Multi-Omics Atlas-Assisted Discovery of Transcription Factors for Selective T Cell State Programming

Authors

H. Kay Chung[1,2,3*], Cong Liu[4*], Alexander N. Jambor[4], Brian P. Riesenberg[2], Ming Sun[1], Eduardo Casillas[1], Brent Chick[5], Jun Wang[4], Shixin Ma[1], Bryan Mcdonald[1], Peixiang He[6,7], Qiyuan Yang[1], Timothy Chen[1], Siva Karthik Varanasi[1], Michael LaPorte[1], Thomas H. Mann[1], Dan Chen[1], Filipe Hoffmann[1], Victoria Tripple[1], Josephine Ho[5], Ukrae H. Cho[2,3], Jennifer Modliszewski[2], April Williams[8], Longwei Liu[6,7], Yingxiao Wang[6], Diana C. Hargreaves[5], Jessica E. Thaxton[2,3], Susan M. Kaech[1#], Wei Wang[4#]

[1]NOMIS Center for Immunobiology and Microbial Pathogenesis, Salk Institute for Biological Studies, La Jolla, CA, USA

[2]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[3]Department of Cell Biology & Physiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[4] Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA, USA.

[5] Molecular and Cell Biology laboratory, Salk Institute for Biological Studies, La Jolla, CA, USA

[6]Alfred E. Mann Department of Biomedical Engineering, University of Southern California, Los Angeles, CA, USA

[7]Department of Bioengineering, Institute of Engineering in Medicine, University of California, La Jolla, CA, USA

[8]The Razavi Newman Integrative Genomics and Bioinformatics Core (IGC), Salk Institute for Biological Studies, La Jolla, CA, USA

* Equal contribution, # Co-corresponding

Transcription factors (TFs) regulate the differentiation of T cells into diverse states with distinct functionalities. To precisely program desired T cell states in viral infections and cancers, we generated a comprehensive transcriptional and epigenetic atlas of nine CD8$^+$ T cell differentiation states for TF activity prediction. Our analysis catalogued TF activity fingerprints of each state, uncovering new regulatory mechanisms that govern selective cell state differentiation. Leveraging this platform, we focused on two critical T cell states in tumor and virus control: terminally exhausted T cells (TEX$_{term}$), which are dysfunctional, and tissue-resident memory T cells (T$_{RM}$), which are protective. Despite their functional differences, these states share significant transcriptional and anatomical similarities, making it both challenging and essential to engineer T cells that avoid TEX$_{term}$ differentiation while preserving beneficial T$_{RM}$ characteristics. Through *in vivo* CRISPR screening combined with single-cell RNA sequencing (Perturb-seq), we validated the specific TFs driving the TEX$_{term}$ state and confirmed the accuracy of TF specificity predictions. Importantly, we discovered novel TEX$_{term}$-specific TFs such as ZSCAN20, JDP2, and ZFP324. The deletion of these TEX$_{term}$-specific TFs in T cells enhanced tumor control and synergized with immune checkpoint blockade. Additionally, this study identified multi-state TFs like HIC1 and GFI1, which are vital for both TEX$_{term}$ and T$_{RM}$ states. Furthermore, our global TF community analysis and Perturb-seq experiments revealed how TFs differentially regulate key processes in T$_{RM}$ and TEX$_{term}$ cells, uncovering new biological pathways like protein catabolism that are specifically linked to TEX$_{term}$ differentiation. In summary, our platform systematically identifies TF programs across diverse T cell states, facilitating the engineering of specific T cell states to improve tumor control and providing insights into the cellular mechanisms underlying their functional disparities.

Introduction

Cell states can be described as the range of cellular phenotypes arising from a defined cell type's interaction with its environment. Within the immune system, T cells possess multiple differentiation states, particularly as naive T cells differentiate into diverse states with different functionality and trafficking patterns in various immune environments, such as tumors and virus infections[1–4]. As transcription factors (TFs) govern cell state differentiation[5,6], understanding the mechanisms by which TFs influence cell states is vital for programming beneficial cell states for therapeutic applications[7,8]. One promising clinical application of cell state engineering is the enhancement of CD8$^+$ cell-based therapies, such as TIL or CAR-

T adoptive transfer therapy. However, identifying the crucial TFs that control different $CD8^+$ T cell states is challenging due to the significant heterogeneity among these states and the similarities in their transcriptomes and gene regulatory networks, even when these cell states are functionally distinct.

The transcriptional similarity between cell states with differing functionalities is exemplified by two key $CD8^+$ T cell states involved in controlling solid tumors and viruses: the beneficial tissue-resident memory (T_{RM}) cell state and the dysfunctional terminally exhausted (TEX_{term}) cell state. Many studies have shown that tumor-infiltrating lymphocytes (TILs) with T_{RM} characteristics correlate with better survival in solid tumor patients[9–13]. Conversely, during persistent antigen stimulation scenarios such as chronic virus infection (e.g. HIV) or tumor, T cells progressively express diverse inhibitory receptors such as PD1, lose memory potential, and eventually effector functions. This process is known as T cell exhaustion (TEX)[14–17], and cells in this trajectory eventually adopt the TEX_{term} cell state. TEX_{term} cells express higher and more diverse inhibitory receptors (e.g., TIM3, and CD101), lack effector and proliferation capacity, and do not respond to immune checkpoint blockade (ICB) such as anti-PD1 antibody treatment[18–20]. Consequently, the expression of TEX_{term} markers is often associated with poor survival of solid tumor patients[21–24]. Despite their distinct functional impact on cancer outcomes, TEX_{term} and T_{RM} cells both preferentially reside in tissues[1,3] and display remarkable similarities in their transcriptional profiles, including key regulatory TFs such as BLIMP1[14,25–27], BHLHE40[28,29], and NR4A2[9,30,31](Fig. 1a,b, Extended Data Fig. 1a-c). These two cell states even exhibit highly correlated open chromatin regions (Extended data Fig. 1d), complicating the precise identification of TFs whose disruption may selectively inhibit TEX_{term} while preserving T_{RM} cell development. Given that many TFs are commonly expressed across different $CD8^+$ T cell states and differentiation trajectories, a sophisticated and precise bioinformatics approach is crucial to pinpoint the bona fide cell-state-specifying TFs essential for T cell programming.

We hypothesized that the challenges of identifying key TFs in control of selective $CD8^+$ T cell differentiation states can be effectively addressed through a systematic comparison of TF activity across the T cell differentiation landscape. Two considerations were postulated for accurate TF activity prediction: first, TF activity is not always proportional to the TF expression level as it is influenced by post-translational modifications, co-factors, and target accessibility[32]. Second, the effect of a TF on target genes propagates over the genetic network, and thus, the global impact of the TF activity needs to be

assessed at the systems level. To address these considerations, a multi-omics atlas-assisted platform was developed, integrating transcriptomic and chromatin data of nine major $CD8^+$ T cell states. This platform analyzes transcriptional networks and genome-wide, "global" influencers in each cell state and identifies cell-state "*selective*" or "*shared*" TFs. Our platform also enables integrative analysis of TF communities and TF target genes (referred to as "regulatees"), providing insight into how the milieu of TFs in each cell state collectively guides specific differentiation states.

Transcriptional and epigenetic atlas generated a transcription factor catalogue predicts to specify certain T cell states

Elucidating the TFs and their regulatory networks that govern each cell state is crucial for accurately programming specific cellular states of interest. Our initial objective was to create a comprehensive catalogue of TF activity across diverse $CD8^+$ T cell states by integrating our TF activity analysis pipeline, Taiji[33–35], with comparative statistical analysis. Taiji builds gene regulatory networks (GRN) in each cell state by linking TF binding sites to target genes utilizing chromatin accessibility and gene expression data where genes and interactions are defined as "node" and "edge", respectively[33,34]. The personalized PageRank algorithm[36] assesses the global influences of TFs, which naturally considers the impact of upstream regulators and downstream target genes, including the feedback, in an iterative manner. In this study, the node weights were determined by gene expression levels, allocating higher ranks to the TFs that regulate more differentially expressed genes. The edge weights were set proportional to the expression levels and binding strength of TFs and open chromatin intensity (Fig. 1c).

Our goal was to calculate global influence of TFs using a personalized PageRank algorithm in each and every cell state, allowing for the identification of TFs that are differentially important across various cell states. Thus, we first generated ATAC-seq and RNA-seq data from 121 $CD8^+$ T cell samples across nine states, derived from well-defined acute and chronic lymphocytic choriomeningitis virus (LCMV) infections (Extended Data Fig. 2). In mice infected with the acute LCMV-Armstrong strain, memory precursor (MP), terminal effector (TE), effector memory (T_{EM}), central memory (T_{CM}), and tissue-resident memory (T_{RM}) are found. In chronic LCMV-Clone 13 infected mice, heterogenous exhaustion cell states are observed such as progenitors of exhaustion (TEX_{prog}), effector-like exhaustion (TEX_{eff}), and terminal exhaustion (TEX_{term}).

Next, we conducted an unbiased comparative analysis using statistical filtering to understand the specificity of TF activity across the major nine $CD8^+$ T cell states (Extended Data Fig. 3a, Extended Data Table 1). A total of 136 TFs were identified as "single-state-specific", with each cell state selectively containing ~12-19 unique TFs (Fig. 1d, Extended Data Fig. 3b). This category included many novel TFs such as *Hoxa7* in Naïve, *Snai1* in T_{RM}, *Hey1* in TEX_{prog}, *Sox8* in TEX_{eff}, and *Zscan20* in TEX_{term} cells. On the other hand, 119 TFs were identified as key regulators in more than one cell state, termed "multi-state TFs" (Fig. 1e). These include TFs, such as TCF-1 and T-bet. TCF-1 is a known driver of Naive-, MP-, and TEX_{prog} states, all of which are multipotent and have high proliferative capacity[3,20]. MP and TE cells share T-bet as well as an additional 22 key TFs that program early activation (40% of TE or MP TFs). Multi-state TFs like *Vax2*, *Batf*, *Irf8,* and *Stat1* were more enriched within the set of exhaustion-associated cell states (TEX_{prog}, TEX_{eff}, and TEX_{term}). This analysis of TF specificity also shows that T_{RM} shares the most TFs with TEX_{term} compared to other cell states (e.g., *Egr2*, *Crem*, *Prdm1*, Extended Data Fig. 3c) in agreement with their high degree of transcriptional and epigenetic similarity (Fig. 1b; Extended Data Fig 1b-d).

Comparative analysis of TFs in TEX_{term} and T_{RM} identifies candidate TFs for state-selective T cell programming

Despite the transcriptional overlap between these two cell states, our Taiji pipeline possesses the predictive power to generate a catalogue of TFs that are specifically active in TEX_{term} cells but not in T_{RM} cells. This may aid in developing better antitumor or antiviral immunotherapies wherein one can engineer T cells away from T cell exhaustion and toward functional effector cell states without negatively impacting their ability to form into T_{RM} cells in tissues and tumors. To this end, we identified 20 and 34 state-specific TFs of T_{RM} or TEX_{term} cells, respectively, and 30 multi-state TFs active in both cell states (Fig. 1f-h; Extended Data Fig. 3a gray boxes; Extended Data Table 2). TEX_{term}-specific TFs included many previously unreported TFs such as *Zscan20*, *Jdp2*, *Zfp324*, *Zfp143*, *Zbtb49*, and *Arid3a* (Fig. 1f). T_{RM}-specific TFs included *Fosb*, *Zfp692*, *Atf4*, *Pbx4*, *Junb*, *Nr4a1*, *and Klf6* (Fig. 1g). Of the TEX_{term} and T_{RM} multi-state TFs, some were well-known to function in the development of *both* cell states, such as *Nr4a2*[12], *Bhlhe40*[28], *Prdm1*[27,37], whereas others such as *Hic1*[38] and *Gfi1*[39] were not, identifying them as new multi-state TFs to consider (Fig. 1h).

To evaluate the predicted TFs governing specific T cell differentiation pathways, we identified dynamic activity patterns of TF groups, termed "TF waves" (Extended Data Fig. 4). Seven TF waves associated with distinctive biological pathways were identified. For example, the T_{RM} TF wave (Fig. 1i) includes several members of the AP-1 family (e.g. *Atf3*, *Fosb*, *Fosl2*, and *Jun*), which aligns well with a recent report of their roles in T_{RM} formation[40]. This wave was uniquely linked to the transforming growth factor beta (TGFβ) response pathway (Extended Data Fig. 4e). Conversely, the TEX TF wave (including TEX_{prog}, TEX_{eff}, and TEX_{term}) was characterized by a distinct set of TFs, such as *Irf8*, *Jdp2*, *Nfatc1*, and *Vax2*, among others, that correlated with pathways related to PD1 and senescence (Fig. 1j, Extended Data Fig. 4e).

TF community analysis reveals biological processes selectively associated with T_{RM} and TEX_{term} cell states

Having cataloged state-specific and multi-state TFs, we next aimed to test the hypothesis that, despite the significant overlap in TF activity between T_{RM} and TEX_{term} cells, distinct TF collaborations regulate specific biological pathways and functions that decide the cell state. Thus, we first constructed TF-TF association network in each cell state and the adjacency matrix based on their shared regulatees (Fig. 2a), i.e. as TFs share a greater number of regulatees, the cooperativity index between two TFs is larger. Although multi-state TFs like HIC1 (Fig. 2b), PRDM1, FLI1, and GFI1 (Extended Data Fig. 5a-c) were predicted to be active in both T_{RM} and TEX_{term} cells, they appeared to associate with distinct sets of TFs in different cell states, suggesting a mechanism for their divergent network in T_{RM} and TEX_{term} cells. Another observation was that some TEX_{term}-specific TFs, including ZFP324, JDP2, and ZSCAN20, were predicted to cooperate with distinct sets of TFs within the TEX_{term} cells, suggesting they control unique cellular processes in TEX_{term} cells (compare Fig. 2c with Extended Data Fig. 5d,e).

Next, the TF-TF association networks were further grouped into distinct "TF neighbor communities" in TEX_{term} and T_{RM} cells, respectively, which are linked to different biological processes (Fig. 2a, Extended Data Table 4). The overall topology of these communities was influenced by multi-state TFs active in both cell states, while cell state-specific TFs created unique interaction patterns between the communities (Fig. 2d-f). Pathway analysis of the regulatees in each community suggested that T_{RM}- or TEX_{term}-specific

TFs within each community controlled different biological processes (Fig. 2g,h). For example, in T_{RM} cells, community 3 was associated with cell-cell adhesion and response to TGFβ, and community 1 was associated with RNA metabolism, but in TEX_{term} cells, community 3 was linked to apoptosis, and community 1 was coupled to catabolism, proteolysis, ubiquitin-proteasome, and autophagy.

The integrative TF network analysis identifies new TEX_{term}-specific pathways

Given that protein degradation and proteasome pathway came out as a strong yet previously uncharacterized biological program in TEX_{term} cells, we investigated the proteasome activity in the terminally exhausted cells. A higher KEGG proteasome gene set score was noted in the $CD8^+$ T cells in TEX_{term} clusters of Non-Small Cell Lung Cancer (NSCLC) patients[41] and murine TILs from MCA-205 sarcomas (Extended Data Figs. 6d,e). Consistent with mRNA enrichment, TEX_{term} $CD8^+$ T cells from two well-established *in vivo* models of T cell exhaustion—chronic LCMV-Clone 13 infection (Fig. 2i) and MCA-205 sarcomas (Fig. 2j)—exhibited the highest proteasome activity. This was monitored by the fluorescent intensity of a proteasome activity probe that we have previously validated in T cells[42]. This heightened proteasome activity was in part by antigen-recognition, as shown by comparing proteasome activity between bystander (OT-1) and tumor-specific (P14) $CD8^+$ T cells, which recognize the LCMV GP33-41 epitope expressed by B16-GP33 tumors (Fig. 2k).

To test whether the increased proteasome activity correlated with T cell dysfunction, TILs with high and low intensities of fluorescent proteasome activity probe were sorted from donor tumor-bearing mice and adoptively transferred into recipient mice with tumors. $Proteasome^{high}$ $CD8^+$ TILs exhibited reduced tumor control compared to $proteasome^{low}$ counterparts (Fig. 2l). This *in vivo* validation demonstrated the accuracy and applicability of the TF-designated communities in defining biological processes predicted to be more active in one cell state over another. In summary, our global TF analysis uncovered differences in biological pathways between TEX_{term} and T_{RM} cells, including some less recognized ones, such as the ubiquitin-proteasome pathway.

Systematic validation of TEX_{term}-driving TF via *in vivo* CRISPR screening coupled with scRNA-seq

The Taiji pipeline enabled comparative analysis of TF activity and curated state-specific TFs for T_{RM} *vs.* TEX_{term} cells (Fig. 1f-h). To evaluate its accuracy, Perturb-seq, which combines *in vivo* CRISPR screening

with single-cell RNA sequencing (scRNA-seq), was conducted in separate infectious environments that promote the differentiation of TEX_{term} and T_{RM} cells, namely chronic and acute LCMV infection (Fig. 3a). Our Perturb-seq guide RNA (gRNA) library targeted 19 TFs, including seven multi-state TFs predicted to impair both TEX_{term} and T_{RM} differentiation and 12 TEX_{term}-specific TFs expected to selectively affect TEX_{term} differentiation. The TEX_{term}-specific TFs included one known TF, *Nfatc1*, and 11 others that had high specificity scores but were not previously linked to TEX_{term} differentiation (gray boxes Fig. 1f,h). The multi-state TFs included two positive controls (*Nr4a2*, *Prdm1*) and five unvalidated multi-state TFs (*Nfil3*, *Hic1*, *Gfi1*, *Ikzf3*, *Stat3*). To ensure comprehensive screening, four gRNAs per target were expressed in two dual-gRNA retroviral (RV) vectors (Extended Data Fig. 7a), along with two control vectors expressing scramble gRNAs (gScramble). This resulted in a library of 40 dual-gRNA vectors, with 72 gRNAs targeting TFs and 4 gScramble controls (Extended Data Table 5). To initiate the screening, $Cas9^+$ P14 $CD8^+$ T cells were transduced with this RV-gRNA library and adoptively transferred into mice infected with LCMV-Clone 13, a model of chronic infection and $CD8^+$ T cell exhaustion. Droplet-based sequencing was performed 18+ days post-transfer to assess sgRNA and transcriptomes of each spleen-derived donor $Cas9^+$ P14 cell (Fig. 3a), analyzing 17,257 cells with unique gene perturbations.

To determine which TFs impaired TEX_{term} cell differentiation, we first employed uniform manifold approximation and projection (UMAP) to visualize single-cell transcriptomes. Four primary clusters were identified consisting of TEX_{prog}, TEX_{eff}, and TEX_{term} cells and those in cell cycle (Fig. 3b; Extended Data Fig. 7b). All cell clusters expressed *Tox* and *Pdcd1*, key exhaustion markers, and TEX_{prog} cells were identified by *Tcf7*, *Slamf6*, and *Sell* expression. TEX_{eff} cells expressed effector markers, including *Cx3cr1*, *Klrd1*, *Klrk1*, *Klf2*, and *Zeb2*[2], while TEX_{term} cells exhibited high expression of inhibitory receptors and well-established exhaustion markers like *Cd101*, *Cd7*, *Cd38*, *Cd39*, *Cxcr6*, and *Nr4a2*. Lastly, the cell cluster was noted for its expression of *Birc5*, *Mki67*, *Stmn1*, and *Tuba1b*.

Next, the impact of TF-depletion was evaluated by analyzing the distribution of each gRNA-positive cell cluster (Fig. 3c,d). CRISPR knockouts (KOs) of all 19 TEX_{term}-specific TFs resulted in a reduced frequency of TEX_{term} cells. Specifically, KOs of multi-state TFs such as *Hic1*, *Stat3*, *Prdm1*, *Ikzf3* (which encodes AIOLOS) reduced TEX_{term} differentiation by 92% to 89%. Notably, depletion of novel TEX_{term}-specific TFs, including *Zfp324*, *Zscan20*, and *Jdp2* (Fig. 3d bold), significantly prevented TEX_{term}

differentiation by 78%, 54%, and 43% respectively. Other novel TEX_{term}-specific TFs, such as *Etv5*, *Arid3a*, *Zfp410*, *Foxd2*, and *Prdm4*, also reduced cell TEX_{term} formation between 25-40%. This Perturb-seq analysis demonstrated the platform's robust ability to identify cell-state-specific TFs, with all tested TFs influencing TEX_{term} state differentiation to some extent, including many TFs previously unknown to promote T cell exhaustion.

To understand how TEX_{term}-specific TF KOs affect $CD8^+$ T cell exhaustion, we used flow cytometry and scRNA-seq to analyze TF KO cells during LCMV-Clone 13 infection (Fig. 3e-i). We tested nine TF KOs, including two known controls (*Prdm1* and *Nfatc1*) and seven newly identified TFs (*Zscan20*, *Nfil3*, *Jdp2*, *Zfp324*, *Ikzf3*, *Stat3*, *Hic1*) that impaired TEX_{term} state differentiation in Perturb-seq. Disrupting these TFs reduced TEX_{term} cell ($PD1^+$ $CX3CR1^-$ $SLAMF6^-$) frequency by 40-75% (Fig. 3f) and decreased expression of inhibitory receptors such as CD101, CD39, and CD38 (Extended Data Fig. 7c,d). All 19 TEX_{term}-TF KOs exhibited a marked decrease in TEX_{term}-signature genes[43], including *Cd7*, *Cxcr6*, *Nr4a2*, and *Entpd1*(Fig. 3i).

