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

H. Kay Chung[1,2,3*#], Cong Liu[4*], Anamika Battu[2], Alexander N. Jambor[4], Brandon M. Pratt[2,9], Fucong Xie[2], Brian P. Riesenberg[2], Eduardo Casillas[1], Ming Sun[1], Elisa Landoni[2], Yanpei Li[2], Qidang Ye[2], Daniel Joo[2], Jarred Green[2], Zaid Syed[2], Nolan J. Brown[2], Mattew Smith[2], Shixin Ma[1], Shirong Tan[1], Brent Chick[5], Victoria Tripple[1], Z. Audrey Wang[2,4], Jun Wang[4], Bryan Mcdonald[1], Peixiang He[6,7], Qiyuan Yang[1], Timothy Chen[1], Siva Karthik Varanasi[1], Michael LaPorte[1], Thomas H. Mann[1], Dan Chen[1], Filipe Hoffmann[1], Josephine Ho[5], Jennifer Modliszewski[2], April Williams[8], Yusha Liu[2], Zhen Wang[11], Jieyuan Liu[11], Yiming Gao[12], Zhiting Hu[11], Ukrae H. Cho[2,3], Longwei Liu[6,7], Yingxiao Wang[6], Diana C. Hargreaves[5], Gianpietro Dotti[2,10], Barbara Savoldo[2,10], Jessica E. Thaxton[2,3], J. Justin Milner[2,10], Susan M. Kaech[1#], Wei Wang[4#]

[1]NOMIS Center for Immunobiology and Microbial Pathogenesis, Salk Institute for Biological Studies, La Jolla, CA, USA
[2]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[3]Department of Cell Biology & Physiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[4] Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA, USA.
[5] Molecular and Cell Biology laboratory, Salk Institute for Biological Studies, La Jolla, CA, USA
[6]Alfred E. Mann Department of Biomedical Engineering, University of Southern California, Los Angeles, CA, USA
[7]Department of Bioengineering, Institute of Engineering in Medicine, University of California, La Jolla, CA, USA
[8]Razavi Newman Integrative Genomics and Bioinformatics Core, Salk Institute for Biological Studies, La Jolla, CA, USA
[9]Department of Pharmacology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
[10]Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
[11]Halıcıoğlu Data Science Institute, University of California San Diego, La Jolla, CA, USA.
[12]Department of Electrical & Computer Engineering, Texas A&M University
* Equal contribution, # Co-corresponding

$CD8^+$ T cells differentiate into diverse states that shape immune outcomes in cancer and chronic infection. To systematically define the transcription factors (TFs) driving these states, we built a comprehensive atlas integrating transcriptional and epigenetic data across nine $CD8^+$ T cell states and inferred TF activity profiles. Our analysis catalogued TF activity fingerprints of each state, uncovering new regulatory mechanisms that govern differentiation of selective cell states. Leveraging this platform, we focused on two transcriptionally similar but functionally opposing states critical in tumor and viral contexts: terminally exhausted T cells (TEX_{term}), which are dysfunctional, and tissue-resident memory T cells (T_{RM}), which are protective. Global TF community analysis revealed distinct state-specific biological pathways and TF-driven networks that govern protective versus dysfunctional T cell states. Through *in vivo* CRISPR screening integrated with single-cell RNA sequencing (*in vivo* Perturb-seq), we delineated TF-driven networks that govern protective versus dysfunctional T cell states. We identified HIC1 and GFI1 as shared regulators of TEX_{term} and T_{RM} differentiation and KLF6 as a unique regulator of T_{RM}. Importantly, we discovered novel TEX_{term} single-state TFs, including ZSCAN20 and JDP2 with no prior known function in T cells. Targeted deletion of these TFs enhanced tumor control and synergized with immune checkpoint blockade. Consistently, their depletion in human T cells reduces the expression of inhibitory receptors and improves effector function. By decoupling exhaustion-selective from protective T_{RM} programs, our platform enables more precise engineering of T cell states, advancing rational design of effective immunotherapies.

Introduction

Cell states can be described as the range of cellular phenotypes arising from a defined cell type's interaction with its environment. Within the immune system, T cells possess multiple differentiation states, particularly as naive T cells differentiate into diverse states with different functionality and trafficking patterns in various immune environments, such as tumors and virus infections[1–4]. As transcription factors (TFs) govern cell state differentiation[5,6], understanding the mechanisms by which TFs influence cell states is vital for programming beneficial cell states for therapeutic applications[7,8]. One promising clinical application of cell state engineering is the enhancement of $CD8^+$ T cell-based therapies, including the adoptive transfer of tumor-infiltrating lymphocytes (TILs) or chimeric antigen receptor-T (CAR-T) cells. However, identifying the crucial TFs that control different $CD8^+$ T cell states is challenging due to the significant heterogeneity among these states and the similarities in their transcriptomes and gene regulatory networks, even when these cell states are functionally distinct.

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

77 across different CD8$^+$ T cell states and differentiation trajectories, a sophisticated and precise
78 bioinformatics approach is crucial to pinpoint the bona fide cell-state-specifying TFs essential for T cell
79 programming.
80
81 We hypothesized that the challenges of identifying key TFs in control of selective CD8$^+$ T cell
82 differentiation states can be effectively addressed through a systematic comparison of TF activity across
83 the T cell differentiation landscape. Two considerations were postulated for accurate TF activity
84 prediction: first, TF activity is not always proportional to the TF expression level, as it is influenced by
85 post-translational modifications, co-factors, and target accessibility[32]. Second, the effect of a TF on
86 target genes propagates over the genetic network, and thus, the global impact of the TF activity needs
87 to be assessed at the systems level. To address these considerations, a multi-omics atlas-assisted platform
88 was developed, integrating transcriptomic and chromatin data of nine major CD8$^+$ T cell states. This
89 platform analyzes transcriptional networks and genome-wide, "global" influencers in each cell state and
90 identifies cell-state "*selective*" or "*shared*" TFs. Our platform also enables integrative analysis of TF
91 communities and TF target genes (referred to as "regulatees"), providing insight into how the milieu of
92 TFs in each cell state collectively guides specific differentiation states.
93
94 **Taiji-based multiomics atlas identifies cell-state–selective and multi-state TFs in CD8$^+$ T cells**
95 Elucidating the TFs and their regulatory networks that govern each cell state is crucial for accurately
96 programming specific cellular states of interest. Our initial objective was to create a comprehensive
97 catalogue of TF activity across diverse CD8$^+$ T cell states by integrating our TF activity analysis
98 pipeline, Taiji[33–35], with comparative statistical analysis. In Taiji, the gene regulatory network (GRN) is
99 a weighted, directed network that models regulatory interactions between TFs and their target genes. In
100 this GRN, each node corresponds to a gene, and its weight is proportional to the gene's expression level.
101 Each edge represents a regulatory interaction and is weighted based on a combination of factors: the
102 TF's predicted binding affinity to the target gene, the chromatin accessibility at the target gene's locus,
103 and the expression levels of both the TF and the target gene (Fig. 1c)[33,34]. To determine the global
104 influence of each TF within the network, Taiji applies a personalized PageRank algorithm, which assigns
105 an "importance" score to each node based on both the quantity and quality of incoming connections.
106 This approach yields a measure of TF activity that reflects the influence of each TF in the broader

107 regulatory landscape, accounting for upstream regulators, downstream targets, and feedback loops
108 through iterative computation.
109
110 With Taiji, we have identified TFs involved in pan-immune lineage commitment, including NK cells,
111 dendritic cells, B cells, and γδ T cells[35]. While earlier studies provided foundational insights into cell
112 differentiation, a more refined analysis is needed to achieve higher resolution to contextualize TF roles
113 within $CD8^+$ T cell type. Therefore, leveraging the improved statistical filtering in our updated version
114 of Taiji, we aimed to quantify the global influence of TFs across all $CD8^+$ T cell states.
115
116 To begin, we analyzed ATAC-seq and RNA-seq datasets from 121 $CD8^+$ T cell samples spanning nine
117 distinct states, using both previously published and newly generated datasets from well-characterized
118 acute and chronic lymphocytic choriomeningitis virus (LCMV) infections[3,9,20,27,36–40] (Extended Data
119 Fig. 2; Extended Data Table 1). In acute LCMV-Armstrong infections, $CD8^+$ T cells differentiate into
120 memory precursor (MP), terminal effector (TE), effector memory (T_{EM}), central memory (T_{CM}), and
121 T_{RM} states. In chronic LCMV-Clone 13 infections, they adopt heterogeneous exhaustion cell states,
122 including progenitors of exhaustion (TEX_{prog}), effector-like exhaustion (TEX_{eff}), and TEX_{term}.
123
124 Next, we conducted an unbiased comparative analysis using statistical filtering to understand the
125 specificity of TF activity across the nine major $CD8^+$ T cell states (Extended Data Fig. 3a, Extended
126 Data Table 2). This identified a total of 255 TFs of which 136 TFs were classified as predominantly
127 '**single-state**' TFs, with each cell state selectively containing ~12-19 unique TFs (Fig. 1d, Extended
128 Data Fig. 3b). This category included many novel TFs such as *Hoxa7* in Naïve, *Snai1* in T_{RM}, *Hey1* in
129 TEX_{prog}, *Sox8* in TEX_{eff}, and *Zscan20* in TEX_{term} cells. In contrast, 119 TFs were classified as key
130 regulators in more than one cell state, termed '**multi-state'** TFs (Fig. 1e), such as *Tcf7* and *Tbx21*. *Tcf7*
131 is a known driver of Naive-, MP-, and TEX_{prog} states, all of which are multipotent and have high
132 proliferative capacity[3,20]. Multi-state TFs like *Vax2*, *Batf*, *Irf8,* and *Stat1* were more enriched within the
133 set of exhaustion-associated cell states (TEX_{prog}, TEX_{eff}, and TEX_{term}). This analysis of cell-state
134 selectivity also shows that T_{RM} shares the most TFs with TEX_{term} compared to other cell states (e.g.,
135 *Egr2*, *Crem*, *Prdm1*, Extended Data Fig. 3c) in agreement with their high degree of transcriptional and
136 epigenetic similarity (Fig. 1b; Extended Data Fig. 1b-d).
137

