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

Authors: H. Kay Chung[1*], Cong Liu[2*], Ming Sun[1], Eduardo Casillas[1], Timothy Chen[1], Brent Chick[3], Jun Wang[2], Shixin Ma[1], Bryan Mcdonald[1], Peixiang He[4], Qiyuan Yang[1], Siva Karthik Varanasi[1], Thomas Mann[1], Dan Chen[1], Filipe Hoffmann[1], Victoria Tripple[1], Yuqing Hang[5], Josephine Ho[3], Ukrae H. Cho[3], April Williams[5], Yingxiao Wang[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] Department of Bioengineering, Institute of Engineering in Medicine, University of California, La Jolla, CA, 92093-0435, USA

[5] 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.

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.

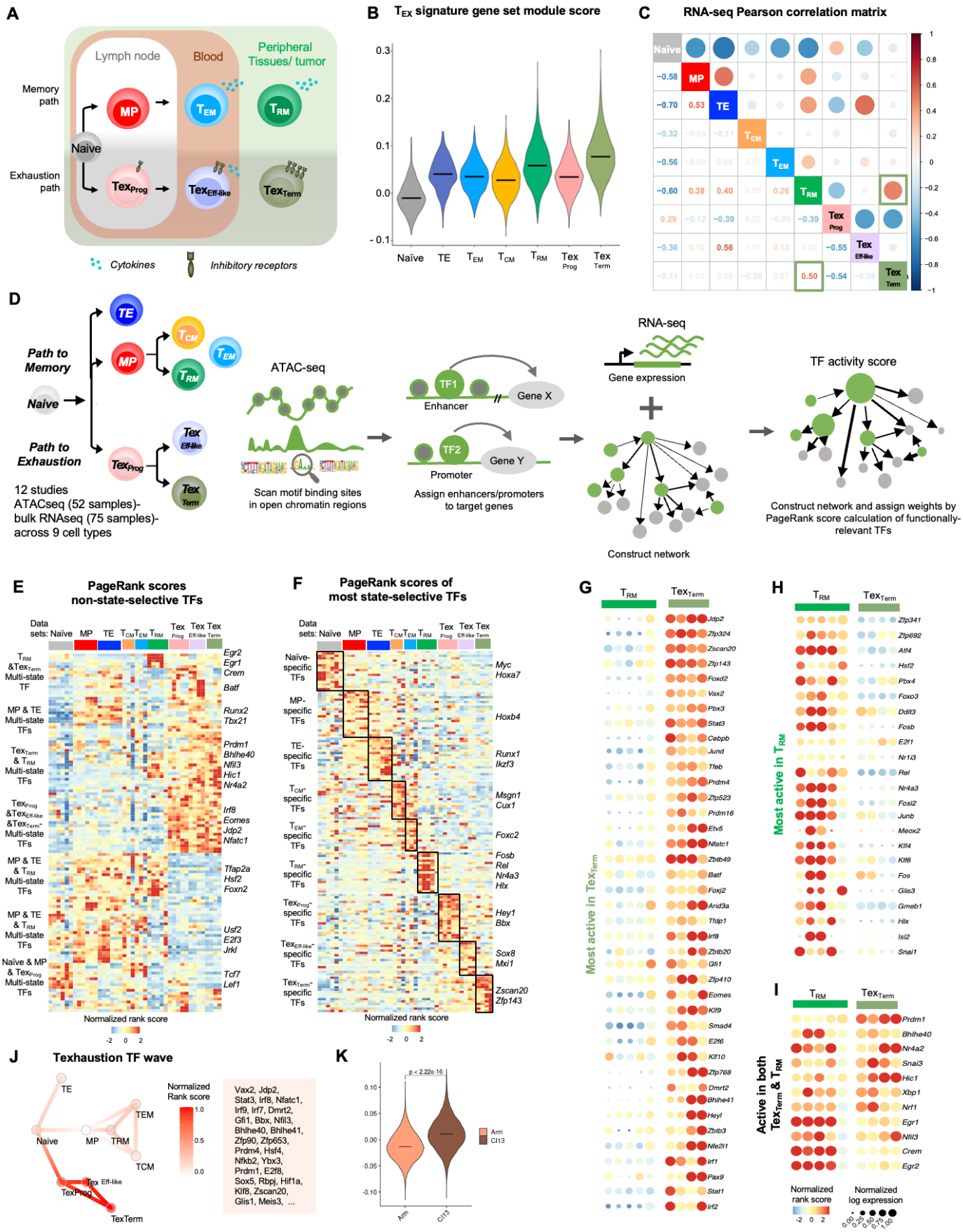
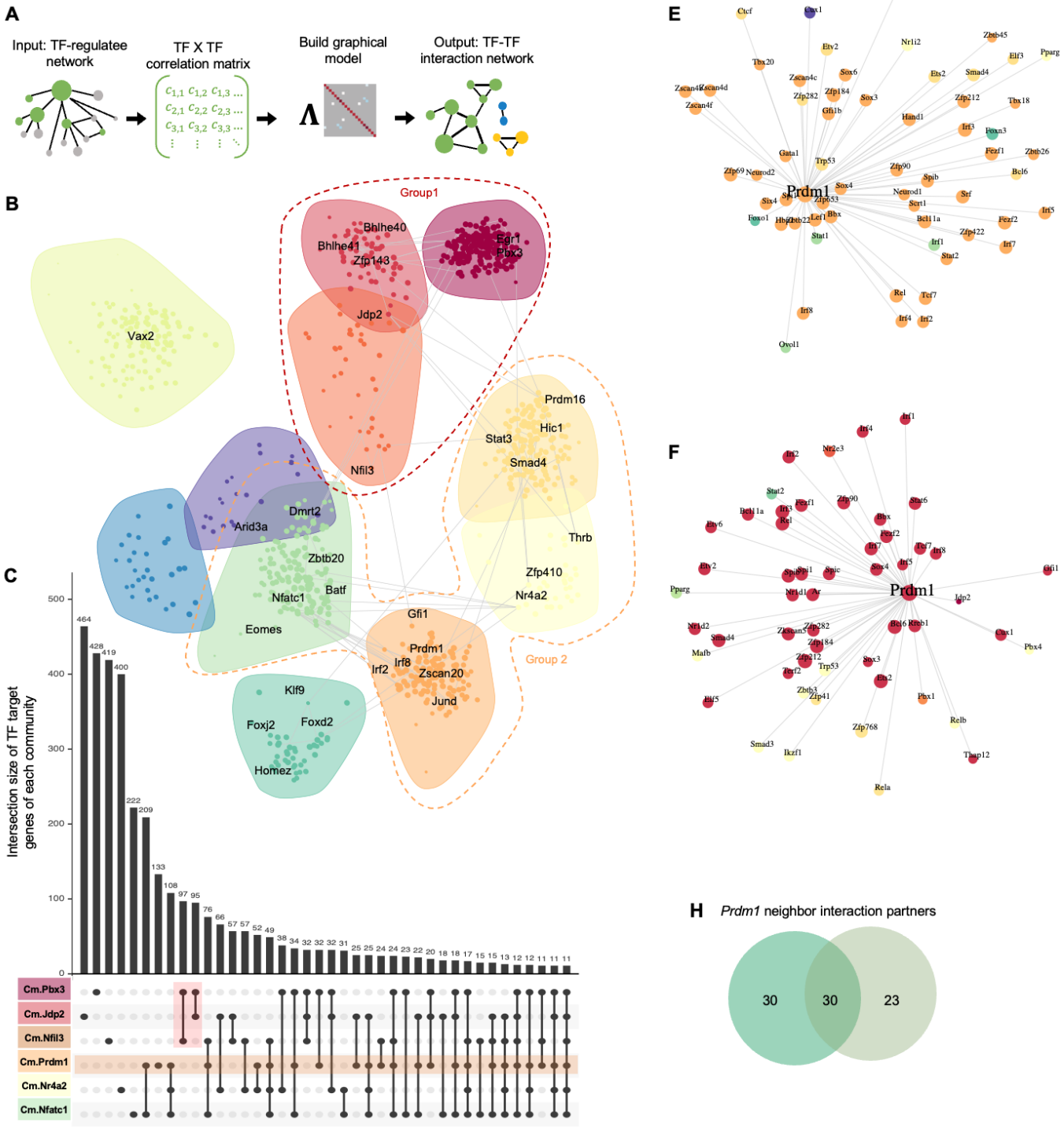

