles *et al.* Nature Immunol 2022, PMID: 36271148) (*left*) and our newly defined clusters (*right*). Cluster 3 in this reanalysis are progenitor CD8 T cells, which included both memory precursor (MP) and pre-exhausted cells (Exh-Pre) defined in the original publication. The justification of this annotation is presented in panels B–D.

**(B)** UMAP showing the progenitor cells. *Left*, cells annotated as Exh-Pre or MP cells in the original publication. The barplot shows the distribution of these annotations within our clusters in our reanalysis for these cells. *Right*, cells annotated as progenitors (cluster 3) in our reanalysis. The barplot shows the distribution of samples of origin for these cells.
**(C)** *Top*, UMAP plots showing the score for the progenitor vs. exhaustion epigenomic signature derived from our bulk ATAC-seq compendium (as in **Fig. 2A,D**), separately for the Arm (*left*) and Cl13 (*right*) samples, for all activated cells (i.e. excluding naive, shown in gray). *Bottom*, violin plot showing the progenitor vs. exhaustion signature score in activated (non-naive) cells and all cluster 3 progenitor cells, separately for Arm and Cl13 infection. This analysis confirms annotation of cells from cluster 3 as progenitors, in both Arm and Cl13 infection.
**(D)** Dotplots showing chromatin accessibility for selected peaks near marker genes in clusters from our reanalysis of the data, separately for the Arm (*left*) and Cl13 (*right*) samples (excluding clusters 1 and 4 that contain only naive cells, and naive cells from other clusters)**,** with the progenitor cluster 3 highlighted. The dot size, fraction of cells with normalized accessibility greater than 0; dot color, mean normalized accessibility in that cluster. Barplots on the right side show the number of cells per cluster. Cluster 3 has enriched activity of progenitor markers *Cxcr5*, *Tcf7*, *Slamf6*, confirming its annotation as progenitors.
**(E)** UMAP from the original paper, colored by our Leiden clusters (*left*), the original annotations (*center*), and the progenitor vs. exhaustion epigenomic signature score (*right*). Both cell subsets annotated as MP and Exh-Pre in the original analysis have the highest progenitor score.
**(F)** CDF plots (*left*) of the progenitor vs. exhaustion epigenomic signature scores for cells defined as progenitors in our annotations (cluster 3) or either Exh-Pre or MP in the original analysis (all other cells are shown in the black dotted line as a reference). In the original analysis, 1,825 cells were annotated as Exh-Pre, and 370 cells annotated as MP. In our analysis, 2,435 cells were annotated as progenitors, including 1,821 cells that were annotated as Exh-Pre (shown in dark blue) and 343 cells annotated as MP (shown in red) in original analysis, and 271 additional cells (shown in light blue); 4 cells were annotated as Exh-Pre (shown in pink) and 27 cells annotated as MP (shown in brown) in the original analysis but not annotated as progenitors in our reanalysis. Venn diagrams (*right*) illustrate the definitions of these groups of cells. The values shown in pink and brown are significantly lower than light blue (FDR < 4.5e−4, two-sided KS test), suggesting that the 271 “light-blue” cells are correctly annotated as progenitors. This analysis also confirms that cells originally annotated as MP are indeed progenitors.