138 While Taiji provides a statistically grounded approach for inferring TF activity, there is no absolute
139 threshold for defining cell state specificity, and some misclassification is expected, particularly for TFs
140 with overlapping functions or modest differences in activity. Still, Taiji is useful to highlight TFs with
141 activity patterns enriched in a given cell state. For instance, although *Eomes* is classified as TEX_{term}
142 single-state TF here, it also functions in effector, T_{EM}, T_{CM}, and T_{RM} differentiation[41–43], illustrating that
143 more accurate classifications require further investigation and resolution, as performed herein for several
144 TFs.
145
146 **Comparative analysis of TFs in TEX_{term} and T_{RM} identifies candidate TFs for state-selective T cell**
147 **programming**
148 Despite the strong transcriptional overlap between TEX_{term} and T_{RM} cells, our Taiji pipeline possesses
149 the predictive power to generate a catalogue of TFs selectively active in either of these two cell states.
150 This may aid in developing better antitumor or antiviral immunotherapies wherein one can engineer T
151 cells away from T cell exhaustion and toward more functional effector cell states without negatively
152 impacting their ability to form into T_{RM} cells in tissues and tumors. Based on statistical criteria (Extended
153 Data Fig. 3a), we identified 20 and 34 TFs as single-state TFs of T_{RM} or TEX_{term} cells, respectively, and
154 30 multi-state TFs active in both cell states (Fig. 1f–h; Extended Data Fig. 3a, blue boxes; Extended
155 Data Table 3). TEX_{term} single-state TFs included many previously unreported TFs such as *Zscan20*,
156 *Jdp2*, *Zfp324*, *Zfp143*, *Zbtb49*, and *Arid3a* (Fig. 1f). T_{RM} single-state TFs included *Fosb*, *Zfp692*, *Atf4*,
157 *Pbx4, Junb, and Klf6* (Fig. 1g). Of the TEX_{term} and T_{RM} multi-state TFs, some were well-known to
158 function in the development of *both* cell states, such as *Nr4a2*[12], *Bhlhe40*[28], *Prdm1*[27,39], whereas others
159 such as *Hic1*[44] and *Gfi1*[45] were not, identifying them as new multi-state TFs to consider (Fig. 1h). We
160 also analyzed previously reported TFs such as cJUN, BATF/BATF3, and TFAP4 that were identified
161 from functional screening of $CD8^+$ T cells[43–48] based on limited phenotypic readouts. These prior screens
162 tended to identify broadly active, multi-state TFs (see Fig. 1). In contrast, our platform enabled a
163 computationally guided, multi-state screen that revealed TFs predicted to have greater state-selective
164 activity (Extended Data Fig. 3d).
165
166 To evaluate the predicted TFs governing selective T cell differentiation pathways, we identified dynamic
167 activity patterns of TF groups, termed "TF waves" (Extended Data Fig. 4). TF waves reveal potential
168 combinations of TFs that govern developmental trajectories coordinately. Seven TF waves associated

169 with distinctive biological pathways were identified. For example, the T_{RM} TF wave (Fig. 1i) includes
170 several members of the AP-1 family (e.g. *Atf3*, *Fosb*, *Fosl2*, and *Jun*), which aligns well with a recent
171 report of their roles in T_{RM} formation[46]. This wave was uniquely linked to the transforming growth factor
172 beta (TGFβ) response pathway (Extended Data Fig. 4e). Conversely, the TEX TF wave (including
173 TEX_{prog}, TEX_{eff}, and TEX_{term}) was characterized by a distinct set of TFs, such as *Irf8*, *Jdp2*, *Nfatc1,* and
174 *Vax2*, among others, that correlated with pathways related to PD1 and senescence (Fig. 1j, Extended
175 Data Fig. 4e).
176
177 **TF community analysis reveals TF-TF networks and biological processes selectively associated**
178 **with T_{RM} or TEX_{term} cell states**
179 To uncover transcriptional programs governing T_{RM} or TEX_{term} differentiation, we constructed TF–TF
180 association networks capturing functional relationships between TFs (Fig. 2a). Analysis of regulatee
181 based adjacency matrices revealed shared and distinct patterns of TF collaboration across the two states.
182 Single-state TFs showed strong intra-state interactions; TEX_{term} single-state TFs like ZSCAN20, JDP2,
183 ZFP324, and IRF8 formed dense intra-state subnetworks in TEX_{term} cells (Fig. 2b; Extended Data Fig.
184 5a). Similarly, T_{RM} single-state TFs such as FOSB, SNAI1, and KLF6 interacted majorly within T_{RM}
185 cells (Extended Data Fig. 5b). Multi-state TFs (e.g., HIC1, PRDM1, FLI1, and GFI1) were active in
186 both states but partnered with distinct TFs depending on the context, indicating they may operate within
187 regulatory networks unique to each cell state (Fig. 2c; Extended Data Fig. 5c).
188
189 We next grouped the TF–TF association networks into distinct "TF neighbor communities" in T_{RM} and
190 TEX_{term} cells (Extended Data Table 5), and found each community linked to specific biological
191 processes (Fig. 2d–f). While multi-state TFs shaped overall community topology, single-state TFs drove
192 unique interaction patterns specific to T_{RM} or TEX_{term} cells within each community. Pathway analysis
193 revealed that these communities regulate divergent programs in each state—for instance, T_{RM}
194 community 3 was associated with cell adhesion and TGFβ response (Fig. 2e, g, h), while TEX_{term}
195 community 3 was linked to apoptosis (Fig. 2f-h). Community 1 in T_{RM} cells controlled RNA metabolism
196 (Fig. 2e, g, h), whereas in TEX_{term} cells, this community was tied to catabolism, proteolysis, and
197 autophagy (Fig. 2f-h).
198

199 To assess the functional relevance of state-enriched pathways, we focused on the proteasome pathway,
200 which emerged as a prominent but previously unrecognized feature of TEX_{term} cells (Fig. 2g, h).
201 Proteasome gene signatures were also enriched in TEX_{term}-like $CD8^+$ T cells from NSCLC patients[47]
202 and murine MCA-205 TILs (Extended Data Fig. 6a,b). Consistently, proteasome activity—measured by
203 a validated fluorescent probe[48]—was highest in TEX_{term} cells from chronic LCMV (Fig. 2i) and in
204 tumor-specific TILs (Fig. 2j, k) relative to bystander OT-1 cells (Fig. 2k). To test if high proteasome
205 activity correlates with dysfunction, we sorted OT-1 cells by proteasome activity probe intensity and
206 adoptively transferred them into B16F10-OVA tumor-bearing mice. $Proteasome^{high}$ cells showed
207 reduced tumor control compared to $proteasome^{low}$ cells (Fig. 2l), a trend also seen in endogenous TILs
208 (Extended Data Fig. 6c). These findings help validate the TF-TF association network and pathway
209 predictions and identify the ubiquitin-proteasome pathway as a functional hallmark of TEX_{term} cells.
210

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

213 The Taiji pipeline enabled comparative analysis of TF activity and curated sets of single-state TFs
214 specific to T_{RM} *vs.* TEX_{term} cells (Fig. 1f-h). To evaluate its accuracy, Perturb-seq, which combines *in*
215 *vivo* CRISPR screening with single-cell RNA sequencing (scRNA-seq), was first conducted in separate
216 infectious environments that promote the differentiation of TEX_{term} and T_{RM} cells, namely chronic and
217 acute LCMV infection (Fig. 3a and 4a). Our Perturb-seq guide RNA (gRNA) library targeted 19 TFs,
218 including seven multi-state TFs predicted to impair both TEX_{term} and T_{RM} differentiation and 12 TEX_{term}
219 single-state TFs expected to selectively affect TEX_{term} differentiation. The TEX_{term}-TFs included one
220 known TF, *Nfatc1*, and 11 others that had high specificity scores but were not previously linked to
221 TEX_{term} differentiation (gray boxes Fig. 1f,h). The multi-state TFs included two positive controls
222 (*Nr4a2*, *Prdm1*) and five multi-state TFs (*Nfil3*, *Hic1*, *Gfi1*, *Ikzf3*, *Stat3*) not previously tested. To ensure
223 comprehensive screening, four gRNAs per target were expressed in two dual-gRNA retroviral (RV)
224 vectors (Extended Data Fig. 7a), along with two control vectors expressing scramble gRNAs
225 (gScramble). This resulted in a library of 40 dual-gRNA vectors, with 72 gRNAs targeting TFs and 4
226 gScramble controls (Extended Data Table 6).
227

228 To initiate the screening of TEX_{term} cells, $Cas9^+$ P14 $CD8^+$ T cells were transduced with this RV-gRNA
229 library and adoptively transferred into mice infected with **LCMV-Clone 13**, a model of **chronic**

230 **infection** and $CD8^+$ T cell exhaustion (*note, recipient mice also expressed Cas9 to prevent rejection of*
231 *donor cells*). Droplet-based sequencing was performed 18+ days post-transfer to assess sgRNA and
232 transcriptomes of each spleen-derived donor $Cas9^+$ P14 $CD8^+$ T (Fig. 3a), analyzing 17,257 cells with
233 unique gRNA expression.
234
235 To determine which TFs impaired TEX_{term} cell differentiation, we first employed uniform manifold
236 approximation and projection (UMAP) to visualize single-cell transcriptomes. Four primary clusters
237 were identified consisting of TEX_{prog}, TEX_{eff}, and TEX_{term} cells and those in cell cycle (Fig. 3b;
238 Extended Data Fig. 7b,c). All clusters expressed *Tox* and *Pdcd1*, key exhaustion markers, and TEX_{prog}
239 cells were identified by *Tcf7*, *Slamf6*, and *Sell* expression. TEX_{eff} cells expressed effector markers,
240 including *Cx3cr1*, *Klrd1*, *Klrk1*, *Klf2*, and *Zeb2*[2], while TEX_{term} cells exhibited high expression of
241 inhibitory receptors and well-established exhaustion markers like *Cd101*, *Cd7*, *Cd38*, *Cd39*, *Cxcr6*, and
242 *Nr4a2*. The cell cycle cluster was noted for its expression of *Birc5, Mki67*, *Stmn1*, and *Tuba1b*.
243
244 Next, we evaluated the impact of individual TF depletion by analyzing the distribution of $gRNA^+$ cells
245 across exhaustion states (Fig. 3c,d). CRISPR KO of the majority of the 19 TEX_{term} driving TFs led to a
246 reduction in TEX_{term} cell frequency. Notably, KOs of multi-state TFs such as *Hic1*, *Stat3*, *Prdm1*, and
247 *Ikzf3* (which encodes AIOLOS) significantly reduced TEX_{term} differentiation by 92% to 89%. Depletion
248 of novel TEX_{term} single-state TFs—including *Zfp324*, *Zscan20*, and *Jdp2*—significantly prevented
249 TEX_{term} differentiation by 78%, 54%, and 43%, respectively (Fig. 3d, bold). Other novel candidates
250 such as *Etv5*, *Arid3a*, *Zfp410*, *Foxd2*, and *Prdm4* also reduced TEX_{term} representation by 25–40%,
251 though some did not reach statistical significance. This Perturb-seq analysis highlights the platform's
252 ability to identify TFs that regulate the TEX_{term} state, with most tested TFs influencing exhaustion to
253 varying degrees.
254
255 To understand how TEX_{term} driving TF KOs affect $CD8^+$ T cell exhaustion, we used flow cytometry and
256 scRNA-seq to analyze TF KO cells during LCMV-Clone 13 infection (Fig. 3e-i; Extended Data Fig. 8).
257 We tested six TF KOs, including known control (*Prdm1*) and five newly identified TFs (*Zscan20*, *Jdp2*,
258 *Zfp324*, *Stat3*, *Hic1*) that impaired TEX_{term} state differentiation in Perturb-seq. Disrupting these TFs
259 reduced TEX_{term} cell ($PD1^+$ $CX3CR1^-$ $SLAMF6^-$) frequency by ~50% (Fig. 3e, f) and decreased
260 expression of inhibitory receptors such as CD101, CD39, and CD38 (Extended Data Fig. 8a,b). All 19