A
Lymph node
Blood
Peripheral Tissues/ tumor
Memory path
Exhaustion path
Naive
MP
TEM
TRM
TexProg
TexEff-like
TexTerm
Cytokines
Inhibitory receptors
B
TEX signature gene set module score
Naïve
TE
TEM
TCM
TRM
Tex Prog
Tex Term
C
RNA-seq Pearson correlation matrix
D
Path to Memory
Path to Exhaustion
12 studies
ATACseq (52 samples)-
bulk RNAseq (75 samples)-
across 9 cell types
ATAC-seq
Scan motif binding sites in open chromatin regions
TF1
Enhancer
Gene X
TF2
Promoter
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
Data sets: Naïve MP TE TCM TEM TRM Tex Prog Tex Eff-like Tex Term
Normalized rank score
F
PageRank scores of most state-selective TFs
Naïve-specific TFs
MP-specific TFs
TE-specific 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
Group 1	Cm. Pbx3	**Pbx3 Egr1** Hinfp Zfp523 E2f4 Tfdp1 Nrf1 E2f6 Etv5	▪ Chromatin modification ▪ Chromosome organization ▪ Chromatin organization
	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	

Common pathways of major communities

Previously reported:

- 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

New pathways:

- **Catabolic process**
- **Cellular carbohydrate metabolic process**
- **Small GTPase-mediated signal**
- **GTPase activity**
- **Response to hypoxia**
- **Response to oxidative stress**

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}.