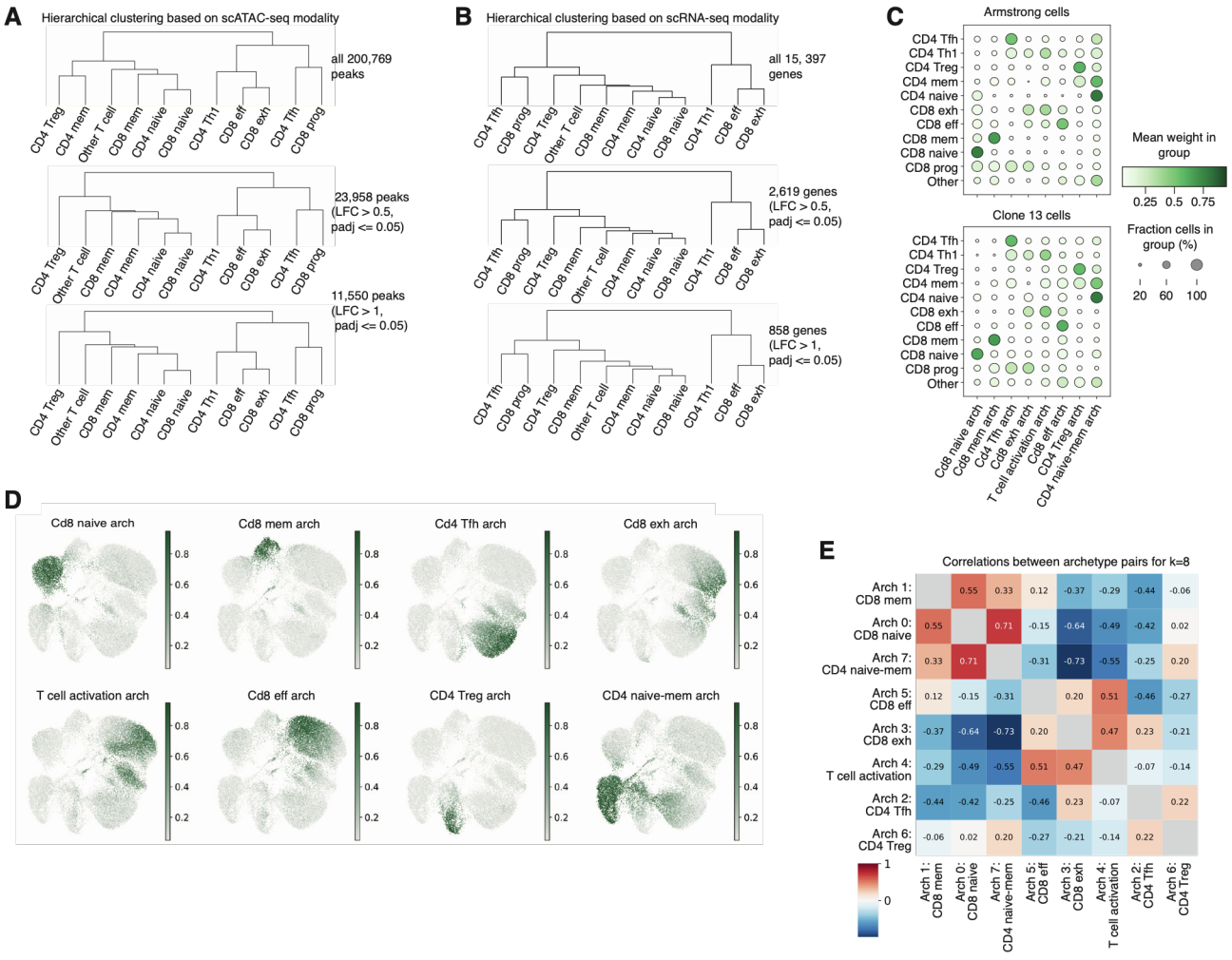

A
Hierarchical clustering based on scATAC-seq modality
all 200,769 peaks
23,958 peaks (LFC > 0.5, padj <= 0.05)
11,550 peaks (LFC > 1, padj <= 0.05)
B
Hierarchical clustering based on scRNA-seq modality
all 15, 397 genes
2,619 genes (LFC > 0.5, padj <= 0.05)
858 genes (LFC > 1, padj <= 0.05)
C
Armstrong cells
Clone 13 cells
Mean weight in group
Fraction cells in group (%)
D
Cd8 naive arch
Cd8 mem arch
Cd4 Tfh arch
Cd8 exh arch
T cell activation arch
Cd8 eff arch
CD4 Treg arch
CD4 naive-mem arch
E
Correlations between archetype pairs for k=8
F
Average accessibility near enriched genes
G
Daniel et al. dataset, day 8 CD8 progenitor cells
Archetype weight
H
Giles et al. dataset, CD8 progenitor cells
Archetype weight
I
Correlation with accessibility
Archetype weight

**Figure S9. Archetypal analysis of the scATAC-seq data (related to Figure 3).**
**(A)** Hierarchical clustering of scATAC-seq data using different sets of peaks (either all peaks, or differentially accessible peaks defined at different levels of stringency) confirms similarity of chromatin accessibility between CD4 Tfh cells and progenitor CD8 T cells (**STAR Methods**).
**(B)** Hierarchical clustering of scRNA-seq data using different sets of genes (either all genes, or differentially expressed genes defined at different levels of stringency) confirms similarity of gene expression between CD4 Tfh cells and progenitor CD8 T cells.
**(C)** Dotplots of archetypal weights for Armstrong and Clone 13 cells. Dot size represents the fraction of cells within a given cell state with strictly positive weight. The color of the dot indicates the average level of archetypal weight in that cell state.
**(D)** UMAP plots showing cell-wise archetypal weights for each of the eight inferred archetypes.
**(E)** Heatmap of pairwise Spearman correlations between archetypes (values of 1.0 on the diagonal not shown), demonstrating that each archetype captures a distinct chromatin accessibility program.
**(F)** CDF plots showing average accessibility of peaks near genes overexpressed in each cell type (LFC > 0, FDR ≤ 0.05) and associated with specific archetypes. Gray lines indicate baseline distributions of average accessibility for all peaks near a gene. Only the archetypes with significantly higher accessibility compared to baseline are shown for each cell state (KS test, FDR ≤ 0.05, and median accessibility difference ≥ 0.03). This analysis illustrates that scATAC-seq archetypes capture regulation of gene expression for the genes overexpressed in functional cell states.
**(G)** CDF plots comparing the distribution of our archetypal weights transferred to the Daniel *et al.* dataset (see reanalysis of this dataset in **Fig. 2, Fig. S7**), focusing on the CD8 progenitor population from Arm day 8 or Cl13 day 8 samples. CD4 Tfh archetype has the strongest weights in CD8 progenitors.
**(H)** CDF plots comparing the distribution of our archetypal weights transferred to the Giles *et al.* dataset (see reanalysis of this dataset in **Fig. S8**), focusing on the CD8 progenitor population. CD4 Tfh archetype has the strongest weights in CD8 progenitors.
**(I)** Scatterplots showing the relationship between archetypal weight and similarity of chromatin accessibility to each archetype. Each point represents one cell. The *x*-axis values represent the archetypal weight in that cell. The *y*-axis values show the Spearman correlation between accessibility profile and the archetypal peak scores over all peaks in that cell. Each cell is colored by its functional annotation, grouped by each of the eight archetypes. Cells with higher archetypal weight tend to have higher correlation of chromatin accessibility with the peak scores of that archetype.

