Multiomics atlas-assisted discovery of transcription factors enables specific cell state programming

Authors: H. Kay Chung[1*], Cong Liu[2*], Eduardo Casillas[1], Bryan Mcdonald[1], Jun Wang[2], Ming Sun[1], Shixin Ma[1], Brent Chick[3], Qiyuan Yang[4], Dan Chen[1], Filipe Hoffmann[1], Siva Karthik Varanasi[1], Victoria Tripple[1], Yuqing Hang[4], Ukrae H. Cho[3], Josephine Ho[3], April Williams[4], Diana Hargreaves[3], Susan M. Kaech[1#], Wei Wang[2#]

Affiliations:

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

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

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

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

* Equal contribution, # Co-corresponding

Abstract: The same types of cells can assume diverse states with varying functionalities. Effective cell therapy can be achieved by specifically driving a desirable cell state, which requires the elucidation of key transcription factors (TFs). Here, we integrated epigenomic and transcriptomic data at the systems level to identify TFs that define different $CD8^+$ T cell states in an unbiased manner. These TF profiles can be used for cell state programming that aims to maximize the therapeutic potential of T cells. For example, T cells can be programmed to avoid a terminal exhaustion state (Tex_{Term}), a dysfunctional T cell state that is often found in tumors or chronic infections. However, Tex_{Term} exhibits high similarity with the beneficial tissue-resident memory T states (T_{RM}) in terms of their locations and transcription profiles. Our bioinformatic analysis predicted *Zscan20,* a novel TF, to be uniquely active in Tex_{Term}. Consistently, *Zscan20* knock-out thwarted the differentiation of Tex_{Term} *in vivo*, but not that of T_{RM}. Furthermore, perturbation of *Zscan20* programs T cells into an effector-like state that confers superior tumor and virus control and synergizes with immune checkpoint therapy. We also identified *Jdp2* and *Nfil3* as powerful Tex_{Term} drivers. In short, our multiomics-based approach discovered novel TFs that enhance anti-tumor immunity, and enable highly effective cell state programming.

One sentence summary: Multiomics atlas enables the systematic identification of cell-state specifying transcription factors for therapeutic cell state programming.

Main

Cell state can be described as the range of cellular phenotypes arising from the interaction of a defined cell type with its environment. Advanced genomics and proteomics technology enable not only precise characterization of cell states but also provide a stunningly high-resolution view of transitions between states. T cells serve as a well-characterized example: these cells exist in different activation states, which arise in response to different stimuli, yet they maintain their T-cell identity (*1–4*). Cell state is directly associated with the function of the given cell type. Cells in the same state will share the pathways that are active, the genes that are expressed, and the functions that are being performed. Therefore effective programming of cell state can be a key for successful cell therapy by generating cells that exert specific therapeutic functions. Transcription factors (TFs) are fundamental regulators of both cell type and cell state differentiation (*5*, *6*). Identification of the key TFs specifying the cell state is pivotal to devising programming strategy but remains a great challenge. Here we report a multiomics approach to systematically identify TFs that define different $CD8^+$ T cell states and that are utilized to program better antiviral and anti-tumor T cell immunity.

Naive $CD8^+$ T cells differentiate into heterogeneous states to offer immune defense against intracellular pathogens and cancer (*1*). In acute infection, T cells assume diverse effector and memory states. After the virus control, highly plastic memory precursor (MP) cells develop into memory T cells that mediate long-term protective immunity at different locations ("Central memory (T_{CM})" cells at secondary lymphoid, "Effector memory (T_{EM})" in blood, "Tissue-resident memory (T_{RM})" at peripheral tissues, Fig. 1A). On the other hand, T cells in chronic virus infections or tumors progressively become dysfunctional T cells and lose memory potential of which phenomenon is known as T cell exhaustion (*7–9*). Similar to acute infection, the $CD8^+$ T cells in the exhaustion trajectory comprise multiple cell states that serve different roles (Fig 1A). Akin to MP cells and T_{CM}, cells in Tex_{Prog} state display plasticity and proliferative capacity and can differentiate into transitory-intermediate "effector-like" states $Tex_{eff\text{-}like}$ (*2*, *3*, *10*). T cells in the chronic antigen presentation further develop into terminally exhausted T cells (Tex_{term}) that express even higher inhibitory receptors and lack effector and proliferation capacity (*10–12*). Understanding the underlying regulatory TF mechanisms of T cell differentiation in multiple contexts is key to programming T cells into the most beneficial state for given therapeutic applications (*13*, *14*). For example, for chronic virus and tumor treatment, the TF program can be harnessed to prevent terminal exhaustion. Tex_{term} cells are epigenetically imprinted to become dysfunctional and do not respond to immune checkpoint therapy such as anti-PD-1. However, Tex_{term} prevention without compromising T_{RM} is challenging because Tex_{term} cells display striking similarities to T_{RM} cells in tissue locations, transcriptional profiles and even open chromatin regions (Fig. 1A-C. fig. S1 (*1*)). Many studies have shown that tumor-infiltrating lymphocytes (TILs) with T_{RM} characteristics are correlated with better patient survival and tumor control (*15–19*). Therefore, Engineering T cells for robust and exclusive prevention of Tex_{term} without compromisation of T_{RM} formation may be an effective strategy toward

adoptive T cell therapy for cancer. While our knowledge of individual T cell states has sharply increased and comparative analysis between exhaustion versus memory has been done (*20*), systems-level understanding of T cell state landscape and cell-state specifying TFs network remains minimal. We previously performed a global analysis of TF specification for hard-wired cell types in embryonic development (*21*, *22*) and hematopoietic lineage (*23*). As different cell states share a large number of same cell type-defining TFs (*24*), an advanced and precise bioinformatics approach is needed to identify TFs that specify cell states. Via our systems biology approach, for the first time, we curated a comprehensive catalog of critical TFs in different states within the same cell. Specifically, we generated an epigenetic and transcription atlas of heterogeneous $CD8^+$ T cells and performed a global analysis of TF activity and their interactions. We demonstrated *in vivo* that this approach enables the discovery of novel TFs (*Zscan20*, *Jdp2*, and *Nfil3*) that programs unwanted Tex_{term}. Among the TFs, *Zscan20* is predicted to be active exclusively in Tex_{term} and of which perturbation drives effector states without compromising the capacity of forming memory cell state. As a result, the T cells that are *Zscan20*-deficient offer greater virus and tumor control. Furthermore, our comprehensive TF analysis revealed several observations about context-specific interactomes and target genes of TFs, and several new and reported pathways associated with Tex_{term} which provide insights to understand underlying molecular mechanism and therapeutic manipulation of T cell differentiation states. This precise pipeline of identification of cell state-specifying TFs can be easily adapted in cell state programming of other types of cells to improve the effectiveness of cell therapy.

Identification of important TFs requires dissecting the transcriptional networks and assessing the genome-wide influences of TFs.

The expression level of a transcription factor (TF) does not always correlate with its activity (*25*). Instead, several mechanisms affect TF activity including TF expression level, post-translational modification, collaborative co-factors, and TF target accessibility. To address this, several methods have been developed to determine the "activity" of a regulator using statistical or machine-learning approaches (*25–28*).These models were able to evaluate the activity of a TF at the local level, i.e., the expression level of their direct target genes, but not at the genome-wide level. The global analysis of TF, in principle, better predicts the cell state change upon perturbing the regulator than its local influence (29, 30) because the effect of a TF on the target gene is propagated over the genetic network and crosstalk amongst each other to form a complex regulatory logic.Therefore, we utilized a method we recently developed called Taiji (21), which can provide the systems-level assessment of a given regulator, to uncover key players in each CD8+ T cell subset. Taiji builds a genetic network by integrating transcriptomic and open chromatin data, upon which it assesses the global rather than local importance of regulators using the Personalized PageRank algorithm. This feature makes Taiji robust and suitable for integrating multiomics data from noisy

genomic measurements. In fact, Taiji clearly outperforms the motif enrichment analysis and the TFA approach(*15*, *21*, *29*).

Transcription and epigenetic atlas generate a transcription factor catalog that specifies heterogeneous T cell state.

To perform an integrated TF activity analysis using the Taiji pipeline(*21*, *30*), we generated and collected ATAC-seq and RNA-seq data in thirty-two $CD8^+$ T cell populations spanning nine states from well-defined LCMV acute and chronic infection (Table S1, Fig. 1D, fig.S2). Taiji integrates aforementioned multiomics data to build gene regulatory networks in which each node is a gene and an edge represents a regulatory interaction. Taiji first scans each open chromatin region, presumably active promoter or enhancer, to identify putative TF binding sites using motifs documented in the CIS-BP database(*31*). These TFs are then linked to their target genes predicted by EpiTensor(*32*). All the regulatory interactions are assembled into a genetic network where we define gene and interactions between genes as "node" and "edge" respectively. Lastly, the personalized PageRank algorithm(*21*) is used to assess the global influences of TFs. In this study, the node weights were determined by the z scores of gene expression levels, allocating higher ranks to the TFs that regulate more differentially expressed genes. The edge weights were set to be proportional to expression levels and binding strength of TFs and open chromatin intensity (Fig. 1D). The average number of nodes and edges of the genetic networks of the landscape of the $CD8^+$ T cell states were 18,041 and 1,613,381, respectively, including 871 (4.83%) TF nodes. On average, each TF regulated 1852 genes, and each gene was regulated by 20 TFs. We identified TFs that have high PageRank scores suggesting high activity in nine states of $CD8^+$ T cell (Fig. 1E,F; fig. S3A; table S1). Interestingly 124 out of the 273 TFs are identified as key TFs for more than one cell state (Fig. 1E; fig. S3B; table S2). These TFs, which we call "multi-taskers", include many well-known key regulators. For instance, TCF7, widely known as a critical proliferation-related protein(*3*, *10*), is identified as Naive-, MP-, and Tex_{Prog}-driving TF. All three cell states are multipotent and have high proliferative capacity. T-bet (encoded by *Tbx21*) is predicted to be highly active in both MP, and TE. T-bet is reported to program cytotoxic T cell terminal differentiation in response to LCMV viral infection in cooperation with another TF ZEB2(*33*). MP and TE shared, not only T-bet, but an additional 21 key TFs that program early activation (30% of TE or MP TFs) (Fig. 1D). Another large number of multi-taskers are observed in exhaustion-associated states, Tex_{Prog}, $Tex_{Eff\text{-}like}$, and Tex_{Term} such as *Vax2*, *Batf*, *Irf8, Stat1*, *Nfatc1*, and *Jdp2* (Fig. 1D). Tex_{Term} also shared 11 multi-taskers with T_{RM}. These include *Nr4a2(18)*, *Bhlhe40(34)*, *Prdm1(35)*, and *Hic1(36)* (Fig 1E). This observation aligns well with the location and transcription similarity between T_{RM} and exhausted T cells (Fig. 1A-C, fig. S1). Our finding of high portion of multi-taskers in each cell state extends this notion of the reuse of TFs between different cell states (*20*, *37*). This calls for the importance of state-defining TFs that are specific to each cell state ("single-taskers") of which information can be harnessed for T cell state program with specificity. Taiji analysis successfully

identified novel single-tasker TFs (Fig. 1F; table S2). Despite the similarity of T_{RM} and Tex_{Term} (Fig. 1A-C. fig. S1), Taiji identified state-selective TFs of T_{RM} and Tex_{Term} (Fig. 1G,H). These single-taskers include a multitude of TFs such as *Zscan20*, *Zbtb49*, *Arid3a*, and *Bhlhe41* that have not yet been reported before wherein their study should further our understanding of TF mechanisms underlying T cell state differentiation. Furthermore, Taiji-predicted multi-taskers of T_{RM} and Tex_{Term} such as *Prdm1*, *Bhlhe40*, *Nr4a2*, *Snai3*, *Hic1*, and *Nfil3* (Fig. 1I) provide insight into understanding the molecular mechanism of similar gene expression and localization between T_{RM} and Tex_{Term}.

Global analysis of the atlas identifies core TF networks that program context-dependent T cell differentiation.

We investigated how TFs requirements and their combinations are changed depending on the differentiation steps. In both acute and chronic infection, naive $CD8^+$ T cells differentiate to early effector cells followed by the differentiation into various $CD8^+$ T cell states in a parallel manner(*1*). In acute infection, naïve cells become either TE or MP. MP can further become diverse types of memory cells including T_{RM} (fig. S4A). In chronic infection, naïve cells generate a counterpart of MP, Tex_{Prog}, which further differentiate into Tex_{Term} (fig. S4B). To identify differentiation-step-specific TFs, PageRank ratios between target and source cell types were calculated. Interestingly, both Naïve → MP and Naïve → Tex_{Prog} transition share similar TF upregulation (*Zbtb32*, *Bhlhe40*, *Batf*) and MP→T_{RM} and Tex_{Prog} → Tex_{Term} share the downregulation of *Tcf7* and upregulation of *Nr4a2* (fig. S4A,B). The shared TF programs suggest that T cell trajectories of $CD8^+$ T cells in acute and chronic infection show not only structural similarity but also their TF regulatory circuits.

High degree of preservation of TFs in different cell states (Fig. 1E,I) and cell state transition (fig. S4) implies TF's context-dependent roles in T cell differentiation. We hypothesized distinct TF networks enable differential outcomes of the multi-taskers. To identify TF connectivity underlying cell state differentiation, we generated unbiased TF clusters based on the normalized PageRank scores across samples (fig. S5; Fig. 1J; table S3). Those combinations of TFs activate in a spatiotemporal manner, which behave like transcriptional waves to orchestrate the developmental progress depending on the immunological context. For example, we curated the transcriptional wave that specifically regulates exhaustion formation (Fig. 1J; table S3). We validated the prediction by analyzing published scRNAseq data (*4*, *38*). The TF members of the cluster show higher expression in CD8+ T cells from chronic LCMV infection compared to acute LCMV infection (Fig. 1K). The TFs are enriched in gene sets associated with cell fate commitment, PD-1 pathway, and chronic inflammation/infection (fig. S5K). Our transcriptional wave analysis also reports distinct clusters TFs bifurcating T cell differentiation in acute vs chronic, resident vs circulating, and MP vs TE and their associated biological pathways (fig. S5D-J; table S3,4). These analyses implicate the importance of differential combinations of TFs for specifically programming distinct cell states.

Integrative analysis of TF networks reveals core biological circuits to drive given cell states.

We hypothesized that effective T cell therapy can be achieved by programming T cells to avoid dysfunctional Tex_{Term} while favoring functional effector states without compromising immunological memory potential. Therefore we sought to identify and perturb TFs that are both highly active in Tex_{Term} and act as core members of TF communities that cooperate with many other TFs to control biological circuits driving terminal exhaustion and/or suppressing effector function. We constructed the network of TFs in Tex_{Term} states (Fig. 2A) and discovered 11 communities (Fig. 2B; table S5). We next curated a list of target genes of TFs (hereafter "regulon") of each community (table S5). Different communities have distinct and shared regulon (Fig. 2C). Regulon of Tex_{Term} TFs communities are commonly associated with biological circuits such as: negative regulation of immune system process, covalent chromatin modification, histone modification, dephosphorylation, FoxO signaling pathway, HBV, PD1-PDL1 pathway, and mitochondria organization (Fig. 2D) of which the associations to T cell exhaustion have previously been reported (*7*, *8*, *39*, *40*). Our network analysis also reveals the connection between Tex_{Term} TFs and pathways that might otherwise have been unappreicated such as cellular catabolic process, GTPase activities, and response to hypoxia/oxidative stress, which provide interesting circuits for future study.

In accordance with the interaction intensities (proximity between communities) between TF communities (Fig. 2B), there are two big groups of communities that share a large number of regulon within each group (Fig. 2C). The first group has community (Cm.) Bhlhe40, Cm.Pbx3, Cm.Nfil3. The regulon of this group are highly enriched in biological pathways associated with epigenetic regulation such as chromatin assembly, methylation, and demethylation (Fig. 2D). The second community group has Cm.Prdm1, Cm.Nfatc1, Cm.Hic1, and Cm.Nr4a2 of which regulon are enriched in calcium ion transport, calcium homeostasis, TNF signaling, regulation of oxidative stress, and phosphates activities (Fig. 2D). Particularly, Cm.Prdm1 shares 86.7% of regulon and a large amount of pathways with other communities (Fig. 2C, orange highlight) implicating the importance of Cm.Prdm1 in orchestrating Tex_{Term} biological circuits.

Next, we further investigated the context-dependent nature of multi-taskers of Tex_{Term} and T_{RM} utilizing community analysis of TFs in Tex_{Term} (Fig. 2) and T_{RM} (fig. S7). We particularly dissected the connection of well-known multi-tasker, Prdm1, a member of the central community in Tex_{Term}, and another multi-taskers of Tex_{Term} and T_{RM}, Nfil3 which has not yet been reported on its role in $CD8^+$ T cell differentiation. Both Prdm1 and Nfil3 show clearly different networks in Tex_{Term} and T_{RM} states (Prdm1: Fig. 2E-G ; Nfil3: fig. S8A-C). For example, Prdm1 is found in the community of Zscan20, Irf8, and Gfi1 in Tex_{Term} state in but Prdm1 is neighboring with Nr4a1, Irf4, and Hsf2 in T_{RM} state (Fig. 2B,E-G; fig. S7). The TF-TF interaction analysis of Prdm1 and Nfil3 reveals that ~ 50 % of interaction partners are distinct

between two states (Fig. 2H; fig. S7C). This result suggest that even multi-taskers can be connected to distinct TF neighbors in different states.

Regulon analysis and in vivo validation reveal new TFs programming Tex_{Term}

From the Tex_{Term} TF network analysis (Fig. 2), Cm. Prdm1 is the hub of regulating biological circuits for overall Tex_{Term}. Cm. Bhlhe40 is the center of group 1 and is particularly associated with cell cycle, demethylation, and chromatin reassembly. Among the members of these two groups, we decided to *in vivo* examine the perturbation TFs regulating terminal exhaustion that prevents terminal differentiation to dysfunctional exhaustion states at the same time promotes effector function without compromising memory formation. We selected *Zscan20* from Cm. Prdm1 and *Jdp2* from Cm. Bhlhe40 that have the most significant p-value activity score within the community (Zscan20 2.97×10^{-9} and Jdp2 3.4×10^{-10}) and no activity in T_{RM} (Fig. 1I). As for the multi-tasker control, we tested new TF, *Nfil3* and well-studied *Prdm1*(fig. S11A).