Lastly, the TEX_{term}-specific TF KOs were classified into two groups based on their effects on TEX_{prog} ($PD1^+$ $CX3CR1^-$ $SLAMF6^+$, Fig. 3g) or TEX_{eff} ($PD1^+$ $CX3CR1^+$ $SLAMF6^-$, Fig. 3h) state differentiation markers. Loss of *Hic1*, *Stat3*, *Nfatc1*, *Zfp324*, or *Zscan20* increased both TEX_{prog}- and TEX_{eff} cell frequencies and signature gene expression, while *Jdp2*, *Ikzf3*, or *Nfil3* mainly increased TEX_{eff} frequencies (Fig. 3g-i). *Prdm1* KO significantly elevated $PD1^{low}$ virus-specific cells (Extended Data Fig. 7e). Importantly, except *Stat3*, depleting TEX_{term}-driving TFs that upregulated TEX_{eff} signatures led to the enrichment of genes upregulated in effector versus naïve T cells, enhanced effector functions (IFNγ and TNFα production), and lowered viral loads in recipient mice (Fig. 3j-l).

TEX_{term}-specific TFs do not disrupt T_{RM} cell differentiation

Our main goal was to identify TFs that specifically regulate TEX_{term} cell differentiation without affecting T_{RM} differentiation, thereby enabling precise programming of $CD8^+$ T cell states. Since nearly all the predicted TEX_{term}-specific TFs impaired TEX_{term} differentiation to some degree (Fig. 3), the next step was to evaluate their effects on T_{RM} differentiation to confirm their selective activity. We used the same Perturb-seq library as before, but this time included only the eight TEX_{term}-specific and six multi-state

TFs that impaired TEX_{term} state development by more than 25% in chronic LCMV infection (Fig. 4a). To assess their impact on memory $CD8^+$ T cell development, we isolated RV-transduced $Cas9^+$ P14 $CD8^+$ T cells from the spleen and small intestine of mice 18 days after acute LCMV-Armstrong infection. We then used 15,211 cells from scRNA-seq to evaluate their impact on the formation of intestinal T_{RM} cells and circulating splenic T_{CM} and T_{EM} cells (Fig. 4a).

The UMAP analysis identified four primary clusters containing cells with features of T_{CM} (*Sell* and *Tcf7),* T_{EM} (*Cx3cr1*, *Klrg1,* and *Klf2)*, and T_{RM} cells (*Cd69, Cd160, and Itgae* (encoding CD103)) as well as a small T_{RM} cell population with lower *Itgae*, but higher *Ifng* and *Irf1* expression designated T_{RM} *Itgae*^{low} (Fig. 4b,c; Extended Data Fig. 8a,b)[27,44]. Examination of the gRNA-positive cells revealed that none of the eight TEX_{term}-specific TF KOs (*Zfp324*, *Irf7*, *Zfp410*, *Nfatc1*, *Zscan20*, *Jdp2*, *Arid3a*, and *Etv5)* negatively affected T_{RM} formation significantly (Fig. 4d,e). In fact, KO of *Etv5* tended to increase the frequency of T_{RM} cells. To evaluate the specificity of the TEX_{term}-specific TFs, we also examined the expression of the T_{RM} gene signatures[9,37] in entire gRNA-positive cells of each TF (Fig. 4f). With the exception of *Etv5* and *Arid3a*, which increased T_{RM}-signature gene expression, perturbation of the TEX_{term}-specific TFs did not substantially alter T_{RM}-signature gene expression. The platform also predicted novel multi-state TFs, including *Hic1* and *Gfi1*. Disruption of these newly identified TFs significantly reduced T_{RM} cell frequency (Fig. 4e) and T_{RM}-signature gene expression (Fig. 4f), mirroring the effects of *Prdm1*, a known T_{RM} TF driver.

To further validate the Perturb-seq data, we individually depleted the TEX_{term}-specific TFs *Zscan20* and *Jdp2* and multi-state TF *Prdm1* in $Cas9^+$ P14 $CD8^+$ T cells, adoptively transferred them into LCMV-Armstrong infected animals, and assessed their differentiation into T_{CM}, T_{EM}, and T_{RM} cells using flow cytometry (Fig. 4g-j). Deletion of *Zscan20* and *Jdp2* did not alter the formation of any memory cell subtypes, while perturbation of *Prdm1* significantly reduced T_{RM} and increased T_{CM} formation, as expected. Altogether, this multi-omics pipeline predicted TEX_{term}-specific TFs that drive TEX_{term} differentiation without affecting T_{RM} cell formation and multi-state TFs that influence both cell states. These results demonstrate the accuracy and predictive power of our approach for pinpointing cell-state-specific and multi-state TFs.

Dual-context Perturb-seq reveals TF regulation switching

To better understand how the same multi-state TFs function differently in TEX_{term}- and T_{RM} cells, we analyzed their predicted effects (gene activation *vs*. repression) on target genes by integrating gene regulation network data with differential gene expression scores from Perturb-seq (Fig. 4k). This revealed that TFs like HIC1 and PRDM1 appear to reverse their regulatory roles depending on the type of infection (Fig. 4l,m; Extended Data Fig. 8 c,d). In acute infections, these TFs were predicted to activate genes linked to T_{RM} or memory cell states (e.g., *Junb*, *Myc*, *Nr4a3*, *Irf4*), while in chronic infections, they repressed the same genes. Consistent with prior reports of the PRDM1's role switching in non-T cells[45,46–48], our report suggests that the reversal of regulatory direction of the "multi-tasker TFs" possibly explains how the same TF enables context-dependent regulation of T cell states.

Disruption of novel TEX_{term}-specific TFs enhances tumor control and synergy with ICB

The above work demonstrated the ability to predict cell-state-specific TF activity and identified TEX_{term}-specific TFs as key targets for engineering T cells to prevent exhaustion while preserving effector and memory cell differentiation, offering a strategy for more effective immunotherapies. Given that T_{RM} cells are associated with better clinical outcomes in solid tumors[9–12,49], we hypothesized that knocking out exhaustion-specific TFs like *Zscan20* or *Zfp324* could be more effective than targeting T_{RM} and TEX_{term} multi-state TFs like *Hic1* (Fig. 4e,f). Utilizing an adoptive therapy model, TF gRNA RV-transduced $Cas9^+$ P14 $CD8^+$ T cells were transferred into mice with established melanoma tumors expressing GP33-41 (Fig. 5a). While disruption of the dual-specific TF *Hic1* failed to control solid tumors, depleting the TEX_{term}-specific TFs, *Zscan20* and *Zfp324* resulted in improved tumor control (Fig. 5b,c). Moreover, *Zscan20* and *Zfp324* gRNA-positive cells more readily became progenitors rather than terminally exhausted TIM3-positive populations (Fig. 5d-f), and their depletion upregulated effector markers (Fig. 5g), consistent with enhanced cytotoxicity observed in chronic infection (Fig. 3k,i).

Tumors with high TEX_{term} cell infiltration often respond poorly to ICB[50]. We tested if P14 $CD8^+$ T cells lacking TEX_{term} -specific TFs (*Zscan20*, *Zfp324* or *Jdp2*) would show improved responsiveness to ICB in B16-GP33 tumor-bearing animals (Extended Data Fig. 9a). Indeed, the combination of *Zscan20*-depletion and ICB (anti-PD1 antibody) synergized to decrease tumor burden (Fig. 5h-j) and enhance mice survival (Extended Data Fig. 9b). Similar synergistic effects were observed when depleting *Zfp324* or *Jdp2* along

with ICB treatment (Fig. 5k-m and Extended Data Fig. 9c). In summary, targeting TFs that selectively regulate TEX_{term} state differentiation is a promising strategy to enhance adoptive T cell therapy by avoiding dysfunctional states while maintaining beneficial ones. This precise pipeline for identifying cell state-specifying TFs can also be adapted to program other cell types for better cell therapy outcomes.

Discussion

Our study introduces a powerful platform for identifying TFs that are pivotal in guiding the differentiation of specific $CD8^+$ T cell states during viral infections and tumor progression. This platform has transformative potential for engineering desirable cell states, which can significantly advance T cell-based therapies, including adoptive cell therapies. By constructing a comprehensive transcriptional and epigenetic atlas from 121 experiments spanning nine distinct $CD8^+$ T cell states, we developed a detailed map of TF activity. This allowed us to predict TF functions within each state, essentially creating a unique TF fingerprint for each cellular context.

Focusing on two critical cell states TEX_{term} (dysfunctional terminally exhausted) and T_{RM} (functional, tissue-resident memory) T cells, we examined similarities and differences of TF activity and their networks in both states and engineered T cells to resist exhaustion while retaining functionality of T_{RM} cells. Using *in vivo* Perturb-seq, we validated TF activity for TEX_{term} and T_{RM} cells in both acute and chronic infection models. While recent CRISPR screenings in $CD8^+$ T cells have identified important TFs in cytotoxicity, memory formation[51–54], cell enrichment[55], and exhaustion[56], a systematic and context-dependent understanding of TF roles across multiple contexts has been lacking. Our study addresses this gap by generating an accurate catalogue of $CD8^+$ T cell state-defining TFs, enabling cost-effective validation of predicted TF activity and selectivity via Perturb-seq. Furthermore, our study offers broader and new insight into context-dependent TF regulation. Previously, differential TF cooperation in different contexts was reported[30,54,57–59]. We extend this by analyzing global TF associations across cell states, revealing how TF communities regulate T cell-specific pathways, including protein catabolism in T cell exhaustion, which aligns with prior research on protein homeostasis[42,60–62]. These TF networks reveal how various cellular processes are differentially controlled between T_{RM} and TEX_{term} cells providing rationale for their different functional capabilities within tissues.

One of the most significant outcomes was the identification of novel TFs driving the TEX_{term} state, such as ZSCAN20, JDP2, and ZFP324, and the discovery of new roles for multi-state TFs like HIC1 and GFI1. Perturbing these TEX_{term}-specific TFs not only prevented T cell exhaustion but also preserved the ability of these cells to differentiate into effector and memory states. This led to significant improvements in tumor control and enhanced synergy with ICB therapy, suggesting a promising strategy for refining antitumor immunity through precise cell-state programming.

Our TF analysis platform effectively addressed critical challenges in the field: preventing T cell exhaustion while preserving the therapeutic benefits of T_{RM} cells for solid tumors and virus control. Depletion of TEX_{term}-specific TFs improved tumor control while perturbing TEX_{term} and T_{RM} multi-state TF reduced it. Our pipeline enhances T cell programming by offering optimized 'recipes' of combinatorial TF perturbations to precisely guide T cell differentiation. In summary, while our study primarily focuses on $CD8^+$ TEX_{term} and T_{RM} cell differentiation, the precise pipeline we developed for identifying cell state-specific TFs can be adapted to program other cell types, ultimately improving the effectiveness of cell therapy across various applications. These findings enhance our understanding of T cell differentiation and suggest potential strategies for boosting antitumor immunity and combating chronic infections.

Figure 1

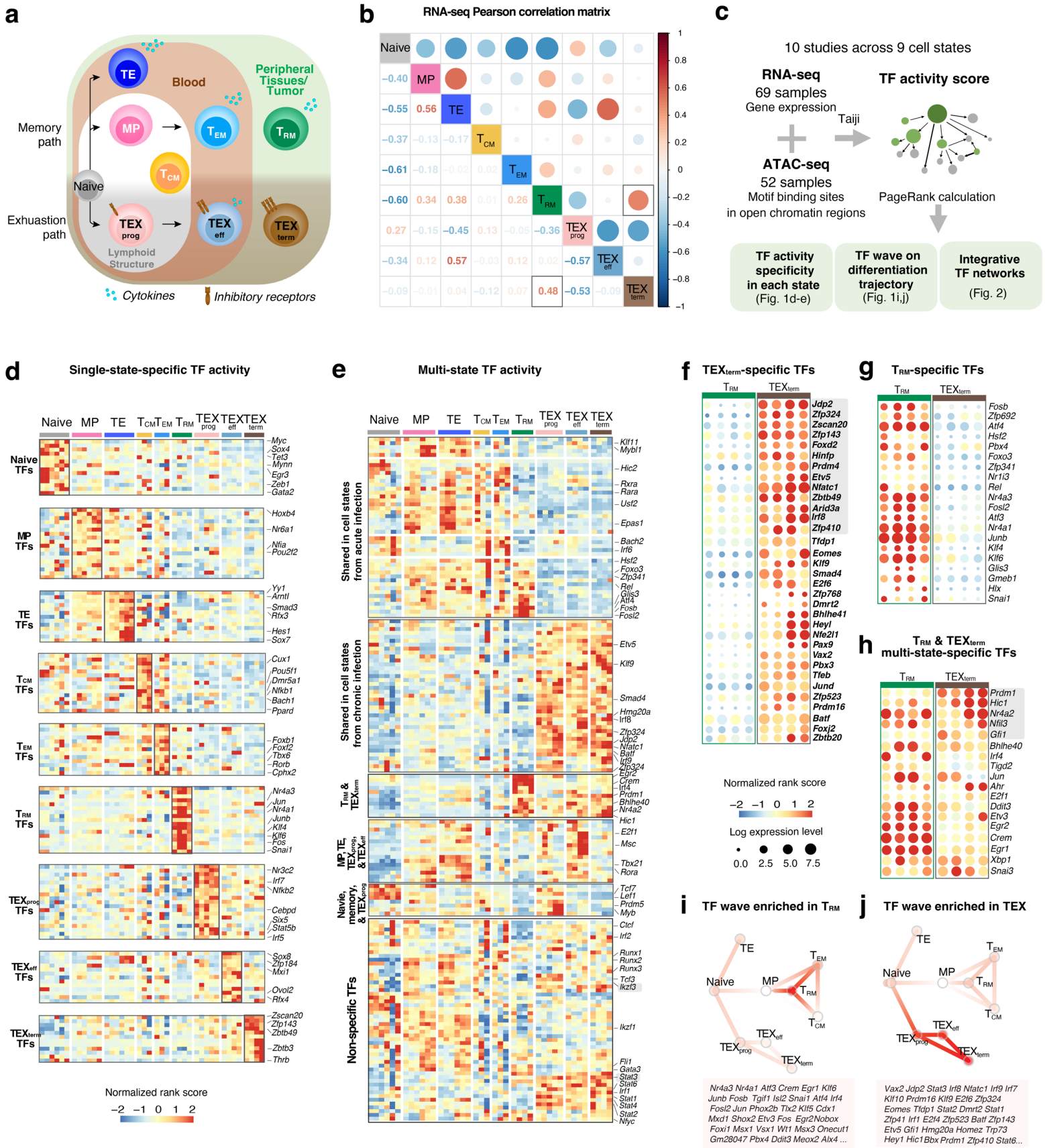


Fig. 1|Transcription and epigenomic atlas of CD8⁺ T cell differentiation states and cell state-specific TFs identification pipeline. a, The diagram outlines several CD8⁺ T cell state trajectories in acute and chronic infection/tumor where various memory and exhaustion states are acquired. Parallel differentiation of T_{RM} and TEX_{term} cells and their similarity in the anatomical location. **b**, Pearson correlation matrix of batch effect corrected RNA-seq. Color and size are both proportional to correlation values, with red being the highest correlation. **c**, Pipeline of our integrative analysis. The matched RNA-seq and ATAC-seq data were used as input for Taiji algorithm to construct a regulatory network and output PageRank score matrix representing the activity of transcription factors (TFs). The downstream analysis included the identification of cell state-specific TFs, as well as the construction of transcriptional waves and TF networks. **d-h**, TFs in rows (z-normalized), samples in columns, and the colors of the cell in the matrix indicate the normalized PageRank scores. Each column represents different data sets. **d**, PageRank scores of 136 single-state-specific TFs. **e**, 173 multi-state TFs. Bubble plots of PageRank scores and expression of TFs that are selectively active in TEX_{term}-(**f**), T_{RM}-(**g**), and multi-state-specific TFs that are active in both TEX_{term} and T_{RM} cells (**h**). TEX_{term}-specific TFs are in bold. Color represents the normalized PageRank score with red displaying high values, and circle size represents the logarithm of gene expression across 5 different data sets. TFs are ordered based on the p-value. **i-j**, Transcription factor waves associated with exhaustion (**i**) or T_{RM} (**j**). Sample numbers for each cell state and the statistics used to curate the TF catalogue are provided in Extended Data Fig. 2a and Extended Data Fig. 3a, respectively.

Figure 2

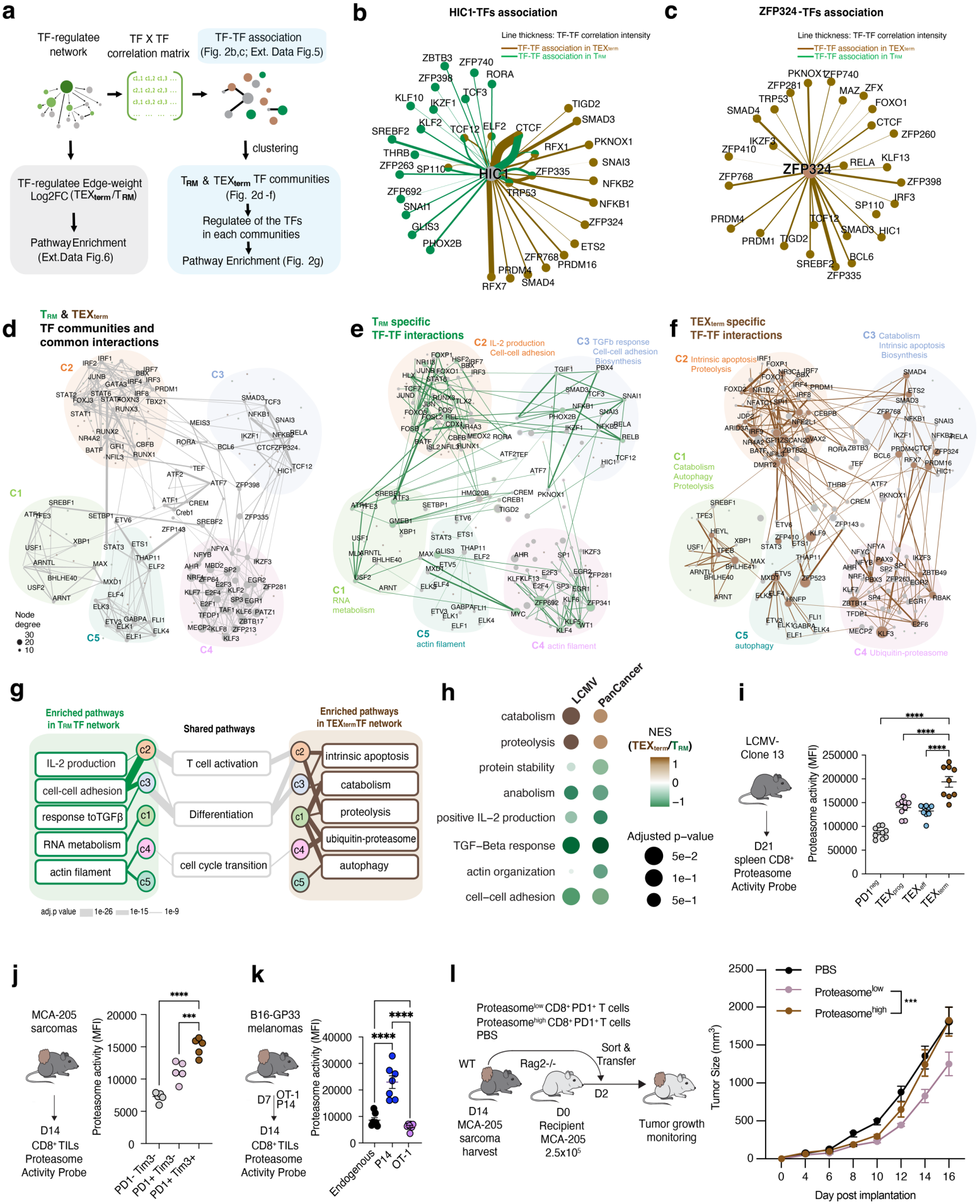


Fig. 2| Global analysis of TF networks in TEX_{term} and T_{RM} cell states. a, TF network analysis in T_{RM} and TEX_{term} cells encompasses both of TF-TF correlation (blue, Fig. 2b-g) and comparative TF-regulatee interaction (gray, Extended Data Fig.6). **b-c**, TF-TF association network plots focus on multi-state TFs like HIC1 **(b)** and TEX_{term}-specific TFs like ZFP324 **(c)** and show unique and shared TF-TF interactions in T_{RM} (green) or TEX_{term} (brown) cells. **d-f**, TF-TF association clustering generates five TF communities between T_{RM} and TEX_{term} cells. The overall community topology is shaped by shared TFs (gray) in both T_{RM} and TEX_{term} cells **(d)**. Specific TF-TF association in T_{RM} **(e)** and TEX_{term} **(f)**. *Legend continued on the next page.*

Fig. 2| Global analysis of TF networks in TEX_{term} and T_{RM} cell states.
g, Summary of the shared and unique pathways enriched in each community in T_{RM} and TEX_{term} cells based on regulatee gene expression. The thickness of the lines reflects p-value. (Genesets: Extended Data Table. 6). **h**, GSEA of TEX_{term} *vs*. T_{RM} in batch-effect corrected LCMV bulk-RNA-seq[3,9,20,37] and pan-cancer scRNA-seq data sets. In both data sets, catabolism and proteolysis are more enriched in TEX_{term} transcriptomes than in T_{RM}. TEX_{term} cells show the highest proteasome activity probe staining in LCMV-Clone 13 infected mice (**i**), or MCA-205 tumor (**j**). **k**, Dual transfer of GP33-41 antigen-specific (P14) and bystander (OT-1) $CD8^+$ T cells in B16-GP33 tumor bearing mice. Proteasome activity probe staining of MCA-205 $CD8^+$ TILs is linked to antigen-specific T cell activation. **l**, Tumor growth of $Rag2^{-/-}$ mice bearing MCA-205 sarcomas infused with $proteasome^{high}$ or $proteasome^{low}$ $CD8^+$ TILs isolated from C57BL/6 mice bearing MCA-205 tumors. **i-l**, Data are presented as the mean ± s.e.m. Ordinary one-way ANOVA (**i-k**) and two-way ANOVA Tukey's multiple comparison test (**l**) were performed. ****P<0.0001, ***P<0.001, **P<0.01, *P<0.05.