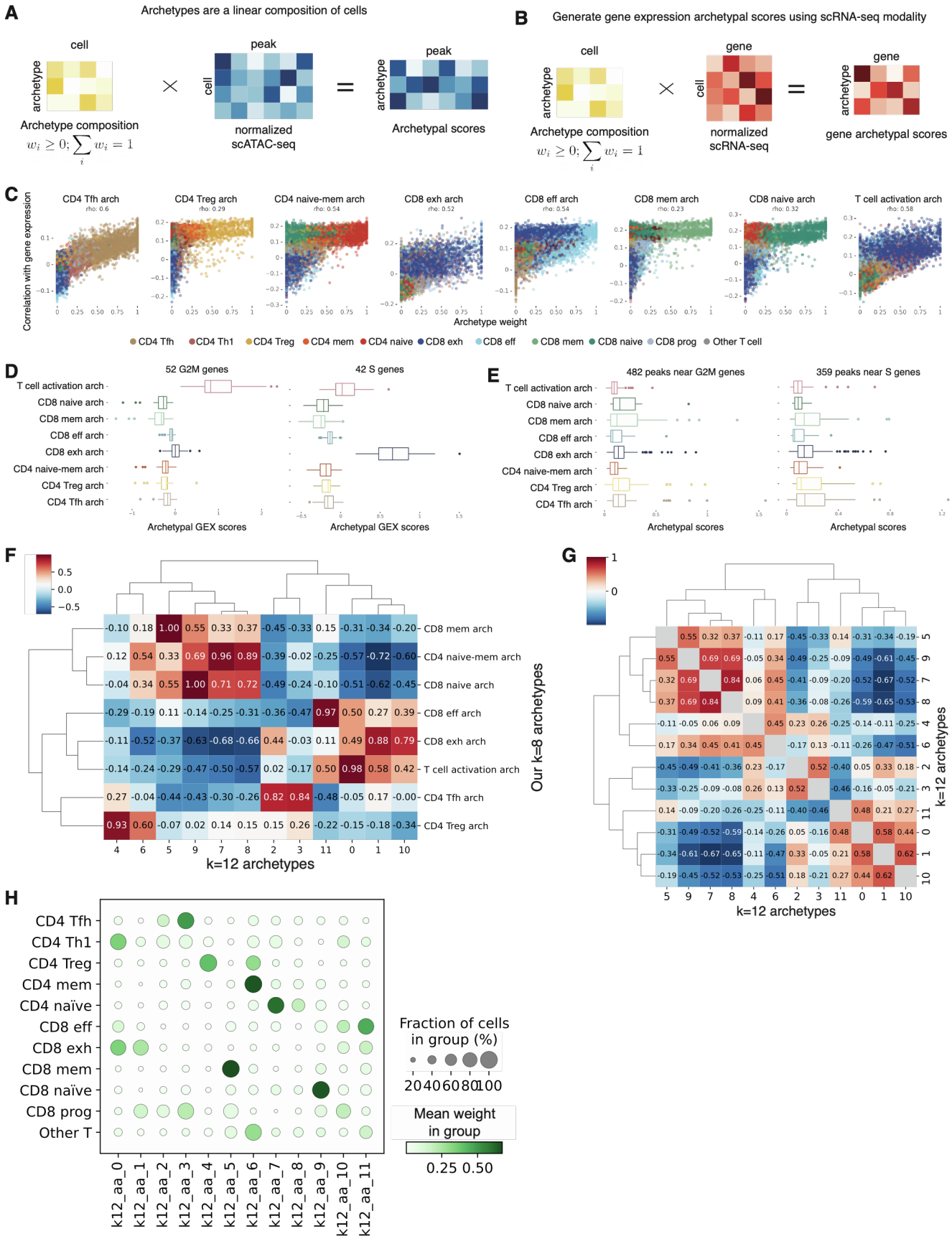

A
Archetypes are a linear composition of cells
cell
archetype
Archetype composition
$w_i \geq 0; \sum_i w_i = 1$
peak
cell
normalized scATAC-seq
peak
archetype
Archetypal scores
B
Generate gene expression archetypal scores using scRNA-seq modality
cell
archetype
Archetype composition
$w_i \geq 0; \sum_i w_i = 1$
gene
cell
normalized scRNA-seq
gene
archetype
gene archetypal scores
C
CD4 Tfh arch
rho: 0.6
CD4 Treg arch
rho: 0.29
CD4 naive-mem arch
rho: 0.54
CD8 exh arch
rho: 0.52
CD8 eff arch
rho: 0.54
CD8 mem arch
rho: 0.23
CD8 naive arch
rho: 0.32
T cell activation arch
rho: 0.58
Correlation with gene expression
Archetype weight
CD4 Tfh
CD4 Th1
CD4 Treg
CD4 mem
CD4 naive
CD8 exh
CD8 eff
CD8 mem
CD8 naive
CD8 prog
Other T cell
D
52 G2M genes
42 S genes
T cell activation arch
CD8 naive arch
CD8 mem arch
CD8 eff arch
CD8 exh arch
CD4 naive-mem arch
CD4 Treg arch
CD4 Tfh arch
Archetypal GEX scores
Archetypal GEX scores
E
482 peaks near G2M genes
359 peaks near S genes
T cell activation arch
CD8 naive arch
CD8 mem arch
CD8 eff arch
CD8 exh arch
CD4 naive-mem arch
CD4 Treg arch
CD4 Tfh arch
Archetypal scores
Archetypal scores
F
0.5
0.0
−0.5
CD8 mem arch
CD4 naive-mem arch
CD8 naive arch
CD8 eff arch
CD8 exh arch
T cell activation arch
CD4 Tfh arch
CD4 Treg arch
k=12 archetypes
Our k=8 archetypes
G
1
0
k=12 archetypes
k=12 archetypes
H
CD4 Tfh
CD4 Th1
CD4 Treg
CD4 mem
CD4 naïve
CD8 eff
CD8 exh
CD8 mem
CD8 naïve
CD8 prog
Other T
k12_aa_0
k12_aa_1
k12_aa_2
k12_aa_3
k12_aa_4
k12_aa_5
k12_aa_6
k12_aa_7
k12_aa_8
k12_aa_9
k12_aa_10
k12_aa_11
Fraction of cells in group (%)
20 40 60 80 100
Mean weight in group
0.25 0.50

**Figure S10. Gene expression patterns associated with scATAC-seq archetypes and additional archetypal analysis (related to Figure 3).**
**(A)** Schematic illustrating how the accessibility profile of each scATAC-seq archetype (an archetypal score for each peak) is defined as a linear combination of cell accessibility profiles (see **STAR Methods**).
**(B)** Schematic showing how gene expression profiles for each scATAC-seq archetype are derived using the cell weights from panel A.
**(C)** Scatterplots showing the relationship between scATAC-seq archetypal weight and the similarity of gene expression to each archetype. (Similar to **Fig. S9I**, but for gene expression instead of chromatin accessibility.) Each point represents one cell. The *x*-axis values represent the scATAC-seq archetypal weight in that cell. The *y*-axis values show the Spearman correlation between the gene expression profile in that cell and the archetypal gene expression profile over all genes. Archetypal gene expression profiles (for scATAC-seq archetypes) were calculated using weighted average of cell expression profiles (see panel B, **STAR Methods**). Each cell is colored by its functional annotation, grouped by each of the eight archetypes. Cells with higher scATAC-seq archetypal weight tend to have higher correlation of gene expression with the archetypal gene expression profile.
**(D)** Boxplots showing gene archetypal scores for genes associated with cell cycle phases. Left: 52 genes related to the G2/M phase; right: 42 genes related to the S phase.
**(E)** Boxplots showing chromatin accessibility archetypal scores for peaks near cell cycle-related genes. Left: 482 peaks near G2/M-phase-associated genes; right: 359 peaks near S-phase-associated genes.
**(F–H)** Results of scATAC-seq archetypal analysis for k = 12.
**(F)** Heatmap showing Spearman correlations between each of our main archetypes (defined for k = 8) and the archetypes defined for k = 12.
**(G)** Heatmap showing Spearman correlations between each pair of archetypes defined for k = 12 (values of 1.0 on the diagonal not shown).
**(H)** Dotplot showing archetypal weights for different cell types for archetypal analysis with k = 12 (similar to **Fig. 3E**).