261 TEX_{term}-TF KOs exhibited a marked decrease in TEX_{term}-signature genes[49], including *Cd7*, *Cxcr6*,
262 *Nr4a2*, and *Entpd1*(Fig. 3g).
263
264 Lastly, the TEX_{term}-driving TF KOs were grouped based on their effects on TEX_{prog} ($PD1^+$ $CX3CR1^-$
265 $SLAMF6^+$, Fig. 3h) or TEX_{eff} ($PD1^+$ $CX3CR1^+$, Fig. 3i) state differentiation. Loss of *Prdm1* and *Stat3*
266 markedly increased the frequency of TEX_{prog} cells and upregulated TEX_{prog} signature genes (Fig. 3g, h)
267 whereas loss of *Hic1*, *Zscan20*, *Zfp324*, or *Jdp2* primarily expanded the TEX_{eff} cell population and
268 effector signature genes (Fig. 3g, i, j). Deletion of the *Zscan20* and *Jdp2* significantly enhanced effector
269 cytokine production (e.g., IFNγ and TNFα) and reduced viral loads in recipient mice (Fig. 3k, l).
270
271 **Depletion of TEX_{term} single-state TFs does not disrupt T_{RM} cell differentiation**
272 A major goal of this work was to identify TFs that selectively repress TEX_{term} cell differentiation without
273 affecting T_{RM} differentiation, thereby enabling more precise programming of $CD8^+$ T cell states. Since
274 nearly all the predicted TEX_{term} single-state TFs impaired TEX_{term} differentiation to some degree (Fig.
275 3), the next step was to evaluate their effects on T_{RM} differentiation to confirm their selective activity.
276 We used the same Perturb-seq library as before, but this time included only the eight TEX_{term} single-
277 state TFs and seven multi-state TFs that impaired TEX_{term} state development by more than 25% in
278 chronic LCMV infection (Fig. 3d). To assess their impact on memory $CD8^+$ T cell development, we
279 isolated RV-transduced $Cas9^+$ P14 $CD8^+$ T cells from the spleen and small intestine (SI) of mice 18 days
280 after **acute LCMV-Armstrong** infection. We then analyzed 15,211 cells using scRNA-seq to determine
281 how these perturbations impacted the formation of intestinal T_{RM} cells, as well as circulating splenic
282 T_{CM} and T_{EM} cells (Fig. 4a).
283
284 The UMAP analysis identified four primary clusters containing cells with features of T_{CM} (*Il7r*, *Tcf7,*
285 *Sell* and *S1pr1),* T_{EM} (*Cx3cr1*, *Klrg1,* and *Klf2)*, and T_{RM} cells (*Cd69, Cd160, and Itgae* (encoding
286 CD103)) as well as a small T_{RM} cell population with lower *Itgae*, but higher *Ifng* and *Irf1* expression
287 designated T_{RM} $Itgae^{low}$ (Fig. 4b,c; Extended Data Fig. 9a,b)[27,50]. Examination of the $gRNA^+$ cells
288 revealed that none of the eight TEX_{term} single-state TF KOs (*Zfp324*, *Irf8*, *Zfp410*, *Nfatc1*, *Zscan20*,
289 *Jdp2*, *Arid3a*, and *Etv5)* negatively affected T_{RM} formation significantly (Fig. 4d,e, **bold** gene symbols).
290 In fact, KO of *Etv5* tended to increase the frequency of T_{RM} cells. To evaluate the specificity of the
291 TEX_{term} single-state TFs, we also examined the expression of the T_{RM} gene signatures[9,39] in the entire

292 population of gRNA$^+$ cells for each TF tested (Fig. 4f). With the exception of *Etv5* and *Arid3a*, which
293 increased T_{RM}-signature gene expression (Fig. 4f, Extended Data Fig. 9c), perturbation of the TEX_{term}
294 single-state TFs did not substantially alter T_{RM}-signature gene expression. The platform also predicted
295 novel multi-state TFs, including *Hic1* and *Gfi1*. Interestingly, disruption of these multi-state TFs
296 significantly reduced T_{RM} cell frequency (Fig. 4e) and T_{RM}-signature gene expression (Fig. 4f, Extended
297 Data Fig. 9c), mirroring the effects of *Prdm1*, a known multi-state TF for T_{RM} and TEX_{term} cells.
298
299 To further validate the Perturb-seq data, we individually depleted the TEX_{term} single-state TFs *Zscan20*
300 and *Jdp2* and multi-state TF *Prdm1* in Cas9$^+$ P14 CD8$^+$ T cells, adoptively transferred them into LCMV-
301 Armstrong infected animals, and assessed their differentiation into T_{CM}, T_{EM}, and T_{RM} cells using flow
302 cytometry (Fig. 4g,h). Deletion of *Zscan20* and *Jdp2* did not alter the formation of any memory cell
303 subtypes, while perturbation of *Prdm1* significantly reduced T_{RM} and increased T_{CM} formation, as
304 expected. Altogether, this multi-omics pipeline predicted TEX_{term} single-state TFs that drive TEX_{term}
305 differentiation without affecting with T_{RM} cell formation and multi-state TFs that influence both cell
306 states. These results demonstrate the accuracy and predictive power of our approach for pinpointing
307 single-state and multi-state TFs.
308
309 **Overexpression of T_{RM}-selective TF, *Klf6*, improves T_{RM} differentiation without accelerating T cell**
310 **exhaustion**
311 To further demonstrate the utility of our cell-state selective TF identification pipeline in discovering
312 novel T_{RM}-associated TFs, we evaluated *Klf6*, which was identified through our Taiji analysis as a T_{RM}
313 single-state TF (Fig. 1g). We hypothesized that overexpressing *Klf6* (*Klf6*-OE) could enhance T_{RM}
314 formation during acute viral infection, but would not worsen terminal exhaustion in chronic infection.
315 Our results confirmed this hypothesis. When empty-vector control and *Klf6*-OE P14 CD8$^+$ T cells were
316 co-transferred, *Klf6*-OE cells robustly outcompeted control cells, resulting in a 15-fold enrichment in
317 the SI compared to controls (Fig. 4i). Furthermore, there were ~42 times more CD69$^+$ CD103$^+$double-
318 positive cells in *Klf6*-OE than in control donor cells, indicating *Klf6*-OE dramatically increased T_{RM}
319 development in the SI (Fig. 4j). Importantly, *Klf6*-OE did not increase terminal exhaustion during
320 chronic infection (Extended Data Fig. 10). This work not only identified KLF6 as a new T_{RM}-driving
321 TF but also confirmed its selectivity.
322

323 **Disruption of novel TEX_{term} single-state TFs enhances tumor control**

324 This platform predicted cell-state-selective TF activity and identified TEX_{term} single-state TFs as targets
325 for engineering T cells that resist exhaustion yet retain effector and memory functions—offering new
326 strategies to improve immunotherapy efficacy. Given that T_{RM} cells are associated with better clinical
327 outcomes in solid tumors[9–12,51], we hypothesized that KO of exhaustion-selective TFs like *Zscan20* could
328 be more effective than targeting T_{RM} and TEX_{term} multi-state TFs like *Hic1*. Utilizing an adoptive cell
329 therapy model, TF gRNA RV-transduced $Cas9^+$ P14 $CD8^+$ T cells were transferred into mice with
330 established melanoma tumors expressing GP33-41 (Fig. 5a). Unlike depletion of the multi-state TF
331 *Hic1*, depleting the TEX_{term} single-state TF, *Zscan20,* resulted in improved tumor control (Fig. 5b,c).
332 Moreover, *Zscan20* $gRNA^+$ cells more readily formed TEX_{prog} cells rather than $TIM3^+$ $CD39^+$ TEX_{term}
333 cells (Fig. 5d,e). To control for inter-mouse variability in antigen load, we co-transferred *Zscan20* or
334 *Hic1* KO with control P14 $CD8^+$ T cells into the same B16-GP33 tumor-bearing mice (Fig. 5f–j;
335 Extended Data Fig. 11). Both KOs significantly increased the frequency of $PD1^+$ $SLAMF6^+$ $TIM3^-$ and
336 decreased the frequency of $TIM3^+$ exhausted cells and TEX_{term} cell state ($PD1^+$ $SLAMF6^-$ $CX3CR1^-$)
337 compared to controls (Extended Data Fig. 11), consistent with their predicted activity in TEX_{term} cells
338 (Fig. 1f, h). However, *Zscan20* KO robustly enhanced effector marker expression (CX3CR1), granzyme
339 B and cytokine production in TILs, whereas *Hic1* KO did not appear to improve effector function to the
340 same degree (Fig. 5g-j). Thus, despite their similar effects on suppressing TEX_{term} cell differentiation in
341 tumors, differences in their ability to promote functional effector-like states may underlie the differential
342 tumor control observed (Fig. 5b,c). Given that *Hic1* functions as a multi-state TF while *Zscan20* as a
343 single-state TF, these findings support the general rationale for targeting state-specific TFs to enable
344 more selective programming of T cell differentiation.

345

346 **Cross-species conservation of cell state–selectivity TFs in human tumors**

347 To evaluate the relevance of our murine findings in human T cells—particularly for applications in
348 immunotherapy—we conducted cross-species validation using publicly available single-cell multi-
349 omics and scRNA-seq datasets from human tumor-infiltrating $CD8^+$ T cells (Fig. 6a; Extended Data Fig.
350 12a; Extended Data Table 1). Leveraging the Taiji TF analysis platform, we mapped murine TEX_{term}-
351 and T_{RM}-associated TFs onto a curated human pan-cancer $CD8^+$ T cell atlas encompassing six tumor
352 types[52–59](GBM, HNSCC, BCC, HCC, RCC, and ccRCC). Human pan-cancer $CD8^+$ T cells were
353 clustered into heterogeneous cell states, including T_{RM} and TEX_{term} clusters (Fig. 6a; Extended Data Fig.

354 12b). This Taiji analysis across tumor types revealed strong cross-species conservation: TEX_{term} single-
355 state TFs such as *JDP2*, *ZNF410*, and *FOXD2* consistently exhibited higher activity in TEX_{term} clusters
356 compared to T_{RM}-like cells (Fig. 6b). Among the 34 murine TEX_{term}-specific TFs, 19 showed conserved
357 activity patterns in human TEX_{term} cells. Similarly, T_{RM}-specific TFs—such as *NR4A1*, *KLF6*, and
358 *FOSB*—displayed enriched activity in the human T_{RM} cluster (Extended Data Fig. 12c). Furthermore,
359 22 of the 30 murine TFs that were active in both TEX_{term} and T_{RM} states showed enriched activity profiles
360 in both cell states in human datasets (Extended Data Fig. 12d). Although a few TFs—such as
361 ZSCAN20—could not be assessed in the Taiji analysis due to missing DNA-binding motifs, we
362 complemented our analysis with comparative RNA expression profiling across 15 tumor types (Fig.
363 6a,c). Notably, 24 of 34 murine TEX_{term} single-state TFs, including *ZSCAN20* and *JDP2*, showed
364 elevated expression in human TEX cells, further supporting their functional relevance across species.
365
366 Given the above correlations between mice and human T cells, we perturbed *ZSCAN20* and *JDP2* to
367 assess the relevance of TEX_{term} single-state TFs in human T cells (Extended Data Fig. 13). Following
368 repeated CD3/CD28 stimulation over 18 days to simulate chronic activation (Fig. 6d), *ZSCAN20*- or
369 *JDP2*-deficient $CD8^+$ T cells exhibited increased expression of CCR7 (naïve/stem cell memory/T_{CM}
370 marker) and decreased levels of inhibitory receptors, including LAG3, PD1, and TIM3 (Fig. 6e, f). These
371 KO cells also produced higher levels of effector cytokines and Granzyme B (Fig. 6g, h), indicating that
372 ZSCAN20 and JDP2 contributed to the regulation of exhaustion-associated features in human $CD8^+$ T
373 cells.
374
375 **Synergistic tumor control by combining ICB therapy with *Zscan20* and *Jdp2* perturbation**
376 **highlights clinical relevance**
377 Tumors with high TEX_{term} cell infiltration often exhibit poor responses to immune checkpoint blockade
378 therapy[60]. We hypothesized that targeting TEX_{term} single-state TFs could enhance ICB efficacy. Among
379 the TEX_{term}-associated TFs, *Zscan20* and *Jdp2* were prioritized for their conservation and functional
380 relevance in human T cells (Fig. 6b-h). To test synergy with ICB, treatment began one day after adoptive
381 transfer of TF-depleted P14 $CD8^+$ T cells (Fig. 6i). The combination of *Zscan20* or *Jdp2* KO with anti-
382 PD-1 therapy significantly reduced tumor burden (Fig. 6j–m) and improved survival (Extended Data
383 Fig. 14). These findings suggest that selectively disrupting TEX_{term} single-state TF represents a
384 promising strategy to enhance T cell therapy by minimizing dysfunctional states while preserving