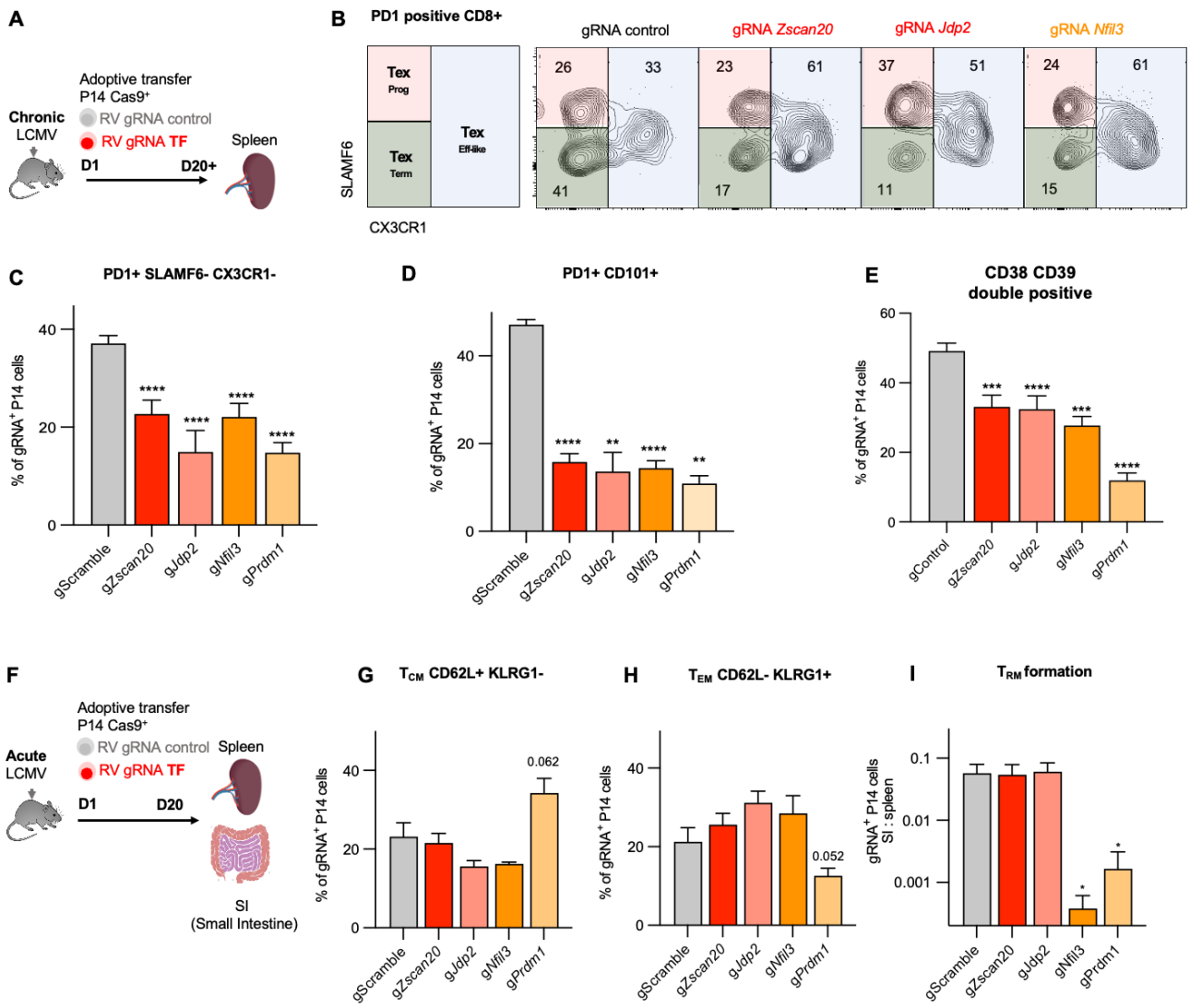


Fig 3. Disruption of Tex_{Term} single-tasker TFs prevents terminal exhaustion without perturbation of memory formation. Three novel TFs (*Zscan20*, *Jdp2*, and *Nfil3*) and one previously reported TF (*Prdm1*) to validate *in vivo*. *Zscan20* and *Jdp2* are highly specific to T cells in exhaustion trajectory 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, $n \geq 6$.