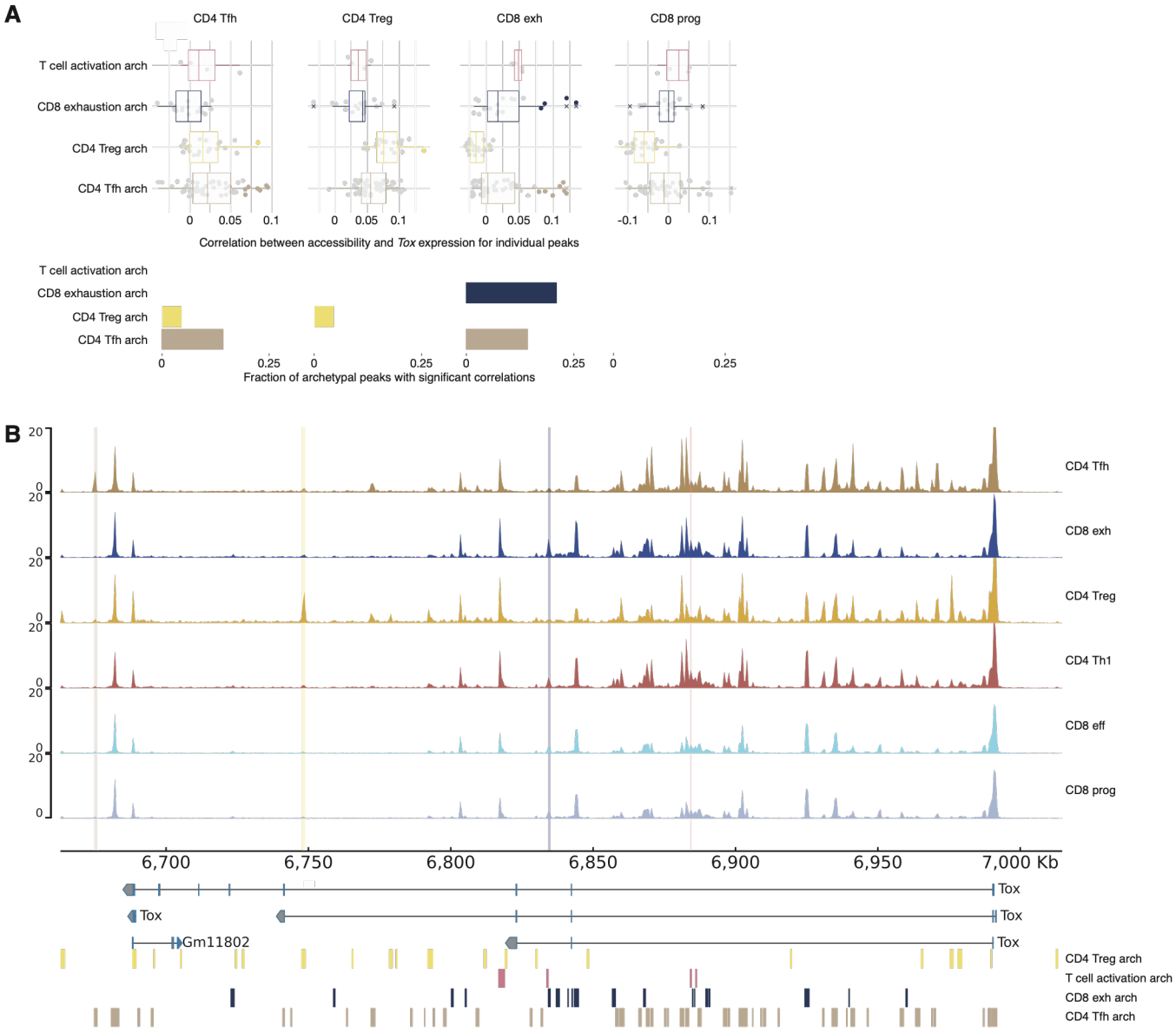



**Figure S11. Archetypal analysis reveals cell state-specific regulation of *Tox* gene expression (related to Figure 3).**

**(A)** Top: Boxplot and scatterplot showing correlations between accessibility of individual peaks near *Tox* and *Tox* gene expression across different cell states. Peaks are grouped by their assigned scATAC-seq archetype. Significant correlations are colored by archetype; non-significant are shown in gray. Bottom: Barplot showing the fraction of peaks within each archetype that significantly correlate with *Tox* expression in each cell population.

**(B)** Pseudobulk scATAC-seq tracks around the *Tox* gene locus for different T cell subsets. Peaks are marked by rectangles at the bottom, colored by their archetype assignment. One representative peak with significant correlation is highlighted for each archetype.