385 beneficial T cell phenotypes. Overall, our cross-species multi-omics and functional perturbation
386 approach underscores the translational potential of Taiji-identified TFs for improving adoptive T cell
387 therapies.
388
389 **Discussion**
390 Our study introduces a powerful platform for identifying TFs that are pivotal in guiding the
391 differentiation of specific $CD8^+$ T cell states during viral infections and tumor progression. This platform
392 has transformative potential for engineering desirable cell states or preventing detrimental ones, offering
393 a powerful strategy to enhance T cell–based therapies, including adoptive cell therapy. By constructing
394 a comprehensive transcriptional and epigenetic atlas from 121 experiments spanning nine distinct $CD8^+$
395 T cell states, we developed a detailed map of TF activity. This allowed us to predict TF activities within
396 each state, essentially creating a unique TF fingerprint for each cellular context. Additionally, to
397 facilitate intuitive and interactive exploration of TF activity and networks analyzed from our multi-
398 omics atlas, we developed an integrated conversational interface, called TaijiChat–a web-based chatbot
399 platform enabling natural language querying of both the study's datasets and related literature (Extended
400 Data Note).

401 Focusing on two critical cell states TEX_{term} (dysfunctional terminally exhausted) and T_{RM} (functional,
402 tissue-resident memory) T cells, we examined similarities and differences of TF activity and their
403 networks in both states and engineered T cells to resist exhaustion while retaining functionality of T_{RM}
404 cells. Using *in vivo* Perturb-seq, we validated TF activity for TEX_{term} and T_{RM} cells in both acute and
405 chronic infection models. While recent CRISPR screenings in $CD8^+$ T cells have identified important
406 TFs in cytotoxicity, memory formation[61–64], cell enrichment[65], and exhaustion[66], a systematic and
407 context-dependent understanding of TF roles across multiple contexts has been lacking. Our study
408 addresses this gap by generating an accurate catalogue of $CD8^+$ T cell state-defining TFs, enabling cost-
409 effective validation of predicted TF activity and selectivity via Perturb-seq. Furthermore, our study
410 offers broader and new insight into context-dependent TF regulation. Previously, differential TF
411 cooperation in different contexts was reported[30,64,67–69]. We extend this by analyzing global TF
412 associations across cell states, revealing how TF communities regulate T cell-specific pathways,
413 including protein catabolism in T cell exhaustion, which aligns with prior research on protein
414 homeostasis[48,70–72]. These TF networks reveal how various cellular processes are differentially

415 controlled between T_{RM} and TEX_{term} cells, providing rationale for their different functional capabilities
416 within tissues.

417 One of the key outcomes of this study was the identification of novel transcription factors with state-
418 specific functions, including ZSCAN20 and JDP2 as TEX_{term} single-state TFs, KLF6 as a T_{RM} single-
419 state TF, and newly uncovered roles for multi-state TFs such as HIC1 and GFI1. Perturbing TEX_{term}
420 single-state TFs not only prevented T cell exhaustion but also preserved the ability of these cells to
421 differentiate into effector and memory states. This led to significant improvements in tumor control.

422 To evaluate the clinical importance of the newly discovered TFs and the catalog of TFs with TEX_{term}
423 and T_{RM} selectivity, we confirmed cross-species conservation of a significant number of TFs using Taiji
424 analysis of a human pan-cancer multi-omics atlas, along with comparative expression analysis across
425 pan-cancer single-cell RNA-seq datasets. Furthermore, we demonstrated enhanced human T cell
426 function following perturbation of the TEX_{term} single-state TFs *ZSCAN20* and *JDP2*. Depletion of these
427 two TFs demonstrates synergistic effects with ICB therapy, resulting in significant tumor regression.
428 These findings highlight a promising strategy for enhancing antitumor immunity through precise cell-
429 state programming.
430

431 Our TF analysis platform addresses key challenges in T cell engineering by enabling the prevention of
432 exhaustion while preserving the beneficial properties of T_{RM} cells in solid tumors and chronic infection.
433 This framework offers a rational approach to T cell programming through optimized engineering of 'TF
434 recipes' that control cell-state differentiation with more precision, robustness and durability. For
435 example, future strategies may involve testing combinatorial TF recipes that combine over expression
436 of T_{RM}-promoting TFs (e.g. *KLF6*) with depletion of TEX_{term} single-state TFs. One could compare this
437 to depletion of multi-state TFs to further test the paradigm presented here. Such combinatorial TF
438 recipes may be further refined or expanded using AI-based models to predict modified cell states, such
439 as scLong[73], ultimately advancing the design of durable and effective T cell therapies.
440

441 In summary, while our study primarily focuses on $CD8^+$ TEX_{term} and T_{RM} cell differentiation to identify
442 potential strategies for boosting antitumor immunity and combating chronic infections, the pipeline we
443 developed for identifying single-state TFs and 'TF recipes' can be adapted to program other cell types,
444 ultimately improving the effectiveness of cell therapy across a range of applications.

445 **References**

446 1. Chung, H. K., McDonald, B. & Kaech, S. M. The architectural design of CD8+ T cell responses in acute and chronic
447 infection: Parallel structures with divergent fates. *J. Exp. Med.* **218**, (2021).
448 2. Kasmani, M. Y. *et al.* Clonal lineage tracing reveals mechanisms skewing CD8+ T cell fate decisions in chronic
449 infection. *J. Exp. Med.* **220**, (2023).
450 3. Beltra, J.-C. *et al.* Developmental Relationships of Four Exhausted CD8 T Cell Subsets Reveals Underlying
451 Transcriptional and Epigenetic Landscape Control Mechanisms. *Immunity* **52**, 825–841.e8 (2020).
452 4. Giles, J. R. *et al.* Shared and distinct biological circuits in effector, memory and exhausted CD8 T cells revealed by
453 temporal single-cell transcriptomics and epigenetics. *Nature Immunology* vol. 23 1600–1613 Preprint at
454 https://doi.org/10.1038/s41590-022-01338-4 (2022).
455 5. Cui, W., Liu, Y., Weinstein, J. S., Craft, J. & Kaech, S. M. An interleukin-21-interleukin-10-STAT3 pathway is
456 critical for functional maturation of memory CD8+ T cells. *Immunity* **35**, 792–805 (2011).
457 6. Kaech, S. M. & Cui, W. Transcriptional control of effector and memory CD8+ T cell differentiation. *Nat. Rev.*
458 *Immunol.* **12**, 749–761 (2012).
459 7. Flynn, J. K. & Gorry, P. R. Stem memory T cells (TSCM)-their role in cancer and HIV immunotherapies. *Clin Transl*
460 *Immunology* **3**, e20 (2014).
461 8. Zhang, Z. *et al.* T Cell Dysfunction and Exhaustion in Cancer. *Front Cell Dev Biol* **8**, 17 (2020).
462 9. Milner, J. J. *et al.* Runx3 programs CD8 T cell residency in non-lymphoid tissues and tumours. *Nature* **552**, 253–257
463 (2017).
464 10. Ganesan, A.-P. *et al.* Tissue-resident memory features are linked to the magnitude of cytotoxic T cell responses in
465 human lung cancer. *Nat. Immunol.* **18**, 940–950 (2017).
466 11. Djenidi, F. *et al.* CD8 CD103 Tumor–Infiltrating Lymphocytes Are Tumor-Specific Tissue-Resident Memory T Cells
467 and a Prognostic Factor for Survival in Lung Cancer Patients. *The Journal of Immunology* vol. 194 3475–3486
468 Preprint at https://doi.org/10.4049/jimmunol.1402711 (2015).
469 12. Corgnac, S., Boutet, M., Kfoury, M., Naltet, C. & Mami-Chouaib, F. The Emerging Role of CD8+ Tissue Resident
470 Memory T (TRM) Cells in Antitumor Immunity: A Unique Functional Contribution of the CD103 Integrin. *Front.*
471 *Immunol.* **9**, 1904 (2018).
472 13. Komdeur, F. L. *et al.* CD103+ tumor-infiltrating lymphocytes are tumor-reactive intraepithelial CD8+ T cells
473 associated with prognostic benefit and therapy response in cervical cancer. *Oncoimmunology* **6**, e1338230 (2017).
474 14. Pauken, K. E. *et al.* Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade.
475 *Science* **354**, 1160–1165 (2016).
476 15. Sen, D. R. *et al.* The epigenetic landscape of T cell exhaustion. *Science* vol. 354 1165–1169 Preprint at
477 https://doi.org/10.1126/science.aae0491 (2016).
478 16. Wherry, E. J., Blattman, J. N., Murali-Krishna, K., van der Most, R. & Ahmed, R. Viral persistence alters CD8 T-cell
479 immunodominance and tissue distribution and results in distinct stages of functional impairment. *J. Virol.* **77**, 4911–
480 4927 (2003).
481 17. Wherry, E. J. *et al.* Molecular signature of CD8+ T cell exhaustion during chronic viral infection. *Immunity* **27**, 670–
482 684 (2007).
483 18. Im, S. J. *et al.* Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. *Nature* **537**, 417–421
484 (2016).
485 19. Miller, B. C. *et al.* Subsets of exhausted CD8 T cells differentially mediate tumor control and respond to checkpoint
486 blockade. *Nat. Immunol.* **20**, 326–336 (2019).
487 20. Hudson, W. H. *et al.* Proliferating Transitory T Cells with an Effector-like Transcriptional Signature Emerge from
488 PD-1 Stem-like CD8 T Cells during Chronic Infection. *Immunity* **51**, 1043–1058.e4 (2019).
489 21. Pu, F. *et al.* TIM-3 expression and its association with overall survival in primary osteosarcoma. *Oncol. Lett.* **18**,
490 5294–5300 (2019).
491 22. Zang, K. *et al.* TIM-3 as a Prognostic Marker and a Potential Immunotherapy Target in Human Malignant Tumors: A
492 Meta-Analysis and Bioinformatics Validation. *Front. Oncol.* **11**, 579351 (2021).
493 23. Liu, F. *et al.* CTLA-4 correlates with immune and clinical characteristics of glioma. *Cancer Cell Int.* **20**, 7 (2020).