The expression of *Zscan20* in human T cells was first reported by Thiesen in 1990(*41*). However, the role of the TFs has not yet been studied. According to our Taiji analysis, *Zscan20* is exclusively active in terminal exhaustion. Our analysis of Zscan20 TF network (fig. S8D) implies its interaction with critical exhaustion-associated TFs that are already published such as Nfatc1, Nr4a1, Nr4a2, and Irf8(*42–45*). The predicted regulatees are specifically upregulated in texterm state and associated with gene sets that are downregulated in KLRG1 high effector and stimulated T cells, DNA repair, T cell activation, lymphocyte differentiation, PDL-1 expression and PD-1 checkpoint pathway in cancer and covalent chromatin modification (fig. S9). All together suggest that blockage of Zscan20 prevents T cell exhaustion differentiation and promotes effector state differentiation.

Pan-exhaustion specific TF, Jdp2 is suggested to control transcription via direct regulation of the modification of histones and the assembly of chromatin or heterodimerization with a component of the AP-1 complex (*46*). The fact that the balance between AP-1 and NFAT activation is reported as one of the key regulatory mechanisms to T cell exhaustion (*42*) implies an important role of Jdp2 in T cell exhaustion. Our Taiji analysis suggests that Jdp2 regulatees suppress the effector-associated genes as they are enriched in naive cells compared to effector cells and biological pathways such as negative regulation of phosphorylation and dephosphorylation (fig. S10). Therefore, depletion of Jdp2 possibly prevents dysfunctional exhaustion states and restores effector function.

To interrogate the new regulators of Tex_{Term} (*Zscan20*, *Jdp2,* and *Nfil3*) *in vivo* we used CRISPR-mediated knockout (KO) utilizing retroviral gRNA transduction in Cas9 transgenic, LCMV virus peptide GP33-41 selective (P14) T cells (Fig. 3A). As we expected, In day 20+ of chronic LCMV infection, perturbation of these TFs and *Prdm1* altered the differentiation pattern of exhaustion subsets. TF gRNA transduced cells show reduced RNA expression score of regulon of Tex_{Term}TFs (fig. S11B). All KOs are skewed away from Tex_{Term} (Fig. 3B,C; fig. S11C,D) and show a reduction in inhibitory receptors such as

CD38, CD39, and CD101 (Fig. 3D,E). Importantly, in acute infection (Fig. 3F), KO of exhaustion-specific TFs (*Zscan20* and *Jdp2*) does not alter the properties of memory formation but KO of T_{RM} & Tex_{Term} multi-tasker (*Nfil3* and *Prdm1*) significantly reduces T_{RM} formation which corresponds to our prediction (Fig. 3G-I). This *in vivo* experiment demonstrates that our systemic approach can identify TF that specifically drives each cell state.

Loss of new Tex_{Term} regulators Jdp2 and Zscan20 contribute to better effector function.

TF regulon analysis suggests that top Jdp2 regulatees are negatively enriched in effector-related genes and dephosphorylation (fig. S10B,C). TF network analysis of Zscan20 suggests that it has tight interaction with Nfatc1, Nr4a1, Irf8 (fig. S8D) and possibly regulates T cell activation (fig. S9C). Along with the bioinformatics analysis, loss of *Jdp2* or *Zscan20* programs the effector state differentiation of $CD8^+$ T cells in chronic infection where effector state are not found otherwise. In chronic LCMV infection T cell adoptive transfer system (Fig. 4A), $CD8^+$ T cells with the perturbation of *Zscan20* or *Jdp2* express higher levels of double positive of KLRG1 and CX3CR1 (Fig. 4B,C). These KO cells release more cytotoxic cytokines (Fig. 4D). Importantly, compared to control, mice transferred with TF KOs P14 cells have lower serum virus levels (Fig. 4E) which indicates that disruption of *Zscan20* or *Jdp2* enables T cells to offer superior virus control. This demonstrates the accuracy and implication of our bioinformatics approach in cell state engineering.

Disruption of novel TF Zscan20 provides better tumor control and synergy with ICB.

Similar to chronic virus control, we expect that Tex_{Term}-specific TF-deficiency confers enhanced tumor control. Tex_{Prog} cells are known to respond to ICB and become effector-like states ($Tex_{Eff\text{-}like}$) of which cells confer moderate anti-tumor immunity but eventually transition to a dysfunctional Tex_{Term} state(*12*). As the loss of *Zscan20* exclusively deters Tex_{Term} differentiation (Fig. 3), perturbation of *Zscan20* will produce more beneficial cell states, Tex_{Prog} and $Tex_{Eff\text{-}like}$ and generate synergy with immune checkpoint therapy. Utilizing an adoptive therapy model, g*Zscan20* RV transduced Cas9 P14 cells were transferred into mice with established melanoma tumors expressing GP33-41 (Fig. 4F). In line with the prevention of Tex_{Term} in the chronic virus infection (Fig. 3A-E), loss of *Zscan20* prevented terminal exhaustion and promoted Tex_{Prog} differentiation (Fig. 4G-I). Perturbation of *Zscan20* enhanced tumor control and mice survival in both anti-PD1 and control treatment conditions (Fig. 4J-L). *Jdp2* KO also improves tumor suppression when it is combined with ICB (fig. S12). However, without ICB treatment, the survival benefit of *Jdp2* disruption is not as great as that of *Zscan20*. As Jdp2 is predicted to be less specific to Tex_{Term}, KO can alter the properties of Tex_{Prog} and $Tex_{Eff\text{-}like}$ in the tumor, possibly making the beneficial effect of KO moderate. Interestingly, T_{RM} and Tex_{Term} multi-tasker, *Nfil3*-deficiency does not improve tumor control (fig S13). This can be explained by the impaired properties of T_{RM} in *Nfil3* KO (Fig. 3I) because a positive prognosis has been correlated with TILs that present qualities of T_{RM}(*16*). Thus,

depleting TFs that selectively program Tex_{Term} such as *Zscan20* without perturbing memory cells including T_{RM} will be a great strategy to enhance adoptive T cell therapy for tumors. This precise pipeline of identification of cell state-specifying TFs can be easily adapted in cell state programming of other types of cells to improve the effectiveness of cell therapy.

Discussion

The same types of cells can assume diverse states with varying functionalities. Single-cell genomics and proteomics enable precise characterization of cell states. As the cells of the same cell state share the same gene expression profile, migration pattern, and functionality, programming cells into a desirable cell state can effectively transform cells equipped with multiple desirable features for cell therapy. The cell state differentiation is regulated by TFs that relay environmental signals through the control of gene expression (*5*, *6*). Therefore investigation of key TFs of each cell state enables efficient and precise regulation of cell programming. Here we generated an epigenomic and transcriptomic atlas to systematically identify TFs that define different $CD8^+$ T cell states and we discovered novel TFs of great importance in improving T cell therapy.

Comprehensive analysis of the multiomics atlas enables the underlying molecular mechanism of the landscape of heterogeneous T cell states. First, we systematically identified cell-state bona fide specific TFs (Fig. 1F) and multitaskers (Fig 1E). We observed that 48% of highly active TFs across the cell states are active in more than one cell state. This high reuse of TFs implies the importance of context-dependency of the TF regulatory network. Indeed, we discovered differential connectivity of TFs and their associated biological pathways in different cell states (Fig. 2E-G) and in along with the differentiation steps (fig. S5). These results illustrate a complex and dynamic network of TFs. Second, our integrative analysis of TF network in Tex_{Term} cell state uncovers cooperation between a group of TFs and revealed biological circuits that have not been appreciated such as cellular catabolic process, GTPase activity, hypoxia, and oxidative stress, which provide interesting pathways for future study (Fig. 2D). Indeed, recently Vignali et al reported the connection of hypoxia and exhaustion that support the accuracy of our bioinformatics analysis (*47*). Furthermore, this systemic analysis indicated shared TFs can be a molecular basis of the parallel differentiation trajectories, Naive → MP → T_{RM} and Naive → Tex_{Prog} → Tex_{Term}, from acute and chronic infection (fig. S4). Lastly, we discovered and *in vivo* validated three different TFs(*Zscan20*, *Jdp2*, and *Nfil3*) and reported their roles in Tex_{Term} for the first time. Our multiomics-based global TF analysis pipeline allows the prediction and investigation of TFs that have not yet been reported due to their low or less significant differential expression.

We demonstrated that single-tasker TFs (e.g. *Zscan20*) are particularly useful in designing therapeutic cell states. We hypothesized that effective T cell therapy can be achieved by programming T cells to avoid dysfunctional Tex_{Term} all the while favoring a functional effector state without compromising immunological memory potential. Therefore we sought to identify TFs to perturb that are

specifically active in Tex_{Term} and possibly regulate genes that suppress effector state transition. Even though many studies have modulated the expression of TFs to improve anti-tumor immunity, they did not aim to precisely block Tex_{Term} or consider the possible compromisation in T_{RM} formation. Our study for the first time aims to specifically prevent the $Tex_{Term,}$ the dysfunctional and terminally differentiated state. For example, Tex_{Term} specific TF, *Zscan20*-deficiency in $CD8^+$ T cells confer enhanced virus and tumor control in the adoptive therapy model of melanoma. Importantly this KO does not have impairment in T_{RM} or other memory formation but has a higher tendency to become effector cells even in the chronic antigen encounter (Fig. 4B,C). Particularly with the combination of immune checkpoint therapy the disruption of *Zscan20* and *Jdp2* creates synergy and removes the tumor (Fig. 4; fig. S13). Compared to a *Nfil3* (T_{RM} and Tex_{Term} multi-tasker), *Zscan20* offers superior tumor control (fig. S13), indicating the importance of identifying and modulating single takers for cell programming.

Taken together we report a rational approach to improve cell therapy by specific cell state engineering assisted by a multiomics atlas. Our pipeline for unbiased identification of TF programs and associated biological circuits can be applied to many other cell types and utilized for therapeutic cell state programming.

References and Notes

1. H. K. Chung, B. McDonald, S. M. Kaech, The architectural design of CD8+ T cell responses in acute and chronic infection: Parallel structures with divergent fates. *J. Exp. Med.* **218** (2021), doi:10.1084/jem.20201730.
2. M. Y. Kasmani, R. Zander, H. K. Chung, Y. Chen, A. Khatun, M. Damo, P. Topchyan, K. E. Johnson, D. Levashova, R. Burns, U. M. Lorenz, V. L. Tarakanova, N. S. Joshi, S. M. Kaech, W. Cui, Clonal lineage tracing reveals mechanisms skewing CD8+ T cell fate decisions in chronic infection. *J. Exp. Med.* **220** (2023), doi:10.1084/jem.20220679.
3. J.-C. Beltra, S. Manne, M. S. Abdel-Hakeem, M. Kurachi, J. R. Giles, Z. Chen, V. Casella, S. F. Ngiow, O. Khan, Y. J. Huang, P. Yan, K. Nzingha, W. Xu, R. K. Amaravadi, X. Xu, G. C. Karakousis, T. C. Mitchell, L. M. Schuchter, A. C. Huang, E. J. Wherry, Developmental Relationships of Four Exhausted CD8+ T Cell Subsets Reveals Underlying Transcriptional and Epigenetic Landscape Control Mechanisms. *Immunity*. **52**, 825–841.e8 (2020).
4. J. R. Giles, S. F. Ngiow, S. Manne, A. E. Baxter, O. Khan, P. Wang, R. Staupe, M. S. Abdel-Hakeem, H. Huang, D. Mathew, M. M. Painter, J. E. Wu, Y. J. Huang, R. R. Goel, P. K. Yan, G. C. Karakousis, X. Xu, T. C. Mitchell, A. C. Huang, E. John Wherry, Shared and distinct biological circuits in effector, memory and exhausted CD8 T cells revealed by temporal single-cell transcriptomics and epigenetics. *Nature Immunology*. **23** (2022), pp. 1600–1613.
5. W. Cui, Y. Liu, J. S. Weinstein, J. Craft, S. M. Kaech, An interleukin-21-interleukin-10-STAT3 pathway is critical for functional maturation of memory CD8+ T cells. *Immunity*. **35**, 792–805 (2011).
6. S. M. Kaech, W. Cui, Transcriptional control of effector and memory CD8+ T cell differentiation. *Nat. Rev. Immunol.* **12**, 749–761 (2012).
7. K. E. Pauken, M. A. Sammons, P. M. Odorizzi, S. Manne, J. Godec, O. Khan, A. M. Drake, Z. Chen, D. R. Sen, M. Kurachi, R. A. Barnitz, C. Bartman, B. Bengsch, A. C. Huang, J. M. Schenkel, G. Vahedi, W. N. Haining, S. L. Berger, E. J. Wherry, Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science*. **354**, 1160–1165 (2016).
8. D. R. Sen, J. Kaminski, R. Anthony Barnitz, M. Kurachi, U. Gerdemann, K. B. Yates, H.-W. Tsao, J. Godec, M. W. LaFleur, F. D. Brown, P. Tonnerre, R. T. Chung, D. C. Tully, T. M. Allen, N. Frahm, G. M. Lauer, E. John Wherry, N. Yosef, W. Nicholas Haining, The epigenetic landscape of T cell exhaustion. *Science*. **354** (2016), pp. 1165–1169.
9. E. J. Wherry, J. N. Blattman, K. Murali-Krishna, R. van der Most, R. Ahmed, Viral persistence alters CD8 T-cell immunodominance and tissue distribution and results in distinct stages of functional impairment. *J. Virol.* **77**, 4911–4927 (2003).
10. W. H. Hudson, J. Gensheimer, M. Hashimoto, A. Wieland, R. M. Valanparambil, P. Li, J.-X. Lin, B. T. Konieczny, S. J. Im, G. J. Freeman, W. J. Leonard, H. T. Kissick, R. Ahmed, Proliferating Transitory T Cells with an Effector-like Transcriptional Signature Emerge from PD-1 Stem-like CD8 T Cells during Chronic Infection. *Immunity*. **51** (2019), pp. 1043–1058.e4.
11. S. J. Im, M. Hashimoto, M. Y. Gerner, J. Lee, H. T. Kissick, M. C. Burger, Q. Shan, J. S. Hale, J.

Lee, T. H. Nasti, A. H. Sharpe, G. J. Freeman, R. N. Germain, H. I. Nakaya, H.-H. Xue, R. Ahmed, Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. *Nature*. **537**, 417–421 (2016).

12. B. C. Miller, D. R. Sen, R. Al Abosy, K. Bi, Y. V. Virkud, M. W. LaFleur, K. B. Yates, A. Lako, K. Felt, G. S. Naik, M. Manos, E. Gjini, J. R. Kuchroo, J. J. Ishizuka, J. L. Collier, G. K. Griffin, S. Maleri, D. E. Comstock, S. A. Weiss, F. D. Brown, A. Panda, M. D. Zimmer, R. T. Manguso, F. S. Hodi, S. J. Rodig, A. H. Sharpe, W. N. Haining, Subsets of exhausted CD8 T cells differentially mediate tumor control and respond to checkpoint blockade. *Nat. Immunol.* **20**, 326–336 (2019).

13. J. K. Flynn, P. R. Gorry, Stem memory T cells (TSCM)-their role in cancer and HIV immunotherapies. *Clin Transl Immunology*. **3**, e20 (2014).

14. Z. Zhang, S. Liu, B. Zhang, L. Qiao, Y. Zhang, Y. Zhang, T Cell Dysfunction and Exhaustion in Cancer. *Front Cell Dev Biol*. **8**, 17 (2020).

15. J. J. Milner, C. Toma, B. Yu, K. Zhang, K. Omilusik, A. T. Phan, D. Wang, A. J. Getzler, T. Nguyen, S. Crotty, W. Wang, M. E. Pipkin, A. W. Goldrath, Runx3 programs CD8 T cell residency in non-lymphoid tissues and tumours. *Nature*. **552**, 253–257 (2017).

16. A.-P. Ganesan, J. Clarke, O. Wood, E. M. Garrido-Martin, S. J. Chee, T. Mellows, D. Samaniego-Castruita, D. Singh, G. Seumois, A. Alzetani, E. Woo, P. S. Friedmann, E. V. King, G. J. Thomas, T. Sanchez-Elsner, P. Vijayanand, C. H. Ottensmeier, Tissue-resident memory features are linked to the magnitude of cytotoxic T cell responses in human lung cancer. *Nat. Immunol.* **18**, 940–950 (2017).

17. F. Djenidi, J. Adam, A. Goubar, A. Durgeau, G. Meurice, V. de Montpréville, P. Validire, B. Besse, F. Mami-Chouaib, 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*. **194** (2015), pp. 3475–3486.

18. S. Corgnac, M. Boutet, M. Kfoury, C. Naltet, F. Mami-Chouaib, 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).

19. F. L. Komdeur, T. M. Prins, S. van de Wall, A. Plat, G. B. A. Wisman, H. Hollema, T. Daemen, D. N. Church, M. de Bruyn, H. W. Nijman, CD103 tumor-infiltrating lymphocytes are tumor-reactive intraepithelial CD8 T cells associated with prognostic benefit and therapy response in cervical cancer. *OncoImmunology*. **6** (2017), p. e1338230.

20. T. A. Doering, A. Crawford, J. M. Angelosanto, M. A. Paley, C. G. Ziegler, E. John Wherry, Network Analysis Reveals Centrally Connected Genes and Pathways Involved in CD8 T Cell Exhaustion versus Memory. *Immunity*. **37** (2012), pp. 1130–1144.

21. K. Zhang, M. Wang, Y. Zhao, W. Wang, Taiji: System-level identification of key transcription factors reveals transcriptional waves in mouse embryonic development. *Sci Adv*. **5**, eaav3262 (2019).

22. J. Wang, C. Liu, Y. Chen, W. Wang, Taiji-reprogram: a framework to uncover cell-type specific

regulators and predict cellular reprogramming cocktails. *NAR Genom Bioinform*. **3**, lqab100 (2021).

23. C. Liu, K. Omilusik, C. Toma, N. S. Kurd, J. T. Chang, A. W. Goldrath, W. Wang, Systems-level identification of key transcription factors in immune cell specification. *PLoS Comput. Biol.* **18**, e1010116 (2022).

24. J. Holmberg, T. Perlmann, Maintaining differentiated cellular identity. *Nat. Rev. Genet.* **13**, 429–439 (2012).

25. T. Schacht, M. Oswald, R. Eils, S. B. Eichmüller, R. König, Estimating the activity of transcription factors by the effect on their target genes. *Bioinformatics*. **30**, i401–7 (2014).

26. M. L. Arrieta-Ortiz, C. Hafemeister, A. R. Bate, T. Chu, A. Greenfield, B. Shuster, S. N. Barry, M. Gallitto, B. Liu, T. Kacmarczyk, F. Santoriello, J. Chen, C. D. A. Rodrigues, T. Sato, D. Z. Rudner, A. Driks, R. Bonneau, P. Eichenberger, An experimentally supported model of the Bacillus subtilis global transcriptional regulatory network. *Mol. Syst. Biol.* **11**, 839 (2015).