Figure 3

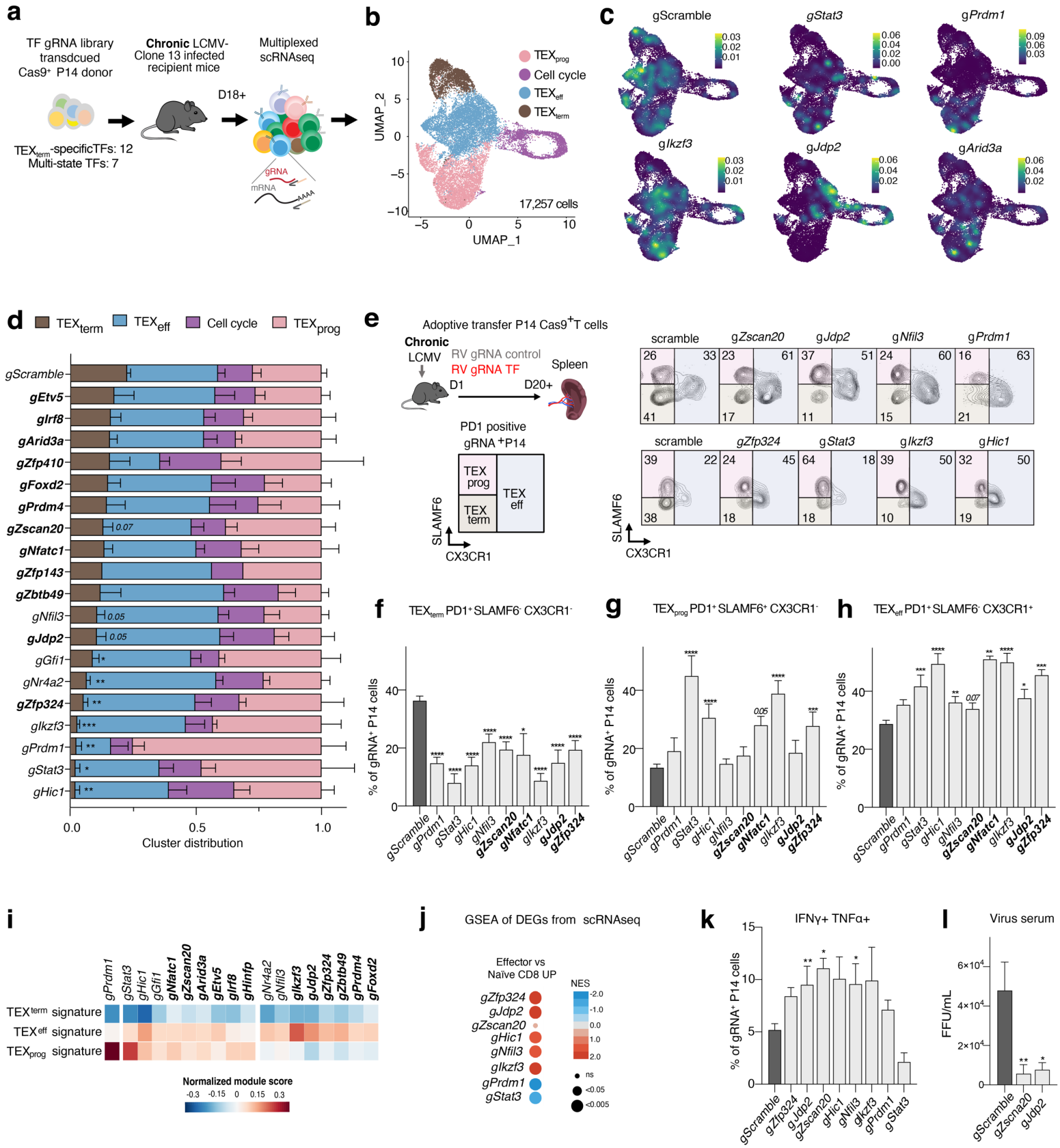


Fig. 3| *in vivo* Perturb-seq validation of TEX_{term}-specific and multi-state driving TFs. a, Schematic of the *in vivo* Perturb-seq strategy. Dual $Cas9^+$ and TCR transgenic (tg) P14 $CD8^+$ T cells that recognize LCMV epitope GP33-41 were transduced with gRNA library expressing retrovirus (RV) and adoptively transferred into mice previously infected one day earlier with LCMV-Clone 13. 18-23 days later, splenic T cells were analyzed by scRNA-seq. **b**, UMAP showing TEX_{prog}, TEX_{eff}, TEX_{term}, and cell cycle clusters. Feature plots of the cluster marker expression are in Extended Data Fig. 7b. **c**, Cells expressing gRNAs for KO of specific genes are denoted on UMAPs and yellow indicates increased cell density. **d**, Cluster distribution of cells with each gRNA KO. TEX_{term}-specific TFs are in bold letters. **e-h**, Individual phenotyping of the TF KOs in LCMV-Clone 13 infected mice based on flow cytometry **(e)**. Quantification of the frequency of TEX subsets: TEX_{term} ($PD1^+$ $SLAMF6^-$ $CX3CR1^-$) **(f)**, TEX_{prog} ($PD1^+$ $SLAMF6^+$ $CX3CR1^-$) (**g**), and TEX_{eff} ($PD1^+$ $SLAMF6^-$ $CX3CR1^+$) (**h**) P14 $CD8^+$ T cells. **i**, TEX_{term}, TEX_{prog}, and TEX_{eff} gene signature analysis in each gRNA KO compared to gScramble. Gene signatures were generated from ref. 46 (Extended Data Table 6). **j**, GSEA of effector-associated gene sets between each TF KOs *vs*. control. **k**, Donor P14 cells with gRNA KO of several TEX_{term}-specific TFs in LCMV-Clone 13 were assessed for IFNγ and TNF production. **i**, TEX_{term}-specific TFs, *Zscan20* KO and *Jdp2* KO group showed lower serum virus levels. Data are presented as the mean ± s.e.m. ****$P<0.0001$, ***$P<0.001$, **$P<0.01$, *$P<0.05$. Two-way ANOVA with Fisher's LSD test. Pooled five biological replicates from three independent experiments (**d**). Data include more than two biological replicates, n ≥ 5. Ordinary one-way ANOVA with Dunnett's multiple comparison test (**f-h, k-l**).

Figure 4

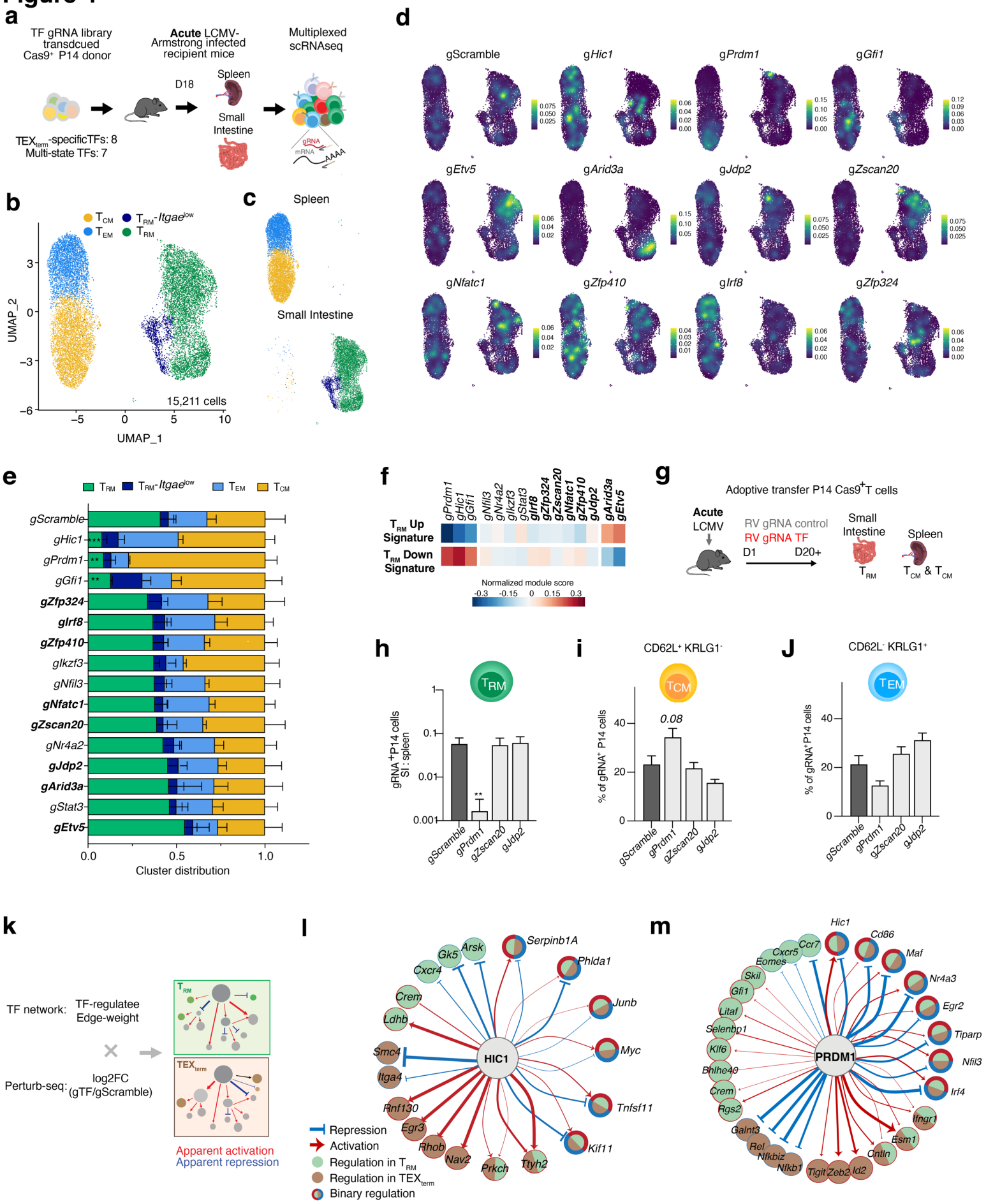


Fig. 4| TEX$_{term}$-specific TFs do not affect T$_{RM}$ cell formation. *Legend on the next page.*

Fig. 4| TEX_{term}-specific TFs do not affect T_{RM} cell formation. a, Schematic of *in vivo* Perturb-seq screening in memory $CD8^+$ T cells during acute LCMV-Armstrong infection. $Cas9^+$ P14 $CD8^+$ T cells were transduced with the TF gRNA library and then transferred into mice previously infected with LCMV-Armstrong one day earlier. The donor cells were then analyzed for T_{RM} cells in the small intestine (SI) and T_{CM} and T_{EM} cells in the spleen. **b**, UMAP showing T_{CM}, T_{EM}, T_{RM} and T_{RM}-*Itgae*^{low} clusters. The cluster marker expression is in Extended Data Fig. 8a. **c**, Differential clustering of cells from different tissue origins. **d**, gRNA KO positive cell density on UMAP. The color bar indicates cell density. **e**, Cluster distribution of cells with each gRNA KO. TEX_{term}-specific TFs are in bold letters. **f**, Normalized expression score of T_{RM} "up and down" gene-signatures[9,37] in cells with each gRNA KO compared to gScramble positive cells. **g-j**, Individual phenotyping of the TF KOs in LCMV-Armstrong infected mice based on flow cytometry. **h**, The ratio of gRNA KO cells in SI:spleens. The frequency of splenic T_{CM} ($CD62L^+$ $KLRG1^-$) (**i**) and T_{EM} ($CD62L^-$ $KLRG1^+$) cells (**j**). **k-m**, Calculation of the apparent regulatory direction by integrating TF-regulated networks with Perturb-seq; blue and red lines represent suppressive and activatory regulation, respectively **(k)**. Regulatory maps of T_{RM} and TEX_{term} multi-state-specific TFs, HIC1 **(l)** and PDRM1 **(m)**. The green nodes represent regulatees in T_{RM} cells, while brown nodes represent those in TEX_{term} cells. Split-colored nodes with red and blue edges indicate predicted opposing regulation of the same regulatee in two cell states. Data are presented as the mean ± s.e.m. ***P<0.001, ** P <0.01, * P <0.05. Two-way ANOVA with Fisher's LSD test. Pooled three biological replicates (**e**). Ordinary one-way ANOVA with Dunnett's multiple comparison test. Data include more than two biological replicates, $n \geq 6$ (**h-j**).

Figure 5

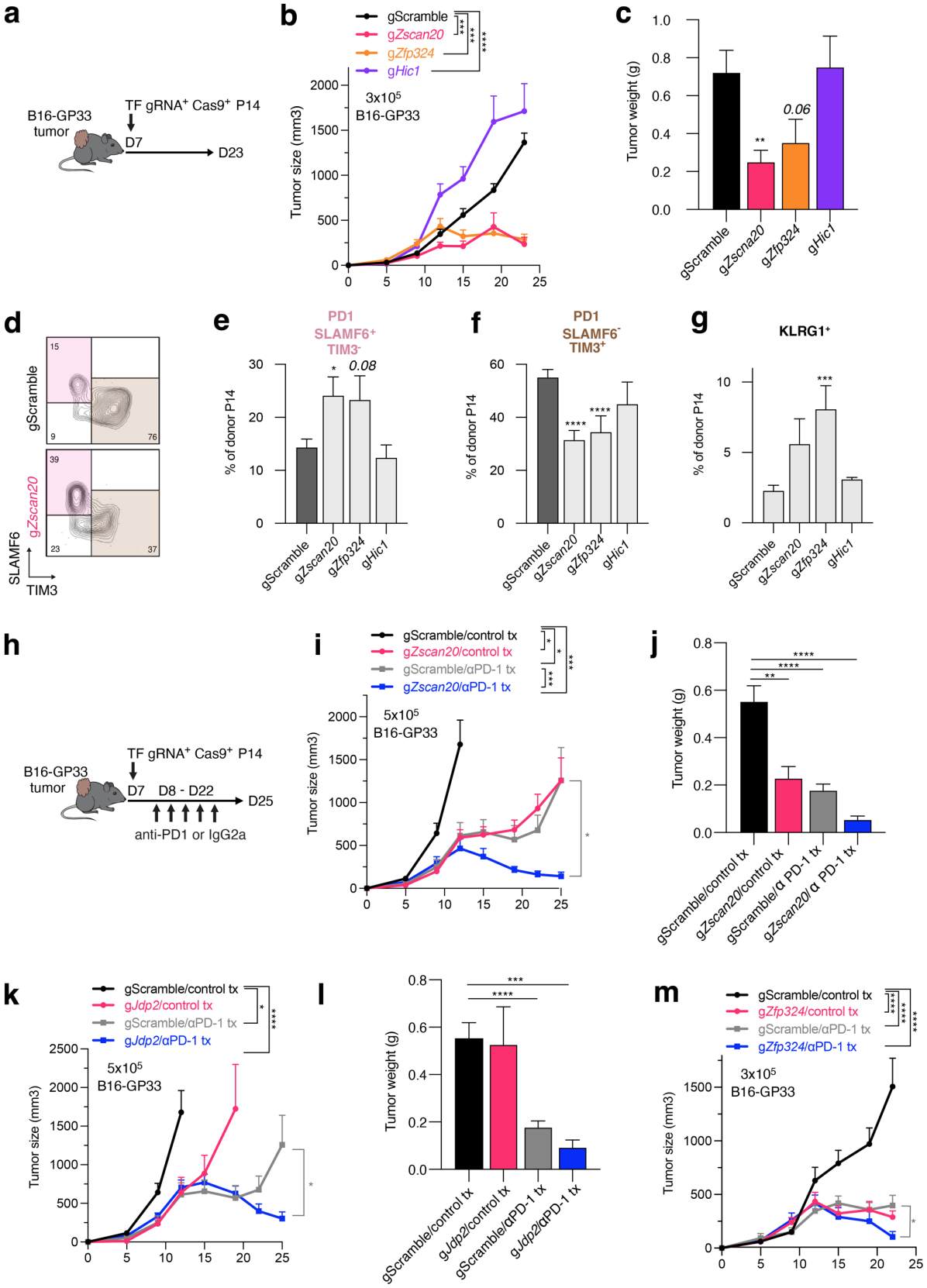


Fig. 5| Disruption of TEX_{term}-specific TFs enhances virus and tumor control. **a-c**, Experimental design **(a)** of gRNA KOs of TEX_{term}-specific TFs (*Zscan20* and *Zfp324*) or the TEX_{term} and T_{RM} multi-state-specific TF (*Hic1*) and control P14 $CD8^+$ T cells were adoptively transferred to melanoma-bearing mice and tumor sizes (**b**) and tumor weights (**c**) were measured. **d,** Representative flow plots of control and gRNA KO $PD1^+$ P14 cells stained with antibodies to SLAMF6 and TIM3**. e-f**, bar graphs of control and gRNA KO cells showing frequency of TEX_{prog} ($PD1^+$ $SLAMF6^+$ $TIM3^-$) **(e)** and TEX_{term} cells ($PD1^+$ $SLAMF6^-$ $TIM3^+$) **(f)**. **g**, Frequency of effector marker $KLRG1^+$ cells. **h**, Experiment setup to test the combination of TF KO and anti-PD1 treatment. Day 7 of B16-GP33 s.c. implantation ($5x10^5$), $Cas9^+$ P14 were adoptively transferred followed by by anti-PD1 or isotype IgG2a control treatment. Tumor growth curve (**i**) and tumor weight (**j**) of *Zscan20* KO. Tumor growth curve (**k**) and tumor weight (**l**) of *Jdp2* KO. **m**, Day 7 of B16-GP33 s.c. implantation ($3x10^5$), $Cas9^+$ P14 of *Zfp324 KO* were adoptively transferred followed by anti-PD1 or isotype IgG2a control treatment. Tumor growth curves were measured. All data include more than three biological replicates, $n \geq 9$. Tumor growth statistical analysis was performed using two-way ANOVA Tukey's multiple comparison test. All other comparisons were performed by pair-wise t-test. Data are presented as the mean ± s.e.m. ****P<0.0001, ***P<0.001, **P<0.01, *P<0.05.

References

1. Chung, H. K., McDonald, B. & Kaech, S. M. The architectural design of CD8+ T cell responses in acute and chronic infection: Parallel structures with divergent fates. *J. Exp. Med.* **218**, (2021).
2. Kasmani, M. Y. *et al.* Clonal lineage tracing reveals mechanisms skewing CD8+ T cell fate decisions in chronic infection. *J. Exp. Med.* **220**, (2023).
3. Beltra, J.-C. *et al.* Developmental Relationships of Four Exhausted CD8+ T Cell Subsets Reveals Underlying Transcriptional and Epigenetic Landscape Control Mechanisms. *Immunity* **52**, 825–841.e8 (2020).
4. Giles, J. R. *et al.* Shared and distinct biological circuits in effector, memory and exhausted CD8 T cells revealed by temporal single-cell transcriptomics and epigenetics. *Nature Immunology* vol. 23 1600–1613 Preprint at https://doi.org/10.1038/s41590-022-01338-4 (2022).
5. Cui, W., Liu, Y., Weinstein, J. S., Craft, J. & Kaech, S. M. An interleukin-21-interleukin-10-STAT3 pathway is critical for functional maturation of memory CD8+ T cells. *Immunity* **35**, 792–805 (2011).
6. Kaech, S. M. & Cui, W. Transcriptional control of effector and memory CD8+ T cell differentiation. *Nat. Rev. Immunol.* **12**, 749–761 (2012).
7. Flynn, J. K. & Gorry, P. R. Stem memory T cells (TSCM)-their role in cancer and HIV immunotherapies. *Clin Transl Immunology* **3**, e20 (2014).
8. Zhang, Z. *et al.* T Cell Dysfunction and Exhaustion in Cancer. *Front Cell Dev Biol* **8**, 17 (2020).
9. Milner, J. J. *et al.* Runx3 programs CD8 T cell residency in non-lymphoid tissues and tumours. *Nature* **552**, 253–257 (2017).
10. Ganesan, A.-P. *et al.* Tissue-resident memory features are linked to the magnitude of cytotoxic T cell responses in human lung cancer. *Nat. Immunol.* **18**, 940–950 (2017).
11. Djenidi, F. *et al.* CD8 CD103 Tumor–Infiltrating Lymphocytes Are Tumor-Specific Tissue-Resident Memory T Cells and a Prognostic Factor for Survival in Lung Cancer Patients. *The Journal of Immunology* vol. 194 3475–3486 Preprint at https://doi.org/10.4049/jimmunol.1402711 (2015).
12. Corgnac, S., Boutet, M., Kfoury, M., Naltet, C. & Mami-Chouaib, F. The Emerging Role of CD8+ Tissue Resident Memory T (TRM) Cells in Antitumor Immunity: A Unique Functional Contribution of the CD103 Integrin. *Front. Immunol.* **9**, 1904 (2018).
13. Komdeur, F. L. *et al.* CD103+ tumor-infiltrating lymphocytes are tumor-reactive intraepithelial CD8+ T cells associated with prognostic benefit and therapy response in cervical cancer. *Oncoimmunology* **6**, e1338230 (2017).
14. Pauken, K. E. *et al.* Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science* **354**, 1160–1165 (2016).
15. Sen, D. R. *et al.* The epigenetic landscape of T cell exhaustion. *Science* vol. 354 1165–1169 (2016)
16. Wherry, E. J., Blattman, J. N., Murali-Krishna, K., van der Most, R. & Ahmed, R. Viral persistence alters CD8 T-cell immunodominance and tissue distribution and results in distinct stages of functional impairment. *J. Virol.* **77**, 4911–4927 (2003).