24. Zhang, M. *et al.* Prognostic Values of CD38+CD101+PD1+CD8+ T Cells in Pancreatic Cancer. *Immunol. Invest.* (2019) doi:10.1080/08820139.2019.1566356.
25. Shin, H. *et al.* A role for the transcriptional repressor Blimp-1 in CD8(+) T cell exhaustion during chronic viral infection. *Immunity* **31**, 309–320 (2009).
26. Behr, F. M. *et al.* Blimp-1 Rather Than Hobit Drives the Formation of Tissue-Resident Memory CD8 T Cells in the Lungs. *Front. Immunol.* **10**, 400 (2019).
27. Milner, J. J. *et al.* Heterogenous Populations of Tissue-Resident CD8+ T Cells Are Generated in Response to Infection and Malignancy. *Immunity* **52**, 808–824.e7 (2020).
28. Li, C. *et al.* The Transcription Factor Bhlhe40 Programs Mitochondrial Regulation of Resident CD8 T Cell Fitness and Functionality. *Immunity* **51**, 491–507.e7 (2019).
29. Wu, J. E. *et al.* In vitro modeling of CD8 T cell exhaustion enables CRISPR screening to reveal a role for BHLHE40. *Sci Immunol* **8**, eade3369 (2023).
30. Chen, J. *et al.* NR4A transcription factors limit CAR T cell function in solid tumours. *Nature* **567**, 530–534 (2019).
31. Mackay, L. K. *et al.* The developmental pathway for CD103(+)CD8+ tissue-resident memory T cells of skin. *Nat. Immunol.* **14**, 1294–1301 (2013).
32. Schacht, T., Oswald, M., Eils, R., Eichmüller, S. B. & König, R. Estimating the activity of transcription factors by the effect on their target genes. *Bioinformatics* **30**, i401–7 (2014).
33. Zhang, K., Wang, M., Zhao, Y. & Wang, W. Taiji: System-level identification of key transcription factors reveals transcriptional waves in mouse embryonic development. *Sci Adv* **5**, eaav3262 (2019).
34. Wang, J., Liu, C., Chen, Y. & Wang, W. Taiji-reprogram: a framework to uncover cell-type specific regulators and predict cellular reprogramming cocktails. *NAR Genom Bioinform* **3**, lqab100 (2021).
35. Liu, C. *et al.* Systems-level identification of key transcription factors in immune cell specification. *PLoS Comput. Biol.* **18**, e1010116 (2022).
36. Milner, J. J. *et al.* Delineation of a molecularly distinct terminally differentiated memory CD8 T cell population. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 25667–25678 (2020).
37. Guan, T. *et al.* ZEB1, ZEB2, and the miR-200 family form a counterregulatory network to regulate CD8 T cell fates. *J Exp Med* **215**, 1153–1168 (2018).
38. Scott-Browne, J. P. *et al.* Dynamic Changes in Chromatin Accessibility Occur in CD8 T Cells Responding to Viral Infection. *Immunity* **45**, 1327–1340 (2016).
39. Mackay, L. K. *et al.* Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes. *Science* **352**, 459–463 (2016).
40. Renkema, K. R. *et al.* KLRG1 Memory CD8 T Cells Combine Properties of Short-Lived Effectors and Long-Lived Memory. *J Immunol* **205**, 1059–1069 (2020).
41. Intlekofer, A. M. *et al.* Effector and memory CD8+ T cell fate coupled by T-bet and eomesodermin. *Nat Immunol* **6**, 1236–1244 (2005).
42. Reiser, J., Sadashivaiah, K., Furusawa, A., Banerjee, A. & Singh, N. Eomesodermin driven IL-10 production in effector CD8 T cells promotes a memory phenotype. *Cell Immunol* **335**, 93–102 (2019).
43. Li, G. *et al.* T-Bet and Eomes Regulate the Balance between the Effector/Central Memory T Cells versus Memory Stem Like T Cells. *PLoS One* **8**, e67401 (2013).
44. Crowl, J. T. *et al.* Tissue-resident memory CD8 T cells possess unique transcriptional, epigenetic and functional adaptations to different tissue environments. *Nature Immunology* vol. 23 1121–1131 Preprint at https://doi.org/10.1038/s41590-022-01229-8 (2022).
45. Chandele, A. *et al.* Formation of IL-7Ralphahigh and IL-7Ralphalow CD8 T cells during infection is regulated by the opposing functions of GABPalpha and Gfi-1. *J. Immunol.* **180**, 5309–5319 (2008).
46. Smith, N. P. *et al.* Resident memory T cell development is associated with AP-1 transcription factor upregulation across anatomical niches. *bioRxiv* (2023) doi:10.1101/2023.09.29.560006.
47. Guo, X. *et al.* Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med* **24**, 978–985 (2018).
48. Riesenberg, B. P. *et al.* Stress-Mediated Attenuation of Translation Undermines T-cell Activity in Cancer. *Cancer Res.* **82**, 4386–4399 (2022).
49. Daniel, B. *et al.* Divergent clonal differentiation trajectories of T cell exhaustion. *Nat. Immunol.* **23**, 1614–1627

(2022).
50. Lin, Y. H. *et al.* Small intestine and colon tissue-resident memory CD8+ T cells exhibit molecular heterogeneity and differential dependence on Eomes. *Immunity* **56**, 207–223.e8 (2023).
51. Komdeur, F. L. *et al.* CD103 tumor-infiltrating lymphocytes are tumor-reactive intraepithelial CD8 T cells associated with prognostic benefit and therapy response in cervical cancer. *OncoImmunology* vol. 6 e1338230 Preprint at https://doi.org/10.1080/2162402x.2017.1338230 (2017).
52. Long, Z. *et al.* Single-cell multiomics analysis reveals regulatory programs in clear cell renal cell carcinoma. *Cell Discov* **8**, 68 (2022).
53. Terekhanova, N. V. *et al.* Epigenetic regulation during cancer transitions across 11 tumour types. *Nature* **623**, 432–441 (2023).
54. Riegel, D. *et al.* Integrated single-cell profiling dissects cell-state-specific enhancer landscapes of human tumor-infiltrating CD8 T cells. *Mol Cell* **83**, 622–636.e10 (2023).
55. Yost, K. E. *et al.* Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat Med* **25**, 1251–1259 (2019).
56. Ma, L. *et al.* Tumor Cell Biodiversity Drives Microenvironmental Reprogramming in Liver Cancer. *Cancer Cell* **36**, 418–430.e6 (2019).
57. Young, M. D. *et al.* Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594–599 (2018).
58. Cillo, A. R. *et al.* Immune Landscape of Viral- and Carcinogen-Driven Head and Neck Cancer. *Immunity* **52**, 183–199.e9 (2020).
59. Zheng, L. *et al.* Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science* **374**, abe6474 (2021).
60. Miller, B. C. *et al.* Author Correction: Subsets of exhausted CD8+ T cells differentially mediate tumor control and respond to checkpoint blockade. *Nat. Immunol.* **20**, 1556 (2019).
61. Cortez, J. T. *et al.* CRISPR screen in regulatory T cells reveals modulators of Foxp3. *Nature* **582**, 416–420 (2020).
62. Schmidt, R. *et al.* CRISPR activation and interference screens decode stimulation responses in primary human T cells. *Science* **375**, eabj4008 (2022).
63. Shifrut, E. *et al.* Genome-wide CRISPR Screens in Primary Human T Cells Reveal Key Regulators of Immune Function. *Cell* **175**, 1958–1971.e15 (2018).
64. McCutcheon, S. R. *et al.* Transcriptional and epigenetic regulators of human CD8+ T cell function identified through orthogonal CRISPR screens. *Nat. Genet.* **55**, 2211–2223 (2023).
65. Chen, Z. *et al.* In vivo CD8+ T cell CRISPR screening reveals control by Fli1 in infection and cancer. *Cell* **184**, 1262–1280.e22 (2021).
66. Zhou, P. *et al.* Single-cell CRISPR screens in vivo map T cell fate regulomes in cancer. *Nature* **624**, 154–163 (2023).
67. Lynn, R. C. *et al.* c-Jun overexpression in CAR T cells induces exhaustion resistance. *Nature* **576**, 293–300 (2019).
68. Seo, H. *et al.* BATF and IRF4 cooperate to counter exhaustion in tumor-infiltrating CAR T cells. *Nat. Immunol.* **22**, 983–995 (2021).
69. Rubin, A. J. *et al.* Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks. *Cell* **176**, 361–376.e17 (2019).
70. Van Der Byl, W. *et al.* The CD8+ T cell tolerance checkpoint triggers a distinct differentiation state defined by protein translation defects. *Immunity* (2024) doi:10.1016/j.immuni.2024.04.026.
71. Widjaja, C. E. *et al.* Proteasome activity regulates CD8+ T lymphocyte metabolism and fate specification. *J. Clin. Invest.* **127**, 3609–3623 (2017).
72. Chang, J. T. *et al.* Asymmetric proteasome segregation as a mechanism for unequal partitioning of the transcription factor T-bet during T lymphocyte division. *Immunity* **34**, 492–504 (2011).
73. Xie, P. *et al.* scLong: A Billion-Parameter Foundation Model for Capturing Long-Range Gene Context in Single-Cell Transcriptomics. (2024) doi:10.21203/rs.3.rs-5420259/v1.

591 **Figure Legends**

592 **Fig. 1| Transcription and epigenomic atlas of CD8$^+$ T cell differentiation states and cell state-**
593 **selective TF identification pipeline. a,** The diagram outlines several CD8$^+$ T cell state trajectories in
594 acute and chronic infection/tumor where various memory and exhaustion states are acquired. Parallel
595 differentiation of T_{RM} and TEX_{term} cells and their similarity in anatomical location. **b,** Pearson correlation
596 matrix of batch effect corrected RNA-seq. Color and size are both proportional to correlation values,
597 with red being the highest correlation. **c,** Pipeline of our integrative analysis. The matched RNA-seq and
598 ATAC-seq data[3,9,20,27,36–40] were used as input for Taiji algorithm to construct a regulatory network and
599 output PageRank score matrix representing the activity of TF. The downstream analysis included the
600 identification of single-state TFs, as well as the construction of transcriptional waves and TF networks.
601 **d-h,** TFs in rows (z- normalized), samples in columns, and the colors of the cell in the matrix indicate
602 the normalized PageRank scores. Each column represents a different data set. **d,** PageRank scores of
603 136 single-state TFs. **e,** 173 multi-state TFs. **f-h**, Bubble plots of PageRank scores and expression of
604 TFs that are selectively active in TEX_{term} state **(f),** T_{RM} state **(g),** and multi-state TFs that are active in
605 both TEX_{term} and T_{RM} cells **(h)**. TEX_{term} single-state TFs are in bold. Color represents the normalized
606 PageRank score with red displaying high values, and circle size represents the logarithm of mRNA
607 expression across 5 different data sets. TFs are ordered based on the p-value. TFs validated *in vivo* are
608 highlighted in gray. **i-j,** Transcription factor waves associated with exhaustion **(i)** or T_{RM} state **(j)**.
609 Sample numbers for each cell state and the statistics used to curate the TF catalogue are provided in
610 Extended Data Fig. 2a and Extended Data Fig. 3a, respectively.