27. S. Aibar, C. B. González-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J.-C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters, S. Aerts, SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods*. **14**, 1083–1086 (2017).

28. A. Maslova, R. N. Ramirez, K. Ma, H. Schmutz, C. Wang, C. Fox, B. Ng, C. Benoist, S. Mostafavi, Immunological Genome Project, Deep learning of immune cell differentiation. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 25655–25666 (2020).

29. B. Yu, K. Zhang, J. Justin Milner, C. Toma, R. Chen, J. P. Scott-Browne, R. M. Pereira, S. Crotty, J. T. Chang, M. E. Pipkin, W. Wang, A. W. Goldrath, Epigenetic landscapes reveal transcription factors that regulate CD8 T cell differentiation. *Nature Immunology*. **18** (2017), pp. 573–582.

30. L. Page, S. Brin, R. Motwani, T. Winograd, The PageRank Citation Ranking: Bringing Order to the Web (1999) (available at http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf).

31. M. T. Weirauch, A. Yang, M. Albu, A. G. Cote, A. Montenegro-Montero, P. Drewe, H. S. Najafabadi, S. A. Lambert, I. Mann, K. Cook, H. Zheng, A. Goity, H. van Bakel, J.-C. Lozano, M. Galli, M. G. Lewsey, E. Huang, T. Mukherjee, X. Chen, J. S. Reece-Hoyes, S. Govindarajan, G. Shaulsky, A. J. M. Walhout, F.-Y. Bouget, G. Ratsch, L. F. Larrondo, J. R. Ecker, T. R. Hughes, Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell*. **158** (2014), pp. 1431–1443.

32. Y. Zhu, Z. Chen, K. Zhang, M. Wang, D. Medovoy, J. W. Whitaker, B. Ding, N. Li, L. Zheng, W. Wang, Constructing 3D interaction maps from 1D epigenomes. *Nature Communications*. **7** (2016), , doi:10.1038/ncomms10812.

33. C. X. Dominguez, R. A. Amezquita, T. Guan, H. D. Marshall, N. S. Joshi, S. H. Kleinstein, S. M. Kaech, The transcription factors ZEB2 and T-bet cooperate to program cytotoxic T cell terminal differentiation in response to LCMV viral infection. *J. Exp. Med.* **212**, 2041–2056 (2015).

34. C. Li, B. Zhu, Y. M. Son, Z. Wang, L. Jiang, M. Xiang, Z. Ye, K. E. Beckermann, Y. Wu, J. W. Jenkins, P. J. Siska, B. G. Vincent, Y. S. Prakash, T. Peikert, B. T. Edelson, R. Taneja, M. H.

Kaplan, J. C. Rathmell, H. Dong, T. Hitosugi, J. Sun, The Transcription Factor Bhlhe40 Programs Mitochondrial Regulation of Resident CD8 T Cell Fitness and Functionality. *Immunity*. **51** (2019), pp. 491–507.e7.

35. L. K. Mackay, M. Minnich, N. A. M. Kragten, Y. Liao, B. Nota, C. Seillet, A. Zaid, K. Man, S. Preston, D. Freestone, A. Braun, E. Wynne-Jones, F. M. Behr, R. Stark, D. G. Pellicci, D. I. Godfrey, G. T. Belz, M. Pellegrini, T. Gebhardt, M. Busslinger, W. Shi, F. R. Carbone, R. A. W. van Lier, A. Kallies, K. P. J. M. van Gisbergen, Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes. *Science*. **352**, 459–463 (2016).

36. J. T. Crowl, M. Heeg, A. Ferry, J. Justin Milner, K. D. Omilusik, C. Toma, Z. He, J. T. Chang, A. W. Goldrath, Tissue-resident memory CD8 T cells possess unique transcriptional, epigenetic and functional adaptations to different tissue environments. *Nature Immunology*. **23** (2022), pp. 1121–1131.

37. N. Novershtern, A. Subramanian, L. N. Lawton, R. H. Mak, W. N. Haining, M. E. McConkey, N. Habib, N. Yosef, C. Y. Chang, T. Shay, G. M. Frampton, A. C. B. Drake, I. Leskov, B. Nilsson, F. Preffer, D. Dombkowski, J. W. Evans, T. Liefeld, J. S. Smutko, J. Chen, N. Friedman, R. A. Young, T. R. Golub, A. Regev, B. L. Ebert, Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*. **144**, 296–309 (2011).

38. Y. Chen, J. Shen, M. Y. Kasmani, P. Topchyan, W. Cui, Single-Cell Transcriptomics Reveals Core Regulatory Programs That Determine the Heterogeneity of Circulating and Tissue-Resident Memory CD8 T Cells. *Cells*. **10** (2021), doi:10.3390/cells10082143.

39. M. M. Staron, S. M. Gray, H. D. Marshall, I. A. Parish, J. H. Chen, C. J. Perry, G. Cui, M. O. Li, S. M. Kaech, The transcription factor FoxO1 sustains expression of the inhibitory receptor PD-1 and survival of antiviral CD8(+) T cells during chronic infection. *Immunity*. **41**, 802–814 (2014).

40. N. E. Scharping, D. B. Rivadeneira, A. V. Menk, P. D. A. Vignali, B. R. Ford, N. L. Rittenhouse, R. Peralta, Y. Wang, Y. Wang, K. DePeaux, A. C. Poholek, G. M. Delgoffe, Mitochondrial stress induced by continuous stimulation under hypoxia rapidly drives T cell exhaustion. *Nat. Immunol.* **22**, 205–215 (2021).

41. H. J. Thiesen, Multiple genes encoding zinc finger domains are expressed in human T cells. *New Biol.* **2** (1990) (available at https://pubmed.ncbi.nlm.nih.gov/2288909/).

42. G. P. Mognol, R. Spreafico, V. Wong, J. P. Scott-Browne, S. Togher, A. Hoffmann, P. G. Hogan, A. Rao, S. Trifari, Exhaustion-associated regulatory regions in CD8 tumor-infiltrating T cells. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E2776–E2785 (2017).

43. G. J. Martinez, R. M. Pereira, T. Äijö, E. Y. Kim, F. Marangoni, M. E. Pipkin, S. Togher, V. Heissmeyer, Y. C. Zhang, S. Crotty, E. D. Lamperti, K. M. Ansel, T. R. Mempel, H. Lähdesmäki, P. G. Hogan, A. Rao, The transcription factor NFAT promotes exhaustion of activated $CD8^+$ T cells. *Immunity*. **42**, 265–278 (2015).

44. H. Seo, J. Chen, E. González-Avalos, D. Samaniego-Castruita, A. Das, Y. H. Wang, I. F. López-Moyado, R. O. Georges, W. Zhang, A. Onodera, C.-J. Wu, L.-F. Lu, P. G. Hogan, A. Bhandoola, A. Rao, TOX and TOX2 transcription factors cooperate with NR4A transcription factors to impose

CD8 T cell exhaustion. *Proceedings of the National Academy of Sciences*. **116** (2019), pp. 12410–12415.

45. J. Chen, I. F. López-Moyado, H. Seo, C.-W. J. Lio, L. J. Hempleman, T. Sekiya, A. Yoshimura, J. P. Scott-Browne, A. Rao, NR4A transcription factors limit CAR T cell function in solid tumours. *Nature*. **567** (2019), pp. 530–534.

46. A. Aronheim, E. Zandi, H. Hennemann, S. J. Elledge, M. Karin, Isolation of an AP-1 repressor by a novel method for detecting protein-protein interactions. *Molecular and Cellular Biology*. **17** (1997), pp. 3094–3102.

47. P. D. A. Vignali, K. DePeaux, M. J. Watson, C. Ye, B. R. Ford, K. Lontos, N. K. McGaa, N. E. Scharping, A. V. Menk, S. C. Robson, A. C. Poholek, D. B. Rivadeneira, G. M. Delgoffe, Hypoxia drives CD39-dependent suppressor function in exhausted T cells to limit antitumor immunity. *Nat. Immunol.* (2022), doi:10.1038/s41590-022-01379-9.

48. K. Zhang, N. Li, R. I. Ainsworth, W. Wang, Systematic identification of protein combinations mediating chromatin looping. *Nat. Commun.* **7**, 1–11 (2016).

49. T. Zhao, H. Liu, K. Roeder, J. Lafferty, L. Wasserman, The huge Package for High-dimensional Undirected Graph Estimation in R (2020), doi:10.48550/arXiv.2006.14781.

50. V. A. Traag, L. Waltman, N. J. van Eck, From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).

51. M. Schönfeld, J. Pfeffer, Fruchterman/Reingold (1991): Graph Drawing by Force-Directed Placement. *Schlüsselwerke der Netzwerkforschung* (2019), pp. 217–220.

52. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **25** (2009), doi:10.1093/bioinformatics/btp324.

53. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. **26** (2010), doi:10.1093/bioinformatics/btp698.

54. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, 1–9 (2008).

55. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

56. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

57. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

58. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell*. **177** (2019), pp. 1888–

1902.e21.

59. H. Wickham, M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. Pedersen, E. Miller, S. Bache, K. Müller, J. Ooms, D. Robinson, D. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, H. Yutani, Welcome to the Tidyverse. *Journal of Open Source Software*. **4** (2019), p. 1686.

Acknowledgements: We thank members of the Kaech and Wang laboratory for advice and assistance with experiments. **Funding:** This work was supported by NIH grants to S.M.K. R37AI066232, R01AI123864, R21AI151986, and R01CA240909 and to W.W. R01AI150282, R01HG009626 (in part). 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 and designed the study. H.K.C and C.L. analyzed the data and wrote the manuscript. H.K.C. designed and performed experiments. E.C., B.M. D.C., B.C., Q.Y., F.H., 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., and A.W. contribute to bioinformatics data analysis. D.H. provide resource and training for ATAC-seqeucning. H.K.C., C.L., E.C., B.M., W.W., and S.M.K. revised and edited the manuscript. W.W. and S.M.K supervised the study. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** ATAC-sequencing data will be deposited GEO database. $CD8^+$ T cell multiomics atlas and analysis result will be shared our website.

Main Figures

Fig 1. Transcription and epigenomic atlas of $CD8^+$ T cell differentiation states and cell-state specific TFs identification pipeline

Fig 2. Global analysis of TF networks in terminal exhaustion

Fig 3. Disruption of Tex_{Term} single-tasker TFs prevents terminal exhaustion without perturbation of memory formation

Fig.4. Disruption of transcription factor *Zscan20* and *Jdp2* provides enhanced virus and tumor control.

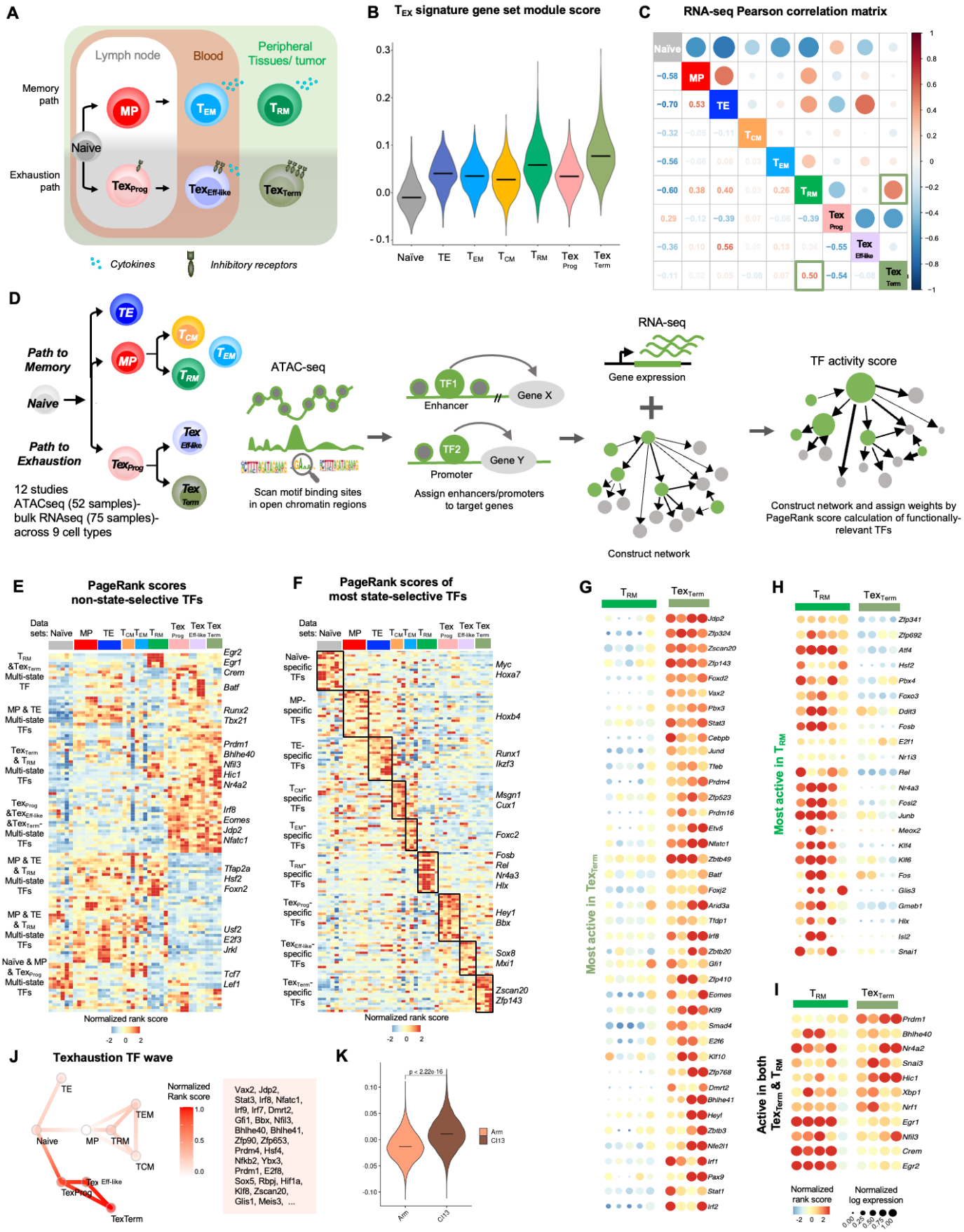
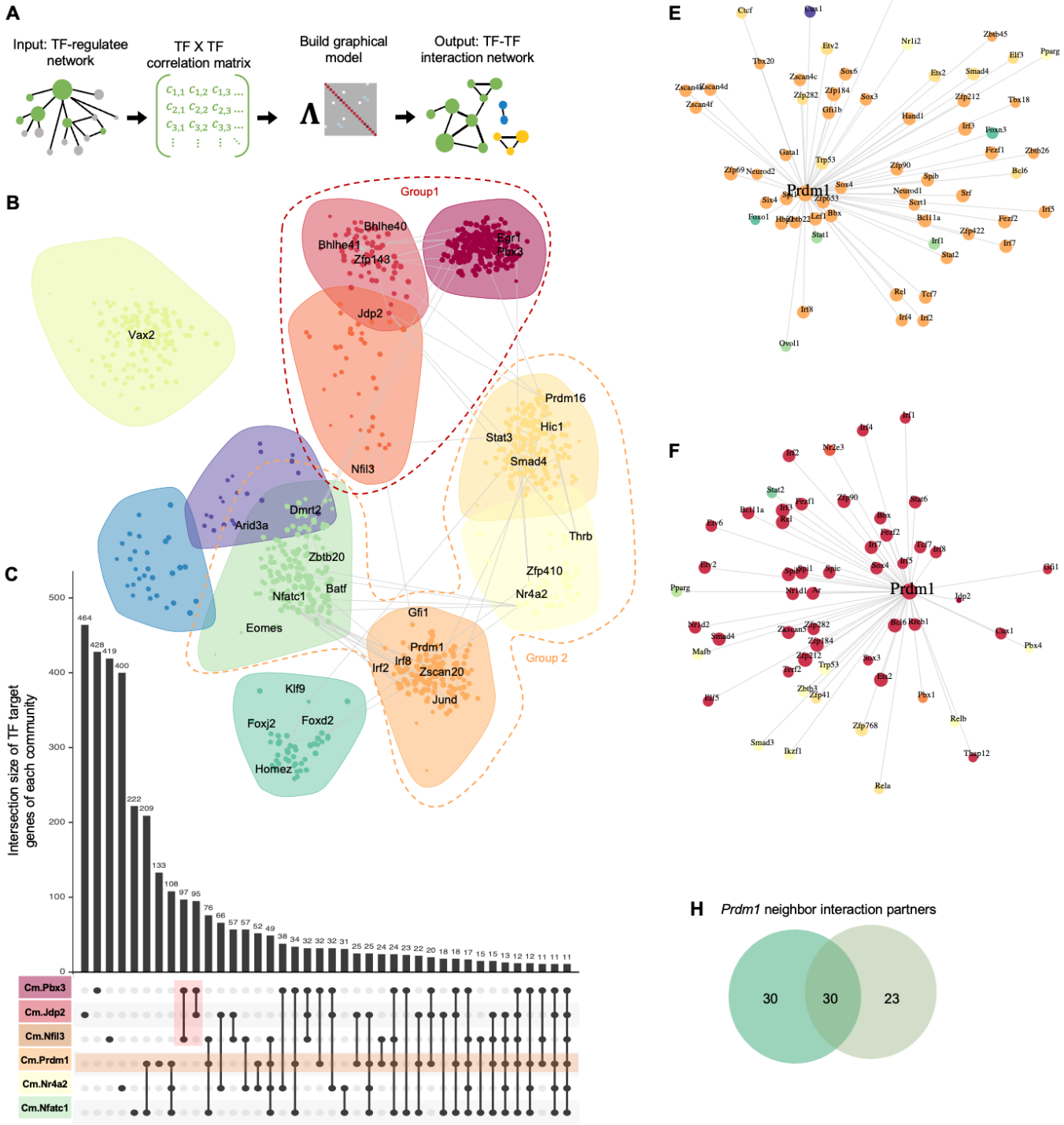