17. Wherry, E. J. *et al.* Molecular signature of CD8+ T cell exhaustion during chronic viral infection. *Immunity* **27**, 670–684 (2007).
18. Im, S. J. *et al.* Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. *Nature* **537**, 417–421 (2016).
19. Miller, B. C. *et al.* Subsets of exhausted CD8 T cells differentially mediate tumor control and respond to checkpoint blockade. *Nat. Immunol.* **20**, 326–336 (2019).
20. Hudson, W. H. *et al.* Proliferating Transitory T Cells with an Effector-like Transcriptional Signature Emerge from PD-1 Stem-like CD8 T Cells during Chronic Infection. *Immunity* vol. 51 1043–1058.e4 (2019)
21. Pu, F. *et al.* TIM-3 expression and its association with overall survival in primary osteosarcoma. *Oncol. Lett.* **18**, 5294–5300 (2019).
22. Zang, K. *et al.* TIM-3 as a Prognostic Marker and a Potential Immunotherapy Target in Human Malignant Tumors: A Meta-Analysis and Bioinformatics Validation. *Front. Oncol.* **11**, 579351 (2021).
23. Liu, F. *et al.* CTLA-4 correlates with immune and clinical characteristics of glioma. *Cancer Cell Int.* **20**, 7 (2020).
24. Zhang, M. *et al.* Prognostic Values of CD38+CD101+PD1+CD8+ T Cells in Pancreatic Cancer. *Immunol. Invest.* (2019) doi:10.1080/08820139.2019.1566356.
25. Shin, H. *et al.* A role for the transcriptional repressor Blimp-1 in CD8(+) T cell exhaustion during chronic viral infection. *Immunity* **31**, 309–320 (2009).
26. Behr, F. M. *et al.* Blimp-1 Rather Than Hobit Drives the Formation of Tissue-Resident Memory CD8 T Cells in the Lungs. *Front. Immunol.* **10**, 400 (2019).
27. Milner, J. J. *et al.* Heterogenous Populations of Tissue-Resident CD8+ T Cells Are Generated in Response to Infection and Malignancy. *Immunity* **52**, 808–824.e7 (2020).
28. Li, C. *et al.* The Transcription Factor Bhlhe40 Programs Mitochondrial Regulation of Resident CD8 T Cell Fitness and Functionality. *Immunity* **51**, 491–507.e7 (2019).
29. Wu, J. E. *et al.* In vitro modeling of CD8 T cell exhaustion enables CRISPR screening to reveal a role for BHLHE40. *Sci Immunol* **8**, eade3369 (2023).
30. Chen, J. *et al.* NR4A transcription factors limit CAR T cell function in solid tumours. *Nature* **567**, 530–534 (2019).
31. Mackay, L. K. *et al.* The developmental pathway for CD103(+)CD8+ tissue-resident memory T cells of skin. *Nat. Immunol.* **14**, 1294–1301 (2013).
32. Schacht, T., Oswald, M., Eils, R., Eichmüller, S. B. & König, R. Estimating the activity of transcription factors by the effect on their target genes. *Bioinformatics* **30**, i401–7 (2014).
33. Zhang, K., Wang, M., Zhao, Y. & Wang, W. Taiji: System-level identification of key transcription factors reveals transcriptional waves in mouse embryonic development. *Sci Adv* **5**, eaav3262 (2019).
34. Wang, J., Liu, C., Chen, Y. & Wang, W. Taiji-reprogram: a framework to uncover cell-type specific regulators and predict cellular reprogramming cocktails. *NAR Genom Bioinform* **3**, lqab100 (2021).
35. Liu, C. *et al.* Systems-level identification of key transcription factors in immune cell specification. *PLoS Comput. Biol.* **18**, e1010116 (2022).
36. Langville, A. N. & Meyer, C. D. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. (Princeton

University Press, 2006).

37. Mackay, L. K. *et al.* Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes. *Science* **352**, 459–463 (2016).

38. Crowl, J. T. *et al.* Tissue-resident memory CD8 T cells possess unique transcriptional, epigenetic and functional adaptations to different tissue environments. *Nature Immunology* vol. 23 1121–1131 https://doi.org/10.1038/s41590-022-01229-8 (2022).

39. Chandele, A. *et al.* Formation of IL-7Ralphahigh and IL-7Ralphalow CD8 T cells during infection is regulated by the opposing functions of GABPalpha and Gfi-1. *J. Immunol.* **180**, 5309–5319 (2008).

40. Smith, N. P. *et al.* Resident memory T cell development is associated with AP-1 transcription factor upregulation across anatomical niches. *bioRxiv* (2023) doi:10.1101/2023.09.29.560006.

41. Guo, X. *et al.* Publisher Correction: Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat. Med.* **24**, 1628 (2018).

42. Riesenberg, B. P. *et al.* Stress-Mediated Attenuation of Translation Undermines T-cell Activity in Cancer. *Cancer Res.* **82**, 4386–4399 (2022).

43. Daniel, B. *et al.* Divergent clonal differentiation trajectories of T cell exhaustion. *Nat. Immunol.* **23**, 1614–1627 (2022).

44. Lin, Y. H. *et al.* Small intestine and colon tissue-resident memory CD8+ T cells exhibit molecular heterogeneity and differential dependence on Eomes. *Immunity* **56**, 207–223.e8 (2023).

45. Prajapati, R. S., Hintze, M. & Streit, A. PRDM1 controls the sequential activation of neural, neural crest and sensory progenitor determinants. *Development* **146**, (2019).

46. Minnich, M. *et al.* Multifunctional role of the transcription factor Blimp-1 in coordinating plasma cell differentiation. *Nat. Immunol.* **17**, 331–343 (2016).

47. Lin, Y., Wong, K. & Calame, K. Repression of c-myc transcription by Blimp-1, an inducer of terminal B cell differentiation. *Science* **276**, 596–599 (1997).

48. Yu, J., Angelin-Duclos, C., Greenwood, J., Liao, J. & Calame, K. Transcriptional repression by blimp-1 (PRDI-BF1) involves recruitment of histone deacetylase. *Mol. Cell. Biol.* **20**, 2592–2603 (2000).

49. Komdeur, F. L. *et al.* CD103 tumor-infiltrating lymphocytes are tumor-reactive intraepithelial CD8 T cells associated with prognostic benefit and therapy response in cervical cancer. *OncoImmunology* vol. 6 e1338230 https://doi.org/10.1080/2162402x.2017.1338230 (2017).

50. Miller, B. C. *et al.* Author Correction: Subsets of exhausted CD8+ T cells differentially mediate tumor control and respond to checkpoint blockade. *Nat. Immunol.* **20**, 1556 (2019).

51. Cortez, J. T. *et al.* CRISPR screen in regulatory T cells reveals modulators of Foxp3. *Nature* **582**, 416–420 (2020).

52. Schmidt, R. *et al.* CRISPR activation and interference screens decode stimulation responses in primary human T cells. *Science* **375**, eabj4008 (2022).

53. Shifrut, E. *et al.* Genome-wide CRISPR Screens in Primary Human T Cells Reveal Key Regulators of Immune Function. *Cell* **175**, 1958–1971.e15 (2018).

54. McCutcheon, S. R. *et al.* Transcriptional and epigenetic regulators of human CD8+ T cell function identified through

orthogonal CRISPR screens. *Nat. Genet.* **55**, 2211–2223 (2023).

55. Chen, Z. *et al.* In vivo CD8+ T cell CRISPR screening reveals control by Fli1 in infection and cancer. *Cell* **184**, 1262–1280.e22 (2021).
56. Zhou, P. *et al.* Single-cell CRISPR screens in vivo map T cell fate regulomes in cancer. *Nature* **624**, 154–163 (2023).
57. Lynn, R. C. *et al.* c-Jun overexpression in CAR T cells induces exhaustion resistance. *Nature* **576**, 293–300 (2019).
58. Seo, H. *et al.* BATF and IRF4 cooperate to counter exhaustion in tumor-infiltrating CAR T cells. *Nat. Immunol.* **22**, 983–995 (2021).
59. Rubin, A. J. *et al.* Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks. *Cell* **176**, 361–376.e17 (2019).
60. Van Der Byl, W. *et al.* The CD8+ T cell tolerance checkpoint triggers a distinct differentiation state defined by protein translation defects. *Immunity* (2024) doi:10.1016/j.immuni.2024.04.026.
61. Widjaja, C. E. *et al.* Proteasome activity regulates CD8+ T lymphocyte metabolism and fate specification. *J. Clin. Invest.* **127**, 3609–3623 (2017).
62. Chang, J. T. *et al.* Asymmetric proteasome segregation as a mechanism for unequal partitioning of the transcription factor T-bet during T lymphocyte division. *Immunity* **34**, 492–504 (2011).

Figure Legends

Fig. 1|Transcription and epigenomic atlas of CD8$^+$ T cell differentiation states and cell state-specific TFs identification pipeline. a, The diagram outlines several CD8$^+$ T cell state trajectories in acute and chronic infection/tumor where various memory and exhaustion states are acquired. Note the parallels in the architecture of T_{RM} and TEX_{term} cell differentiation and their anatomical locations. **b**, Pearson correlation matrix of batch effect corrected RNA-seq. Color and size are both proportional to correlation values, with red being the highest correlation. **c**, Pipeline of our integrative analysis. The matched RNA-seq and ATAC-seq data were used as input for Taiji algorithm to construct a regulatory network and output PageRank score matrix representing the activity of transcription factors (TFs). The downstream analysis included the identification of cell state-specific TFs, as well as the construction of transcriptional waves and TF networks. **d-h**, TFs in rows (z-normalized), samples in columns, and the colors of the cell in the matrix indicate the normalized PageRank scores. Each column represents different data sets. **d**, PageRank scores of 136 single-state-specific TFs. **e**, 173 multi-state TFs. Bubble plots of PageRank scores and expression of TFs that are selectively active in TEX_{term}-(**f**), T_{RM}-(**g**), and multi-state-specific TFs that are active in both TEX_{term} and T_{RM} cells (**h**). TEX_{term}-specific TFs are in bold. Color represents the normalized PageRank score with red displaying high values, and circle size represents the logarithm of gene expression across 5 different data sets. TFs are ordered based on the p-value. **i-j**, Transcription factor waves associated with exhaustion (**i**) or T_{RM} (**j**). Sample numbers for each cell state and the statistics used to curate the TF catalogue are provided in Extended Data Fig. 2a and Extended Data Fig. 3a, respectively.

Fig. 2| Global analysis of TF networks in TEX_{term} and T_{RM} cell states. a, TF network analysis in T_{RM} and TEX_{term} cells encompasses both of TF-TF correlation (blue, Fig. 2b-g) and comparative TF-regulatee interaction (gray, Extended Data Fig.6). **b-c**, TF-TF association network plots focus on multi-state TFs like HIC1 **(b)** and TEX_{term}-specific TFs like ZFP324 **(c)** and show unique and shared TF-TF interactions in T_{RM} (green) or TEX_{term} (brown) cells. **d-f**, TF-TF association clustering generates five TF communities between T_{RM} and TEX_{term} cells. **d**, The overall community topology is shaped by shared TFs (gray) in both T_{RM} and TEX_{term} cells. **e-f**, T_{RM} and TEX_{term} cells show differential TF-TF interactions within each community and between communities. Top 10% of interactions are shown. The line thickness represents interaction intensity. **g**, Summary of the shared and unique pathways enriched in each community in T_{RM} and TEX_{term} cells based on regulatee gene expression. The thickness of the lines reflects p-value. (Genesets:

Extended Data Table. 6). **h**, GSEA of TEX_{term} *vs*. T_{RM} in batch-effect corrected LCMV bulk-RNA-seq[3,9,20,37] and pan-cancer scRNA-seq data sets. In both data sets, catabolism and proteolysis are more enriched in TEX_{term} transcriptomes than in T_{RM}. TEX_{term} cells show the highest proteasome activity probe staining in LCMV-Clone 13 infected mice (**i**), or MCA-205 tumor (**j**). **k**, Dual transfer of GP33-41 antigen-specific (P14) and bystander (OT-1) $CD8^+$ T cells in B16-GP33 tumor bearing mice. Proteasome activity probe staining of MCA-205 $CD8^+$ TILs is linked to antigen-specific T cell activation. **l**, Tumor growth of $Rag2^{-/-}$ mice bearing MCA-205 sarcomas infused with $proteasome^{high}$ or $proteasome^{low}$ $CD8^+$ TILs isolated from C57BL/6 mice bearing MCA-205 tumors. **i-l**, Data are presented as the mean ± s.e.m. Ordinary one-way ANOVA (**i-k**) and two-way ANOVA Tukey's multiple comparison test (**l**) were performed. ****P<0.0001, ***P<0.001, **P<0.01, *P<0.05.

Fig. 3| *in vivo* Perturb-seq validation of TEX_{term}-specific and multi-state driving TFs. a, Schematic of the *in vivo* Perturb-seq strategy. Dual $Cas9^+$ and TCR transgenic (tg) P14 $CD8^+$ T cells that recognize LCMV epitope GP33-41 were transduced with gRNA library expressing retrovirus (RV) and adoptively transferred into mice previously infected one day earlier with LCMV-Clone 13. 18-23 days later, splenic T cells were analyzed by scRNA-seq. **b**, UMAP showing TEX_{prog}, TEX_{eff}, TEX_{term}, and cell cycle clusters. Feature plots of the cluster marker expression are in Extended Data Fig. 7b. **c**, Cells expressing gRNAs for KO of specific genes are denoted on UMAPs and yellow indicates increased cell density. **d**, Cluster distribution of cells with each gRNA KO. TEX_{term}-specific TFs are in bold letters. **e-h**, Individual phenotyping of the TF KOs in LCMV-Clone 13 infected mice based on flow cytometry **(e)**. Quantification of the frequency of TEX subsets: TEX_{term} ($PD1^+$ $SLAMF6^-$ $CX3CR1^-$) **(f),** TEX_{prog} ($PD1^+$ $SLAMF6^+$ $CX3CR1^-$) (**g**), and TEX_{eff} ($PD1^+$ $SLAMF6^-$ $CX3CR1^+$) (**h)** P14 $CD8^+$ T cells. **i**, TEX_{term}, TEX_{prog}, and TEX_{eff} gene signature analysis in each gRNA KO compared to gScramble. Gene signatures were generated from ref. 46 (Extended Data Table 6). **j**, GSEA of effector-associated gene sets between each TF KOs *vs*. control. **k**, Donor P14 cells with gRNA KO of several TEX_{term}-specific TFs in LCMV-Clone 13 were assessed for IFNγ and TNF production. **l**, TEX_{term}-specific TFs, *Zscan20* KO and *Jdp2* KO group showed lower serum virus levels. Data are presented as the mean ± s.e.m. ****P<0.0001, ***P<0.001, **P<0.01, *P<0.05. Two-way ANOVA with Fisher's LSD test. Pooled five biological replicates from three independent experiments. (d). Data include more than two biological replicates, n ≥ 5. Ordinary one-way ANOVA with Dunnett's multiple comparison test (**f-h, k-l**).

Fig. 4| TEX_{term}-specific TFs do not affect T_{RM} cell formation. a, Schematic of *in vivo* Perturb-seq screening in memory $CD8^+$ T cells during acute LCMV-Armstrong infection. $Cas9^+$ P14 $CD8^+$ T cells were transduced with the TF gRNA library and then transferred into mice previously infected with LCMV-Armstrong one day earlier. The donor cells were then analyzed for T_{RM} cells in the small intestine (SI) and T_{CM} and T_{EM} cells in the spleen. **b**, UMAP showing T_{CM}, T_{EM}, T_{RM} and T_{RM}-*Itgae*^{low} clusters. The cluster marker expression is in Extended Data Fig. 8a. **c**, Differential clustering of cells from different tissue origins. **d**, gRNA KO positive cell density on UMAP. The color bar indicates cell density. **e**, Cluster distribution of cells with each gRNA KO. TEX_{term}-specific TFs are in bold letters. **f**, Normalized expression score of T_{RM} "up and down" gene-signatures[9,37] in cells with each gRNA KO compared to gScramble positive cells. **g-j**, Individual phenotyping of the TF KOs in LCMV-Armstrong infected mice based on flow cytometry. **h**, The ratio of gRNA KO cells in SI:spleens. The frequency of splenic T_{CM} ($CD62L^+$ $KLRG1^-$) (**i**) and T_{EM} ($CD62L^-$ $KLRG1^+$) cells (**j**). **k-m**, Calculation of the apparent regulatory direction by integrating TF-regulated networks with Perturb-seq; blue and red lines represent suppressive and activatory regulation, respectively **(k)**. Regulatory maps of T_{RM} and TEX_{term} multi-state-specific TFs, HIC1 **(l)** and PDRM1 **(m)**. The green nodes represent regulatees in T_{RM} cells, while brown nodes represent those in TEX_{term} cells. Split-colored nodes with red and blue edges indicate predicted opposing regulation of the same regulatee in two cell states. Data are presented as the mean ± s.e.m. ***P<0.001, ** P <0.01, * P <0.05. Two-way ANOVA with Fisher's LSD test. Pooled three biological replicates (**e**). Ordinary one-way ANOVA with Dunnett's multiple comparison test. Data include more than two biological replicates, n ≥ 6 (**h-j**).

Fig. 5| Disruption of TEX_{term}-specific TFs enhances virus and tumor control. a-c, Experimental design **(a)** of gRNA KOs of TEX_{term}-specific TFs (*Zscan20* and *Zfp324*) or the TEX_{term} and T_{RM} multi-state-specific TF (*Hic1*) and control P14 $CD8^+$ T cells were adoptively transferred to melanoma-bearing mice and tumor sizes (**b**) and tumor weights (**c**) were measured. **d,** Representative flow plots of control and gRNA KO $PD1^+$ P14 cells stained with antibodies to SLAMF6 and TIM3**. e-f**, bar graphs of control and gRNA KO cells showing frequency of TEX_{prog} ($PD1^+$ $SLAMF6^+$ $TIM3^-$) **(e)** and TEX_{term} cells ($PD1^+$ $SLAMF6^-$ $TIM3^+$) **(f)**. **g**, Frequency of effector marker $KLRG1^+$ cells. **h**, Experiment setup to test the

combination of TF KO and anti-PD1 treatment. Day 7 of B16-GP33 s.c. implantation ($5x10^5$), $Cas9^+$ P14 were adoptively transferred followed by by anti-PD1 or isotype IgG2a control treatment. Tumor growth curve (**i**) and tumor weight (**j**) of *Zscan20* KO. Tumor growth curve (**k**) and tumor weight (**l**) of *Jdp2* KO. **m**, Day 7 of B16-GP33 s.c. implantation ($3x10^5$), $Cas9^+$ P14 of *Zfp324 KO* were adoptively transferred followed by anti-PD1 or isotype IgG2a control treatment. Tumor growth curves were measured. All data include more than three biological replicates, $n \geq 9$. Tumor growth statistical analysis was performed using two-way ANOVA Tukey's multiple comparison test. All other comparisons were performed by pair-wise t-test. Data are presented as the mean ± s.e.m. ****$P<0.0001$, ***$P<0.001$, **$P<0.01$, *$P<0.05$.

Methods

Dataset acquisition: The analyzed $CD8^+$ T cell samples were collected from ten datasets, including this study (Extended Data Fig. 2). In total, 121 experiments, including 52 ATAC-seq and 69 RNA-seq were harnessed to generate the paired samples and used as input for the Taiji pipeline. The $CD8^+$ T cell samples span nine subtypes: naive cells (Naive), terminal effector cells (TE), memory precursor cells (MP), tissue-resident memory cells (T_{RM}), effector memory cells (T_{EM}), central memory cells (T_{CM}), progenitor exhausted cells (TEX_{prog}), intermediate exhausted cells (TEX_{eff}) and terminal exhausted cells (TEX_{term}).

TF regulatory networks construction and visualization: To perform integrative analysis of RNA-seq and ATAC-seq data, we developed Taiji v2.0. which allows visualization of several downstream analysis-TF wave, TF-TF association, TF community analysis. Epitensor was used for the prediction of chromatin interactions. Putative TF binding motifs were curated from the latest CIS-BP database[63]. In this analysis, 695 TFs were identified as having binding sites centered around ATAC-seq peak summits. The average number of nodes (genes) and edges (interactions) of the genetic regulatory networks across $CD8^+$ T cell states were 15,845 and 1,325,694, respectively, including 695 (4.38%) TF nodes. On average, each TF regulated 1,907 genes, and each gene was regulated by 22 TFs.

Identification of cell-state-specific/non-specific TFs: We first identified universal TFs with mean PageRank across 9 cell states ranked as top 10% and Coefficients of variation (CV) less than 0.5. In total, 54 universal TFs were identified and shown in Extended Data Table 1. The remaining 641 TFs were candidate TFs for cell-state-specific TFs. To identify cell-state-specific TFs, we divided the samples into two groups: target group and background group. Target group included all the samples belonging to the cell state of interest and the background group comprised the remaining samples. We then performed the normality test using Shapiro-Wilk's method to determine whether the two groups were normally distributed, and we found that the PageRank scores of most samples (90%) follow a log-normal distribution. Based on the log-normality assumption, an unpaired t-test was used to calculate the P-value. A P-value cutoff of 0.05 and log2 fold change cutoff of 0.5 were used for calling lineage-specific TFs In total, 255 specific TFs were identified (Extended Data Table 1). Depending on whether the TF appeared in multiple cell states, they could be further divided into multi-state specific TFs (Extended Data Table 1, Fig 1e) and 136 single-state exclusive TFs (Extended Data Table 1, Fig. 1d). Out of 255 specific TFs, 84

TFs appear in TEX_{term} or T_{RM}. To identify the truly distinctive TFs between TEX_{term} and T_{RM}, we performed a second round of unpaired t-test only between TEX_{term} and T_{RM} (Extended Data Table 2). The same cut-offs, i.e. P-value of 0.05 and log2 fold change of 0.5, were applied to select TEX_{term} single-taskers and T_{RM} single-taskers. Thirty out of 84 TFs that did not pass the cut-off were identified as TEX_{term} and T_{RM} multi-taskers. The full workflow was summarized in Extended Data Fig. 3a.