611

612 **Fig. 2| Global analysis of TF networks in TEX_{term} and T_{RM} cell states. a,** TF network analysis in T_{RM}
613 and TEX_{term} cells encompasses both TF-TF association (blue, Fig. 2b-g). **b-c,** TF-TF association
614 network plots focus on TEX_{term} single-state TFs like ZSCAN20 **(b)** and multi-state TFs like HIC1 **(c)**
615 showing predicted TF-TF network in T_{RM} (green) or TEX_{term} (brown) cells. **d-f,** TF-TF association
616 clustering generates five TF communities in T_{RM} and TEX_{term} cells. The overall community topology is
617 shaped by shared TFs (gray) in both T_{RM} and TEX_{term} cells **(d)** while the TF-TF associations exclusive
618 in T_{RM} **(e)** and TEX_{term} **(f)** cells are shown as green and brown lines, respectively. **g,** Summary of the
619 shared and unique pathways enriched in each community in T_{RM} and TEX_{term} cells. The thickness of the
620 lines reflects p-value. (Gene sets: Extended Data Table. 8). **h**, GSEA of indicated pathways in TEX_{term}

621 vs. T_{RM} cells using batch-effect corrected LCMV bulk-RNA-seq[3,9,20,27,36–40] and human pan-cancer
622 scRNA-seq data sets[47,59,74]. **i-k,** Flow cytometry analysis of proteasome activity. TEX_{term} cells show the
623 highest proteasome activity in LCMV-Clone 13 (chronic strain) infected mice **(i),** or MCA-205 tumors
624 **(j). k,** Dual transfer of GP33-41 antigen-specific (P14) $CD8^+$ T cells and bystander (OT-1) $CD8^+$ T cells
625 in B16-GP33 tumor bearing mice. **l,** Evaluation of B16F1-OVA tumor growth by OVA antigen-specific
626 OT-1 cells with differential proteasome activity. C57BL/6 mice bearing tumors were infused with
627 $proteasome^{high}$ or $proteasome^{low}$ sorted OT-1 cells, which were pre-stimulated for 7 days with B16F1-
628 OVA tumor cells. Tumor growth was monitored every two days. Data are presented as the mean ±s.e.m.
629 Ordinary one-way ANOVA **(i-k),** and two-way ANOVA Tukey's multiple comparison test **(l)** were
630 performed. ****P<0.0001, ***P<0.001, **P<0.01, *P<0.05.
631
632 **Fig. 3| *In vivo* Perturb-seq validation of TEX_{term}-driving TFs. a**, Schematic of the *in vivo* Perturb-
633 seq strategy. Dual $Cas9^+$ and $P14^+$ TCR transgenic $CD8^+$ T cells that recognize LCMV epitope GP33-
634 41 were transduced with retroviruses (RVs) expressing the gRNA library, adoptively transferred into
635 mice previously infected one day earlier with LCMV-Clone 13, and analyzed 18-23 days later by
636 scRNA-seq. **b**, UMAP showing TEX_{prog}, TEX_{eff}, TEX_{term}, and cell cycle clusters. Feature plots and
637 heatmap of the cluster marker mRNA expression are in Extended Data Fig. 7b,c. **c**, Kernel density of
638 each TF gRNA-expressing cells. **d**, Cluster distribution of cells for each TF $gRNA^+$ cells. TEX_{term} single-
639 state TFs are indicated in bold. Data represent five pooled replicates from three independent
640 experiments. Values are shown as mean ± s.e.m. Statistical analysis was performed using two-way
641 ANOVA followed by Fisher's LSD test, with comparisons made relative to the gScramble control.
642 Statistical results for TEX_{term} and TEX_{prog} populations are shown; additional statistical comparisons are
643 provided in Extended Data Table 7. ****P < 0.0001, ***P < 0.001, **P < 0.01, *P < 0.05. **e**, Individual
644 phenotyping of the TF KOs in LCMV-Clone 13 infected mice based on flow cytometry and
645 representative flow plots. **f**, Quantification of the frequency of TEX_{term} ($PD1^+$ $SLAMF6^-$ $CX3CR1^-$)
646 subsets within P14 $CD8^+$ T cells. **g**, TEX_{term}, TEX_{prog}, and TEX_{eff} gene signature analysis in each gRNA
647 KO compared to gScramble. Gene signatures were generated from ref. [49] (Extended Data Table 8). **h**,
648 Quantification of the frequency of TEX_{prog} ($PD1^+$ $SLAMF6^+$ $CX3CR1^-$) cells. **i**, Quantification of the
649 frequency of TEX_{eff} ($PD1^+$ $CX3CR1^+$) cells. **j**, GSEA of an effector-associated gene set between each
650 TF KOs *vs*. control. **k-l**, P14 $CD8^+$ T cells lacking the indicated TFs were compared to gScramble
651 controls for IFNγ and TNFα production **(k)** or virus titers **(l)** in LCMV-Clone 13 infected mice. Statistic

652 analysis for **f, h, i, k, l**: Ordinary one-way ANOVA with Dunnett's multiple comparison test versus
653 gScramble (n ≥ 6 from ≥ 3 biological replicates). Data are presented as the mean ± s.e.m. ****P<0.0001,
654 ***P<0.001, **P<0.01, *P<0.05.

655

656 **Fig. 4| Functional validation of TFs with distinct roles in TEX_{term} and T_{RM} differentiation. a,**
657 Schematic of *in vivo* Perturb-seq screening in memory $CD8^+$ T cells during acute LCMV-Armstrong
658 infection. The donor cells were then analyzed for T_{RM} cells in the small intestine (SI) and T_{CM} and T_{EM}
659 cells in the spleen. **b,** UMAP showing T_{CM}, T_{EM}, T_{RM} and T_{RM}- *Itgae*^{low} clusters. **c,** Differential clustering
660 of cells from different tissue origins. **d,** Kernel density plot of $gRNA^+$ cells on UMAP. **e,** Cluster
661 distribution of cells with each gRNA. TEX_{term} single-state TFs are in bold letters. Data represent three
662 replicates. Values are shown as mean ± s.e.m. Statistical analysis was performed using two-way
663 ANOVA followed by Fisher's LSD test, with comparisons made relative to the gScramble control.
664 Statistical results for T_{RM} populations are shown; additional statistical comparisons are provided in
665 Extended Data Table 5. ****P < 0.0001, ***P < 0.001, **P < 0.01, *P < 0.05. **f,** Normalized expression
666 score of T_{RM} "up and down" gene-signatures[9,39] in cells with each gRNA KO compared to gScramble
667 cells. **g-h,** Individual TF KO phenotyping. **g,** The ratio of gRNA KO cells in SI:spleen. **h,** The frequency
668 of splenic T_{CM} ($CD62L^+$ $KLRG1^-$) and T_{EM} ($CD62L^-$ $KLRG1^+$) cells. **i-j,** P14 $CD8^+$ T cells were
669 transduced with RVs expressing the T_{RM} single-state TF *Klf6* or empty-control vectors and co-
670 transferred at ~1:1 ratios into mice infected with acute LCMV-Armstrong one day earlier. **i,**
671 Representative plots of donor P14 cells pre- and post-transfer, with the latter isolated from the SI. **j,**
672 Quantification of donor P14 cells expressing T_{RM} markers ($CD69^+$ $CD103^+$) in the SI. Data are presented
673 as the mean ± s.e.m. ***P<0.001, ** P <0.01, * P <0.05. Ordinary one-way ANOVA with Dunnett's
674 multiple comparison test versus gScramble (n ≥ 5 from ≥ 2 biological replicates, **g-h**). Co-transfer
675 experiments (**i–j**) were analyzed using paired t-tests (n ≥ 6 from ≥ 2 biological replicates).

676

677 **Fig. 5 | Distinct roles of TEX_{term} single-state TF *Zscan20* and multi-state TF *Hic1* in tumor control.**
678 **a–c**, Experimental design (**a**) for adoptive transfer of P14 $CD8^+$ T cells with CRISPR-mediated KO of
679 either the TEX_{term} single-state TF (*Zscan20*) or a multi-state TF (*Hic1*, active in both TEX_{term} and T_{RM})
680 into B16-GP33 melanoma-bearing mice. Tumor volumes were measured twice per week (**b**) and weights
681 were assessed at day 23 (**c**). **d**, Representative flow cytometry plots of $PD1^+$ P14 $CD8^+$ T cells stained

682 for SLAMF6 and TIM3, with quantification of TEX_{prog} ($PD1^+$ $SLAMF6^+$ $TIM3^-$) and exhausted ($PD1^+$
683 $SLAMF6^-$ $TIM3^+$) subsets. **e**, Frequency of $PD1^+$ $CD39^+$ double-positive cells. **f-j**, Co-transfer
684 experiments in which $Cas9^+$ P14 $CD8^+$ T cells expressing gRNA targeting *Zscan20* or *Hic1* were mixed
685 with control (scramble gRNA) and transferred into B16-GP33 tumor-bearing mice. Experimental design
686 (**f**). Representative flow cytometry plots of effector marker CX3CR1 and the cytotoxic cytokine
687 Granzyme B (GzmB) in TF KO and control cells (**g**). Quantification of $CX3CR1^+$ (**h**), $GzmB^+$ (**i**), and
688 $IFN\gamma^+$ $TNF\alpha^+$ (**j**) population in *Zscan20* KO vs. control cells (left), and *Hic1* KO vs. control cells (right).
689 Statistical analysis was performed using two-way ANOVA with Tukey's multiple comparisons test (**b**)
690 and one-way ANOVA with Dunnett's test versus gScramble (**c**). Data are from $n \geq 9$ from ≥2 biological
691 replicates (**b-e**). Co-transfer experiments (**g–j**) were analyzed using paired t-tests ($n \geq 6$ from ≥2
692 biological replicates). Data are presented as mean ± s.e.m. ****$P < 0.0001$, ***$P < 0.001$, **$P < 0.01$,
693 *$P < 0.05$.
694
695 **Fig. 6 | Targeting conserved TEX_{term} single-state TFs enhances human T cell effector function and**
696 **immunotherapy responses**
697 **a,** Human pan-cancer single-cell multi-omics and scRNA-seq datasets[52–59] were curated to evaluate TF
698 mRNA expression across diverse T cell states and to infer TF activity using the scTaiji algorithm. **b,** For
699 each sample, paired scRNA-seq and scATAC-seq data were used to construct a regulatory network and
700 compute PageRank-based TF activity scores with scTaiji. Shown are normalized PageRank scores of
701 TEX_{term} single-state TFs identified in this study (Fig. 1f) with known binding motifs conserved in the
702 human genome. **c,** mRNA expression of TEX_{term} single-state TFs across TEX_{term} and T_{RM} cell states in
703 human pan-cancer datasets. TFs with conserved cell-state-selectivity in human and mice are highlighted
704 in bold (**b,c**). **d,** Human PBMCs with *ZSCAN20* or *JDP2* KO were stimulated with anti-CD3/CD28
705 antibodies to induce chronic activation and mimic T cell exhaustion. Cells were expanded every 2–3
706 days at a density of 1×10^6 cells/mL. **e,** KO and control human $CD8^+$ T cells were analyzed for
707 CCR7(central memory/stem-like marker) and LAG3 (exhaustion marker) by flow cytometry as shown
708 in representative flow plots and cumulative bar graphs. **f,** TF KO and control human $CD8^+$ T cells were
709 analyzed for inhibitory receptors $PD1^+$ $TIM3^+$ double positive cells by flow cytometry as shown in
710 representative flow plots and cumulative bar graphs. **g,** Frequency of $IFN\gamma^+$ $TNF\alpha^+$ double-positive cells
711 and and IL-2 producing cells. **i,** Polyfunctionality analysis based on the number of cytokines produced
712 per cell. **i,** Experimental design for testing the synergy between TEX_{term} single-state TF KO and anti–