A
Lymph node
Blood
Peripheral Tissues/ tumor
Memory path
Exhaustion path
Naive
MP
TEM
TRM
TexProg
TexEff-like
TexTerm
Cytokines
Inhibitory receptors
B
TEX signature gene set module score
Naïve
TE
TEM
TCM
TRM
Tex Prog
Tex Term
C
RNA-seq Pearson correlation matrix
D
Path to Memory
Path to Exhaustion
12 studies
ATACseq (52 samples)-
bulk RNAseq (75 samples)-
across 9 cell types
ATAC-seq
Scan motif binding sites in open chromatin regions
Enhancer
Promoter
Gene X
Gene Y
Assign enhancers/promoters to target genes
RNA-seq
Gene expression
Construct network
TF activity score
Construct network and assign weights by PageRank score calculation of functionally-relevant TFs
E
PageRank scores non-state-selective TFs
Normalized rank score
F
PageRank scores of most state-selective TFs
Normalized rank score
G
Most active in TexTerm
H
Most active in TRM
I
Active in both TexTerm & TRM
Normalized rank score
Normalized log expression
J
Texhaustion TF wave
Normalized Rank score
Vax2, Jdp2, Stat3, Irf8, Nfatc1, Irf9, Irf7, Dmrt2, Gfi1, Bbx, Nfil3, Bhlhe40, Bhlhe41, Zfp90, Zfp653, Prdm4, Hsf4, Nfkb2, Ybx3, Prdm1, E2f8, Sox5, Rbpj, Hif1a, Klf8, Zscan20, Glis1, Meis3, ...
K
p < 2.22e-16
Arm
Cl13

Fig 1. Transcription and epigenomic atlas of CD8$^+$ T cell differentiation states and cell-state specific TFs identification pipeline (**A-C**) Parallel differentiation of TRM and TexTerm and their similarity in the transaction. (**A**) CD8$^+$ T cell state trajectories in acute and chronic infection/tumor where various memory and exhaustion states are assumed. (**B**) Module score of exhaustion signature gene in each cluster of cell state. Exhaustion signature genes are upregulated in both T_{RM} and Tex_{Term}. (**C**) Pearson correlation matrix of batch effect corrected RNA-seq (fig. S2). Color and size are both proportional to correlation values. (**D**) Pipeline of 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 scores matrix representing the activity of transcription factors (TFs). The downstream analysis included the identification of cell-type-specific TFs, differentiation-specific TFs, and the construction of temporal transcriptional waves. (**E**) 124 multi-tasker TFs. Multi-taskers are active in more than one T cell state. (**F**) PageRank scores of 151 bona fide cell-state-specific TFs. TFs in rows (z-normalized), samples in columns, and the color of the cell in the matrix indicates the normalized PageRank scores. Bubble plots of TFs that are selectively active in Tex_{Term} (**G**), T_{RM} (**H**), and multi-taskers that are active in both Tex_{Term} and T_{RM} (**I**). Color represents the normalized PageRank score and circle size represents the logarithm of gene expression. TFs are ordered based on the p-value. (**J**) Transcription factor waves associated with exhaustion. Circles represent specific cell state. Color indicates normalized PageRank scores with red displaying high values. TFs members of the exhaustion wave are shown. (**K**) The expression score of TF members of exhaustion wave are calculated in the single-cell RNA sequencing cell clusters of CD8$^+$ T cells from LCMV acute or chronic infection from GSE199565.



D

	Communities	TF members	Pathways of each group	Common pathways of major communities: Previously reported:	New pathways:
Group 1	Cm. Pbx3	**Pbx3 Egr1** Hinfp Zfp523 E2f4 Tfdp1 Nrf1 E2f6 Etv5	▪ Chromatin modification ▪ Chromosome organization ▪ Chromatin organization	▪ Chromatin remodeling ▪ Histon modification ▪ Dephosphorylation ▪ Negative regulation of phosphorylation ▪ PD-L1-PD-1 pathway ▪ Response to TGF beta ▪ FoXO signaling pathway ▪ Cell cycle regulation ▪ Mitochondria organization ▪ Hepatitis B ▪ Herpes simplex virus 1	▪ **Catabolic process** ▪ **Cellular carbohydrate metabolic process** ▪ **Small GTPase-mediated signal** ▪ **GTPase activity** ▪ **Response to hypoxia** ▪ **Response to oxidative stress**
	Cm. Bhlhe40	**Bhlhe40 Jdp2** Zfp143 Xbp1 Tfeb Mxd1			
	Cm. Nfil3	**Nfil3** Cebpb Atf4 Etv3			
Group 2	Cm. Hic1	**Hic1** Stat3 Prdm4 Snai3 Zfp324 Trp73 Prdm16 Smad4	▪ Calcium ion homeostasis ▪ Calcium ion transfer ▪ Response to TNF or TNF signaling		
	Cm. Prdm1	**Zscan20 Prdm1** Gfi1 Jund Irf8			
	Cm. Nr4a2	**Nr4a2** Zfp410 Nr1i3			
	Cm. Nfat1	**Nfatc1** Batf Eomes Arid3a			

Fig 2. Global analysis of TF networks in terminal exhaustion (**A**) Pipeline of TF interaction network analysis. (**B**) Communities of TFs in Tex_{Term}. Communities consist of Tex_{Term} TFs and their interaction partners. Dots represent each TF and Tex_{Term} TFs with high interaction scores are labeled. Gray lines represent the interactions between TexTerm-specific TFs. (**C**) Intersection size of the target of TFs (regulon) of communities. Top1000 target of TFs for each community were selected by the average edge weight. Community (Cm.) Prdm1 shares the most regulatees with other communities (orange highlight). Cm.Pbx3, Cm.Bhlhe40, and Cm.Nfil3 also have a large number of common regulatees (red highlight). (**D**) Table of TF communities in Tex_{Term} with regulatees of each community and associated biological pathways. TF-TF interaction analysis of T_{RM} and Tex_{Term} multi-tasker, Prdm1. Prdm1 and interaction partner TFs in (**E**) Tex_{Term} or (**F**) T_{RM} state are visualized. (**H**) Venn Diagrams show Prmd1 interacting partner TFs overlapping in T_{RM} and Tex_{Term}.

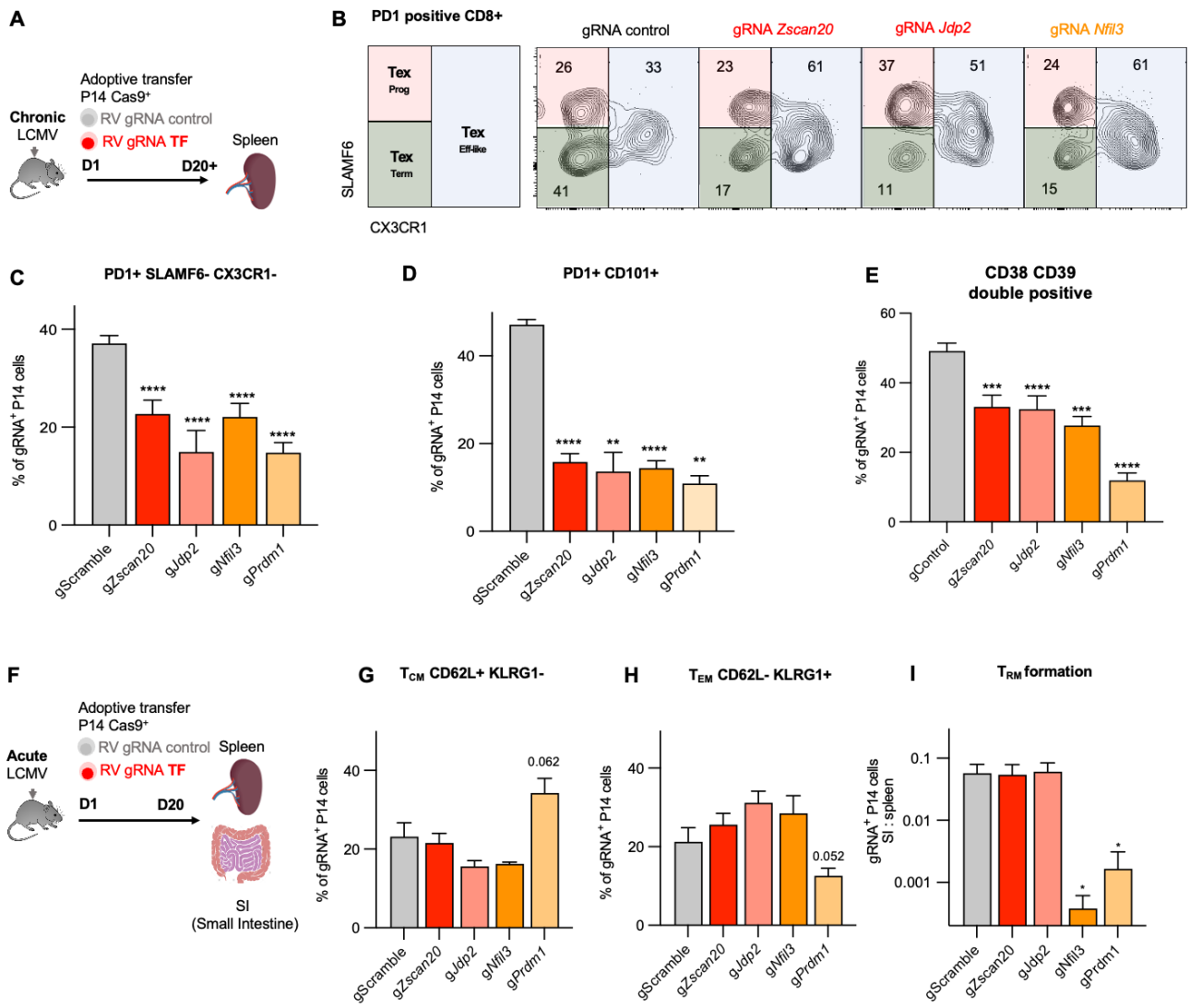


Fig 3. Disruption of Tex_{Term} single-tasker TFs prevents terminal exhaustion without perturbation of memory formation. Three novel TFs (*Zscan20*, *Jdp2*, and *Nfil3*) and one previously reported TF (*Prdm1*) to validate *in vivo*. *Zscan20* and *Jdp2* are highly specific to T cells in exhaustion trajectory and have low activity in T_{RM}. *Nfil3* and *Prdm1* are multitaskers that have high TF activity in both Tex_{Term} and T_{RM}. (**A**) LCMV (epitope: GP33-41)-specific TCR expressing T cells (P14 cells) from Cas9 transgenic mice were transduced with gRNA targeting TFs and adoptively transferred to Day 1 LCMV chronic strain infected animals. (**B**) Representative flow plot of PD1 positive P14 T cells gated with SLAMF6 and CX3CR1. Quantification of (**C**) TexTerm PD1+ SLAMF6- CX3CR1- and inhibitory receptor, PD1+ CD101+ (**D**) and CD39+CD38+ positive cells. CRISPR knock-outs of all four TFs fail to adopt TexTerm states. Deletion of Tex_{Term} exclusive TF genes (*Zscan20*, *Jdp2*) does not compromise memory formation while deletion of Tex_{Term} & T_{RM} multitasker TF genes (*Nfil3*, *Prdm1*) significantly deters the process. (**F**) TF KOs were tested in the acute LCMV infection model to profile memory formation. Percentage of (**G**) T_{CM} ($CD62L^+$ $KLRG1^-$), (H) T_{EM} ($CD62L^-$ $KLRG1^+$) among gRNA positive P14 cells in the spleen. (**I**) The ratio of gRNA containing P14 cell number of small intestines (SI) and that of spleens. Perturbation of multitasker TFs (*Nfil3* & *Prdm1*) significantly compromises T_{RM} formation. Statistic summary on each bar is from unpaired t-test between gScramble vs TF gRNA. Data include more than three biological replicates, $n \geq 6$.

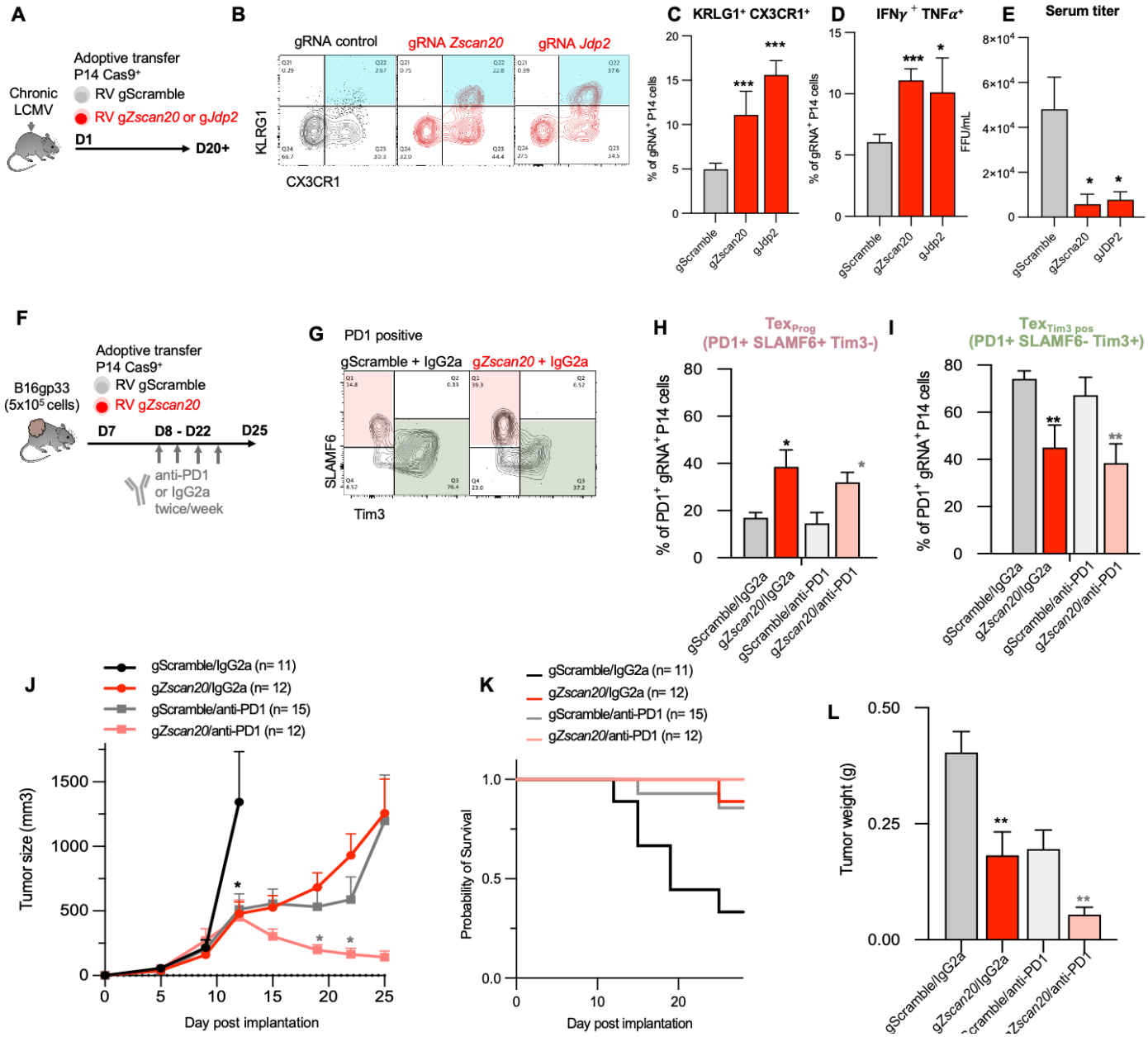


Fig.4. Disruption of transcription factor *Zscan20* and *Jdp2* provides enhanced virus and tumor control. In LCMV chronic strain-infected mice (**A-E**), (**A**) Experiment timeline. Cas9+ P14 T cells are transduced with gRNA *Zscan20*, gRNA *Jdp2*, or scramble gRNA. (**B**) g*Zscan20* and g*Jdp2* RV group shows higher effector-associated markers (CX3CR1, KLRG1). (**C**) Quantification of (**B**). (**D**) T cells with gRNA *Zscan20* or gRNA *Jdp2* are more polyfunctional and release both IFNγ and TNFα. (**E**) g*Zscan20* and g*Jdp2* RV group shows lower serum virus level. (**F**) *Zscan20* KO or control P14 cells were adoptively transferred to melanoma-bearing mice followed by anti-PD1 or isotype IgG2a control treatment. (**G**) P14 cells from Isotype control-treated mice show that *Zscan20* perturbation reduces exhaustion. Flow plot of PD1 positive P14 cells gated with SLAMF6 (Tex_{Prog} marker) and TIM3 (Tex_{Term} marker in cancer). (**H**) Tex_{Prog} Quantification of (**G**). (**I**) Tex_{Term} Quantification of (**G**). (**J**) Tumor growth (**K**) survival curve and (**K**) tumor weight show that *Zscan20* KO reduces tumor burden and elongates survival with both IgG2a and anti-PD1 treatment groups. With a combination of anti-PD1, disruption of *Zscan20* significantly suppress the tumor growth. All data include more than three biological replicates, n ≥ 9. Data are expressed as mean ± SEM. Statistical analysis was performed using Student's t test (two-tailed) comparing TF gRNA vs scramble gRNA within no anti-PD1 treatment groups (black) or anti-PD1 group (gray).

Materials and Methods

Dataset acquisition: The analyzed 42 $CD8^+$ T cell samples were collected from ten datasets including this study (fig. S2A). In total, 139 experiments including 64 ATAC-seq and 75 RNA-seq were harnessed to generate the paired samples and used as input of 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 ($Tex_{Eff\text{-}like}$) and terminal exhausted cells (Tex_{Term}).

TF regulatory networks construction and visualization: Taiji v1.1.0 with default parameters was used for the integrative analysis of RNA-seq and ATAC-seq data (https://github.com/Taiji-pipeline/Taiji). The motif file was downloaded directly from the CIS-BP database(*31*). There were 871 mouse motifs in the analysis.

Identification of 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 type 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 log-normal distribution. Based on 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 (table S1).

Identification of differentiation-step-specific TFs: We used a framework we recently developed called Taiji-reprogram(*22*) to identify differentiation-step-specific TFs, we first calculated PageRank ratios between target and source cell types. We then defined ratio.abs as the reciprocal of ratio if the ratio was smaller than 1, or otherwise, as the ratio itself. A higher PageRank ratio.abs represents that the TF behaves quite differently between target and source. We selected the top 30 TFs based on PageRank score ratio.abs as candidate TFs. We then calculated the product of PageRank score ratio.abs of three candidate TFs for all combinations. All the products were transformed to z-scores and p-value of 0.001 was used as cut-off to select candidate recipes. Finally we ranked the TFs based on the frequency of candidate TFs in all the candidate recipes.

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 of all, we performed the principal component analysis (PCA) for dimension reduction of TF score matrix. We retained the first 10 principal components for further clustering analysis, which explained more than 70% variance (fig. S4B). 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 (fig. S4C). 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 (listed in fig. S4D-G).