Identification of transcriptional waves: To identify the TFs which show similar temporal activity patterns, we clustered the TFs based on the normalized PageRank scores across samples. First, we performed principal component analysis (PCA) for dimensionality reduction of the TF score matrix. We retained the first 10 principal components for further clustering analysis, which explained more than 70% variance (Extended Data Fig. 4b; left panel). We used the k-means algorithm for clustering analysis. To find the optimal number of clusters and similarity metric, we performed the Silhouette analysis to evaluate the clustering quality using five distance metrics: Euclidean distance, Manhattan distance, Kendall correlation, Pearson correlation, and Spearman correlation (Extended Data Fig. 4b; right panel). Pearson correlation was the most appropriate distance metric since the average Silhouette width was the highest among all five-distance metrics. Based on these analyses, we identified 7 distinct dynamic patterns of TF activity during immune cell development. We further performed functional enrichment analysis to identify GO terms for these clusters.

Regulatee-TF interaction in TEX_{term} and T_{RM} context: Taiji's outputs cell state-specific network between TFs and their targets. We first combined the cell state-important TFs (PageRank mean is top 20% in each cell state, Extended Data Table 3) and cell state-specific TFs (Extended Data Table 2) in TEX_{term} and T_{RM} respectively. In total, 159 TFs for TEX_{term} and 170 TFs for T_{RM} were selected. We then combined TEX_{term}/T_{RM} samples' network by taking the mean value of edge weight for each TF-regulatee pair. Next, the top 500 regulatees per TF based on edge weight were kept for the following analysis. Then we took the union of the regulatees from TEX_{term} and T_{RM}. The absent edge weight of TF-regulatee pairs were set to 0. Thus, we calculated the log2 fold change of edge weight between TEX and T_{RM} with the 1e-5 as the pseudo number to avoid the zero division. In total, 17,483 regulatees, 192 TFs were included. We used the k-means algorithm for clustering regulatees. To find the optimal number of clusters and similarity metric, we performed similar Silhouette analysis as mentioned above. We selected K as 16 considering

both silhouette width and balanced gene set size. Pheatmap function in R was used for visualization in Extended Data Fig. 6.

TF community construction and visualization: Inspired by DBPNet[64], which is a framework to identify cooperations between DNA-binding proteins using Chromatin immunoprecipitation followed by sequencing (ChIP-seq) and Hi-C data, we constructed TF interaction network based on Taiji's output, which is a TF-regulatee network. For each context, we first combined the cell state-important TFs and cell state-specific TFs. In total, 159 TFs for TEX_{term} and 170 TFs for T_{RM} were selected. We then combined TEX_{term}/T_{RM} samples' network by taking the mean value of edge weight for each TF-regulatee pair. Next, regulatees with low variation across TFs (standard deviation <= 1) were removed, then correlation matrix between TFs is calculated by taking account of the SpeArmstrongan's correlation of edge weight for each TF-regulatee pair. R package "huge"[65] is used to build a graphical model and construct the graph. We employed the Graphical lasso algorithm and the shrunken ECDF (empirical cumulative distribution function) estimator. We used a lasso penalty parameter lambda equal to 0.052 to control the regularization. We chose this value based on the local minimum point on the sparsity-lambda curve. When lambda = 0.05, around 15% of TF-TF pairs are considered as connected in the network. To estimate the false discovery rate, we generated a null model by random shuffling the edge weight of TF-regulatee pair across TFs. When the same algorithm is applied to this dataset, the chosen cutoff identifies zero interaction, suggesting that the method with cutoff equal to 0.05 has a very low false discovery rate. Communities were detected using Leiden algorithm[66] with modularity as objective function and resolution as 0.9 since it reached the highest clustering modularity. In total, we identified 5 communities for each context. Network visualization was performed by graph with Fruchterman-Reingold layout algorithm[67].

Pathway enrichment analysis: The enriched functional terms in this study were analyzed by the R package clusterProfiler 4.0.5. We used GO, KEGG, MSigDb databases for annotation. For gene set enrichment analysis (GSEA), the genes were first ranked by the mean edge weight in corresponding samples and H, C5, C6, and C7 collections from MSigDb were used for annotation. A cutoff of P-value < 0.05 was used to select the significantly enriched GO terms and KEGG pathways.

Heuristic score calculation via the integration of TF regulatory network and Perturb-seq:

We reasoned that (1) log2FC change in expression due to TF knockout, and (2) TF-gene regulatory edge weights could be combined to provide heuristic scores for a TF's regulatory effect on a target gene. For each guide RNA knockout, Seurat's FindMarkers() function was used to quantify log2FC of gene's expression with respect to the gScramble condition. Heuristic scores were calculated for each TF-gene pair by multiplying the gene log2FC with the corresponding edge weight from the Taiji analysis. Regulatees of a TF were annotated as high-confidence if the magnitude log2FC of the regulate exceeded 0.58 (corresponding to a fold-change of 1.5 or its reciprocal) and if the edge weight belonged to the upper quantile of all edge weights attributed to the TF. The sign of the log2FC was used to determine if the TF activated or repressed each target gene.

Mice and Infections: C57BL/6/J, OT-1 (C57BL/6-Tg(TcraTcrb)1100Mjb/J), B6.Cg-*Rag2*$^{tm1.1Cgn}$/J, CD45.1 (B6.SJL-*PtprcaPepcb*/BoyJ) mice were purchased from Jackson Laboratories. P14 mice (Pircher et al., 1987) mice have been previously described. Cas9 P14 mice were generated by crossing P14 mice with B6(C)-Gt(ROSA)26Sorem1.1(CAG-cas9*,-EGFP)Rsky/J (Jackson Laboratories). Animals were housed in specific-pathogen-free facilities at the Salk Institute and University and at the University of North Carolina at Chapel Hill. All animal experiments were approved by the Institutional Animal Care and Use Committee (IACUC). Mice were infected with $2x10^5$ PFU LCMV-Armstrong by intraperitoneal injection or $2x10^6$ PFU LCMV-Clone13 by retro-orbital injection under anesthesia.

Viral titers: LCMV Fluorescence Focus Unit titration was performed seeding Vero cells at a density of 30k cells/100µl in a 96 well flat bottom plate in DMEM + 10% FBS + 2% HEPES + 1% Pen/Strep. On the next day, tissues were homogenized on ice, spun down at 1000g for 5 min at 4°C and supernatants or serum were diluted in 10-fold steps. Diluted samples were added to Vero cells and incubated at 37°C, 5% CO_2 for ~20h. Subsequently, inocula were aspirated and wells were incubated with 4% PFA for 30 min at RT before washing with PBS. VL-4 antibody (BioXCell) was conjugated using the Invitrogen AF488 conjugation kit and added to the wells in dilution buffer containing 3% BSA and 0.3% Triton (Thermo Fisher Scientific) in PBS). Cells were incubated at 4°C overnight before washing with PBS and counting foci under the microscope. FFU was calculated per this formula: FFU/ml=Number of plaques/ (dilution× volume of diluted virus added to the plate).

Cell Isolation: Spleens were mechanically dissociated with 1mL syringe plungers over a 70um nylon strainer. Spleens were incubated in ammonium chloride potassium (ACK) buffer for 5 minutes. For isolation of small intestinal IEL, Peyer's patches were first removed by dissection. Intestines were longitudinally cut and then cut into 1cm pieces and washed in PBS. Pieces were incubated in 30mL HBSS with 10% FBS, 10mM HEPES, and 1mM dithioerythritol with vigorous shaking at 37C for 30 minutes. Supernatants were collected, washed, and further isolated using a 40/67% discontinuous percoll density centrifugation for 20 minutes at room temp with no brakes.

Cell lines and *in vitro* cultures: B16-GP33 melanoma cell line was cultured in DMEM (Invitrogen) with 10% fetal bovine serum, 1% penicillin-streptomycin and 250 μg/ml G418 (Invitrogen #10131027). MCA-205 (Sigma) tumor line was maintained in RPMI supplemented with 10% FBS, 300 mg/L L-glutamine, 100 units/mL penicillin, 100 μg/mL streptomycin, 1mM sodium pyruvate, 100μM NEAA, 1mM HEPES, 55μM 2-mercaptoethanol, and 0.2% Plasmocin mycoplasma prophylactic (InvivoGen). All the tumor cell lines were used for experiments when in exponential growth phase. For *in vitro* T cell culture, P14 splenocytes were activated in RPMI 1640 medium (Invitrogen) containing 10% fetal bovine serum and 1% penicillin-streptomycin, 2mM L-glutamine, 0.1 mg/ml GP33, beta-mercaptoethanol 50 mM and 10 U/ml IL-2.

Tumor engraftment and treatment of tumor-bearing mice: For tumor engraftment, 5×10^5 B16-GP33 tumor cells (Fig. 6h-l) and 3×10^5 B16-GP33 tumor cells (Fig. 6a-c,m) were injected subcutaneously in 100 μl PBS. Tumors were measured every 2-3 days post-tumor engraftment for indicated treatments and calculated. Tumor volume was calculated by volume = (length × width^2)/2. For antibody-based treatment, tumor-bearing mice were treated with anti-PD1 antibody (200 μg per injection, clone RMP1-14, BioXcell) twice per week from day 7 post-tumor implantation. All experiments were conducted according to the Salk Institute and University of North Carolina Animal Care and Use Committee.

Tumor digestion and cell isolation for Fig. 4: Tumors were minced into small pieces in RPMI containing 2% FBS, DNase I (0.5 μg/ml, Sigma-Aldrich), and collagenase (0.5 mg/ml, Sigma-Aldrich) and kept for digestion for 30 min at 37˚C, followed by filtration with 70 μm cell strainers (VWR). Filtered cells were incubated with ACK lysis buffer (Invitrogen) to lyse red blood cells, mixed with excessive RPMI 1640

medium (Invitrogen) containing 10% fetal bovine serum and 1% penicillin-streptomycin, and centrifuged at 400g for 5 min to obtain single-cell suspension.

Proteasome activity analysis: For experiments involving the Proteasome Activity Probe (R&D systems), cells of interest were incubated with the probe at concentration of 2.5mM for 2 hours at 37C in PBS. Samples were washed and then stained with Zombie NIR viability dye (Biolegend) in PBS at 4C for 15 minutes. Samples were then stained with some variation of the following antibodies for 30 minutes in FACS Buffer on ice: CD45-BV510 (BD Biosciences), CD45.2-BV510 (Biolegend), CD45.1-PE-Cy7 (Invitrogen), CD4-APC Fire 810 (Biolegend), CD11b-Alexa Fluor 532 (Invitrogen), CD8-Spark NIR 685 (Biolegend), CD44-Brilliant Violet 785 (Biolegend), CD62L-BV421 (BD Biosciences), PD1-BB700 (BD Biosciences), TIM-3-BV711 (Biolegend), LAG-3-APC-eFluor 780 (Invitrogen), SlamF6-APC (Invitrogen), CD39-Superbright 436 (Invitrogen), CX3CR1-PE/Fire 640 (Biolegend), CD69-PE-Cy5 (Biolegend), GITR-BV650 (BD Biosciences), CD27-BV750 (BD Biosciences). Samples were collected on a Cytek Northern Lights and analyzed via Cytek SpectraFlo software.

Tumor experiment for proteasome assay: MCA-205 fibrosarcomas (2.5 x 10^5) were established via subcutaneous injection into the right flank of C57BL/6 mice. After 12-14 days of tumor growth, spleens, draining lymph nodes, and tumors from groups of mice were harvested and tumors were processed using the Mouse Tumor Dissociation Kit and gentleMACS dissociator (Miltenyi Biotec) according to manufacturer's protocol. For purification experiments, samples were pre-enriched using the EasySep Mouse $CD8^+$ T Cell Isolation Kit (Stemcell Technologies) according to manufacturer's protocol and stained with Live-or-Dye PE Fixable Viability Stain (Biotium) and CD8a-APC (Invitrogen) and live $CD8^+$ cells were sorted using the FACSAria II cell sorter. $CD8^+$ spleen and pooled TIL samples were washed in PBS and frozen for RNA-seq analysis. For adoptive cellular therapy experiments, B16-GP33 melanomas were established subcutaneously by injecting 5.0×10^5 cells into the right flank of CD45.1 pepboy mice and tumor-bearing hosts were irradiated with 5 Gy 24 hours prior to T cell transfer. After 7 days of tumor growth, 1.5×10^6 CD45.2 OT-1 T cells and 1.5×10^6 CD45.1/CD45.2 P14 T cells were infused in 100 μL PBS via tail vein into tumor bearing mice. Tumors were harvested 14 days after adoptive cell transfer and CD8 TILs were analyzed for proteasome activity. All experiments were conducted according to the University of North Carolina at Chapel Hill Animal Care and Use Committee.

Proteasome^{high}/Proteasome^{low} T cell adoptive transfer experiment: MCA-205 fibrosarcomas (2.5 x 10^5) were established via subcutaneous injection into the right flank of C57BL/6 mice. After 14 days of tumor growth, Live $CD45^+$ $CD8^+$ $CD44^+$ $PD1^+$ T cells were sorted from tumors based on proteasome activity (high vs low) using the FACSAria II cell sorter. 2.5 x 10^4 cells were then injected into the 2 day MCA-205 bearing RAG2-/- hosts (n=5 per group) and tumor growth was monitored every other day starting on day 4. Tumor growth was measured every other day with calipers, and survival was monitored with an experimental endpoint of tumor growth > 300 mm^2. All experiments were conducted according to the University of North Carolina at Chapel Hill Animal Care and Use Committee.

Retrovirus transduction and adoptive transfer: For gRNA retrovirus vector over-expression, 293T cells were transfected with Eco-helper and MSCV gRNA vectors. 48 h and 72 h later, supernatant containing retroviral particles was ready for transduction. P14 donor splenocytes were *in vitro* activated by 0.1 mg/ml GP33 and 10 U/ml IL-2 at 37°C for 24h, then spin-transduced (1500 g) with fresh RV supernatant from 293T cells for 90 min at 30°C in the presence of 5 μg/ml polybrene. Right after viral transduction 0.5-10^6 Cas9 expressing GP33-specific P14 TCR transgenic CD8+ T cells were transferred into C57BL/6 mice (retro orbital) that were implanted with 5 x 10^5 B16-GP33 cells 7-8 days ago or were infected with either LCMV-Clone 13 or Armstrong 1 day ago.

Flow cytometry, cell sorting, and antibodies: Both single cell suspensions were incubated with Fc receptor-blocking anti-CD16/32 (BioLegend) on ice for 10 min before staining. Cell suspensions were first stained with Red Dead Cell Stain Kit (ThermoFisher) for 10 min on ice. Surface proteins were then stained in FACS buffer (PBS containing 2% FBS and 0.1% sodium azide) for 30 min at 4˚C. To detect cytokine production *ex-vivo*, cell suspensions were re-suspended in RPMI 1640 containing 10% FBS, stimulated by 50 ng/ml PMA and 3 μM Ionomycin in the presence 2.5 μg/ml Brefeldin A (BioLegend #420601) for 4 h at 37°C. Cells were processed for surface marker staining as described above. For intracellular cytokine staining, cells were fixed in BD Cytofix/Cytoperm (BD #554714) for 30 min at 4 °C, then washed with 1× Permeabilization buffer (Invitrogen #00-8333-56). For transcription factor staining, cells were fixed in Foxp3 / Transcription Factor Fixation/Permeabilization buffer (Invitrogen #00-5521-00) for 30 min at 4 °C, then washed with 1× Permeabilization buffer. Cells were then stained with

intraceulluar antibodies for 30 min at 4 °C. Samples were processed on LSR-II flow cytometer (BD Biosciences) and data were analyzed with FlowJo V10 (TreeStar). Cells were sorted either on FACSAria™ III sorter or Fusion sorter (BD Biosciences). The following antibodies against mouse proteins were used: anti-CD8a (53-6.7), anti-PD1 (29F.1A12), anti-CX3CR1 (SA011F11), anti-SLAMF6 (13G3), anti-CD38 (90), anti-CD39 (24DMS1), anti-CD101 (Moushi101), anti-KRLG1 (2F1), anti-CD69 (H1.2F3), anti-CD103 (M290), anti-CD62L(MEL-14), anti-Tim3 (RMT3-23), anti-Ly5.1 (A20), anti-Ly5.2 (104), anti-IFN-γ (XMG1.2), anti-TNF-α (MP6-XT22). These antibodies were purchased from Invitrogen, Biolegend, Cell Signaling, or eBiosciences.

Perturb seq screening using the retroviral transcriptional factor library:

D*ual-guide direct-capture retroviral sgRNA vector.* To generate dual-guide sgRNA vector (MSCV-hU6-mU6-SV40-EGFP), we replaced the hU6 RNA scaffold region of the previously described retroviral sgRNA vector MG-guide[68] with an additional scaffold[69] and murine U6 promoter.

Dual-guide direct-capture retroviral library construction. For the curated gene list containing 21 TFs, a total of four gRNA sequences distributed on two individual constructs were designed for each gene. To construct the library, a customized double-strand DNA fragment pool containing 80 oligonucleotides targeting those 19 TFs and 4 scramble gRNAs (each oligonucleotide contains two guides targeting the same gene) (Extended Data Table 5) was ordered from IDT. The dual-guide library was generated using an In-Fusion (Takara) reaction. In brief, the gRNA containing DNA fragment pool was combined in MG-guide vector linearized with BpiI (Thermo Fisher). The construct was then transformed into Stellar competent cells (Takara) and amplified, and the resulting intermediate, individual construct was assessed for quality using Sanger sequencing. Then, individual dual-gRNA vectors were combined. For quality control, sgRNA skewing was measured using the MAGeCKFlute[70] to monitor how closely sgRNAs are represented in a library.

*In vivo screening.*Retrovirus was produced by co-transfecting the dual-guide, direct-capture retroviral library with pCL-Eco in HEK293 cells. At 48 h and 72 h after transfection, the supernatant was collected and frozen at −80 °C. Cas9-expressing P14 cells were transduced to achieve 20–30% transduction efficiency. Transduced cells were sorted based on the expression of EGFP, and an aliquot of 1×10^6

transduced P14 cells was saved as input. Transduced P14 cells (5×10^4) were then transferred intravenously to either LCMV-Clone 13 or LCMV-Armstrong strain infected C57BL/6 PuroR Cas9 expressing recipient mice at day 1 post-infection. 25 LCMV Clone13 strain infected recipient mice were used for 5 biological replicates, and ten LCMV-Armstrong strain infected recipient mice were used for three biological replicates. 18+ days or later, donor-derived total P14 cells were sorted and pooled for Perturb-seq analysis. Sorted $EGFP^+$ P14 cells were resuspended and diluted in 10% FBS RPMI at a concentration of 1×10^6 cells per ml. Both the gene expression library and the CRISPR screening library were prepared using a Chromium Next GEM Single Cell 5′ kit with Feature Barcode technology for CRISPR Screening (10x Genomics). In brief, the single-cell suspensions were loaded onto the Chromium Controller according to their respective cell counts to generate 10,000 single-cell gel beads in emulsion per sample. Each sample was loaded into four separate channels. Chromium Next GEM Single Cell 5' Kit v2, PN-1000263, Chromium 5' Feature Barcode Kit, PN-1000541, 5' CRISPR Kit, PN-1000451, Chromium Next GEM Chip K Single Cell Kit, PN-1000287, Dual Index Kit TT Set A, PN-1000215, Dual Index Kit TN Set A, PN-1000250 (10x Genomics) in total were used for each reaction. The resulting libraries were quantified and quality checked using TapeStation (Agilent). Samples were diluted and loaded onto a NovaSeq (Illumina) using 100 cycle kit to obtain a minimum of 20,000 paired-end reads (26x91 bp) per cell for the gene expression library and 5,000 paired-end reads per cell for the CRISPR screening library, yielding an average of 42639, 36739, and 53413 reads aligned from cells from *in vivo* LCMV-Clone 13, *in vivo* LCMV-Armstrong infection, and *in vitro* donor respectively.

Data analysis. Alignments and count aggregation of gene expression and sgRNA reads were completed using Cell Ranger (v.7.0.1). Gene expression and sgRNA reads were aligned using the cell Ranger multi count command with default settings. Gene expression reads were aligned to the mouse genome (mm10 from ENSEMBL GRCm38 loaded from 10x Genomics). The median average of 4, 2, and 33 unique molecular identifiers (UMIs) were detected from cells from *in vivo* LCMV-Clone 13 and LCMV-Armstrong infection, and *in vitro* donor respectively. Droplets with sgRNA UMI passing of default Cell Ranger CRISPR analysis Protospacer UMI threshold were used in further analysis. The filtered feature matrices were imported into Seurat (v.4.3.0)[71] to create assays for a Seurat object containing both gene expression and CRISPR guide capture matrices. Cells detected with sgRNAs targeting two or more genes were then removed to avoid interference from multi-sgRNA-transduced cells. Low-quality cells with <200

detected genes, >10% mitochondrial reads, and <5% ribosomal reads were discarded. A total of 17,257 cells (Clone 13) and 15,211 cells (Armstrong) were passed quality filtering and were used for downstream analysis. Count data was normalized by a global-scaling normalization method and linear transformed.[72] Cluster-specific genes were identified using the FindAllMarkers function of Seurat. We utilized Nebulosa[73] to recover signals from sparse features in single-cell data and made gRNA density plots with scCustomize[74] based on kernel density estimation. In each biological replicate (Clone 13 n=5 and Armstrong =3), the percent cluster distribution of cells with each TF gRNA vector was calculated. Among two gRNA vectors per target TF, the gRNA vector with higher TEX_{term} reduction was shown in Fig. 3d and utilized for Perturb-seq in LCMV-Armstrong infection (Extended Data Table 5). Two-way ANOVA with Fisher's LSD test was performed to determine statistical significance. DEG was calculated using the MAST model[75], the results were then used as inputs for GSEA to evaluate the effect on selected pathways. Genes with p-value < 0.05 were considered as DEG.