713 PD-1 treatment. Cas9^{+} P14 CD8^{+} T cells (with or without TF KO) were adoptively transferred into B16-
714 GP33 tumor-bearing mice on day 7 post-implantation, followed by anti–PD-1 or isotype IgG2a
715 treatment. **j–k,** Tumor growth curves (**j**) and tumor weights (**k**) of *Zscan20* KO vs. control. **l–m,** Tumor
716 growth curves (**l**) and tumor weights (**m**) of *Jdp2* KO vs. control. Data are presented as mean ± s.e.m. n
717 ≥ 6 from ≥2 biological replicates. One-way ANOVA with Dunnett's test versus gScramble (**e-g**); two-
718 way ANOVA with Dunnett's multiple comparisons test versus gScramble (**h**); two-way ANOVA with
719 Tukey's multiple comparisons test (**j,l**); one-way ANOVA with Tukey's test (**k, m**) ****$P < 0.0001$,
720 ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$.
721

722 **Methods**

723 **Dataset acquisition for murine CD8$^+$ T cell state multiomics atlas:** CD8$^+$ T cell samples were
724 collected from ten datasets, including those generated in this study (Extended Data Fig. 2a). In total, we
725 analyzed 121 experiments, comprising 52 ATAC-seq and 69 RNA-seq datasets, which were integrated
726 to generate paired samples and served as input for the Taiji pipeline. The samples encompassed nine
727 distinct CD8$^+$ T cell subtypes: naive cells (Naive), terminal effector cells (TE), memory precursor cells
728 (MP), tissue-resident memory cells (T_{RM}), effector memory cells (T_{EM}), central memory cells (T_{CM}),
729 progenitor exhausted cells (TEX_{prog}), intermediate exhausted cells (TEX_{eff}) and terminal exhausted cells
730 (TEX_{term}). Cell states were defined based on established surface marker combinations and LCMV
731 (lymphocytic choriomeningitis virus)-specific tetramers, including IL7R, KLRG1, PD1, SLAMF6,
732 CD101, Tim3, CD69, CD103, H2-Db LCMV gp 33-41 and H2-Db LCMV gp 276-286 or congenic
733 markers for P14 (T cell receptor (TCR) specific for the LCMV GP33–41 peptide CD8$^+$ T cells), in the
734 context of either acute (LCMV-Armstrong) or chronic (LCMV-Clone 13) infection models. A complete
735 summary of dataset sources, accession numbers, infection conditions, and corresponding cell state
736 definitions (sorting gates) is provided in Extended Data Table 1 and Extended Data Fig. 2a.

737

738 **TF regulatory networks construction and visualization:** To perform integrative analysis of RNA-seq
739 and ATAC-seq data, we developed Taiji v2.0. which allows visualization of several downstream
740 analysis-TF wave, TF-TF association, TF community analysis. Epitensor was used for the prediction of
741 chromatin interactions. Putative TF binding motifs were curated from the latest CIS-BP database[75]. In
742 this analysis, 695 TFs were identified as having binding sites centered around ATAC-seq peak summits.
743 The average number of nodes (genes) and edges (interactions) of the genetic regulatory networks across
744 CD8$^+$ T cell states were 15,845 and 1,325,694, respectively, including 695 (4.38%) TF nodes. On
745 average, each TF regulated 1,907 genes, and each gene was regulated by 22 TFs.

746

747 **Identification of single-state/multi-state TFs:** We first identified universal TFs with mean PageRank
748 across 9 cell states ranked as top 10% and Coefficients of variation (CV) less than 0.5. In total, 54
749 universal TFs were identified and shown in Extended Data Table 1. The remaining 641 TFs were
750 candidate TFs for single-state TFs. To identify single-state TFs, we divided the samples into two groups:
751 target group and background group. Target group included all the samples belonging to the cell state of
752 interest and the background group comprised the remaining samples. We then performed the normality

753 test using Shapiro-Wilk's method to determine whether the two groups were normally distributed, and
754 we found that the PageRank scores of most samples (90%) follow a log-normal distribution. Based on
755 the log-normality assumption, an unpaired t-test was used to calculate the P-value. A P-value cutoff of
756 0.05 and log2 fold change cutoff of 0.5 were used for calling lineage-specific TFs In total, 255 specific
757 TFs were identified (Extended Data Table 2). Depending on whether the TF appeared in multiple cell
758 states, they could be further divided into multi-state TFs (Extended Data Table 2, Fig 1e) and 136 single-
759 state exclusive TFs (Extended Data Table 2, Fig. 1d). Out of 255 single-state TFs, 84 TFs appear in
760 TEX_{term} or T_{RM}. To identify the truly distinctive TFs between TEX_{term} and T_{RM}, we performed a second
761 round of unpaired t-test only between TEX_{term} and T_{RM} (Extended Data Table 3). The same cut-offs, i.e.
762 P-value of 0.05 and log2 fold change of 0.5, were applied to select TEX_{term} single-taskers and T_{RM} single-
763 taskers. Thirty out of 84 TFs that did not pass the cut-off were identified as TEX_{term} and T_{RM} multi-
764 taskers. The full workflow was summarized in Extended Data Fig. 3a.
765
766 **Identification of transcriptional waves:** Combined with prior knowledge on the T cell differentiation
767 path, TF waves are combinations of TFs which are particularly active in certain differentiation stages,
768 revealing possible mechanisms of how TF activities are coordinated during differentiation. To be more
769 specific, we clustered the TFs based on the normalized PageRank scores across samples. First, we
770 performed principal component analysis (PCA) for dimensionality reduction of the TF score matrix. We
771 retained the first 10 principal components for further clustering analysis, which explained more than
772 70% variance (Extended Data Fig. 4b; left panel). We used the k-means algorithm for clustering
773 analysis. To find the optimal number of clusters and similarity metric, we performed the Silhouette
774 analysis to evaluate the clustering quality using five distance metrics: Euclidean distance, Manhattan
775 distance, Kendall correlation, Pearson correlation, and Spearman correlation (Extended Data Fig. 4b;
776 right panel). Pearson correlation was the most appropriate distance metric since the average Silhouette
777 width was the highest among all five distance metrics. Based on these analyses, we identified 7 distinct
778 dynamic patterns of TF activity during immune cell development. We further performed functional
779 enrichment analysis to identify GO terms for these clusters.
780
781 **TF-TF collaboration network analysis and visualization:** To build the TF-TF association networks,
782 we first defined a set of relevant TFs for each context (TEX_{term} and T_{RM}) by combining cell state-
783 important and single-state TFs, resulting in 159 TFs for TEX_{term} and 170 for T_{RM}. The analysis was

784 based on a TF-regulatee network derived from Taiji, where we first consolidated sample networks by
785 averaging the edge weights for each TF-regulatee pair. To reduce noise, regulatees with low variation
786 across all TFs (standard deviation ≤ 1) were removed. Subsequently, a TF-TF correlation matrix was
787 generated by calculating the Spearman's correlation of edge weights for each TF pair across their
788 common regulatees. From this matrix, we constructed a graphical model using the R package "huge"[76],
789 which employs the Graphical Lasso algorithm and a shrunken ECDF (empirical cumulative distribution
790 function) estimator. An edge between two TFs was established if their correlation was deemed
791 significant by the model, controlled by a lasso penalty parameter (lambda) of 0.052. This value was
792 chosen as it represents a local minimum on the sparsity-lambda curve, resulting in approximately 15%
793 of TF-TF pairs being connected. To validate this method, we estimated the false discovery rate by
794 generating a null model through random shuffling of the TF-regulatee edge weights. Applying our
795 algorithm to this null data identified zero interactions, confirming that our approach has a very low false
796 discovery rate.
797
798 **TF community construction and visualization:** Following the construction of the TF-TF association
799 networks, we identified functionally related TF communities within each network. We applied the
800 Leiden algorithm[77], using modularity as the objective function and setting the resolution parameter to
801 0.9, as this value achieved the highest clustering modularity in our analysis. This procedure identified
802 five distinct communities for each context (TEX_{term} and T_{RM}). The final networks, with their detected
803 communities, were visualized using the Fruchterman-Reingold layout algorithm[78] to spatially represent
804 the TF-TF association structure.
805
806 **Pathway enrichment analysis:** The enriched functional terms in this study were analyzed by the R
807 package clusterProfiler 4.0.5. We used GO, KEGG, MSigDb databases for annotation. For gene set
808 enrichment analysis (GSEA), the genes were first ranked by the mean edge weight in corresponding
809 samples and H, C5, C6, and C7 collections from MSigDb were used for annotation. A cutoff of P-value
810 < 0.05 was used to select the significantly enriched GO terms and KEGG pathways.
811
812 **Heuristic score calculation via the integration of TF regulatory network and Perturb-seq:**
813 We reasoned that (1) log2FC change in expression due to TF KO, and (2) TF-gene regulatory edge
814 weights could be combined to provide heuristic scores for a TF's regulatory effect on a target gene. For

each guide RNA KO, Seurat's FindMarkers() function was used to quantify log2FC of gene's expression with respect to the gScramble condition. Heuristic scores were calculated for each TF-gene pair by multiplying the gene log2FC with the corresponding edge weight from the Taiji analysis. Regulatees of a TF were annotated as high-confidence if the magnitude log2FC of the regulate exceeded 0.58 (corresponding to a fold-change of 1.5 or its reciprocal) and if the edge weight belonged to the upper quantile of all edge weights attributed to the TF. The sign of the log2FC was used to determine if the TF activated or repressed each target gene.

Human TF activity and cell-state selectivity analysis:

Taiji-based analysis of human multiomic datasets. Transcription factor (TF) activity was inferred using the Taiji pipeline applied to matched single-cell RNA-seq (scRNA-seq) and ATAC-seq (scATAC-seq) datasets from various human cancers, including ccRCC (n=2; PRJNA768891, GSE240822), GBM (GSE240822), BCC (GSE123814, EGAS00001006141), HNSCC (GSE139324, EGAS00001006141), HCC (GSE125449, EGAS00001006141), and RCC (PMID: 30093597; EGAS00001006141). Cell types were annotated using canonical marker gene expression and categorized into six major CD8+ T cell states: T_Ex (exhausted), T_ExProg (progenitor exhausted), T_Eff (effector), T_RM (tissue-resident memory), T_CM/Naive (central memory/naive), and proliferating. PageRank scores derived from Taiji were log-transformed, averaged within each T cell category, and then standardized using z-score normalization. Results were visualized with a focus on TF activity in T_RM and T_Ex populations.

TF expression comparison across human CD8+ T cell states. To assess TF expression across diverse T cell states, raw count matrices from a published pan-cancer CD8+ T cell atlas (GSE156728) were reprocessed using Seurat's standard workflow. The dataset encompassed T cells from 11 tumor types, including BC, BCL, CHOL, ESCA, FTC, MM, OV, PACA, RC, THCA, and UCEC. Cell type annotations provided by the original study were retained and mapped to the following broad categories: T_RM, T_Ex, T_EM (effector memory), T_m (memory), T_n (naive), and T_k (cycling). Seurat's AverageExpression function was used to compute average logCPM expression for each TF in each T cell category, followed by z-score normalization. Data visualization emphasized comparisons between T_RM and T_Ex subsets.