TF community construction and visualization: Inspired by DBPNet (*48*), 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 the TF interaction network based on Taiji's output, which is TF-regulatee network. We first combined TexTerm 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. And then $n \times n$ correlation matrix (n is the total TF number) is calculated by taking account of the Spearman's correlation of edge weight for each TF-regulatee pair. R package "huge"(*49*) is used to build graphical model and construct graph. We employed the Graphical lasso algorithm and the shrunken ECDF (empirical cumulative distribution function) estimator. We used a lasso penalty λ equal to 0.2 to control the regularization. We chose this value because around 5% of TF-TF pairs have a correlation score > 0.2. 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 0.2 has a very low false discovery rate. Communities were detected using Leiden algorithm (*50*) with modularity as objective function and resolution as 1.8. Network visualization was performed by igraph with Fruchterman-Reingold layout algorithm (*51*).

GO and KEGG pathway enrichment analysis: The enriched functional terms in this study were analyzed by R package clusterProfiler_4.0.5. . For gene set enrichment analysis (GSEA) in fig. S9-11, the genes were first ranked by the mean log normalized expression values in TexTerm samples and then CD8_TCELL related pathways were plotted using running scores. A cutoff of P-value < 0.05 was used to select the significantly enriched GO terms and KEGG pathways.

Mice and Infections: C57BL/6/J 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 all experimental studies were approved and performed in accordance with guidelines and regulations implemented by the Salk Institute Animal Care

and Use Committee. Mice were infected with $2x10^5$ PFU LCMV-Armstrong by intraperitoneal injection or $2x10^6$ PFU LCMV-Clone13 by retro-orbital injection under anesthesia.

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). 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, BME 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. 4; fig. S13) and 3×10^5 B16-gp33 tumor cells (fig. S14) were injected subcutaneously in 100µl PBS. Tumors were measured every 2-3 days post tumor engraftment or 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 GK1.5, BioXcell) twice per week from day 7 post tumor implantation. All experiments were conducted according to the Salk Institute Animal Care and Use Committee.

Tumor digestion and cell isolation: 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.

Retrovirus transduction and adoptive transfer: For gRNA retrovirus vector over-expression, 293T cells were transfected with Eco-helper and MSCV gRNA vectors. 48 hr and 72 hr 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-PD-1 (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), anti-GZMB (GB11). These antibodies were purchased from Invitrogen, Biolegend, Cell Signaling, or eBiosciences.

ATAC-Seq library preparation and sequencing: ATAC-seq was performed as previously described (Corces et al., 2017). 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 H_2O) 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 (*52*, *53*)with parameters "bwa mem -M -k 32". ATAC-seq peaks were called using MACS2 (*54*) program using parameters "callpeaks -qvalue 5.0e-2 –shift -100 –extsize 200". Differentially accessible regions were identified using DESeq2 (*55*) . Batch effect was removed using limma (*56*). Heatmap visualization of ATAC-seq data was performed using pheatmap.

Single-cell RNA seq analysis Data analysis: Analysis was primarily performed in R (v 3.6.1) using the package Seurat (v 3.1) (*57*, *58*), with the package tidyverse (v 1.2.1) (*59*) 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.

Statistical analyses: Statistical tests for flow cytometry data were performed using Graphpad Prism 7. p-values were calculated using either two-tailed unpaired Student's t-tests or one-way ANOVA while correcting for multiple comparisons via the Tukey method. Linear regressions were performed using the ordinary least squares method in R (v 3.6.1). Boxplots and violin plots were compared pairwise in R using the Wilcoxon test with Holm-Sidak correction. All data were presented as the mean ± SEM. The *P* values were represented as follows: ****$P < 0.0001$, ***$P < 0.001$, **$P < 0.01$, and *$P < 0.05$.

Data and code availability: ATAC-sequencing data from this paper will be deposited in the GEO database. All other raw data and scripts are available from the corresponding author upon request.

Supplementary Figures

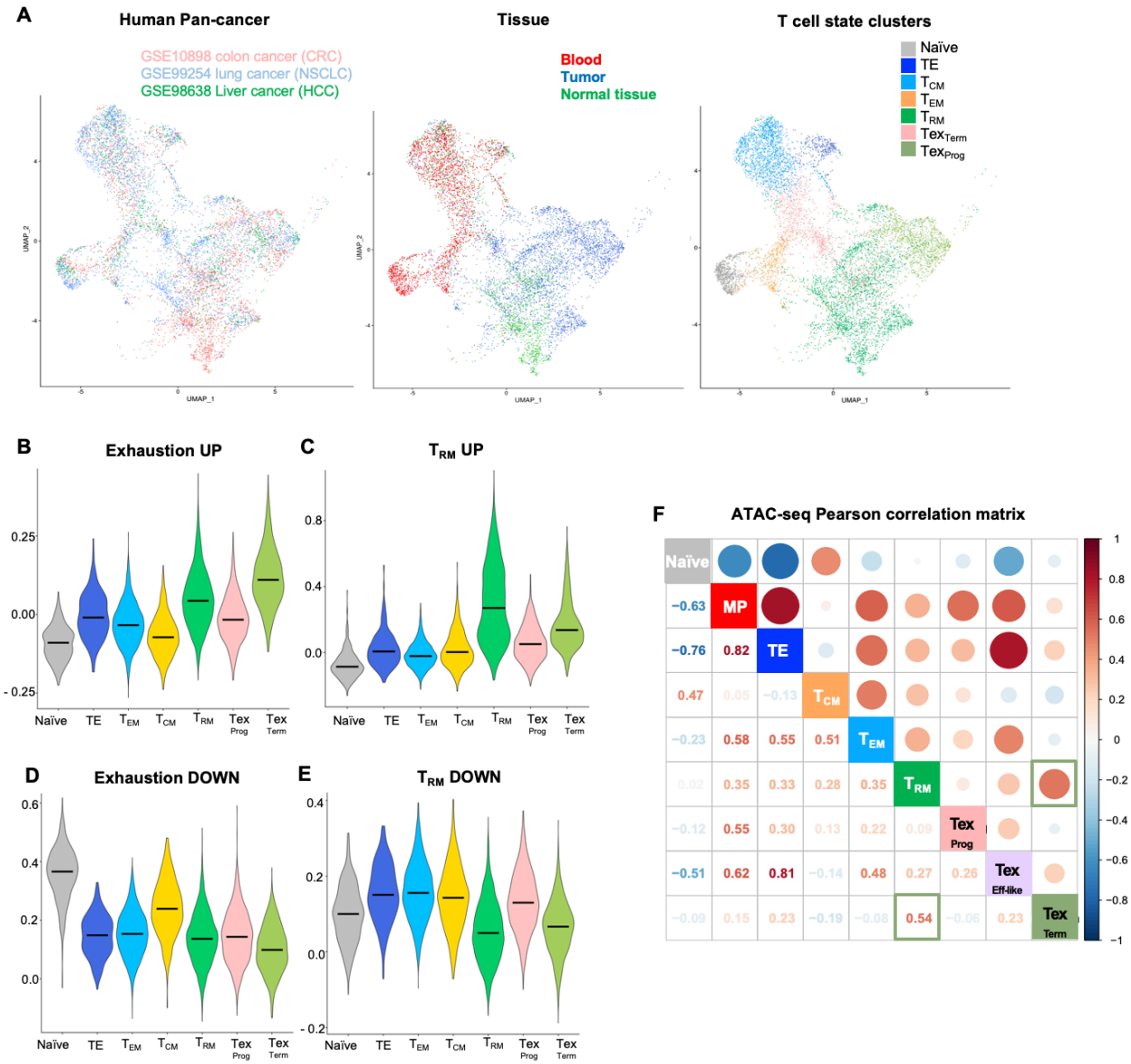


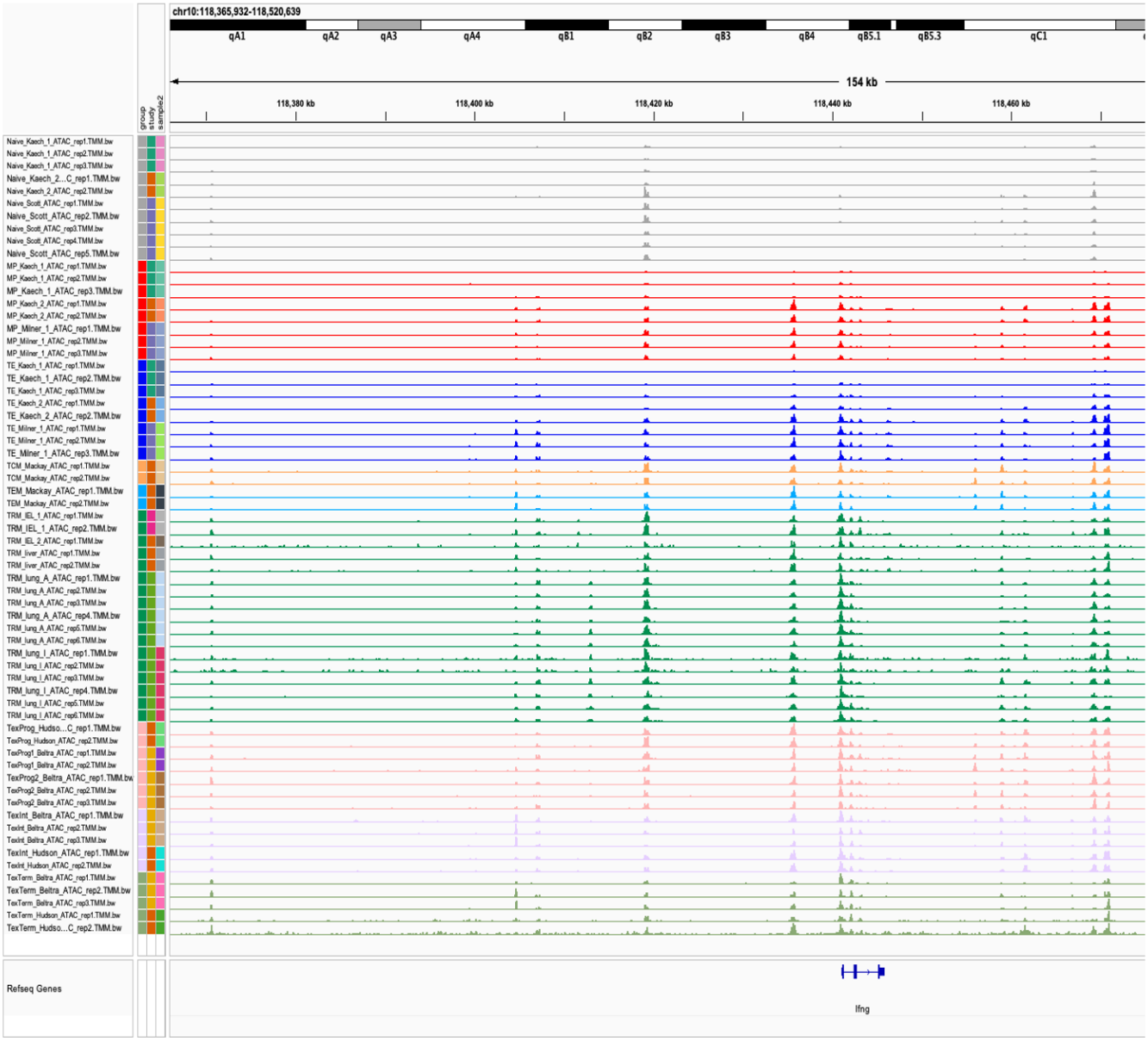
Fig. S1. Parallel differentiation of T_{RM} and Tex_{Term} and their similarity in the transaction.
(**A**) UMAP from Single-cell RNA sequencing data of T cells from blood, tumor, normal tissues of CRC, NSCLC, and HCC patients. Unbiased clustering demonstrate multiple T cell states that are previously reported in LCMV infection and tumor. (**B-E**) Module score of genes that are up regulated in exhausted T cells (**B**), up regulated in T_{RM} (**C**), down regulated in exhausted T cells (**D**), down regulated in T_{RM} (**E**), in each cluster of cell state. (**F**) Pearson correlation matrix of batch effect corrected ATAC-seq (fig. S2). Color and size are both proportional to correlation values.

A

10 datasets | 9 subsets | 121 experiments | 52 ATAC-seq | 75 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}, $Tex_{Eff\text{-}like}$, Tex_{Term}
Beltra(2020)	GSE149876	12 RNA + 11 ATAC	Tex_{Prog}, $Tex_{Eff\text{-}like}$, Tex_{Term}
Chung(2022)	This study	19 ATAC	Naïve, TE, MP, T_{CM}, T_{EM}, T_{RM}, Tex_{Prog}, $Tex_{Eff\text{-}like}$, Tex_{Term}

B



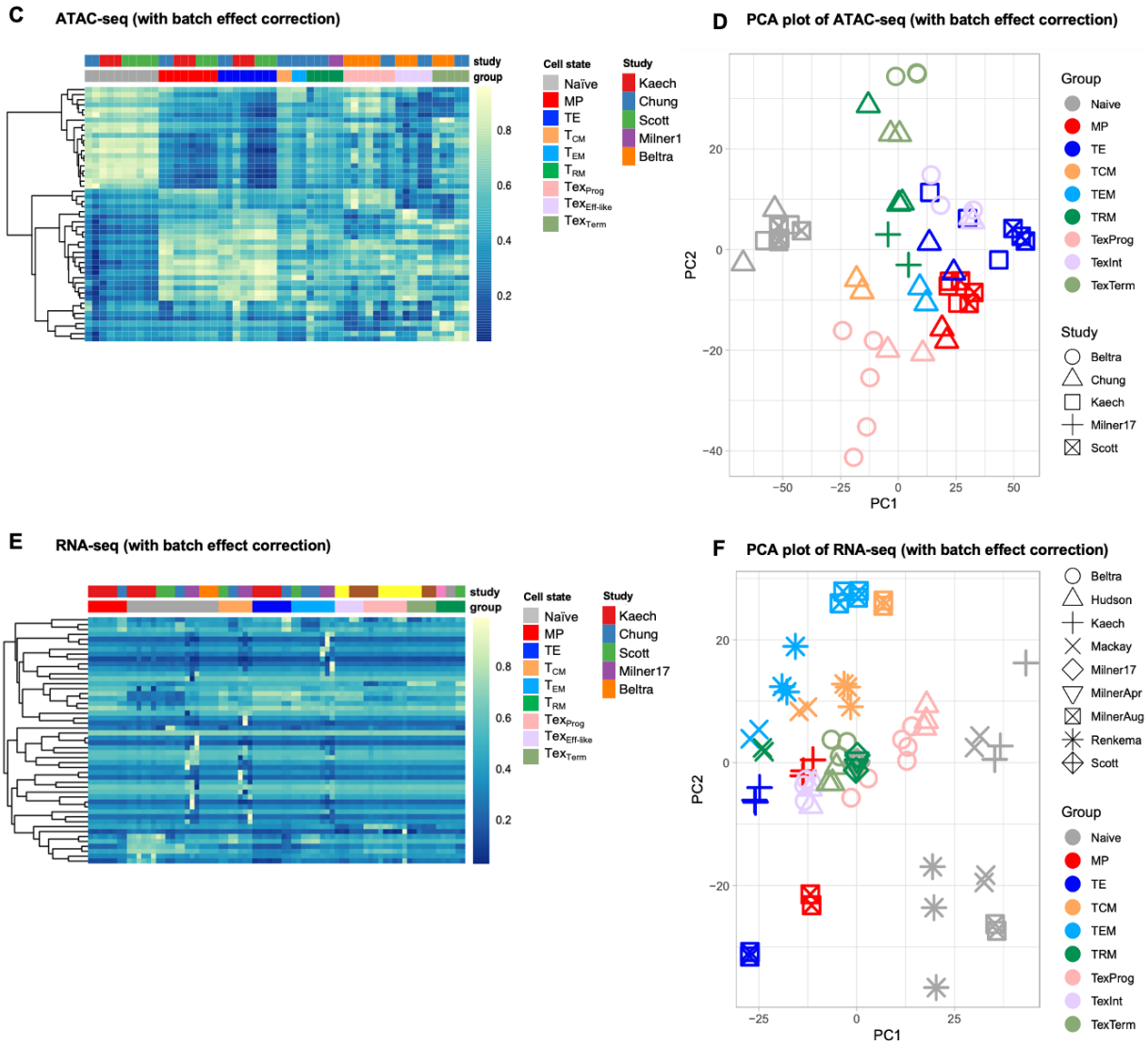


Fig. S2. 121 experiments from multiple data sets are utilized to generate an epigenetic and transcriptional atlas of CD8$^+$ T cells in chronic and acute antigen exposure. (A) Dataset summary. (**B**) Cchromatin accessibility of *Ifng* locus from ATAC-seq data that are used in this study. (**C, D**) Batch effect-corrected ATAC-seq data and top1000 most variable peak regions were selected for visualization. (**C**) Heatmap of batch effect corrected ATAC-seq with rows clustered using kmeans by k=50 and columns arranged by cell state and study**.** (**D**) PCA shows the ATAC-seq samples tend to group together by cell states rather than study. (**E**) Heatmap of batch effect corrected RNA-seq with rows clustered using kmeans by k=50 and columns arranged by cell state and study**.** (**F**) PCA shows the RNA-seq samples tend to group together by cell states rather than study.