ATAC-Seq library preparation and sequencing: ATAC-seq was performed as previously described[76]. Briefly, 5,000-50,000 viable cells were washed with cold PBS, collected by centrifugation, then lysed in resuspension buffer (RSB) (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2) supplemented with 0.1% NP40, 0.1% Tween-20, and 0.01% digitonin. Samples were incubated on ice for 3 min, then washed out with 1 ml RSB containing 0.1% Tween-20. Nuclei were pelleted by centrifugation at 500g for 10 min at 4°C then resuspended in 50 ul transposition mix (25ul 2x TD buffer, 2.5 ul transposase (100 nM final), 16.5 ul PBS, 0.5 ul 1% digitonin, 0.5 ul 10% Tween-20, 5 ul H2O) and incubated at 37°C for 30 min in a thermomixer with 1000 RPM mixing. DNA was purified using a Qiagen MinElute PCR cleanup kit, then PCR amplified using indexed oligos. The optimal number of amplification cycles for each sample was determined by qPCR. Libraries were size selected using AmpureXP beads and sequenced using an Illumina NextSeq500 for 75bp paired-end reads.

ATAC-Seq analysis: Paired-end 42-bp, or paired-end 75-bp reads were aligned to the M. musculus mm10 genome using BWA[77,78]with parameters "bwa mem -M -k 32". ATAC-seq peaks were called using MACS2[79] program using parameters "callpeaks -qvalue 5.0e-2 –shift -100 –extsize 200". Differentially accessible regions were identified using DESeq2[80] . Batch effect was removed using limma[81]. Heatmap visualization of ATAC-seq data was performed using pheatmap.

Single-cell RNA sequencing metadata analysis: Analysis was primarily performed in R (v 3.6.1) using the package Seurat (v 3.1)[71,82], with the package tidyverse (v 1.2.1)[83] used to organize data and the package ggplot2 (v 3.2.1) used to generate figures. scRNA-seq data from GSE10898, GSE99254, GSE98638, GSE199565, and GSE181785 were filtered to keep cells with a low percentage of mitochondrial genes in the transcriptome (< 5%) and between 200 and 3000 unique genes to exclude poor quality reads and doublets. Cell cycle scores were regressed when scaling gene expression values and T cell receptor genes were regressed during the clustering process, which was performed with the Louvain algorithm within Seurat and visualized with UMAP. To quantify the gene expression patterns, we used Seurat's module score feature to score each cluster based on its per cell expression of TFs.

To obtain Extended Data Fig. 6d, raw single-cell count data and cell annotation data was downloaded from NCBI GEO (GSE99254)[41]. Count data was normalized and transformed by derivation of the residuals from a regularized negative binomial regression model for each gene (SCT normalization method in Seurat[87], v4.1.1), with 5000 variable features retained for downstream dimensionality reduction techniques. Integration of data was performed on the patient level with Canonical Correlation Analysis as the dimension reduction technique[84]. Principal component analysis and UMAP dimension reduction were performed, with the first 50 PCs utilized in UMAP generation. Cells were clustered utilizing the Louvain algorithm with multi-level refinement. The data was subset to CD8+ T cells, which were identified utilizing the labels provided by Guo et al[66]. Cell type labels were confirmed via 1) SingleR[85] (v1.8.1) annotation utilizing the ImmGen[86] database obtained via celda(1.10), 2) cluster marker identification, and 3) cell type annotation with the ProjecTILs T cell atlas[7] (v2.2.1). After sub-setting to $CD8^+$ T cells, cells were again normalized utilizing SCT normalization, with 3000 variable features retained for dimension reduction. Due to the low number of cells on the per-patient level, HArmstrongony[87](v1.0) was utilized to integrate the data at the patient level, rather than Seurat. PCA and UMAP dimensionality reduction were performed as above.

Statistical analyses: Statistical tests for flow cytometry data were performed using Graphpad Prism 10. *p-values* were calculated using either two-tailed unpaired Student's t-tests, one-way ANOVA or two-way ANOVA as indicated in each figure. Linear regressions were performed using the ordinary least squares

method in R (v 3.6.1). All data were presented as the mean ± s.e.m. The *P* values were represented as follows: ****$P < 0.0001$, ***$P < 0.001$, **$P < 0.01$, and *$P < 0.05$.

Data availability

ATAC-sequencing data from this paper will be deposited in the GEO database. Taiji v2.0 output of this study (TF activity atlas, TF-TF interaction maps, and TF activity on genome browser view) will be available at our $CD8^+$ T cell TF atlas portal (https://github.com/apotato123/CD8-TF-Atlas-Portal). All other raw data are available from the corresponding author upon request.

Code availability

All scripts and Taiji v2.0 package will be uploaded at GitHub (https://github.com/cong-003/Taiji2).

Methods references

63. Weirauch, M. T. *et al.* Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell* vol. 158 1431–1443 https://doi.org/10.1016/j.cell.2014.08.009 (2014).
64. Zhang, K., Li, N., Ainsworth, R. I. & Wang, W. Systematic identification of protein combinations mediating chromatin looping. *Nat. Commun.* **7**, 1–11 (2016).
65. Zhao, T., Liu, H., Roeder, K., Lafferty, J. & Wasserman, L. The huge Package for High-dimensional Undirected Graph Estimation in R. (2020) doi:10.48550/arXiv.2006.14781.
66. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
67. Schönfeld, M. & Pfeffer, J. Fruchterman/Reingold (1991): Graph Drawing by Force-Directed Placement. *Schlüsselwerke der Netzwerkforschung* 217–220 https://doi.org/10.1007/978-3-658-21742-6_49 (2019).
68. Bailis, W. *et al.* Author Correction: Distinct modes of mitochondrial metabolism uncouple T cell differentiation and function. *Nature* **573**, E2 (2019).
69. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
70. Wang, B. *et al.* Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nat. Protoc.* **14**, 756–780 (2019).
71. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* vol. 177 1888–1902.e21 https://doi.org/10.1016/j.cell.2019.05.031 (2019).
72. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
73. Alquicira-Hernandez, J. & Powell, J. E. Nebulosa recovers single-cell gene expression signals by kernel density estimation. *Bioinformatics* **37**, 2485–2487 (2021).
74. Marsh, S. E. scCustomize: custom visualizations & functions for streamlined analyses of single cell sequencing. *https://doi. org/10.5281/zenodo*.
75. Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
76. McDonald, B. *et al.* Canonical BAF complex activity shapes the enhancer landscape that licenses CD8 T cell effector and memory fates. *Immunity* **56**, 1303–1319.e5 (2023).
77. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, (2009).
78. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, (2010).
79. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, 1–9 (2008).
80. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
81. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

82. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
83. Wickham, H. *et al.* Welcome to the Tidyverse. *Journal of Open Source Software* vol. 4 1686 https://doi.org/10.21105/joss.01686 (2019).
84. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
85. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
86. Heng, T. S. P., Painter, M. W. & Immunological Genome Project Consortium. The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008).
87. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
88. Nakanishi, Y., Lu, B., Gerard, C. & Iwasaki, A. CD8+ T lymphocyte mobilization to virus-infected tissue requires CD4+ T-cell help. *Nature* **462**, 510–513 (2009).
89. Gilchuk, P. *et al.* A distinct lung-interstitium-resident memory CD8(+) T cell subset confers enhanced protection to lower respiratory tract infection. *Cell Rep.* **16**, 1800–1809 (2016).
90. Pizzolla, A. *et al.* Resident memory CD8+ T cells in the upper respiratory tract prevent pulmonary influenza virus infection. *Sci Immunol* **2**, (2017).

Acknowledgments

We thank members of the Kaech and Wang laboratory for their advice and assistance with experiments. We thank Dr. Justin Milner (University of North Carolina Chapel Hill) for helpful discussions, reagents, and LCMV-infected mice. This work was supported by NIH grants to S.M.K. R37AI066232, R01AI123864, R21AI151986, and R01CA240909, to W.W. R01AI150282, R01HG009626 (in part) to H.K.C K01EB034321, to J.E.T R01CA248359 and R01CA244361, to D.C.H AI151123, to Y.W. EB029122 and GM140929, and to L.L. K01EB035649. H.K.C. is a Damon Runyon Fellow supported by Damon Runyon Cancer Research Foundation grant [DRG-2374-19].

Author contributions

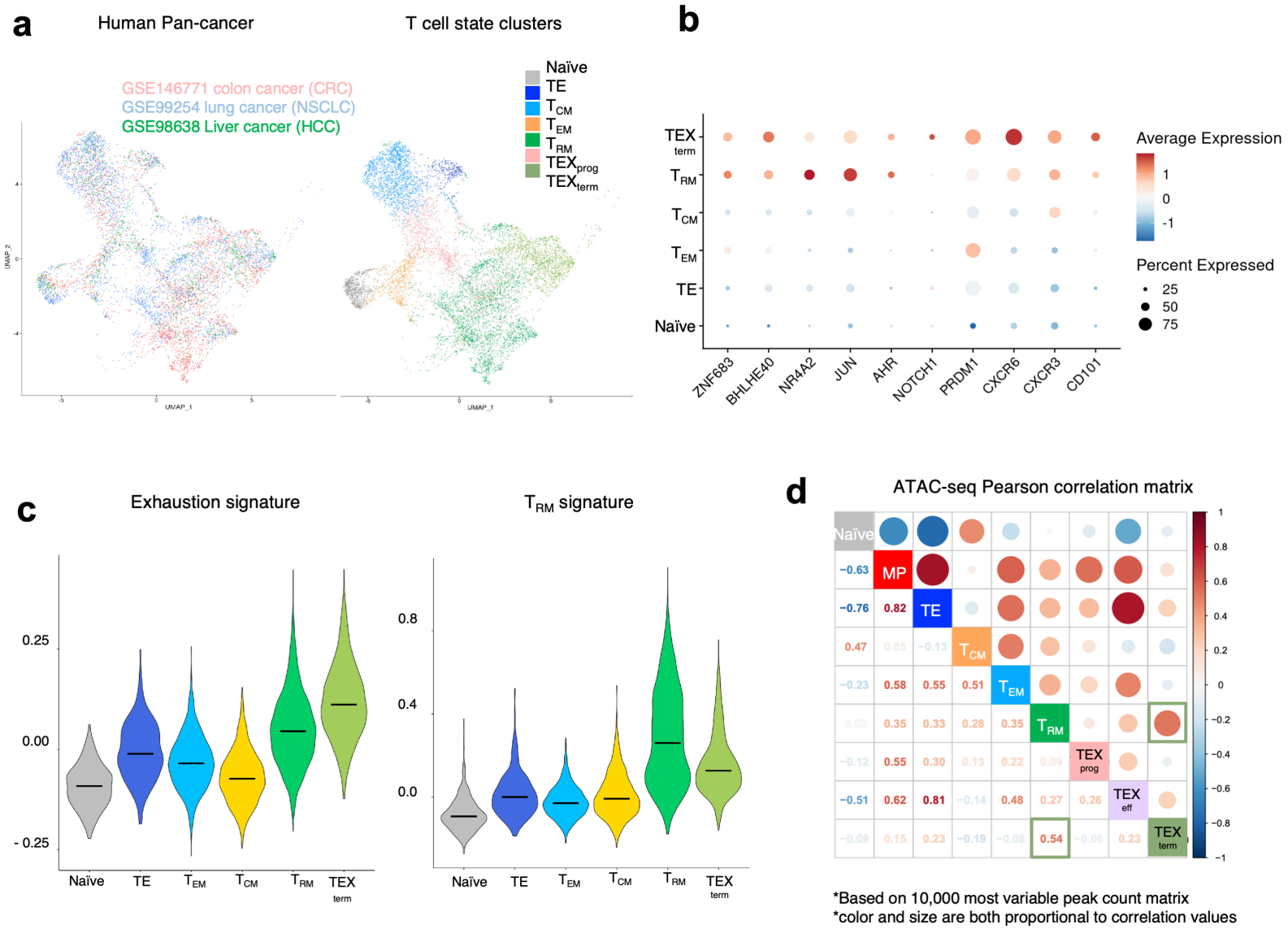
H.K.C., S.M.K., and W.W. conceptualized, designed, and supervised the study. H.K.C. designed and performed experiments. H.K.C, C.L., A.J., and M.S. analyzed the data. B.P.R. and J.E.T. performed proteasome activity experiments, provided resources, and reviewed the manuscripts. E.C., B.M. D.C., B.C., Q.Y., F.H., T.H.M., S.K.V., V.T., U.H.C., and J.H. assist experiments. J.W., B.M., M.S., Y.H., L.H., U.H.C., J. M., and A.W. contribute to bioinformatics data analysis. J.H. and D.H. provided resources and training for ATAC-sequencing. H.K.C, and C.L. wrote the manuscript. H.K.C., W.W., and S.M.K. edited and revised the manuscript.

Competing interests

S.M.K. is SAB member for Pfizer, EvolveImmune Therapeutics, Arvinas and Affini-T, advisor for Barer Institute of Raphael Holdings and Academic Editor at Journal of Experimental Medicine. The remaining authors declare no conflict of interest.

Additional information

Supplementary Information is available for this paper. Correspondence and requests for materials should be addressed to Susan M. Kaech (skaech@salk.edu) and Wei Wang (wei-wang@ucsd.edu).



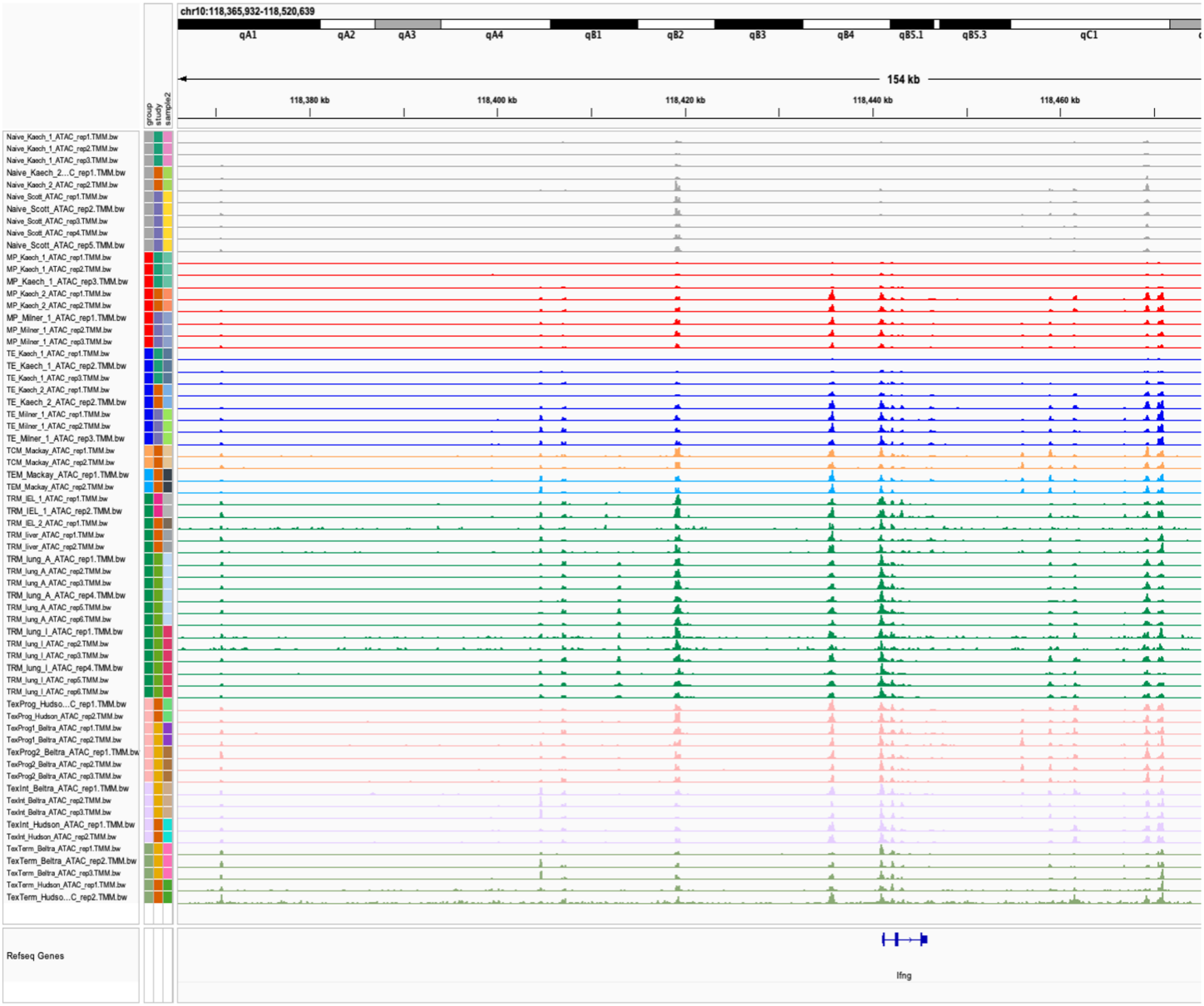
Extended Data Fig. 1 | Parallel differentiation of T_{RM} and TEX_{term} and their transactional and epigenetic similarity. a, UMAP from scRNA-seq data of T cells from blood, tumor, and normal tissues of CRC, NSCLC, and HCC patients. Unbiased clustering identified several T cell states previously reported in LCMV infection and tumors. **b**, T_{RM} markers show higher expression in TEX_{term} cluster from Pan-cancer scRNA-seq in a. **c**, Exhaustion[17] and T_{RM}-signature genes[37] are upregulated in both T_{RM} and TEX_{term} clusters. **d**, Pearson correlation matrix of batch-corrected ATAC-seq. Color and size are both proportional to correlation values.

a

10 datasets | 9 subsets | 121 experiments | 52 ATAC-seq | 69 RNA-seq

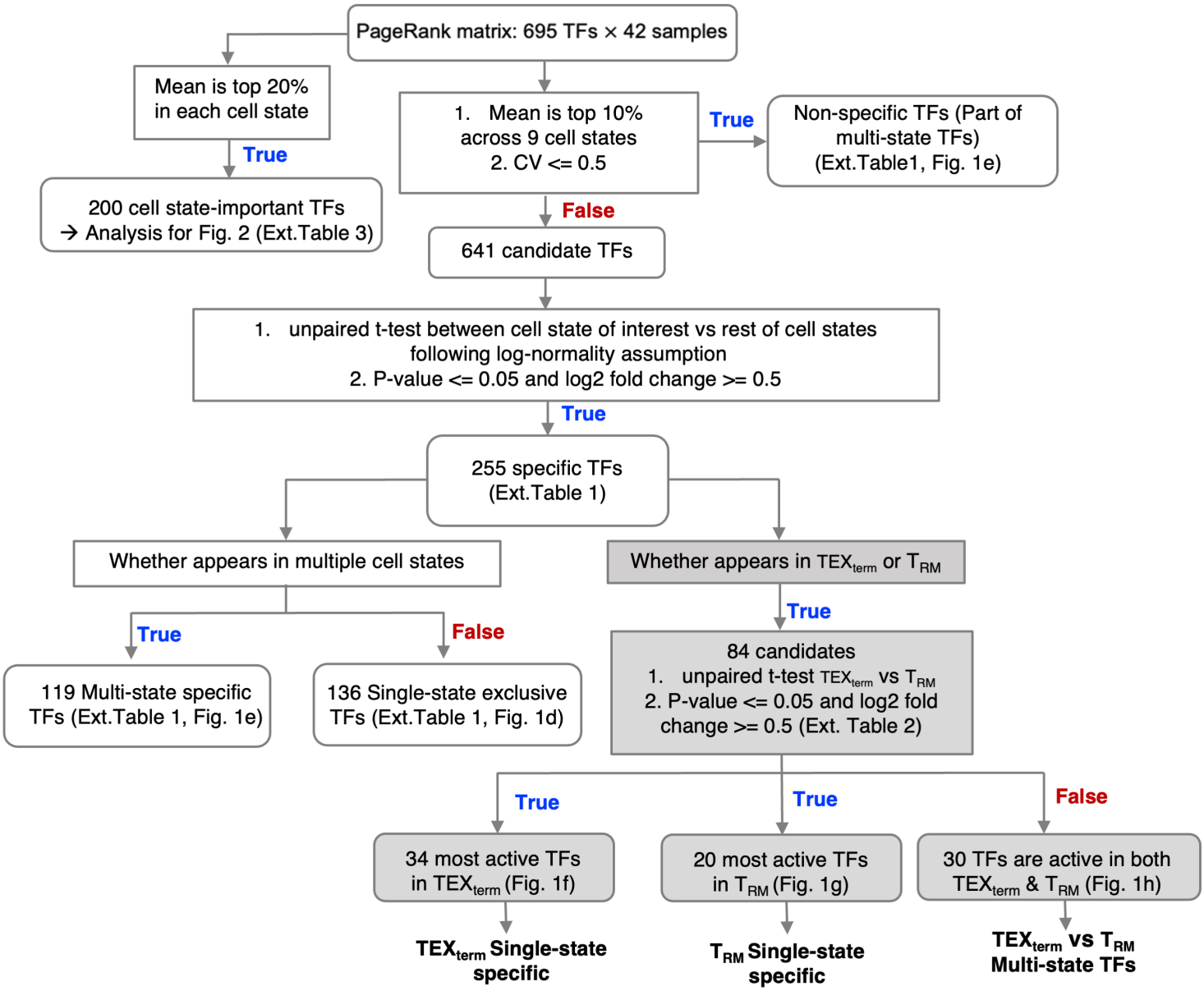
Dataset	Accession	# of experiments	Subsets
Milner(2017)	GSE107281	2 RNA + 2 ATAC	T_{RM}
Milner(2020.Apr)	GSE147080	2 RNA	T_{RM}
Milner(2020.Aug)	GSE157072	12 RNA	Naïve, MP, TE, T_{CM}, T_{EM}
Kaech(2018)	GSE111902	9 RNA + 9 ATAC	Naïve, TE, MP
Scott(2016)	GSE88987	4 RNA + 11 ATAC	Naïve, TE, MP
Mackay(2016)	GSE70813	10 RNA	Naïve, T_{CM}, T_{EM}, T_{RM}
Renkema(2020)	GSE152841	9 RNA	Naïve, T_{CM}, T_{EM}
Hudson(2019)	PRJNA497086	9 RNA	TEX_{prog}, TEX_{eff}, TEX_{term}
Beltra(2020)	GSE149876	12 RNA + 11 ATAC	TEX_{prog}, TEX_{eff}, TEX_{term}
Chung(2024)	This study	19 ATAC	Naïve, TE, MP, T_{CM}, T_{EM}, T_{RM}, TEX_{prog}, TEX_{eff}, TEX_{term}

b



Extended Data Figure 2 | Composition of multi-omic atlas. **a**, A total of 121 experiments across multiple data sets were utilized to generate an epigenetic and transcriptional atlas of $CD8^+$ T cells under chronic and acute antigen exposure. **b**, Batch-effect corrected ATAC-seq across 9 $CD8^+$ T cell states.