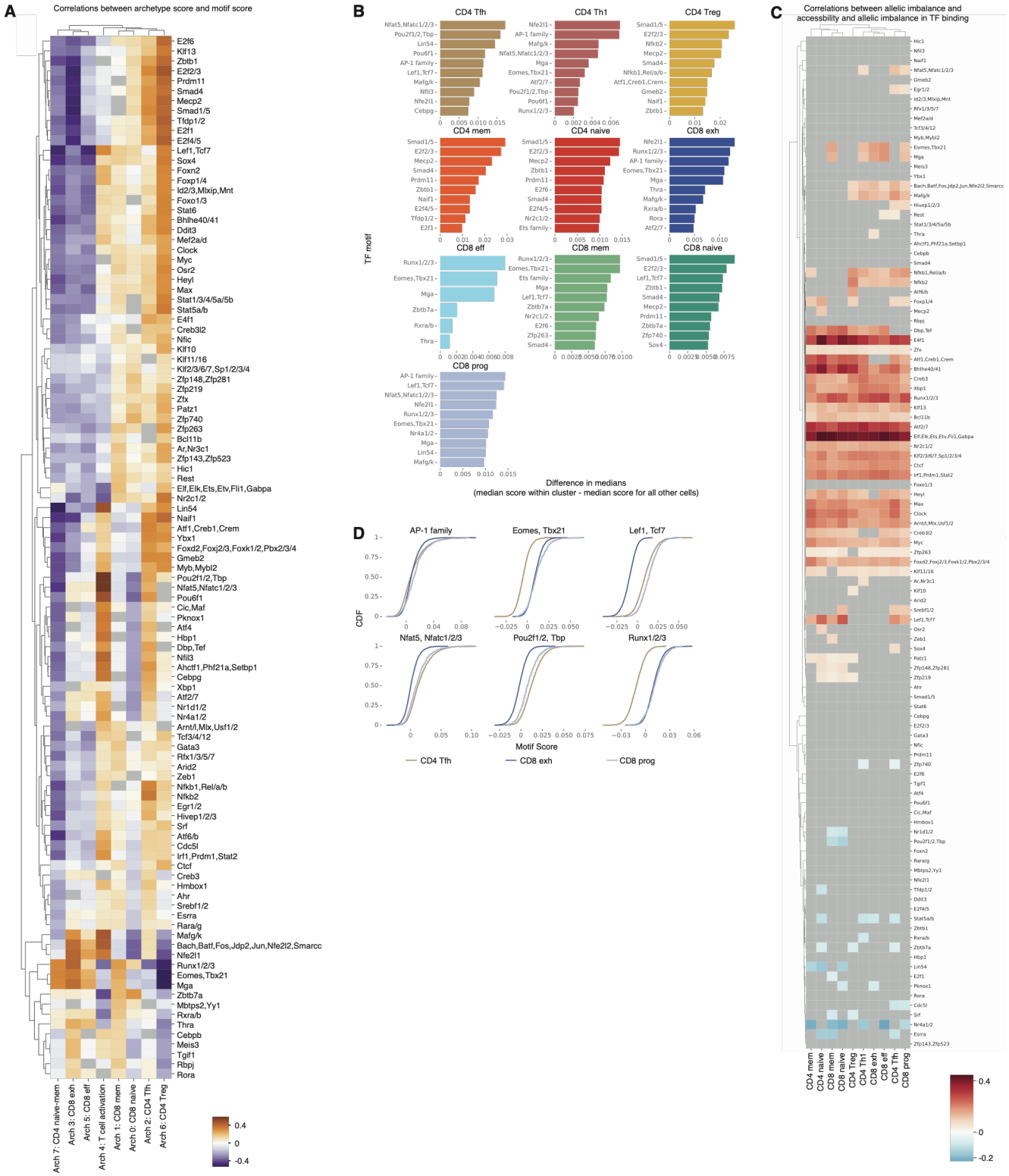



**Figure S12. Transcription factors associated with distinct cell states and archetypes (related to Figure 4).**

**(A)** Clustered heatmap showing Spearman correlations between archetypal weights and motif scores.
**(B)** Motifs most significantly enriched in each functional cell state compared with all other cells. The barplot shows the difference in medians (median motif score in cells of the selected cell state minus the median score for all other cells) for the top 10 motifs per cell state. All enrichments are significant (FDR ≤ 0.05, LFC > 0, one-sided KS test, with Benjamini-Hochberg correction).
**(C)** Clustered heatmap showing Spearman correlations between allelic imbalance of motif matches and chromatin accessibility across functional T cell states. Non-significant correlations (FDR > 0.05) are shown in gray.
**(D)** CDF plots showing the distributions of motif scores for selected motifs in three cell populations: CD4 Tfh cells, exhausted CD8 T cells, and progenitor CD8 T cells. This analysis illustrates that motifs most active in progenitors are a combination of motifs most active in CD8 exhaustion and CD4 Tfh cells.