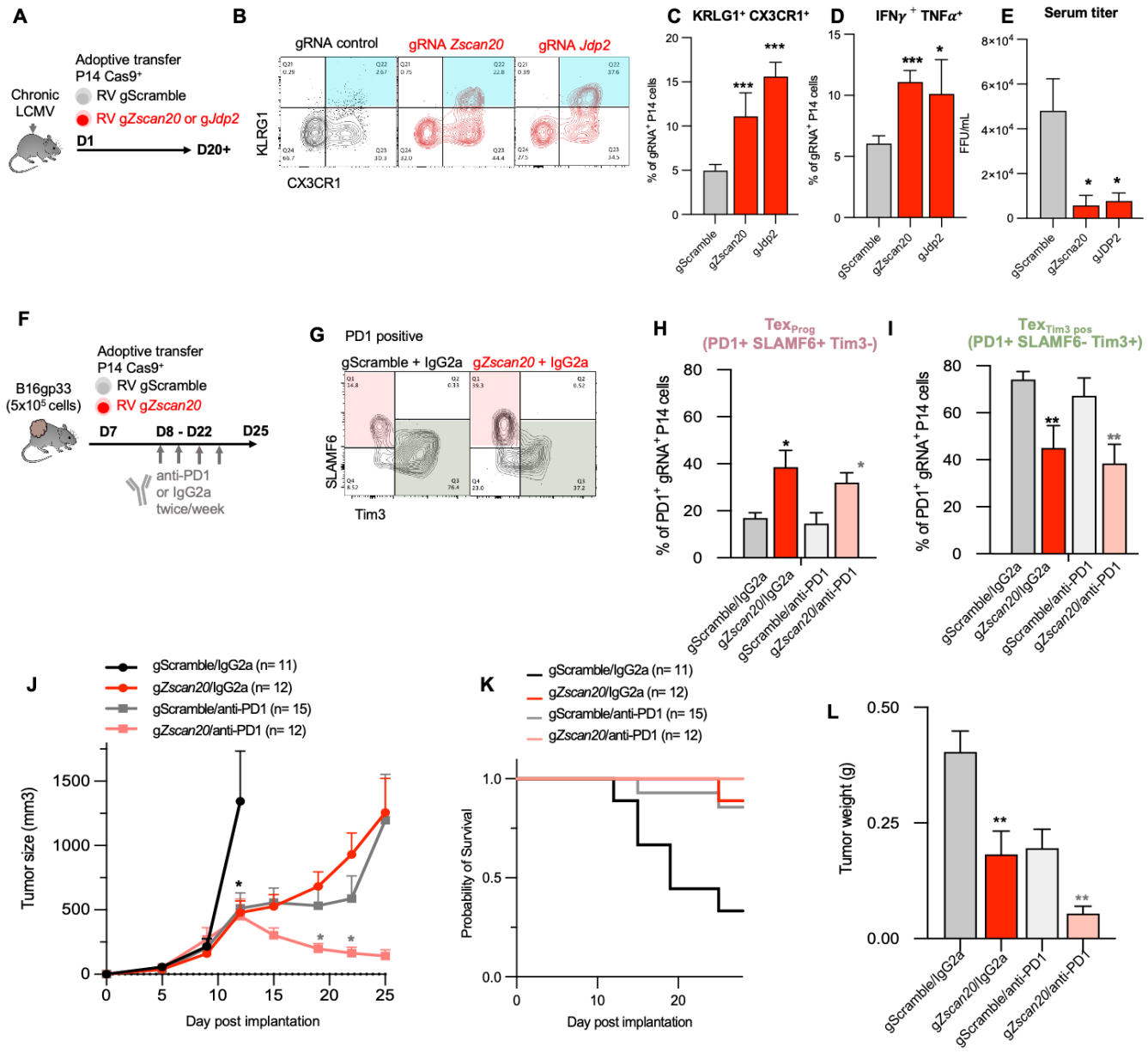


Fig.4. Disruption of transcription factor *Zscan20* and *Jdp2* provides enhanced virus and tumor control. In LCMV chronic strain-infected mice (**A-E**), (**A**) Experiment timeline. Cas9+ P14 T cells are transduced with gRNA *Zscan20*, gRNA *Jdp2*, or scramble gRNA. (**B**) g*Zscan20* and g*Jdp2* RV 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).