A

	Naive	MP	TE	T_{CM}	T_{EM}	T_{RM}	Tex_{Prog}	Tex_{Int}	Tex_{Term}
# of specific TFs	34	59	61	30	26	44	67	66	58
# of unique TFs	17	19	18	16	13	17	20	14	15
Previously reported TFs examples	*Tcf7* *Lef1*	*Tcf7* *Myb* *Runx2*	*Tbx21* *Irf4* *Runx2* *Runx1* *Prdm1*	*Lef1* *Tcf7* *Runx2* *Fosb*	*Prdm1* *Runx2*	*Prdm1* *Bhlhe40* *Nr4a2* *Hic1* *Fosb* *Junb* *Fos* *Jun*	*Tcf7* *Eomes* *Nr4a2* *Myb* *Batf* *Jund*	*Nfatc1* *Prdm1* *Nr4a2* *Batf* *Bhlhe40* *Jund*	*Nfatc1* *Prdm1* *Nr4a2* *Batf* *Bhlhe40* *Jund*
New TFs examples	*Rarg*	*E2f2* *Gils3*	*E2f2* *Atf6*	*Fosl2* *Rara*	*Fosl2* *Atf4*	*Fosl2* *Nfil3* *Atf4* *Vax2*	*Jdp2* *Irf9* *Vax2*	*Jdp2* *Irf9* *Vax2*	*Hic1* *Jdp2* *Nfil3* *Vax2*
New, exclusive TFs examples	*Hoxa7*	*Hoxb4*	*Ikzf3*	*Cux1*	*Meis3* *Zfp398*	*Snai1*	*Hey1*	*Sox8* *Arid3c*	*Zscan20* *Zfp143* *Stat3* *Arid3a*

B

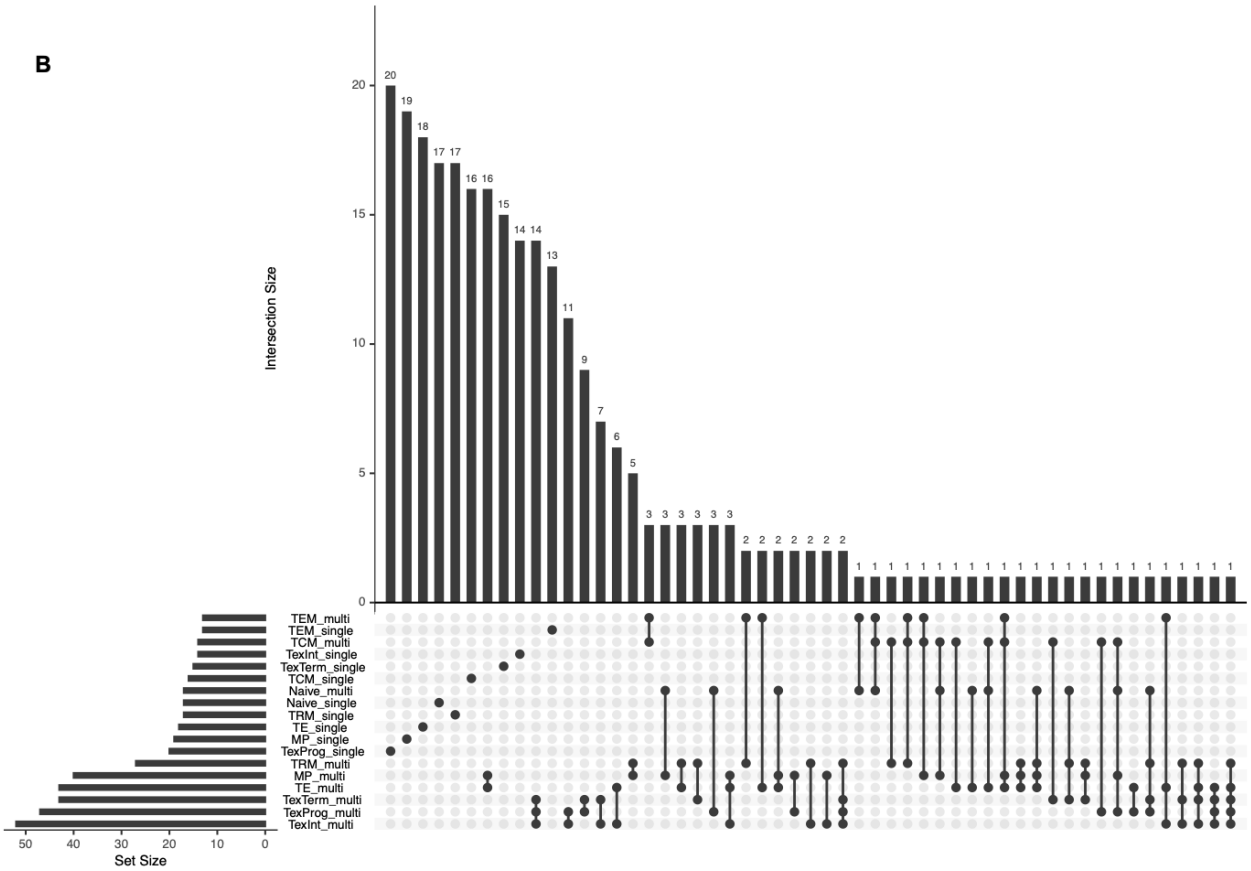


Fig. S3. Catalog of key TFs in 9 different CD8$^+$ T cell states. (**A**) Summary table of identified key TFs in 9 cell states. (**B**) UpSetR plot shows the intersection size between multi-taskers along with single-taskers size.

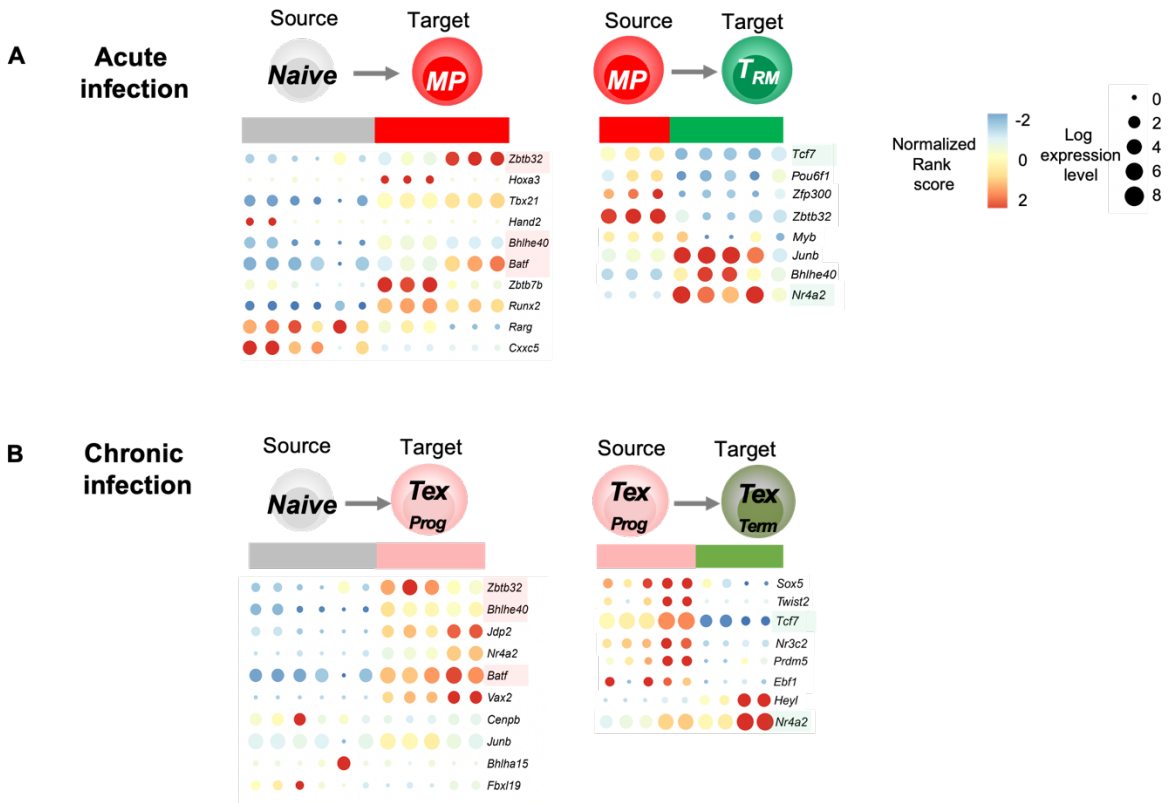


Fig. S4. Systemic analysis of driving TFs in each differentiation step.

(**A**) Differentiation-step-specific TFs are identified for the cell state transition from Naive to MP and MP to T_{RM} in the acute infection. Dots represent the labeled source and target cell state data. Circle size represents the logarithm of gene expression, and the color represents normalized PageRank score in each state. (**B**) The parallel differentiation step in chronic infection. Up and downregulation of TFs in Naive to Tex_{Prog} transition and Tex_{Prog} to Tex_{Prog} transition are shown.

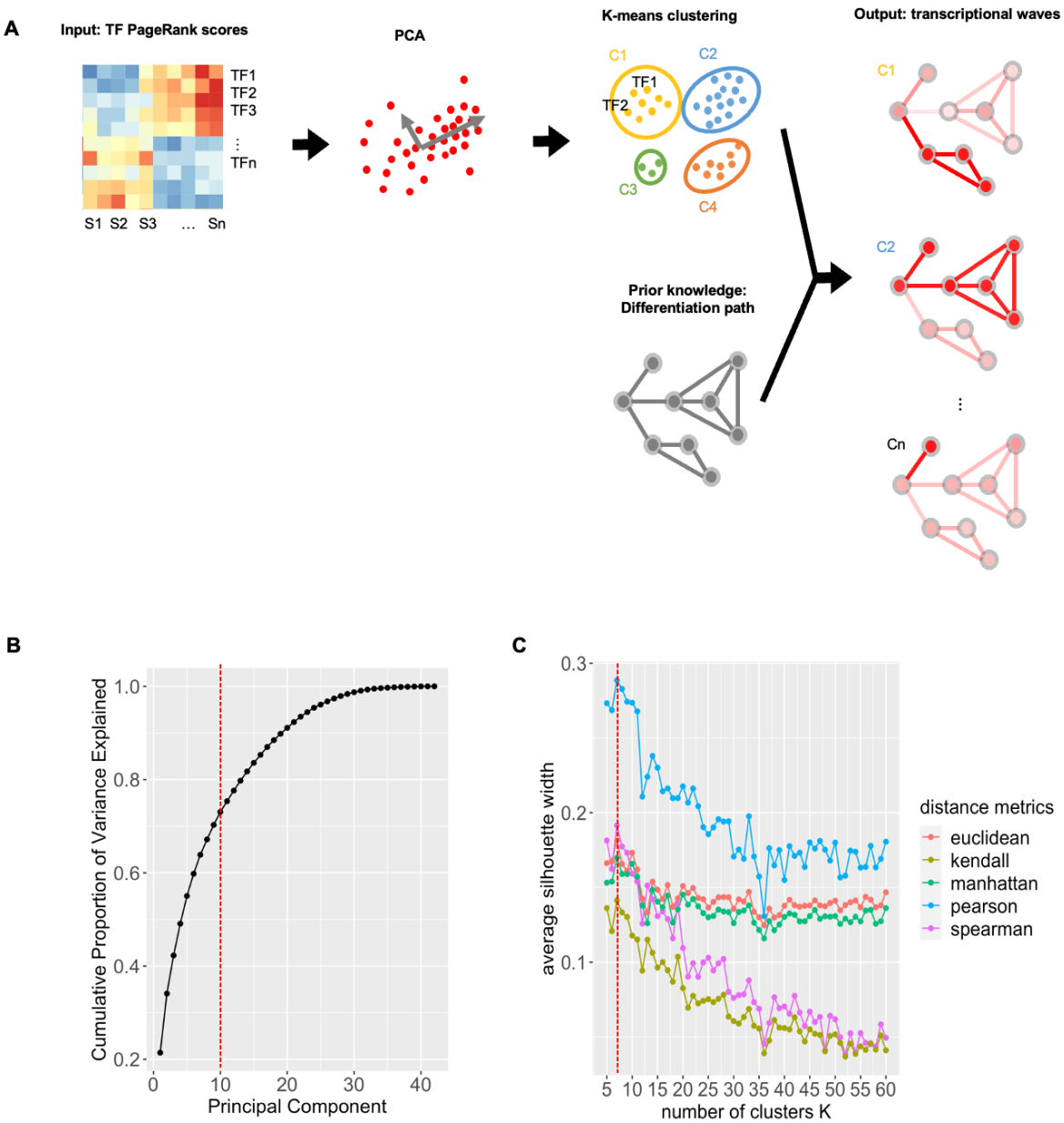
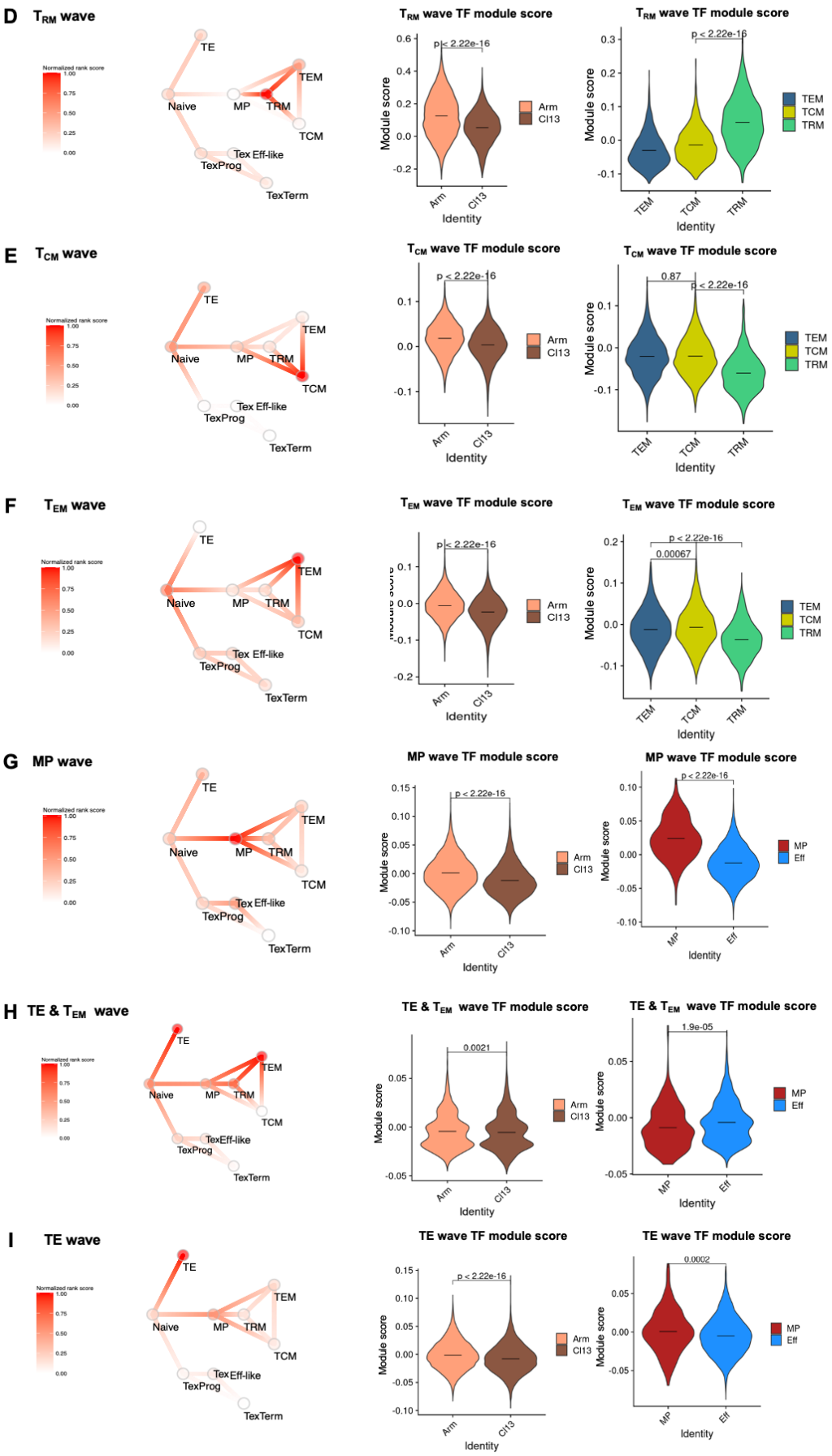


Fig. S5. Transcription factor wave analysis (**A**) Analysis pipeline. (**B,C**) Selecting algorithms and parameters for clustering analysis for transcription factor wave. (**B**) Plotting the cumulative proportion of variance explained against the number of principal components (PCs). The first 20 PCs were kept for the following K-means clustering, which explained ~70% variance. (**C**) Selecting the best distance metric and number of clusters according to the Silhouette metric. The Pearson correlation was chosen and k=7 was the ideal cluster number.


D
T_{RM} wave
T_{RM} wave TF module score
T_{RM} wave TF module score
E
T_{CM} wave
T_{CM} wave TF module score
T_{CM} wave TF module score
F
T_{EM} wave
T_{EM} wave TF module score
T_{EM} wave TF module score
G
MP wave
MP wave TF module score
MP wave TF module score
H
TE & T_{EM} wave
TE & T_{EM} wave TF module score
TE & T_{EM} wave TF module score
I
TE wave
TE wave TF module score
TE wave TF module score
Normalized rank score
TE
Naive
MP
TRM
TEM
TCM
Tex Eff-like
TexProg
TexTerm
TexEff-like
Module score
Identity
Arm
Cl13
TEM
TCM
TRM
MP
Eff
p < 2.22e-16
0.87
0.00067
0.0021
1.9e-05
0.0002