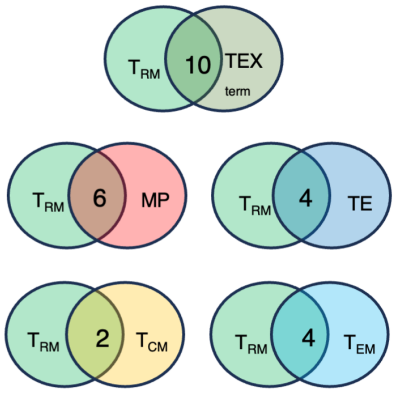
a



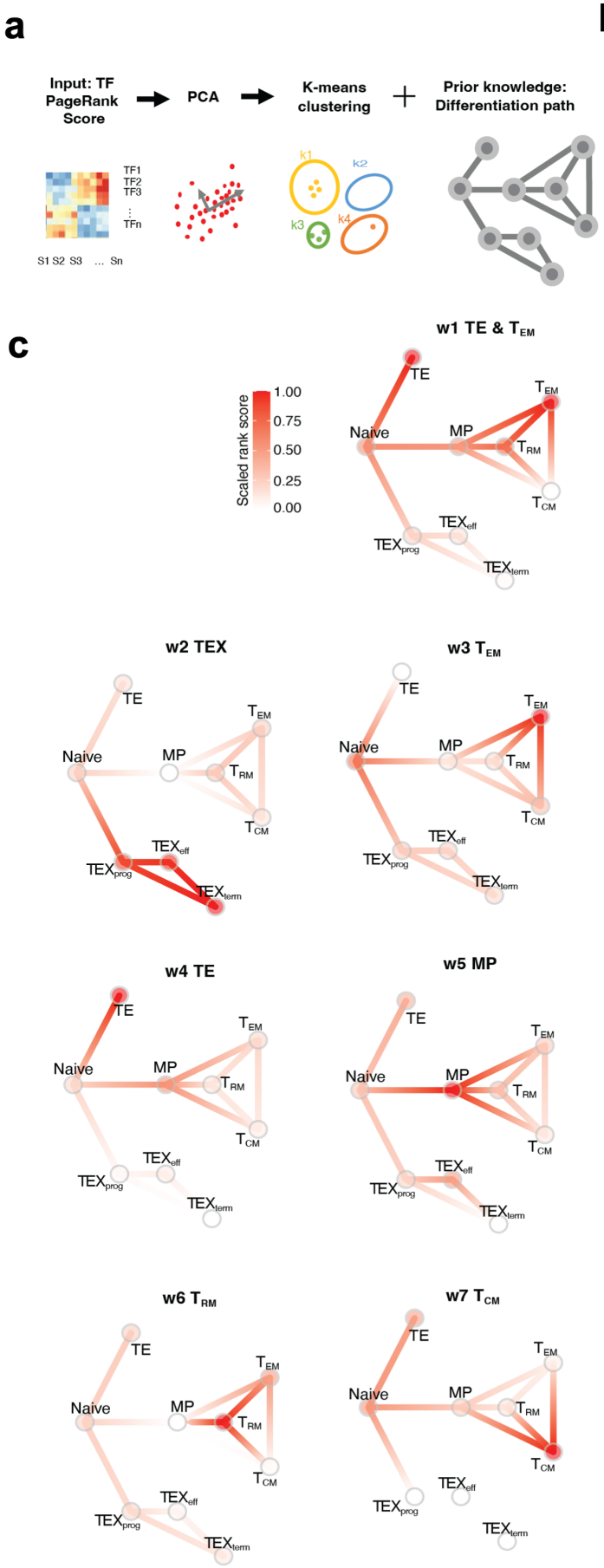
b

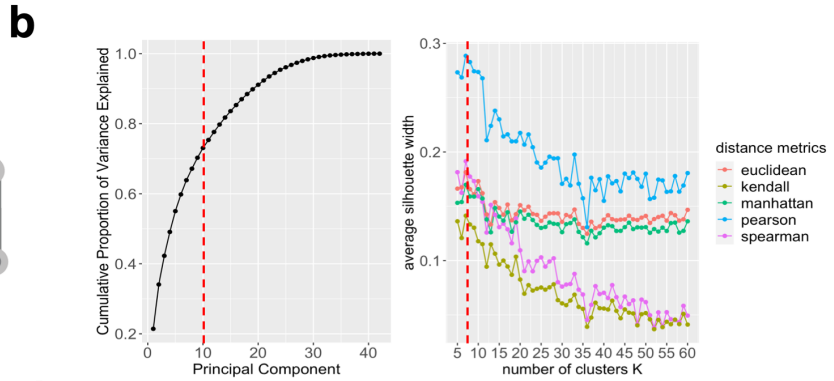
	# of TFs active in each cell state	# of state- specific TFs	New, specific TFs examples
Non-specific TFs	54	NA	
Naive	31	14	*Hoxa7*
MP	57	18	*Hoxb4*
TE	57	15	*Amtl*
T_{CM}	28	15	*Cux1*
T_{EM}	25	13	*Meis3, Zfp398*
T_{RM}	43	17	*Snai1*
TEX_{prog}	63	19	*Hey1*
TEX_{eff}	62	13	*Sox8*
TEX_{term}	51	12	*Zscan20, Zfp143, Arid3a*

c



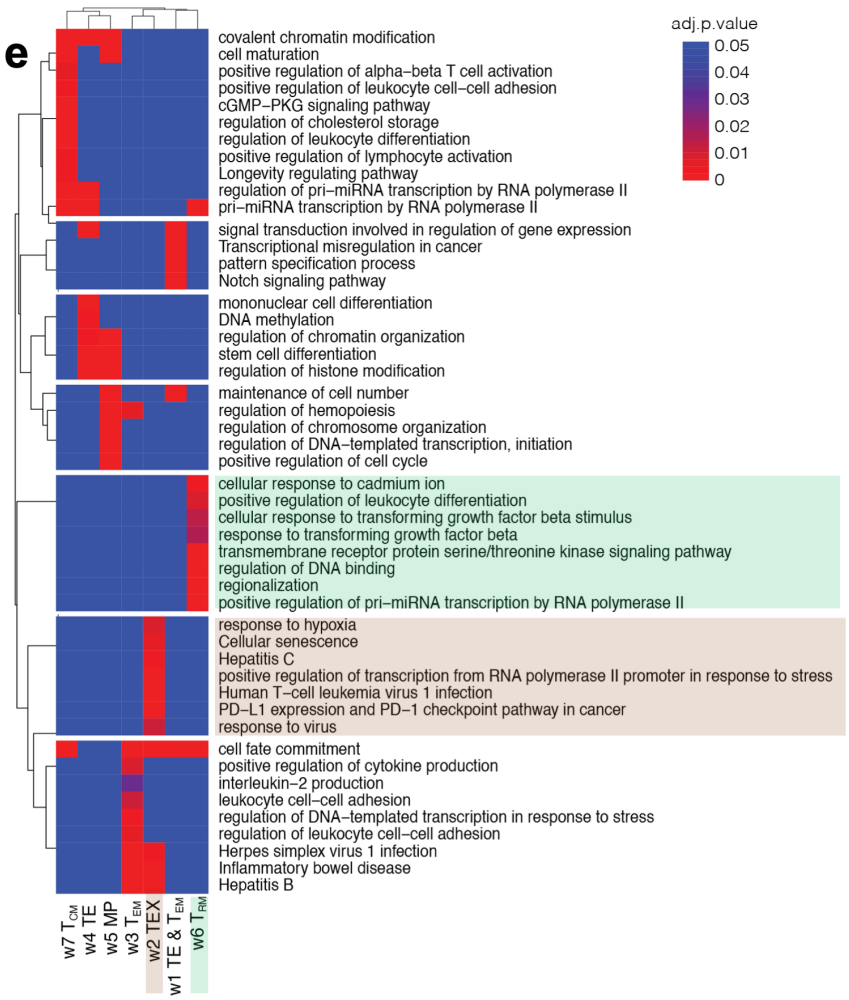
Extended Data Fig. 3 | Cataloging key TFs in 9 different CD8⁺ T cell states. **a**, Logic flow of the unbiased PageRank comparison to catalogue cell-state-specific/non-specific TFs. **b**, The number of TFs catalogued in each cell state. **c**, Venn Diagrams showing TFs shared with the T_{RM} cell state.



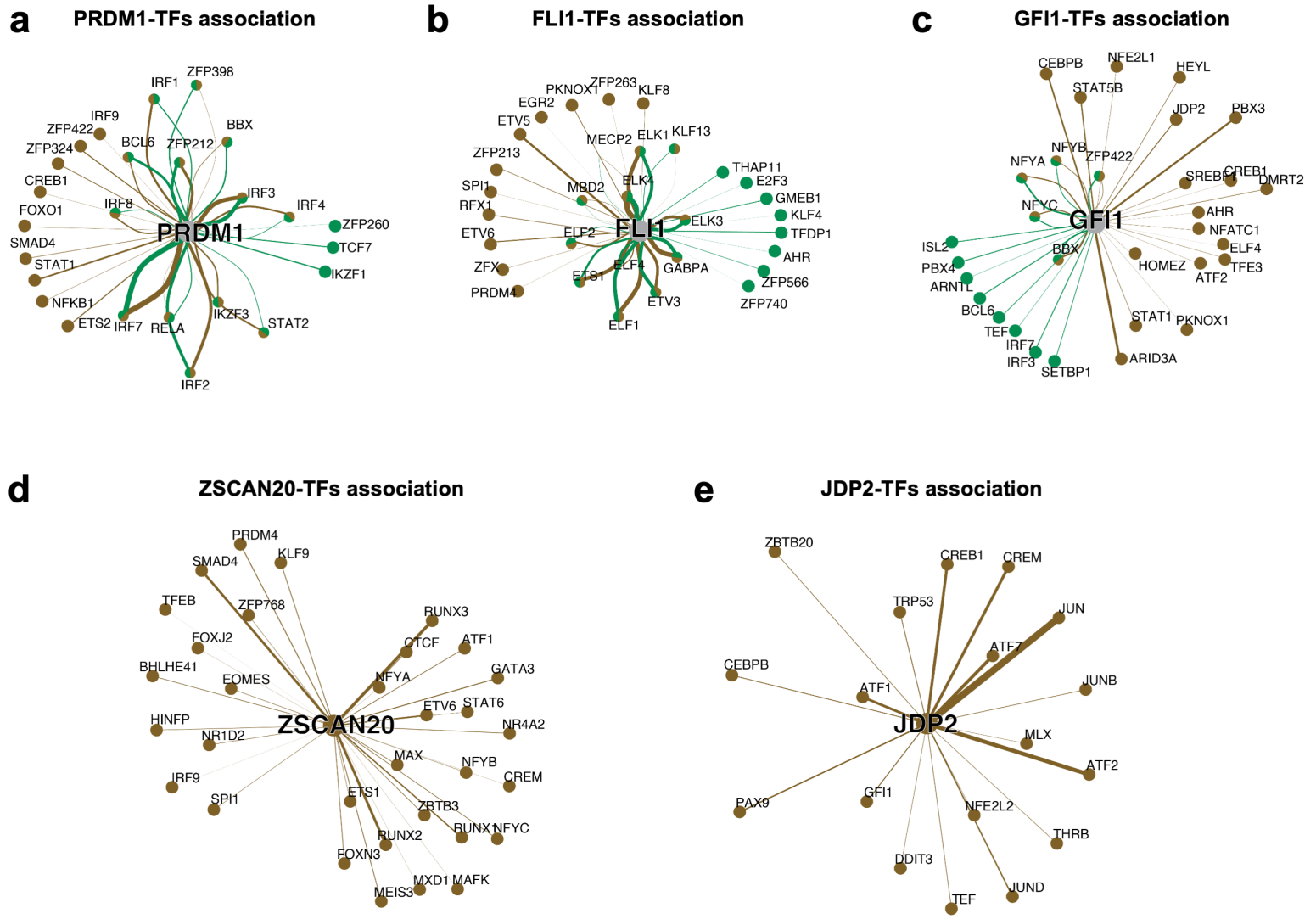


d

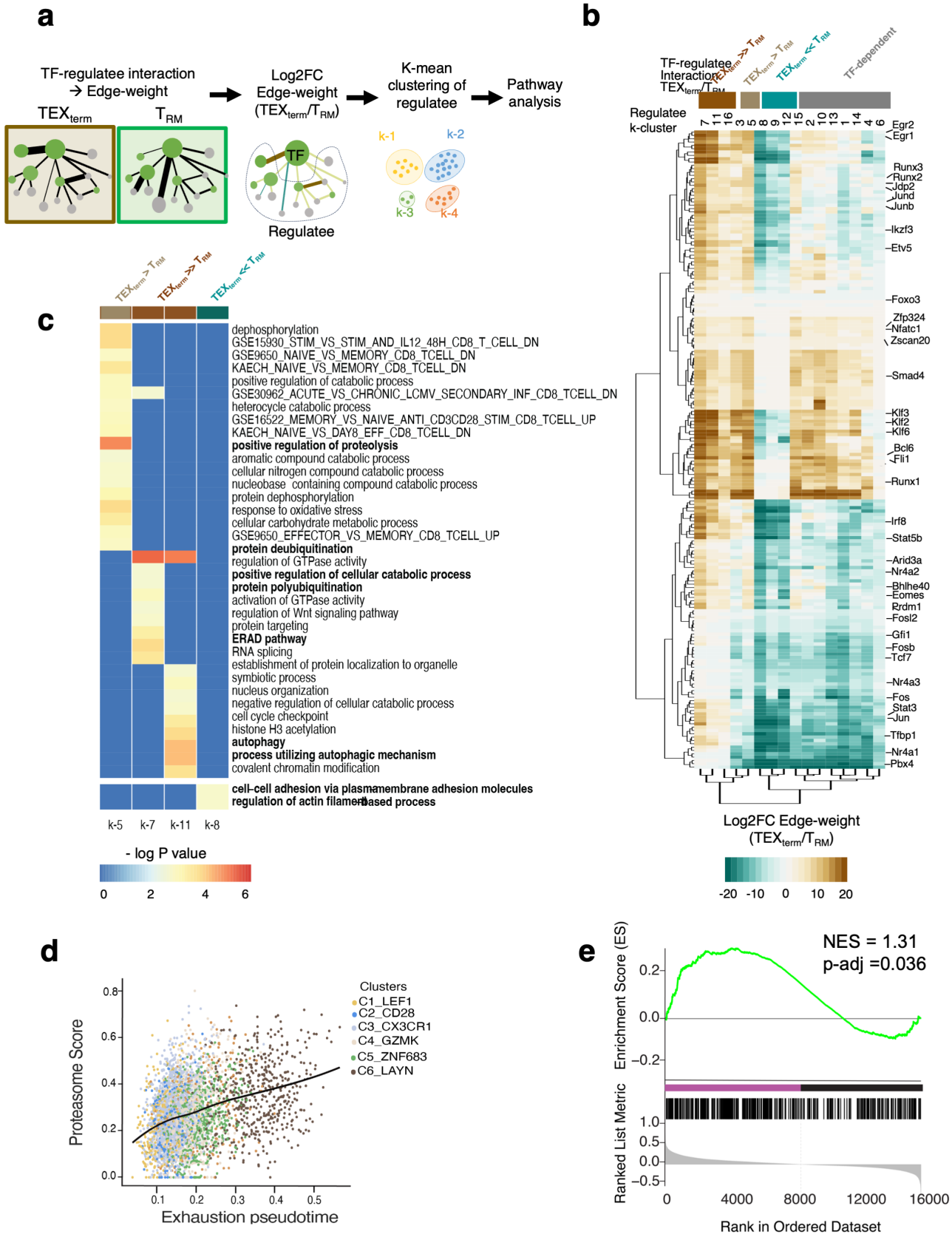
Wave	TFs (Top 30)
T_{RM} **wave**	*Nr4a3 Nr4a1 Atf3 Crem Egr1 Klf6 Junb Fosb Tgif1 Isl2 Snai1 Atf4 Irf4 Fosl2 Jun Phox2b Tlx2 Klf5 Cdx1 Mxd1 Shox2 Etv3 Fos Egr2 Nobox Foxi1 Msx1 Vsx1 Wt1 Msx3 Onecut1 Gm28047*
T_{CM} **wave**	*Elk4 Foxo3 Plag1 Zfp341 Usf2 Rel Rreb1 Zbtb33 Pou4f1 Zbtb6 Sp4 Erf Yy1 Foxk1 Zbtb26 Zfp691 Epas1 Cic Rara Arid2 Srebf2 Bach1 Glis3 Zfp398 Jrkl Nfat5 Sox12 Zscan26 Zfp770*
T_{EM} **wave**	*Sp140 Atf6 Lcor Arntl E2f2 Naif1 Zfp69 Vezf1 Tbx21 Zfp580 Lbx2 Zzz3 Mybl1 Zfp189 Zfp212 Zfp260 Creb3l1 Lcorl Sp110 Nr6a1 Fli1 Ets1 Zkscan1 Hoxb1 Zfp932 Gabpa Hmg20b Mlxipl*
MP wave	*Lef1 Hoxa7 Sox4 Nfix Gata2 Esr1 Tcf7 Foxj3 Lyl1 Pbx1 Cebpa Tcf12 Foxj1 Hand2 Cebpe Gata1 Cxxc5 Ets2 Hoxb3 Pax5 Thra Tcf7l2 Mecom Tal1 Zfp281 Zbtb18 Myc Zfp202 Zfp647*
TE & T_{EM} **wave**	*Sall4 Insm1 Ascl1 Sox17 Nr2f2 Nhlh1 Tfap2c Klf15 Gata4 Tfap2b Tal2 Gata6 Atf7 Esr2 Neurog2 Foxi3 Six1 Ehf Elf5 Sox3 Zfp454 Nhlh2 Foxa3 Prdm14 Gli3 Sp8 Tcfl5 Six4 Foxp2*
TE wave	*Nr1d1 Tfap2a Foxo4 Zfp652 Nr1h2 Gli1 Zbtb48 Glis2 Smad3 E2f5 Phf21a Clock Hmbox1 Tead2 Bcl6b Bach2 Nr1i3 Zfp768 Rarb Taf1 Nr2c2 Zfp296 Zscan10 Nr1i2 Grhl2 Cux2 Tgif2*
TEX wave	*Vax2 Jdp2 Stat3 Irf8 Nfatc1 Irf9 Irf7 Klf10 Prdm16 Klf9 E2f6 Zfp324 Eomes Tfdp1 Stat2 Dmrt2 Stat1 Zfp41 Irf1 E2f4 Zfp523 Batf Zfp143 Grhl1 Stat5b Cebpb Spic Zfp128 Tfeb Osr2 Pax9*