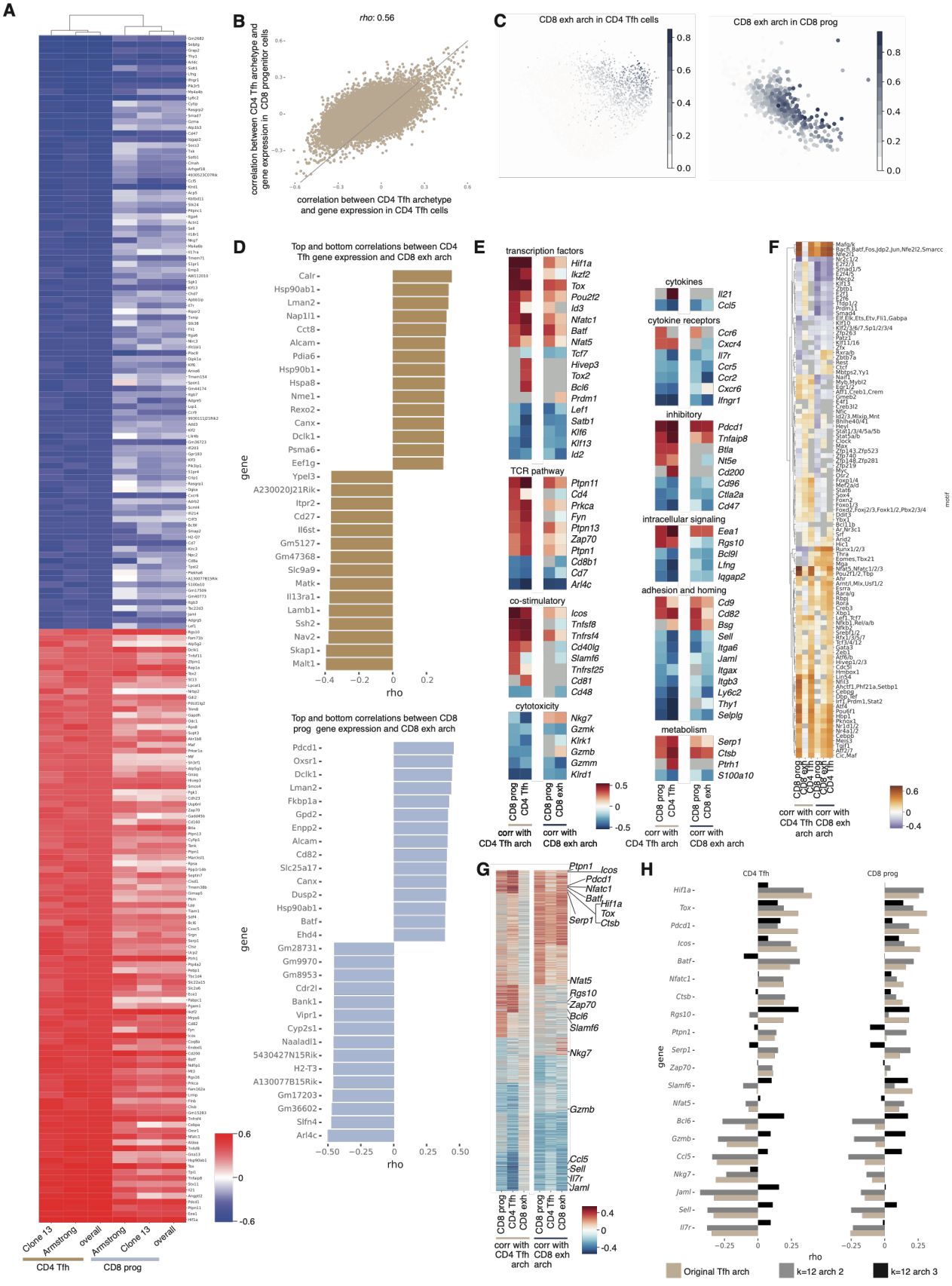

A
Clone 13
Armstrong
overall
Armstrong
Clone 13
overall
CD4 Tfh
CD8 prog
B
rho: 0.56
correlation between CD4 Tfh archetype and gene expression in CD8 progenitor cells
correlation between CD4 Tfh archetype and gene expression in CD4 Tfh cells
C
CD8 exh arch in CD4 Tfh cells
CD8 exh arch in CD8 prog
D
Top and bottom correlations between CD4 Tfh gene expression and CD8 exh arch
Calr
Hsp90ab1
Lman2
Nap1l1
Cct8
Alcam
Pdia6
Hsp90b1
Hspa8
Nme1
Rexo2
Canx
Dclk1
Psma6
Eef1g
Ypel3
A230020J21Rik
Itpr2
Cd27
Il6st
Gm5127
Gm47368
Slc9a9
Matk
Il13ra1
Lamb1
Ssh2
Nav2
Skap1
Malt1
gene
rho
Top and bottom correlations between CD8 prog gene expression and CD8 exh arch
Pdcd1
Oxsr1
Dclk1
Lman2
Fkbp1a
Gpd2
Enpp2
Alcam
Cd82
Slc25a17
Canx
Dusp2
Hsp90ab1
Batf
Ehd4
Gm28731
Gm9970
Gm8953
Cdr2l
Bank1
Vipr1
Cyp2s1
Naaladl1
5430427N15Rik
H2-T3
A130077B15Rik
Gm17203
Gm36602
Slfn4
Arl4c
gene
rho
E
transcription factors
Hif1a
Ikzf2
Tox
Pou2f2
Id3
Nfatc1
Batf
Nfat5
Tcf7
Hivep3
Tox2
Bcl6
Prdm1
Lef1
Satb1
Klf6
Klf13
Id2
TCR pathway
Ptpn11
Cd4
Prkca
Fyn
Ptpn13
Zap70
Ptpn1
Cd8b1
Cd7
Arl4c
co-stimulatory
Icos
Tnfsf8
Tnfrsf4
Cd40lg
Slamf6
Tnfrsf25
Cd81
Cd48
cytotoxicity
Nkg7
Gzmk
Klrk1
Gzmb
Gzmm
Klrd1
CD8 prog
CD4 Tfh
CD8 exh
corr with CD4 Tfh arch
corr with CD8 exh arch
cytokines
Il21
Ccl5
cytokine receptors
Ccr6
Cxcr4
Il7r
Ccr5
Ccr2
Cxcr6
Ifngr1
inhibitory
Pdcd1
Tnfaip8
Btla
Nt5e
Cd200
Cd96
Ctla2a
Cd47
intracellular signaling
Eea1
Rgs10
Bcl9l
Lfng
Iqgap2
adhesion and homing
Cd9
Cd82
Bsg
Sell
Itga6
Jaml
Itgax
Itgb3
Ly6c2
Thy1
Selplg
metabolism
Serp1
Ctsb
Ptrh1
S100a10
F
motif
corr with CD4 Tfh arch
corr with CD8 exh arch
G
Ptpn1
Icos
Pdcd1
Nfatc1
Batf
Hif1a
Tox
Serp1
Ctsb
Nfat5
Rgs10
Zap70
Bcl6
Slamf6
Nkg7
Gzmb
Ccl5
Sell
Il7r
Jaml
H
CD4 Tfh
CD8 prog
gene
rho
Hif1a
Tox
Pdcd1
Icos
Batf
Nfatc1
Ctsb
Rgs10
Ptpn1
Serp1
Zap70
Slamf6
Nfat5
Bcl6
Gzmb
Ccl5
Nkg7
Jaml
Sell
Il7r
Original Tfh arch
k=12 arch 2
k=12 arch 3