J

Wave	TFs
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 Pbx4 Ddit3 Meox2 Alx4 Hmx1 Klf4 Dlx1 Elf3 Hmx2 En2 Hesx1 Mlx Dlx4 Kdm2b Gmeb1 Zfp879 Zfp692 Etv1 Nkx1-2 Hoxa1 Sohlh2 Atf1 Foxn3 Hlx Prrx2 Dlx6 Hmx3 Xbp1 Hes7 Tfec Cxxc4 Nfyc Arx Hsf5 Zfp105 Klf7 Lhx9 Arid5a Dpf3 Gbx2 Pou4f2 Hsfy2 Zfp184 Sox8 Dbx2 Zfp13 Gsx2
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 Tet3 Kmt2a Zbtb7b Fbxl19 Zbtb37 Ppard Hoxb4 Nfya Zbtb1 Tbx6 Nr2c1 Mnt E2f3 Rxra Zfp335 Creb3l2 Pbx2 Cenpb Dmrta1 Dnajc21 Mef2a Mecp2 Zbtb12 Rest Hoxa3 Runx2 Zfp777 Maz Rfx7 Arid5b Terf2 Esrra Mafb Mafa Ubp1 Runx3 Mef2d Mlxip Nfatc2 Foxp4 Nfkb1 Zfp146 Elf4 Mga Nr1h3 Pparg Foxk2 Nfic Crebl2 Kdm2a Gmeb2 Nr1d2 Srf Ikzf3 Prdm11 Cbfb Pou2f2 Zfp787 Zbtb45 Zfp12 Zfp384 Max Zbtb22 Tef Mbd2 Foxd2 Rxrb Tcf3 Crebzf Klf2 Nfatc3 Jund
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 Zfp300 Hes1 Zfp524 Irf3 Cphx3 Lhx1 Sp100 Rora Rbak Sox15 Rfx3 Arid3b Runx1 Hoxc6 Ikzf1 Stat4 E2f7 Zfp322a Gata3 Sp3 Klf11 Klf13 Dbx1 Pou3f4 Hes6 Cxxc1 Nfia Zbtb32 Prrx1 Esx1 Zfp113 Bcl11b Elk1 Zfp3 Zfp64 Zfp317 Elf1 Zfp263 Mafk Zkscan6 Pitx3 Tigd2 Zfp46 Arnt Mynn Rorc Rela Zbtb17 Foxp1 Zkscan5 Zfp422 Zfx Cdc5l Thap11 Zkscan17 Snai3 Zfp771
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 Pknox1 Zfp661 Sp6 Tfap4 Trp53 Zik1 Spi1 Bcl11a Hic2 Gfi1b Klf1 Spib Hhex Zfp566 Prdm9 Klf3 Hes5 Zbtb7a Zfp248 Cebpd Cux1 Zfp110 Ebf1 Rfx2 Tcf4 Ybx1 Plagl1 Tfdp2 Pou5f1 Mitf Bcl6 Nfe2
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 Pknox2 Lhx6 Esrrg Tbx2 Egr4 Creb5 Lin28b Lin28a Cphx2 Duxbl1 Hnf4a Snai2 Foxb1 Foxf1 Gata5 Fezf2 Zfp57 Nr5a2 Nfib Tead4 Tead1 Rxrg Zfp42 Six2 Rfx6 Pax8 Hnf4g Msantd3 Tcf21 Hoxa5 Zscan4f Nr5a1 T Zfp711 Foxs1 Neurod2 Rhox8 Thap1 Hey2 Dmrtb1 Sox9 Mycn Nkx3-2 Sox11 Otx2 Pgr Sox1 Scrt1 Dmrta2 Mesp2 Prdm6 Foxe1 Hoxc9 Tbx15 Sox7 Sox18 Sp7 Foxl1 Zscan4d Sox30 Nfe2l2 Pitx2 Sox13 Bsx Nkx2-3 Sox21 Irx3 Meis2 Mafg Hoxc8 Onecut2 Tbx4 Gli2 Dmbx1 Prox1 Neurod4 Zfp423 Pou3f3 Twist1 Tbx19 Npas2 Hdx Irx5 Tbr1 Hoxc4 Tbx5 Hnf1b Foxc2 Hoxa9 Rhox9 Rorb Hoxb13 Hoxb2 Dmrt3 Msgn1 Pax3 Gcm2 Erg Hoxb5 Sox6 Hoxc5 Onecut3 Pou3f2 Meis1 Hoxa6 Tbx10 Bhlhe23 Aire Lhx4 Foxh1 Hmga2 Zfp354a Pax7 Isl1 Tbx1 Foxq1 Hoxa4 Six3 Rfx8 Pou3f1 Zfp354b Nfyb Pou6f2 Tbp Rfx4
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 Sp9 Fosl1 Bhlha15 Ar Zfp174 Gbx1 Myrf Tcf7l1 Myod1 Nkx2-5 Pou2f1 Irf5 Hoxd8 Ppara Mypop Nanog Zic5 Hoxb7 Msx2 Hsf2 Olig3 Nkx6-2 Mzf1 Rarg Prox2 Zfp383 Pou1f1 Dbp Rfx1 Ovol1 Nr2f6 Sox10 Hbp1 Sebox Zfp513 Maff Stat5a Foxp3 Zscan29 Cphx1 Vdr Foxc1 Irf6 Zeb1 Hnf1a Creb1 Zfp784 Sp1 Ctcf Etv4 Mef2c Cebpg Foxo1 Mxi1 Foxn2 Elk3 Zfp740 Barhl1 Patz1 Sp2 Egr3 Hoxa2 Myb Pax6 Ctcfl Ahctf1 Tet1 Srebf1
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 Etv5 Gfi1 Hmg20a Homez Trp73 Zbtb20 Lin54 Bbx Zfp410 Stat6 Zfp287 Arid3a Nfil3 Msc Mybl2 Bhlhe40 Nr4a2 Nrf1 Irf2 Lhx2 Arnt2 Hey1 Hic1 Hinfp Bhlhe41 Smad4 Foxj2 Zfp90 Zfp653 Maf Zfp637 Pbx3 Ovol2 Nfe2l1 Heyl Zbtb3 E2f1 Prdm4 Batf3 Foxm1 Hsf4 Nfkb2 Ybx3 Usf1 Prdm1 Hlf Tfe3 Atf2 Zfp282 Zbtb49 Zfp558 Mtf1 Zfp449 Trps1 Relb Klf12 Nr3c1 E2f8 Sox5 Rfx5 Otx1 Twist2 Zfp213 Zbtb14 Six5 Rbpj Ahr Hsf1 Foxf2 Elf2 Nr3c2 Pou6f1 Arid3c Prdm5 Zfp14 Etv6 Hif1a Klf8 Tfcp2 Thrb Zbtb2 Irx2 Zbtb43 Zscan20 Glis1 Meis3 Atf6b Setbp1 Zfp276

K

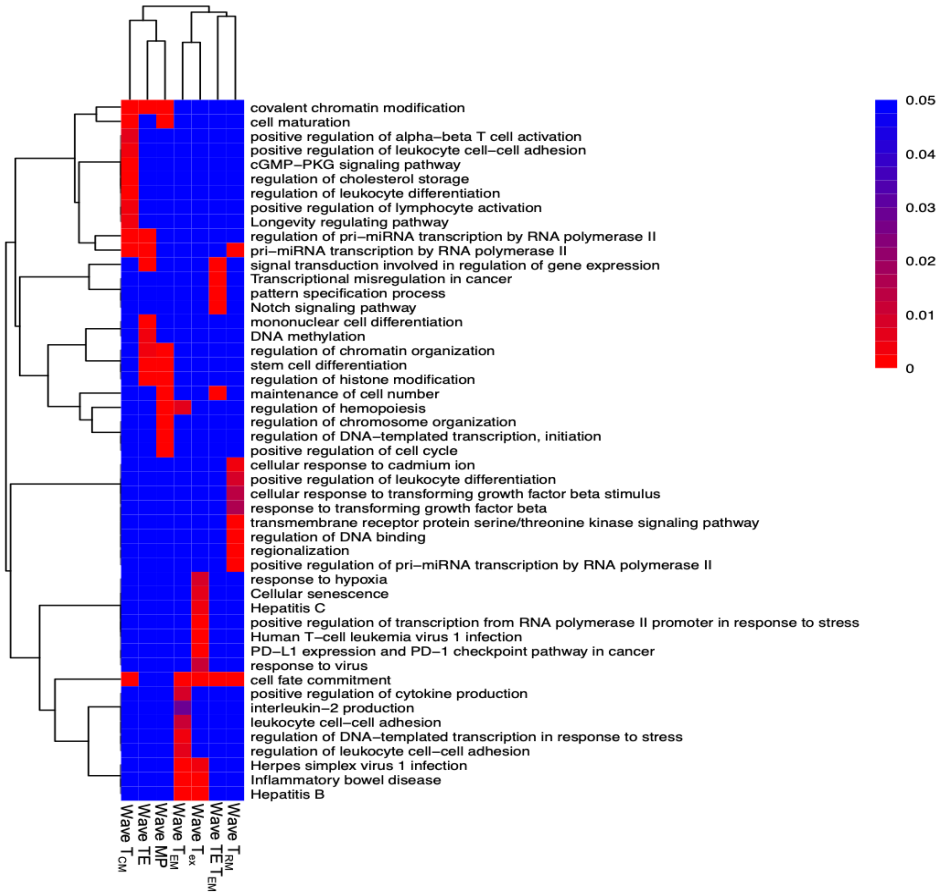


Fig. S5. Transcription factor wave analysis (**D-I**) The six transcriptional waves. Exhaustion wave is depicted in Fig. 1J. Circles represent specific cell state. Color indicates normalized PageRank scores with red displaying high values. The gene expression score of the all TF members of each wave are calculated in the associated single-cell RNA sequencing cell clusters of $CD8^+$ T cells from LCMV acute or chronic infection. Scores of MP, TE, acute, chronic clusters are calculated from GSE199565 (*4*) and scores of T_{CM}, T_{EM}, and T_{RM} are calculated from GSE181785(*38*). (**J**) TF members of each wave are listed. (**K**) Heatmap of biological pathways enriched in each transcription wave. Color represents p value.

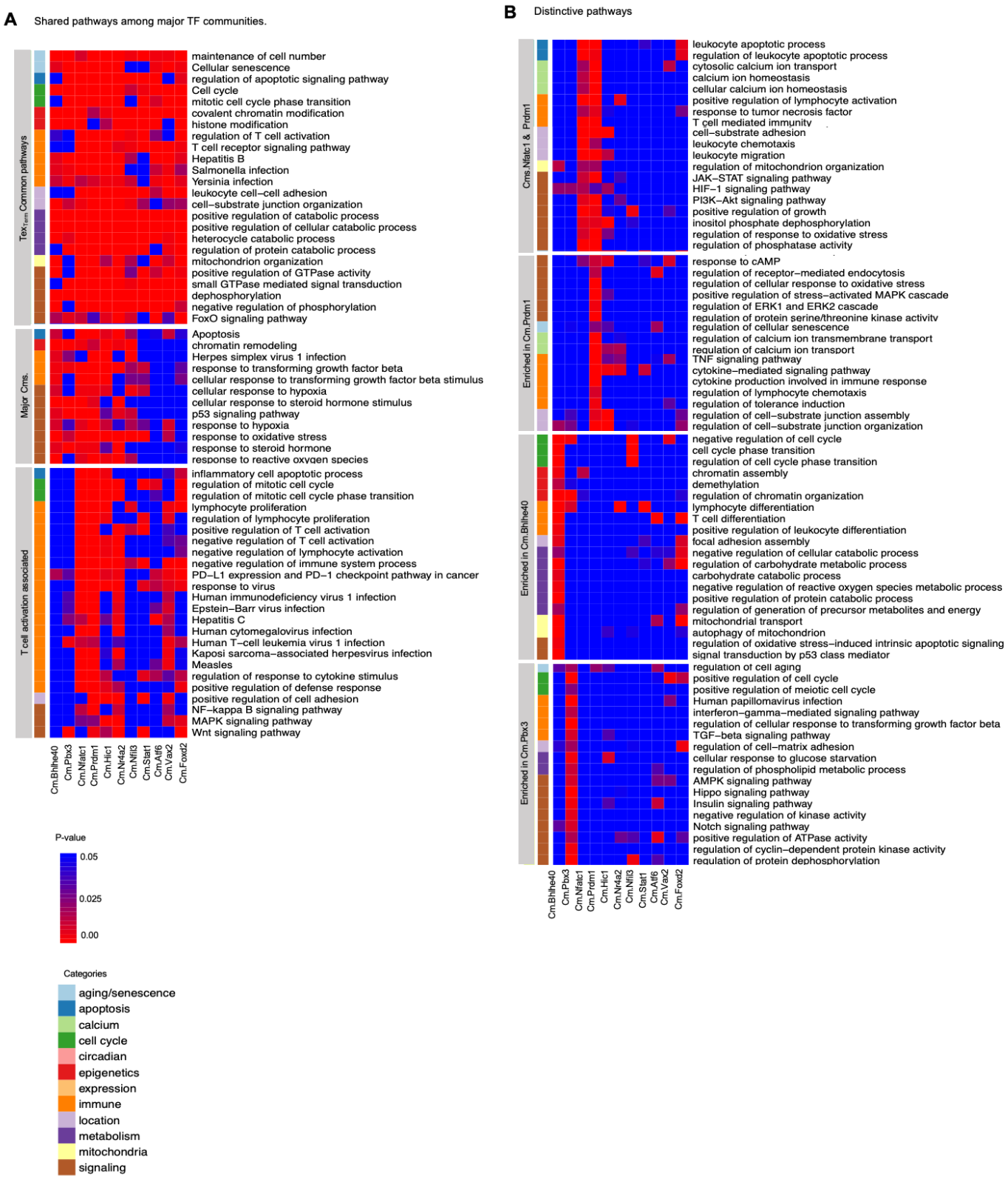


Fig. S6. Heat map of biological pathways enriched in each community in Tex$_{Term}$ (**A,B**) p-value color scale and color codes of the pathways are shown below. (**A**) Shared biological pathways among TF communities of Tex$_{Term}$. (**B**) Biological pathways that are distinct to different communities.

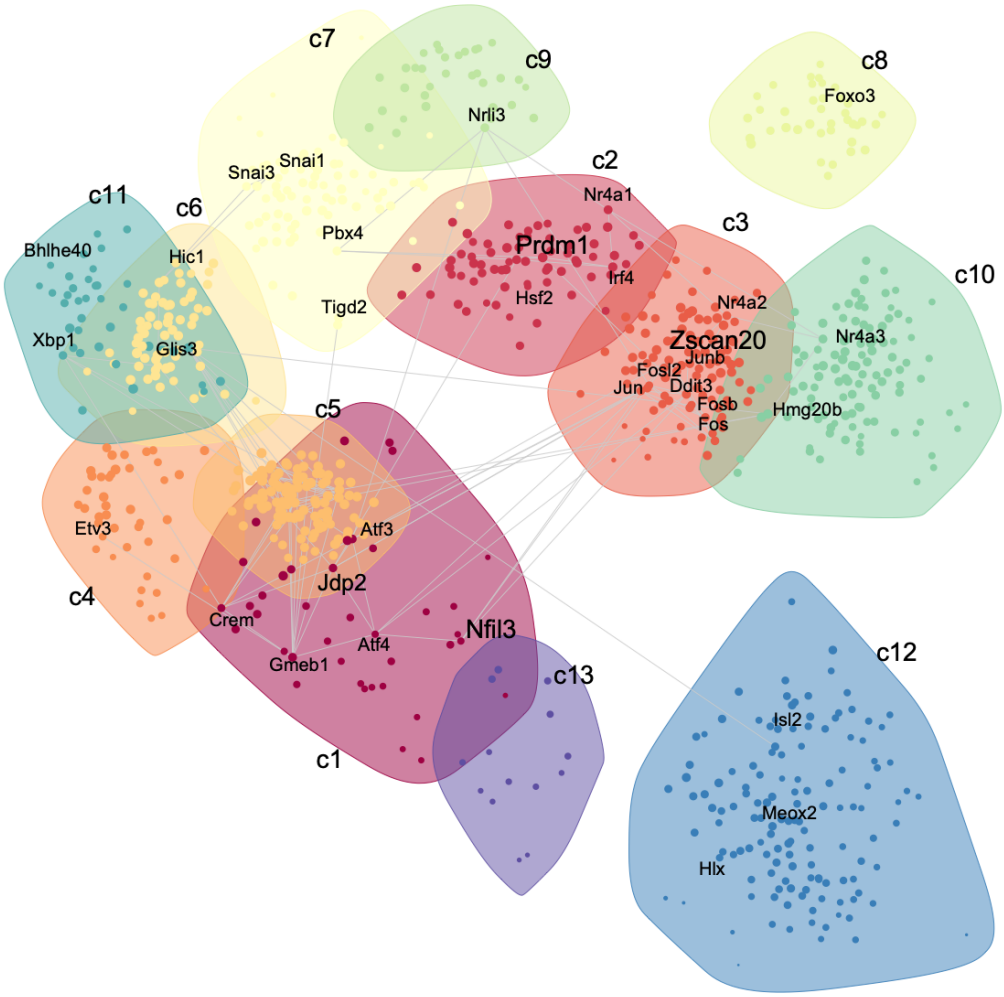


Fig. S7. Network analysis of TFs for T_{RM} formation

Communities of TFs in T_{RM}. Communities consist of T_{RM} TFs and their interaction partners. Dots represent each TF. Dot size represents the node degree and color represents each community. T_{RM}-specific TFs are labeled. Gray lines represent the interactions between T_{RM}-specific TFs.

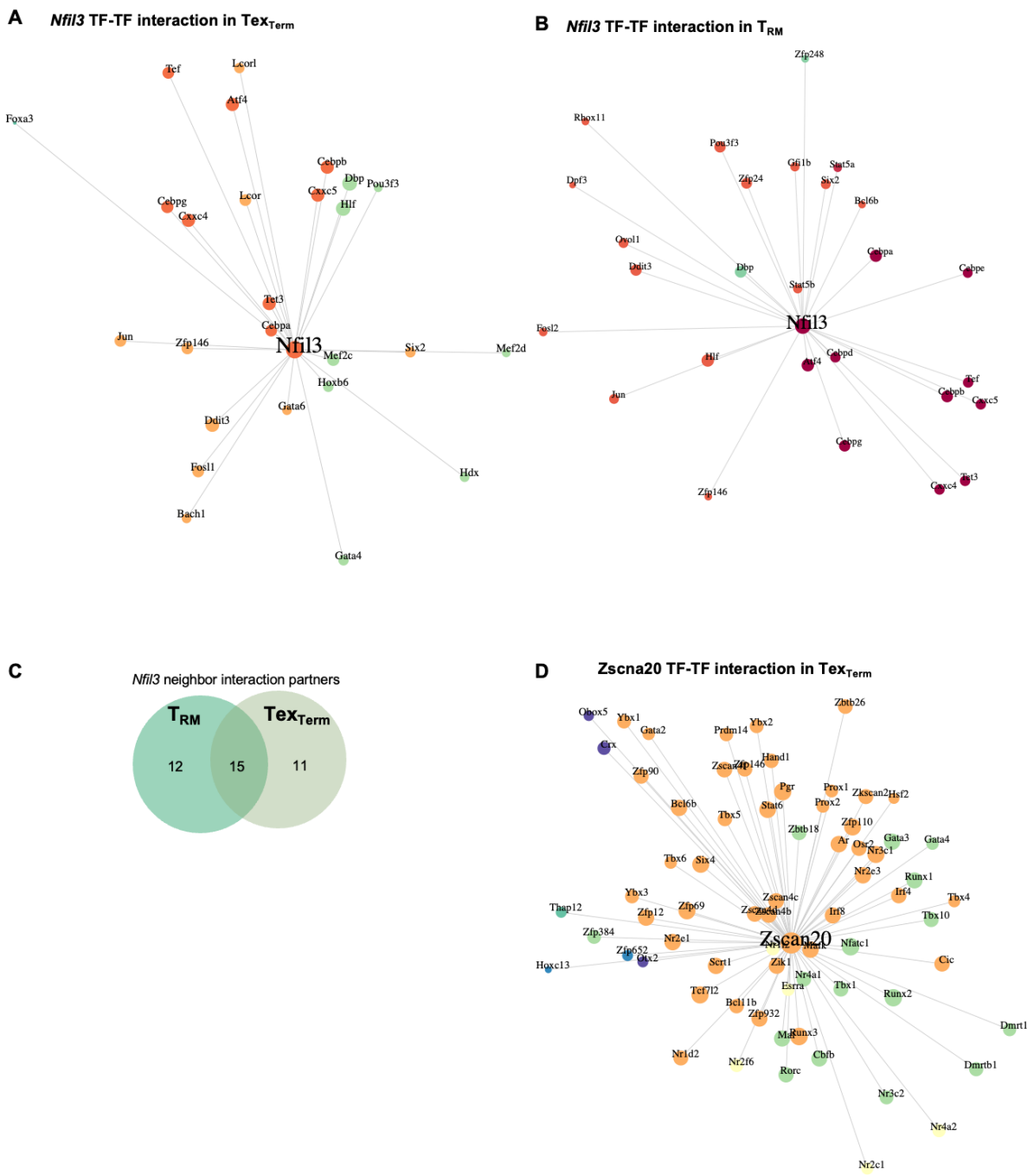


Fig. S8. TF-TF interaction analysis of Nfil3 and Zscan20

Shared or distinct Nfil3 interaction TFs in Tex$_{Term}$ and T$_{RM}$. Nfil3 and partner TF interactions of (**A**) Tex$_{Term}$ and (**B**) T$_{RM}$ are shown. Each color of the dots represents cluster where interaction partner belongs to and dot size represents the node degree.(**C**) Venn Diagrams show Nfil3 interacting partner TFs overlapping in Tex$_{Term}$ and T$_{RM}$. **(D)** Zscan20 and partner TF interactions in Tex$_{Term}$ are predicted and visualized.