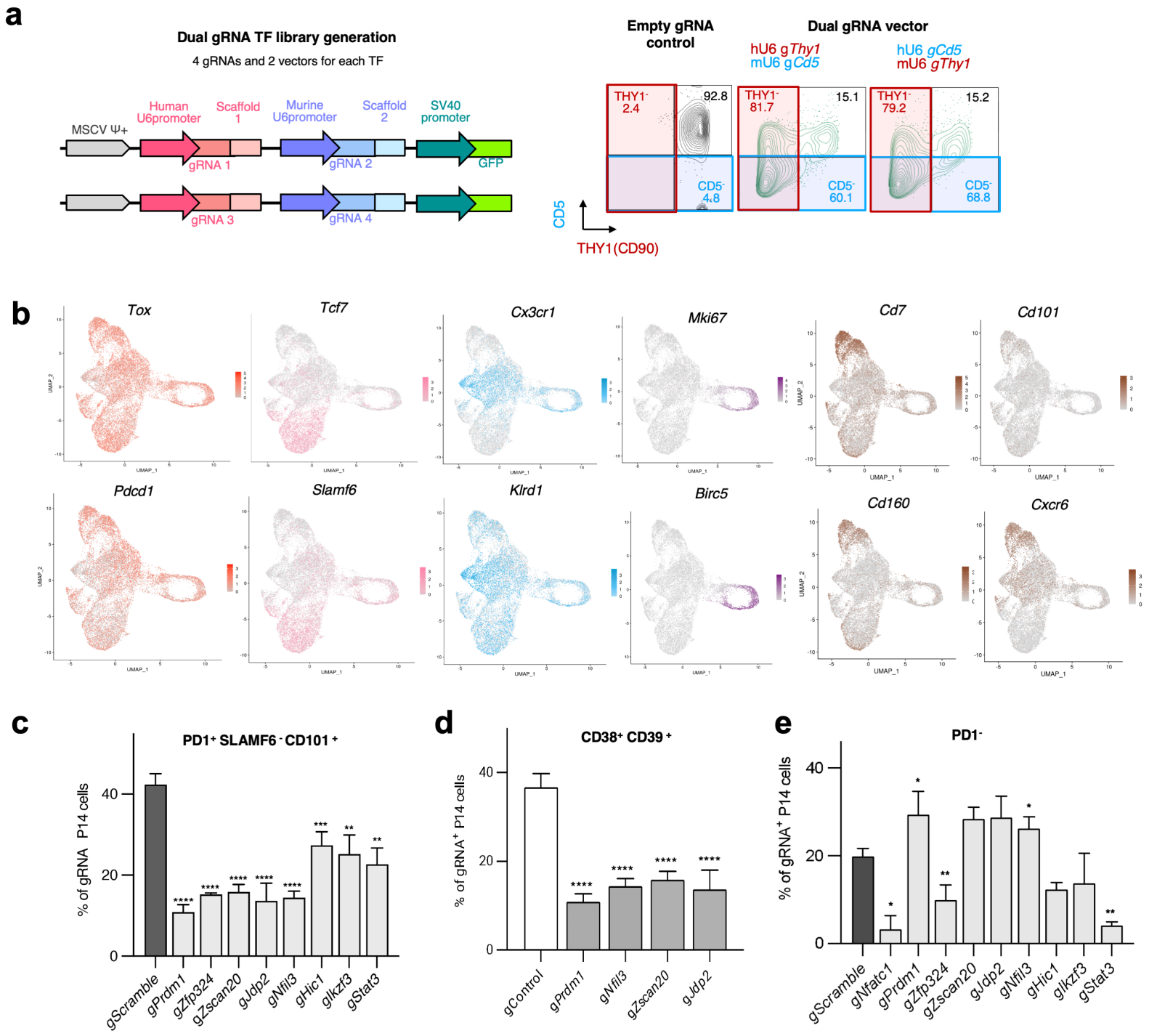
Extended Data Fig. 4 | TF wave analysis. a, Schematic of the analysis pipeline. **b**, Selection of algorithms and parameters for TF wave analysis. The Pearson correlation was chosen for the distance metric, with k=7 chosen as the optimal cluster number. **c**, The seven TF waves. Circles represent specific cell states. Red color indicates normalized PageRank scores. **d**, List of TF members in each wave **e**, Heatmap of biological pathways enriched in each TF wave, with a red-blue color scale indicating the p-value.



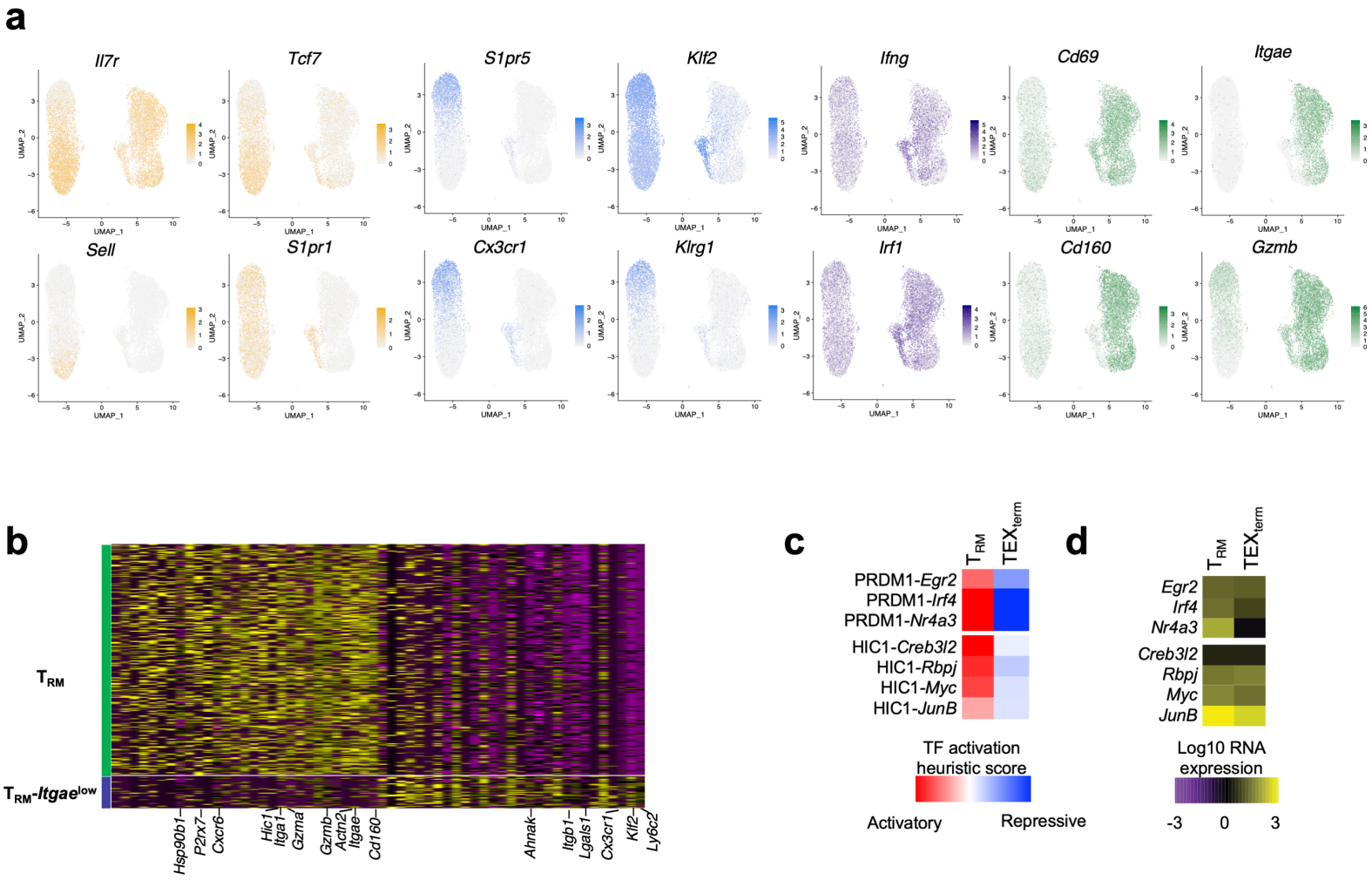
Extended Data Fig. 5 | TF-TF association in T_{RM} or TEX_{term} a-e, PRDM1(**a**), FLI1(**b**), GFI1 (**c**), ZSCAN20 (**d**) and JDP2 (**e**) and TF association networks. Line color indicates interaction specificity to T_{RM} (green) or TEX_{term} (brown), and line thickness represents interaction intensity.



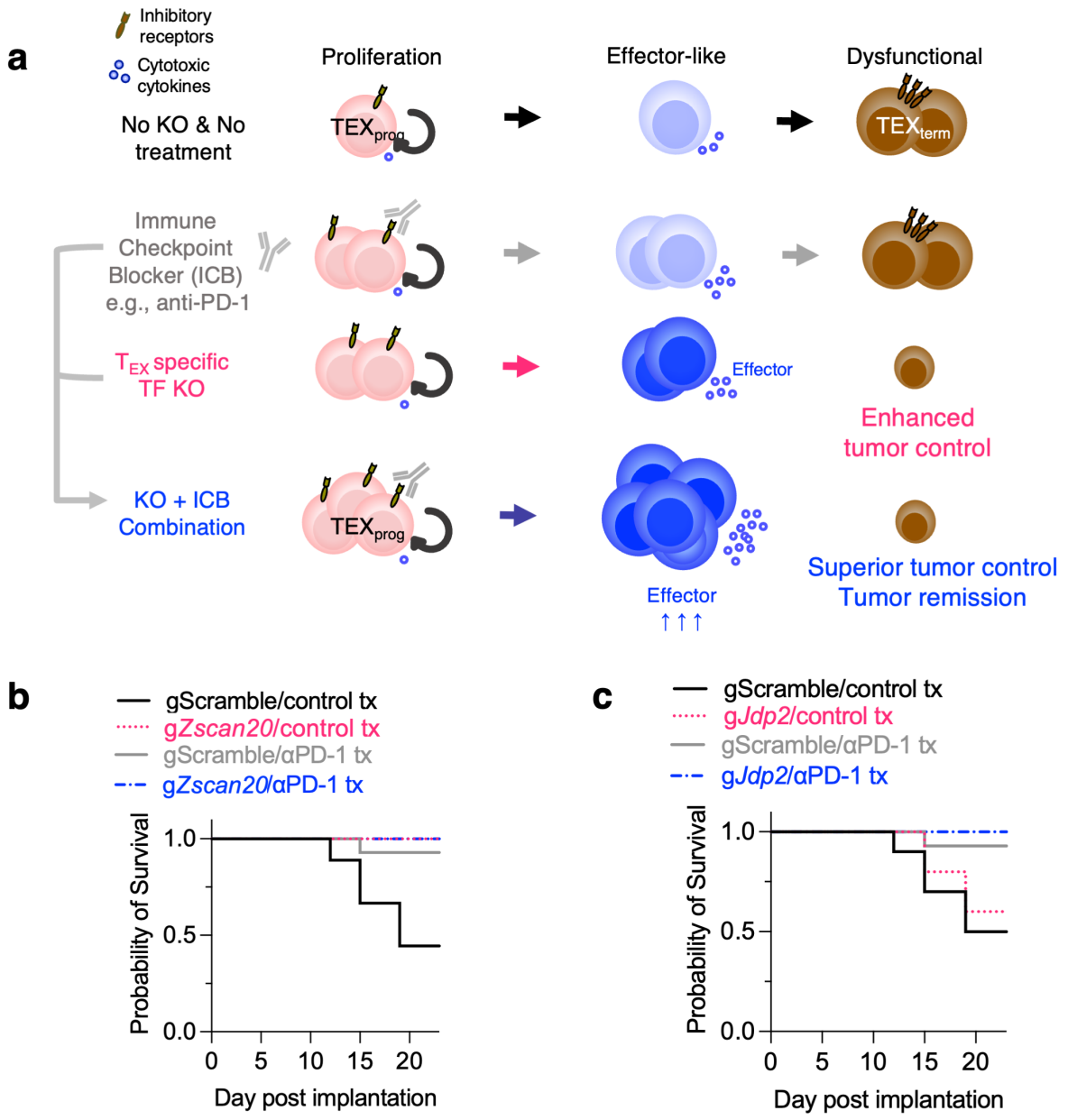
Extended Data Fig. 6 | Global TF-TF and TF-regulatees analysis revealed biological pathways unique in T_{RM} or TEX_{term}. **a**, Pipeline of a comparative analysis of the TF-regulatee interaction intensity between T_{RM} or TEX_{term}. **b**, Heatmap of the log2 fold-change (log2FC) of edge weight (TF-regulatee interaction) between TEX_{term} and T_{RM}. Rows represent regulatee clusters, and columns represent TFs. The brown color denotes the edge weight in TEX_{term} than T_{RM} cells. The green color indicates the converse. **c**, Pathway enrichment in each K-means cluster of regulatees. **d**, Retrospective pseudotime analysis of $CD8^+$ TIL scRNA-seq data categorized by differentiation status, showing correlation of the KEGG proteasome gene score (M10680) with exhaustion. Data from non-small cell lung cancer patients (n=14). **e**, Proteasome-mediated ubiquitin-dependent protein catabolic process gene set (GO:0043161) enrichment analysis of RNA-seq from $CD8^+$ splenocytes (black) or TILs (purple), n=3/3.



Extended Data Fig. 7 | *in vivo* Perturb-seq and individual validation identified several novel TFs for TEX_{term} differentiation. a, Retroviral dual-gRNA vector design and validation. **b**, Feature plot of differentiation markers (Pan-exhaustion: red, TEX_{prog}: pink, TEX_{eff}: blue, Cell cycle: purple, TEX_{term}: brown). **c-e**, Quantification of the frequency of $PD1^+$ $SLAFM6^-$ $CD101^+$ P14 $CD8^+$ cells (**c**), Inhibitory receptors $CD38^+$ $CD39^+$ double positive P14 $CD8^+$ cells (**d**), $PD1^-$ P14 $CD8^+$ cells (**e**). Ordinary one-way ANOVA with Dunnett's multiple comparison test. Data are from more than two biological replicates, n ≥ 5. ****P<0.0001, ***P<0.001, **P<0.01, *P<0.05. Data are presented as the mean ± s.e.m.



Extended Data Fig. 8 | Perturb-seq and TF regulation switching in T_{RM} and TEX_{term}. **a**, Feature plots of differentiation markers (T_{CM}: red, T_{EM}: blue, T_{RM}-*Itgae*^{low}: navy, T_{RM}: green). **b**, Heatmap of differentially expressed genes between T_{RM}-*Itgae*^{low} and T_{RM} clusters. **c**, A heatmap of the heuristic scores for TF and regulatee interactions, representing the intensity of TF activation and the direction of regulation (Activatory: red, Repressive: blue). **d**, RNA expression levels of target genes regulated by TFs.



Extended Data Fig. 9 | Tumor control by TEX_{term}-specific TF deficient T cells. **a,** Model illustrating the synergy between TEX_{term} driver TF deficiency and immune checkpoint therapy. **b**, Survival of melanoma-bearing mice receiving *Zscan20* KO or control P14 cells, followed by anti-PD1 or isotype IgG2a treatment. Mice were implanted with 5 X 10^5 B16-GP33 cells. **c**, Survival of tumor-bearing mice with *Jdp2* KO P14 transfer.

Extended Data Figure Legends

Extended Data Fig. 1 | Parallel differentiation of T_{RM} and TEX_{term} and their transactional and epigenetic similarity. a, UMAP from scRNA-seq data of T cells from blood, tumor, and normal tissues of CRC, NSCLC, and HCC patients. Unbiased clustering identified several T cell states previously reported in LCMV infection and tumors. **b**, T_{RM} markers show higher expression in TEX_{term} cluster from Pan-cancer scRNA-seq in a. **c**, Exhaustion[17] and T_{RM}-signature genes[37] are upregulated in both T_{RM} and TEX_{term} clusters. **d**, Pearson correlation matrix of batch-corrected ATAC-seq. Color and size are both proportional to correlation values.

Extended Data Figure 2 | Composition of multi-omic atlas. **a**, A total of 121 experiments across multiple data sets were utilized to generate an epigenetic and transcriptional atlas of $CD8^+$ T cells under chronic and acute antigen exposure. **b**, Batch-effect corrected ATAC-seq across 9 $CD8^+$ T cell states.

Extended Data Fig. 3 | Cataloging key TFs in 9 different $CD8^+$ T cell states. **a**, Logic flow of the unbiased PageRank comparison to catalogue cell-state-specific/non-specific TFs. **b**, The number of TFs catalogued in each cell state. **c**, Venn Diagrams showing TFs shared with the T_{RM} cell state.

Extended Data Fig. 4 | TF wave analysis. a, Schematic of the analysis pipeline. **b**, Selection of algorithms and parameters for TF wave analysis. The Pearson correlation was chosen for the distance metric, with k=7 chosen as the optimal cluster number. **c**, The seven TF waves. Circles represent specific cell states. Red color indicates normalized PageRank scores. **d**, List of TF members in each wave **e**, Heatmap of biological pathways enriched in each TF wave, with a red-blue color scale indicating the p-value.

Extended Data Fig. 5 | TF-TF association in T_{RM} or TEX_{term} a-e, PRDM1(**a**), FLI1(**b**), GFI1 (**c**), ZSCAN20 (**d**) and JDP2 (**e**) and TF association networks. Line color indicates interaction specificity to T_{RM} (green) or TEX_{term} (brown), and line thickness represents interaction intensity.

Extended Data Fig. 6 | Global TF-TF and TF-regulatees analysis revealed biological pathways unique in T_{RM} or TEX_{term}. a, Pipeline of a comparative analysis of the TF-regulatee interaction intensity

between T_{RM} or TEX_{term}. **b**, Heatmap of the log2 fold-change (log2FC) of edge weight (TF-regulatee interaction) between TEX_{term} and T_{RM}. Rows represent regulatee clusters, and columns represent TFs. The brown color denotes the edge weight in TEX_{term} than T_{RM} cells. The green color indicates the converse. **c**, Pathway enrichment in each K-means cluster of regulatees. **d**, Retrospective pseudotime analysis of $CD8^+$ TIL scRNA-seq data categorized by differentiation status, showing correlation of the KEGG proteasome gene score (M10680) with exhaustion. Data from non-small cell lung cancer patients (n=14). **e**, Proteasome-mediated ubiquitin-dependent protein catabolic process gene set (GO:0043161) enrichment analysis of RNA-seq from $CD8^+$ splenocytes (black) or TILs (purple), n=3/3.

Extended Data Fig. 7 | *in vivo* Perturb-seq and individual validation identified several novel TFs for TEX_{term} differentiation. a, Retroviral dual-gRNA vector design and validation. **b**, Feature plot of differentiation markers (Pan-exhaustion: red, TEX_{prog}: pink, TEX_{eff}: blue, Cell cycle: purple, TEX_{term}: brown). **c-e**, Quantification of the frequency of $PD1^+$ $SLAFM6^-$ $CD101^+$ P14 $CD8^+$ cells (**c**), Inhibitory receptors $CD38^+$ $CD39^+$ double positive P14 $CD8^+$ cells (**d**), $PD1^-$ P14 $CD8^+$ cells (**e**). Ordinary one-way ANOVA with Dunnett's multiple comparison test. Data are from more than two biological replicates, n ≥ 5. ****P<0.0001, ***P<0.001, **P<0.01, *P<0.05. Data are presented as the mean ± s.e.m.

Extended Data Fig. 8 | Perturb-seq and TF regulation switching in T_{RM} and TEX_{term}. a, Feature plots of differentiation markers (T_{CM}: red, T_{EM}: blue, T_{RM}-*Itgae*^{low}: navy, T_{RM}: green). **b**, Heatmap of differentially expressed genes between T_{RM}-*Itgae*^{low} and T_{RM} clusters. **c**, A heatmap of the heuristic scores for TF and regulatee interactions, representing the intensity of TF activation and the direction of regulation (Activatory: red, Repressive: blue). **d**, RNA expression levels of target genes regulated by TFs.

Extended Data Fig. 9 | Tumor control by TEX_{term}-specific TF deficient T cells. a, Model illustrating the synergy between TEX_{term} driver TF deficiency and immune checkpoint therapy. **b**, Survival of melanoma-bearing mice receiving *Zscan20* KO or control P14 cells, followed by anti-PD1 or isotype IgG2a treatment. Mice were implanted with 5 X 10^5 B16-GP33 cells. **c**, Survival of tumor-bearing mice with *Jdp2* KO P14 transfer.

Extended Data Tables

Extended Data Table 1 | TF PageRank scores of multi-state and state-specific TFs

Extended Data Table 2 | TF PageRank scores of T_{RM}- and TEX_{term}-specific TFs and multi-state TFs

Extended Data Table 3 | Catalog of cell-state important TFs

Extended Data Table 4 | Catalog of TF communities in T_{RM} and TEX_{term}

Extended Data Table 5 | TF gRNA library

Extended Data Table 6 | Gene signatures

Extended Data Note

The comparative landscape of TF and regulatee interaction in T_{RM} and TEX_{term}

Our global TF community analysis reveals several biological pathways that are differentially associated with either T_{RM} or TEX_{term} cell states (Fig. 2f-h). Interestingly, many core pathways are jointly regulated by more than one TF community. For example, the ubiquitin-proteasome pathway is associated with communities 1 and 4 of TEX_{term}. We postulated that this is due to shared regulatees across different communities and thus sought to profile the global landscape of TF and regulatee interaction strength in T_{RM} or TEX_{term} contexts. Unbiased K-means clustering based on the TF and its regulatee interaction (Taiji edge-weights) in TEX_{term} and T_{RM} identified 16 clusters of genes with distinctive comparative interaction patterns (Extended Data Fig. 6a,b). Regulatees in k-8, 9, and 12, "$TEX_{term} < T_{RM}$ group", interact more strongly with TFs in T_{RM} than TEX_{term}. Genes in the $TEX_{term} < T_{RM}$ group (k-8) are enriched in cell-cell adhesion and actin filament-based processes (Extended Data Fig. 6c), which promote intraepithelial migration and the subsequent formation of $CD8^+$ T_{RM} cells in the epithelium[31,88–90]. Conversely, regulatees in k-7,11,16, 3, and 5 clusters, "$TEX_{term} > T_{RM}$ group", have higher edge-weights in TEX_{term} than T_{RM}. The $TEX_{term} > T_{RM}$ group (k-11 and k-7) associated with signatures related to proteostasis such as protein deubiquitination, polyubiquitination, endoplasmic reticulum-associated protein degradation (ERAD), and autophagy (Extended Data Fig. 6c), which are also reported in TF community analysis (Fig. 2g).