**Figure S13. Tfh archetypal gradient in Tfh cells and progenitor CD8 T cells (related to Figure 5).**
**(A)** Heatmap showing Spearman correlations between Tfh archetypal weights and gene expression in Tfh cells and progenitor CD8 T cells. Genes selected for this plot are the top 100 most positively and top 100 most negatively correlated with the Tfh archetype across all Tfh cells. Correlations are computed separately for each cell state, as well as within Armstrong-only and Clone 13-only subsets. The strong similarity of these transcriptional associations across both cell populations and both infection models confirms a shared Tfh-like differentiation gradient in both Tfh and CD8 progenitor cells.
**(B)** Scatterplot comparing gene expression correlations with the Tfh archetype in CD8 progenitor cells vs. Tfh cells. This analysis further illustrates that genes with most strongly positively or negatively correlated expression are similar between Tfh cells and CD8 progenitors, suggesting a similar Tfh-like differentiation gradient across both cell populations.
**(C)** UMAP plots highlighting CD8 exhaustion archetypal weights in Tfh cells (*left*) and progenitor CD8 T cells (*right*).
**(D)** Barplots showing the top 15 most correlated and anti-correlated gene expression with the CD8 exhaustion archetype in Tfh cells (*top*) and progenitor CD8 T cells (*bottom*).
**(E)** Heatmap showing Spearman correlations between archetypal weights and normalized gene expression within progenitor CD8 T cells and Tfh cells for selected genes from different functional categories (similar to **Fig. 5C**). Left two columns: correlations with the Tfh archetypal weights for progenitor CD8 T cells and Tfh cells. Right two columns: correlations with CD8 exhaustion archetypal weights for progenitor and exhausted CD8 T cells. The gene expression values were averaged over neighbors in the scRNA-seq kNN graph (k = 20). Non-significant correlations (FDR > 0.05) are shown in gray.
**(F)** Clustered heatmap of correlations between archetypal weights and transcription factor motif scores for all motifs, highlighting the Tfh and CD8 exhaustion archetypes across three selected cell populations.
**(G)** Heatmap showing Spearman correlations between Tfh or CD8 exhaustion archetypal weights and normalized gene expression within Tfh, progenitor and exhausted CD8 T cells. Genes with the strongest positive or negative correlations between exhausted CD8 T cells and CD8 exhaustion archetype, Tfh cells and Tfh archetype, and CD8 progenitors and either archetype were selected (total 1,131 genes). The gene expression values were averaged over neighbors in the scRNA-seq kNN graph (k = 20).
**(H)** Barplots comparing Spearman correlations between expression of functionally important T cell genes and the Tfh archetype across different resolutions of archetypal analysis (k = 8 vs. k = 12 archetypes), evaluated within Tfh and CD8 progenitor cells. The analysis compares the primary Tfh archetype (from the analysis for k = 8) against archetypes 2 and 3 from the archetypal analysis for k = 12 which showed the strongest association with Tfh cells (see **Fig. S10F–H**). Genes are ordered by their correlation with the primary Tfh archetype (defined for k = 8) within Tfh cells. This analysis confirms that for k = 12, archetype 2 captures the same Tfh differentiation gradient as the Tfh archetype in the primary analysis for k = 8.