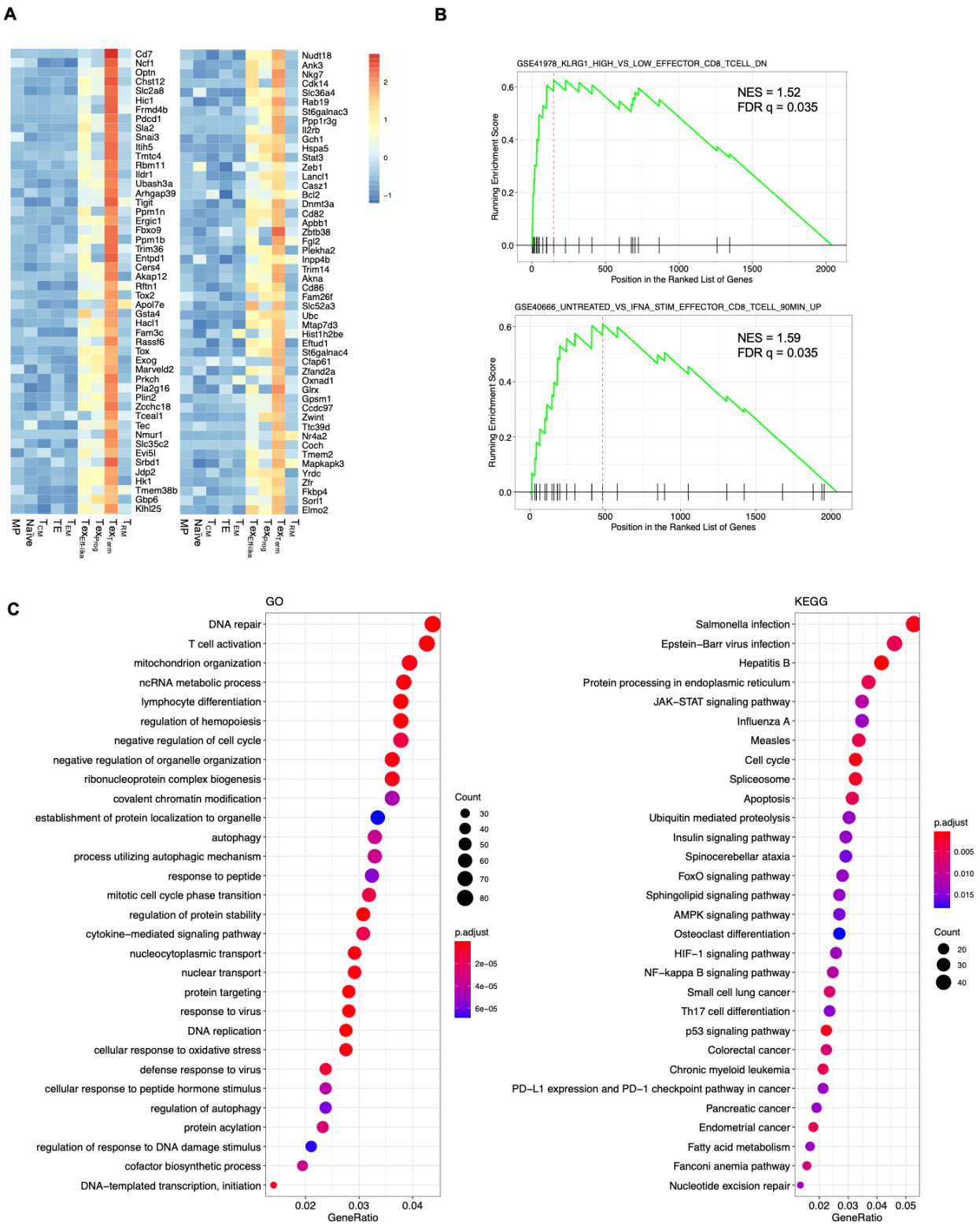


Fig. S9. Regulatee analysis – Zscan20 (A) Top 100 regulatees of Zscan20 ranked by the mean gene expression difference between TexTerm and TE samples. (B) Regulatees of Zscan20 are enriched in immunological genesets associated to downregulated genes in KRLG1 high effector or activated T cells. (C) Enriched GO terms (left) and KEGG pathways (right) of Zscan20 regulatees.

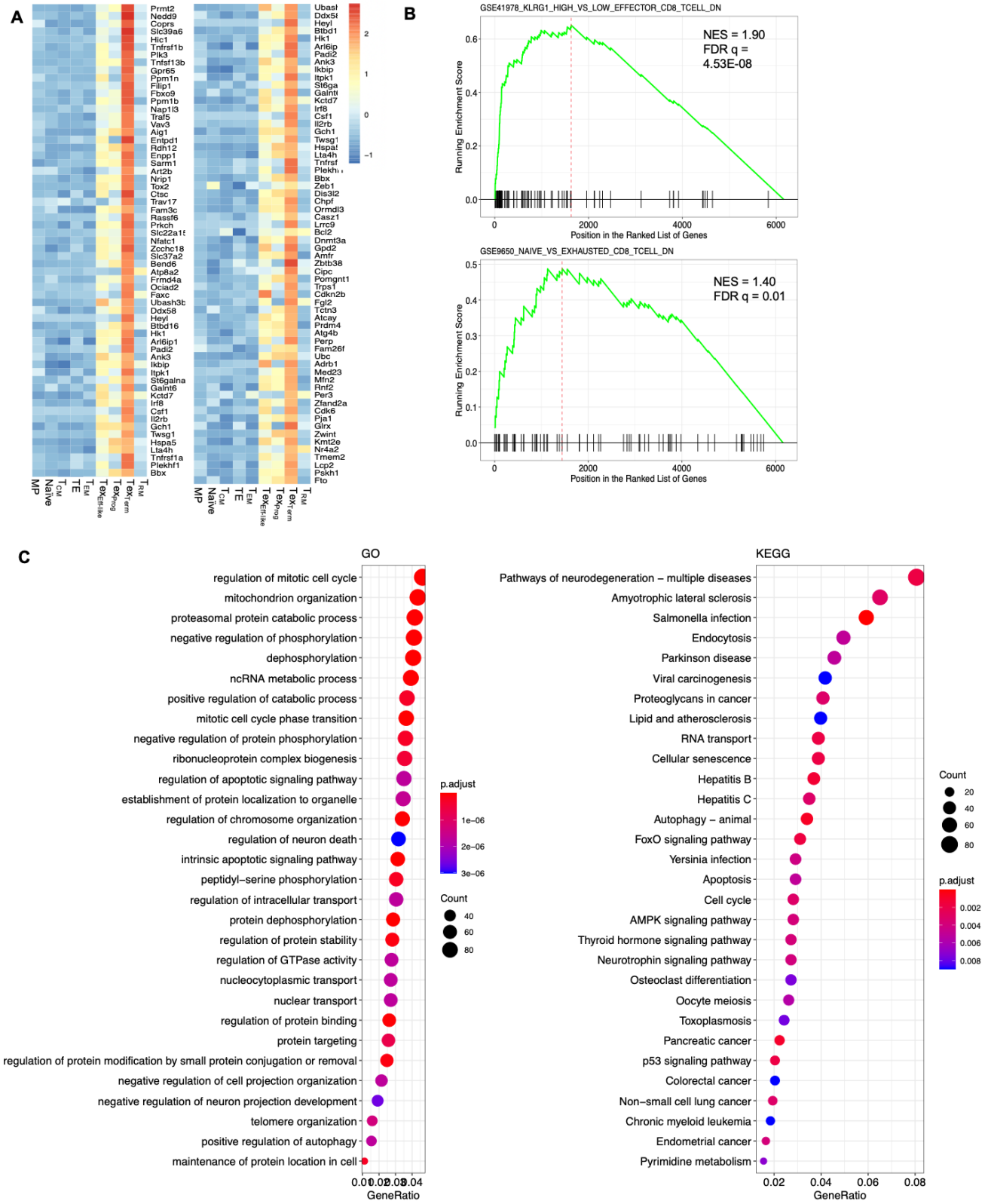


Fig. S10. Regulatee analysis – Jdp2

(A) Top 100 regulatees of Jdp2. (B) Regulatees of Jdp2 are enriched in immunological genesets associated to downregulated genes in effector states. (C) Enriched GO terms (left) and KEGG pathways (right) of Jdp2 regulatees.

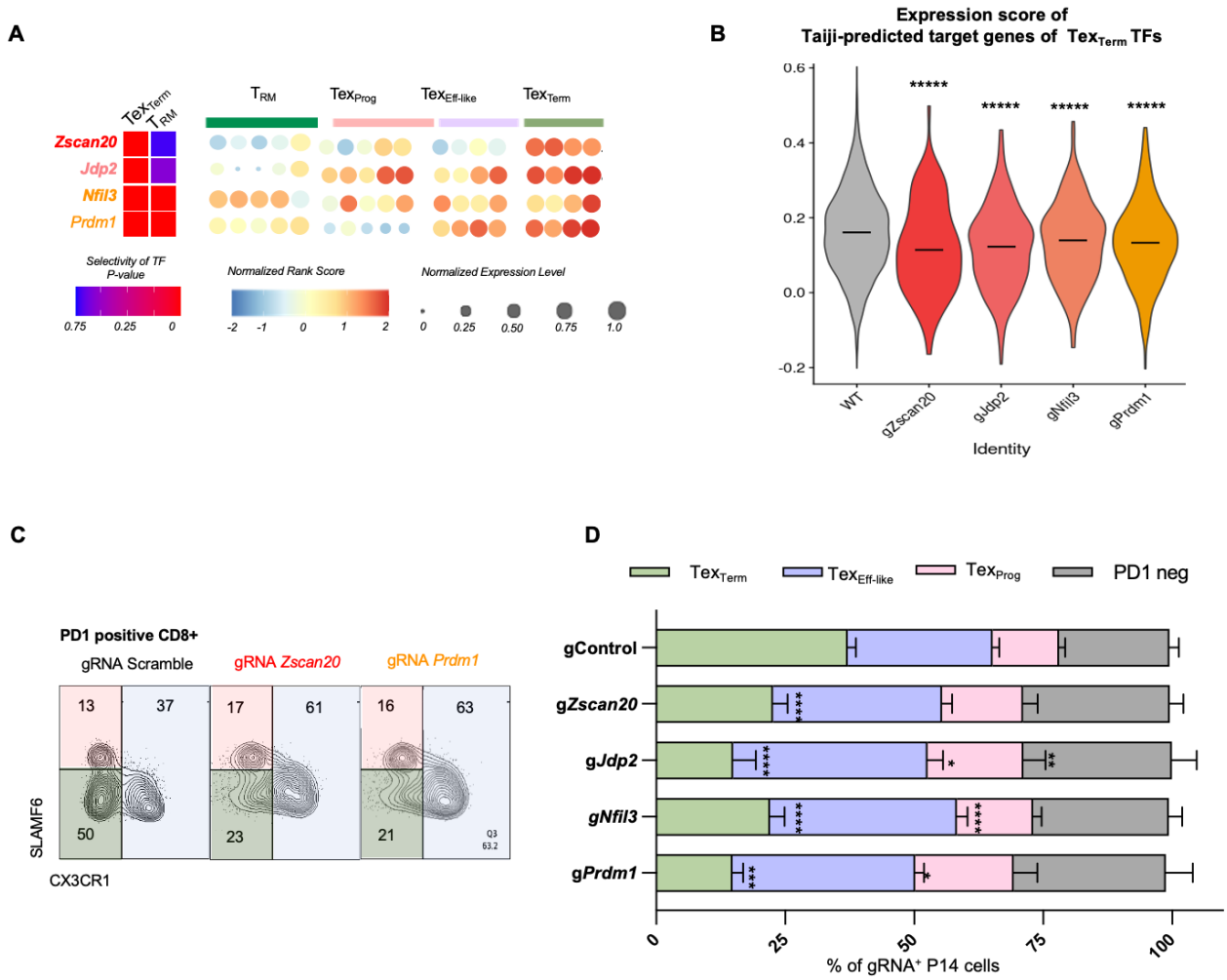


Fig. S11. Tex_{Term} driver TF KOs express lower terminal exhaustion markers and inhibitory receptors. (**A**) Three novel TFs (*Zscan20*, *Jdp2*, and *Nfil3*) and one previously reported TF (*Prdm1*) to validate *in vivo*. *Zscan20* and *Jdp2* are highly specific to T cells in exhaustion trajectory and have low activity in T_{RM}. *Nfil3* and *Prdm1* are multitaskers that have high TF activity in both Tex_{Term} and T_{RM}. (**B**) Module score of Tex_{Term} – TF regulon expression is calculated from scRNAseq of scramble gRNA or TF gRNA RV transduced cells from LCMV day 20. (**C**) Flow plot of PD1 positive P14 cells with gRNA. CD8+ Tex subsets are defined below. Tex_{Prog}: PD1+ CX3CR1- SLAMF6+, $Tex_{Eff\text{-}like}$: PD1+ CX3CR1+, Tex_{Term}: PD1+ CX3CR1- SLAMF6-. (**D**) Percentage of PD1 negative/Tex_{Prog}/$Tex_{Eff\text{-}like}$/Tex_{Term} of gRNA marker positive donor P14 CD8 T cells. Data include more than three biological replicates, $n \geq 9$. ****$P < 0.0001$, ***$P < 0.001$, **$P < 0.01$, and *$P < 0.05$.

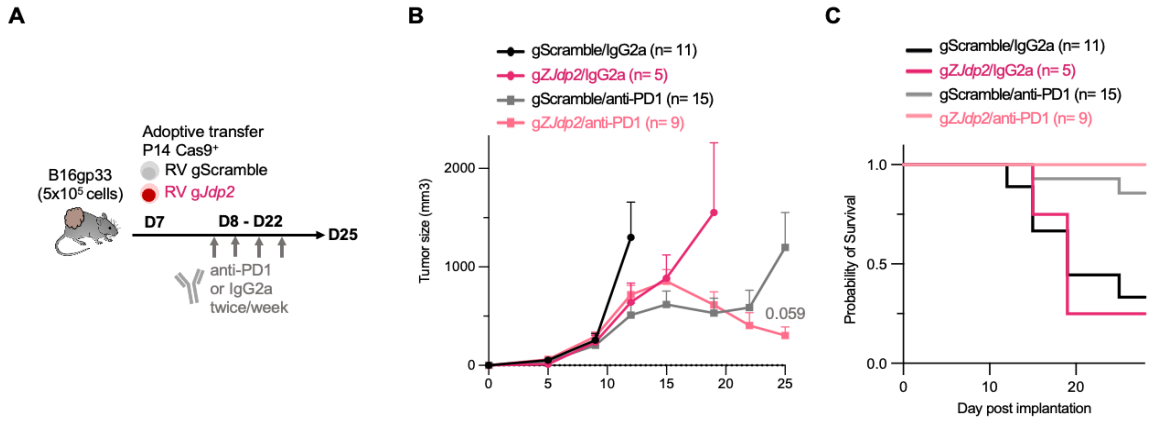


Fig. S12. Tumor control by T cells with *Jdp2* deficiency.

(**A**) *Jdp2* KO or control P14 cells were adoptively transferred to melanoma-bearing mice that are implanted with 5 X 10^5 B16-gp33 cells) followed by anti-PD1 or isotype IgG2a control treatment. (**B**) Tumor growth (**C**) survival curve and show improved tumor control only when *Jdp2* KO is combined with anti-PD1 treatment. Data are expressed as mean ± SEM. Statistical analysis was performed using Student's t test (two-tailed) comparing TF gRNA vs scramble gRNA within anti-PD1 group (gray).

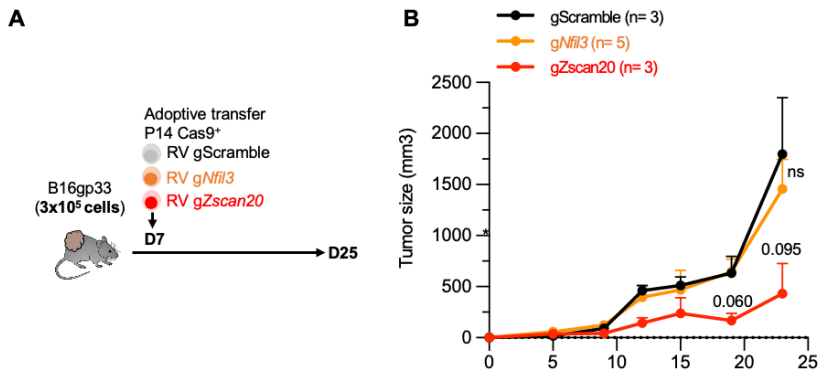


Fig. S13. Tumor control by T cells with T_{RM} and Tex_{Term} multitasker, Nfil3 deficiency.
(**A**) *Nfil3* KO, *Zscan20* KO, or control P14 cells were adoptively transferred to melanoma-bearing mice that are implanted with 3 X 10^5 B16-gp33 cells. (**B**) Tumor growth show that only mice with *Zscan20* KO T cells show tumor reduction but Nfil3 KO does not improve tumor control. Data are expressed as mean ± SEM. Statistical analysis was performed using Student's t test (two-tailed) comparing TF gRNA vs scramble gRNA.