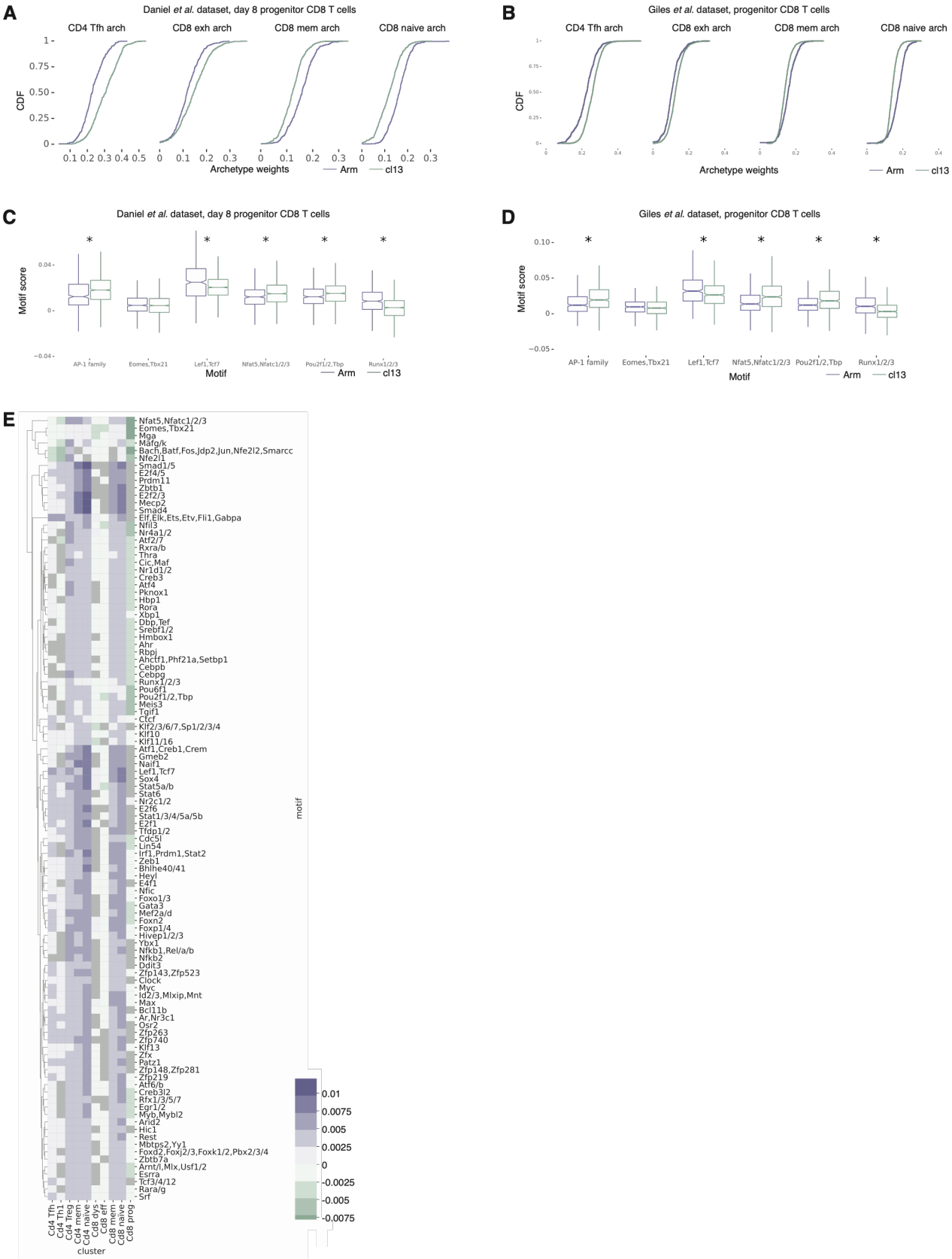

A
Daniel et al. dataset, day 8 progenitor CD8 T cells
CD4 Tfh arch
CD8 exh arch
CD8 mem arch
CD8 naive arch
CDF
Archetype weights
Arm
cl13
B
Giles et al. dataset, progenitor CD8 T cells
CD4 Tfh arch
CD8 exh arch
CD8 mem arch
CD8 naive arch
CDF
Archetype weights
Arm
cl13
C
Daniel et al. dataset, day 8 progenitor CD8 T cells
Motif score
AP-1 family
Eomes,Tbx21
Lef1,Tcf7
Nfat5,Nfatc1/2/3
Pou2f1/2,Tbp
Runx1/2/3
Motif
Arm
cl13
D
Giles et al. dataset, progenitor CD8 T cells
Motif score
AP-1 family
Eomes,Tbx21
Lef1,Tcf7
Nfat5,Nfatc1/2/3
Pou2f1/2,Tbp
Runx1/2/3
Motif
Arm
cl13
E
Nfat5,Nfatc1/2/3
Eomes,Tbx21
Mga
Maf/g/k
Bach,Batf,Fos,Jdp2,Jun,Nfe2l2,Smarcc
Nfe2l1
Smad1/5
E2f4/5
Prdm11
Zbtb1
E2f2/3
Mecp2
Smad4
Elf,Elk,Ets,Etv,Fli1,Gabpa
Nfil3
Nr4a1/2
Atf2/7
Rxra/b
Thra
Clc,Maf
Nr1d1/2
Creb3
Atf4
Pknox1
Hbp1
Rora
Xbp1
Dbp,Tef
Srebf1/2
Hmbox1
Ahr
Rbpj
Ahctf1,Phf21a,Setbp1
Cebpb
Cebpg
Runx1/2/3
Pou6f1
Pou2f1/2,Tbp
Meis3
Tgif1
Ctcf
Klf2/3/6/7,Sp1/2/3/4
Klf10
Klf11/16
Atf1,Creb1,Crem
Gmeb2
Naif1
Lef1,Tcf7
Sox4
Stat5a/b
Stat6
Nr2c1/2
E2f6
Stat1/3/4/5a/5b
E2f1
Tfdp1/2
Cdc5l
Lin54
Irf1,Prdm1,Stat2
Zeb1
Bhlhe40/41
Heyl
E4f1
Nfic
Foxo1/3
Gata3
Mef2a/d
Foxn2
Foxp1/4
Hivep1/2/3
Ybx1
Nfkb1,Rel/a/b
Nfkb2
Ddit3
Zfp143,Zfp523
Clock
Myc
Id2/3,Mlxip,Mnt
Max
Bcl11b
Ar,Nr3c1
Ogr2
Zfp263
Zfp740
Klf13
Zfx
Patz1
Zfp148,Zfp281
Zfp219
Atf6/b
Creb3l2
Rfx1/3/5/7
Egr1/2
Myb,Mybl2
Arid2
Hic1
Rest
Mbtps2,Yy1
Foxd2,Foxj2/3,Foxk1/2,Pbx2/3/4
Zbtb7a
Arnt/l,Mlx,Usf1/2
Esrra
Tcf3/4/12
Rara/g
Srf
Cd4 Tfh
Cd4 Th1
Cd4 Treg
Cd4 mem
Cd4 naive
Cd8 exh
Cd8 eff
Cd8 mem
Cd8 naive
Cd8 prog
cluster
motif
0.01
0.0075
0.005
0.0025
0
-0.0025
-0.005
-0.0075

**Figure S14. Validation of differential regulatory dynamics between Armstrong and Clone 13 responses (related to Figure 6).**

**(A)** CDF plots of our archetypal weights transferred to the Daniel *et al.* dataset (see our reanalysis of this dataset presented in **Fig. 2, Fig. S7**), comparing Armstrong and Clone 13 cells within the CD8 progenitor population. All distributions are significantly different (two-sided KS test, FDR < 2.4e−13).

**(B)** CDF plots of our archetypal weights transferred to the Giles *et al.* dataset (see our reanalysis of this dataset presented in **Fig. S8**), comparing Armstrong and Clone 13 cells wtihin the CD8 progenitor population (Leiden cluster 3). All distributions are significantly different (two-sided KS test, FDR < 7.6e−19).

Together, the analysis in panels A–B confirms transferability of our archetypes to other datasets and independently validates that cells in Clone 13 undergo accelerated activation compared with Armstrong, because they tend to have larger Tfh and CD8 exhaustion archetypal weights.

**(C–D)** Boxplots comparing the selected TF motif scores between Armstrong and Clone 13 cells in the CD8 progenitor population in (C) the Daniel *et al.* and (D) the Giles *et al.* datasets. Significant difference marked with asterisk (two-sided KS test, FDR < 0.05). These results independently validate the differential TF motif activity between Arm and Cl13 observed in our main analysis (see **Fig. 6F**).

**(E)** Heatmap comparing the motif scores of selected motifs between Armstrong and Clone 13 cells for all subpopulations in our dataset. Color represents the difference in medians between the Armstrong score and the Clone 13 score. Non-significant differences are shown in gray (FDR > 0.05, two-sided KS test).