845 **Mice and infections:** C57BL/6/J, OT-1 (C57BL/6-Tg(TcraTcrb)1100Mjb/J), B6.Cg-*Rag2*$^{tm1.1Cgn}$/J,
846 CD45.1 (B6.SJL-*PtprcaPepcb*/BoyJ) mice were purchased from Jackson Laboratories. P14 mice
847 (Pircher et al., 1987) mice have been previously described. Cas9 P14 mice were generated by crossing
848 P14 mice with B6(C)-Gt(ROSA)26Sorem1.1(CAG-cas9*,-EGFP)Rsky/J (Jackson Laboratories).
849 Animals were housed in specific-pathogen-free facilities at the Salk Institute and University and at the
850 University of North Carolina at Chapel Hill. All animal experiments were approved by the Institutional
851 Animal Care and Use Committee (IACUC). Mice were infected with $2x10^5$ PFU LCMV-Armstrong by
852 intraperitoneal injection or $2x10^6$ PFU LCMV-Clone13 by retro-orbital injection under anesthesia.
853
854 **Viral titers:** LCMV Fluorescence Focus Unit titration was performed seeding Vero cells at a density of
855 30k cells/100µl in a 96 well flat bottom plate in DMEM + 10% FBS + 2% HEPES + 1% Pen/Strep. On
856 the next day, tissues were homogenized on ice, spun down at 1000g for 5 min at 4°C and supernatants
857 or serum were diluted in 10-fold steps. Diluted samples were added to Vero cells and incubated at 37°C,
858 5% CO_2 for ~20h. Subsequently, inocula were aspirated and wells were incubated with 4% PFA for 30
859 min at RT before washing with PBS. VL-4 antibody (BioXCell) was conjugated using the Invitrogen
860 AF488 conjugation kit and added to the wells in dilution buffer containing 3% BSA and 0.3% Triton
861 (Thermo Fisher Scientific) in PBS. Cells were incubated at 4°C overnight before washing with PBS and
862 counting foci under the microscope. FFU was calculated per this formula: FFU/ml=Number of plaques/
863 (dilution× volume of diluted virus added to the plate).
864
865 **Cell isolation:** Spleens were mechanically dissociated with 1mL syringe plungers over a 70um nylon
866 strainer. Spleens were incubated in ammonium chloride potassium (ACK) buffer for 5 minutes. For
867 isolation of small intestinal IEL, Peyer's patches were first removed by dissection. Intestines were
868 longitudinally cut and then cut into 1cm pieces and washed in PBS. Pieces were incubated in 30mL
869 HBSS with 10% FBS, 10mM HEPES, and 1mM dithioerythritol with vigorous shaking at 37C for 30
870 minutes. Supernatants were collected, washed, and further isolated using a 40/67% discontinuous percoll
871 density centrifugation for 20 minutes at room temp with no brakes.
872
873 **Cell lines and *in vitro* cultures:** B16-GP33 melanoma cell line were cultured in DMEM (Invitrogen)
874 with 10% fetal bovine serum, 1% penicillin-streptomycin and 250 µg/ml G418 (Invitrogen #10131027).
875 MCA-205 (Sigma) tumor line was maintained in RPMI supplemented with 10% FBS, 300 mg/L L-

876 glutamine, 100 units/mL penicillin, 100 μg/mL streptomycin, 1mM sodium pyruvate, 100μM NEAA,
877 1mM HEPES, 55μM 2-mercaptoethanol, and 0.2% Plasmocin mycoplasma prophylactic (InvivoGen).
878 All the tumor cell lines were used for experiments when in the exponential growth phase. For *in vitro* T
879 cell culture, splenocytes were activated in RPMI 1640 medium (Invitrogen) containing 10% fetal bovine
880 serum and 1% penicillin-streptomycin, 2mM L-glutamine, 0.1 mg/ml GP33, beta-mercaptoethanol 50
881 mM and 10 U/ml IL-2.
882
883 **Tumor engraftment and treatment of tumor-bearing mice (Fig. 5, 6):** 3×10^5 B16-GP33 (Fig. 5),
884 5×10^5 B16-GP33 tumor cells (Fig. 6) were injected subcutaneously in 100 μl PBS. Around 0.5-1x10^6
885 $Cas9^+$ P14 T cells with CD45.1 markers were transferred to tumor on day 7 without pre-radiation of
886 tumor-bearing mice. Tumors were measured every 2-3 days post-tumor engraftment for indicated
887 treatments and calculated. Tumor volume was calculated by volume = (length × $width^2$)/2. For antibody-
888 based treatment, tumor-bearing mice were treated with anti-PD1 antibody (200 μg per injection, clone
889 RMP1-14, BioXcell) twice per week from day 7 post-tumor implantation. Tumor growth was measured
890 twice per week with calipers. Survival events were recorded each time a mouse reached the endpoint
891 (tumor volume ≥1,500 mm^3). Tumor weights were measured on day 23 for Fig. 5 and on day 25 for Fig.
892 6. All experiments were conducted according to the Salk Institute Animal Care and Use Committee.
893
894 **Tumor digestion and cell isolation for Fig. 5:** Tumors were minced into small pieces in RPMI
895 containing 2% FBS, DNase I (0.5 μg/ml, Sigma-Aldrich), and collagenase (0.5 mg/ml, Sigma-Aldrich)
896 and kept for digestion for 30 min at 37˚C, followed by filtration with 70 μm cell strainers (VWR).
897 Filtered cells were incubated with ACK lysis buffer (Invitrogen) to lyse red blood cells, mixed with
898 excessive RPMI 1640 medium (Invitrogen) containing 10% fetal bovine serum and 1% penicillin-
899 streptomycin, and centrifuged at 400g for 5 min to obtain a single-cell suspension.
900
901 **Proteasome activity analysis:** For experiments involving the Proteasome Activity Probe (R&D
902 systems), cells of interest were incubated with the probe at concentration of 2.5mM for 2 hours at 37C
903 in PBS. Samples were washed and then stained with Zombie NIR viability dye (Biolegend) in PBS at
904 4C for 15 minutes. Samples were then stained with some variation of the following antibodies for 30
905 minutes in FACS Buffer on ice: CD45-BV510 (BD Biosciences), CD45.2-BV510 (Biolegend), CD45.1-
906 PE-Cy7 (Invitrogen), CD4-APC Fire 810 (Biolegend), CD11b-Alexa Fluor 532 (Invitrogen), CD8-

907 Spark NIR 685 (Biolegend), CD44-Brilliant Violet 785 (Biolegend), CD62L-BV421 (BD Biosciences),
908 PD1-BB700 (BD Biosciences), TIM-3-BV711 (Biolegend), LAG-3-APC-eFluor 780 (Invitrogen),
909 SlamF6-APC (Invitrogen), CD39-Superbright 436 (Invitrogen), CX3CR1-PE/Fire 640 (Biolegend),
910 CD69-PE-Cy5 (Biolegend), GITR-BV650 (BD Biosciences), CD27-BV750 (BD Biosciences). Samples
911 were collected on a Cytek Northern Lights and analyzed via Cytek SpectraFlo software.
912
913 **Tumor experiment for proteasome assay (Fig. 2):** MCA-205 fibrosarcomas (2.5 x 10^5) were
914 established via subcutaneous injection into the right flank of C57BL/6 mice. After 12-14 days of tumor
915 growth, spleens, draining lymph nodes, and tumors from groups of mice were harvested and tumors
916 were processed using the Mouse Tumor Dissociation Kit and gentleMACS dissociator (Miltenyi Biotec)
917 according to manufacturer's protocol. For purification experiments, samples were pre-enriched using
918 the EasySep Mouse $CD8^+$ T Cell Isolation Kit (Stemcell Technologies) according to manufacturer's
919 protocol and stained with Live-or-Dye PE Fixable Viability Stain (Biotium) and CD8a-APC (Invitrogen)
920 and live $CD8^+$ cells were sorted using the FACSAria II cell sorter. $CD8^+$ spleen and pooled TIL samples
921 were washed in PBS and frozen for RNA-seq analysis. For adoptive cellular therapy experiments, B16-
922 GP33 melanomas were established subcutaneously by injecting 5.0×10^5 cells into the right flank of
923 CD45.1 mice and tumor-bearing hosts were irradiated with 5 Gy 24 hours prior to T cell transfer. In
924 contrast, mice used in the experiments shown in Fig. 5 and Fig. 6 were not irradiated prior to T cell
925 transfer. After 7 days of tumor growth, 1.5×10^6 CD45.2 OT-1 T cells and 1.5×10^6 CD45.1/CD45.2 P14
926 T cells were infused in 100 μL PBS via tail vein into tumor bearing mice. Tumors were harvested 14
927 days after adoptive cell transfer and CD8 TILs were analyzed for proteasome activity. All experiments
928 were conducted in accordance with the guidelines of the University of North Carolina at Chapel Hill
929 Animal Care and Use Committee.
930
931 **Proteasome^{high}/Proteasome^{low} T cell adoptive transfer experiment (Fig. 2 and Extended Data Fig.**
932 **6c):** For the adoptive transfer experiment involving proteasome^{high} and proteasome^{low} tumor-specific
933 OT-1 T cells (Fig. 2l), whole splenocytes from OT-1 mice were activated with 1 μg/mL OVA_257–264
934 peptide and expanded for 7 days in the presence of 200 U/mL rhIL-2 (NCI). On day 7, OT-1 cells were
935 FACS-sorted based on proteasome activity to isolate proteasome^{high} and proteasome^{low} OT-1
936 populations. A total of 2.5×10^5 sorted OT-1 cells were injected into C57BL/6 mice bearing B16F1-
937 OVA melanomas. Tumors were established by subcutaneous injection of 3×10^5 B16F1-OVA cells into

938 the right flank 7 days prior to T cell transfer. Recipient mice were preconditioned with 5 Gy total body
939 irradiation 24 hours before adoptive transfer. Tumor growth was measured every other day with calipers.
940 For Extended Data Fig. 6c, MCA-205 fibrosarcomas (2.5 x 10^5) were established via subcutaneous
941 injection into the right flank of C57BL/6 mice. After 14 days of tumor growth, Live $CD45^+$ $CD8^+$ $CD44^+$
942 $PD1^+$ T cells were sorted from tumors based on proteasome activity (high vs low) using the FACSAria
943 II cell sorter. 2.5 x 10^4 cells were then injected into the 2 day MCA-205 bearing RAG2-/- hosts (n=5 per
944 group) and tumor growth was monitored every other day starting on day 4. All experiments were
945 conducted in accordance with the guidelines of the University of North Carolina at Chapel Hill Animal
946 Care and Use Committee.
947
948 **Retrovirus transduction and adoptive transfer:** For overexpression of the gRNA retrovirus vector,
949 293T cells were transfected with the Eco-helper and MSCV gRNA vectors. 48 h and 72 h later,
950 supernatant containing retroviral particles was ready for transduction. Donor P14 splenocytes were *in*
951 *vitro* activated by 0.1 mg/ml GP33 and 10 U/ml IL-2 at 37°C for 24h, then spin-transduced (1500 g)
952 with fresh RV supernatant from 293T cells for 90 min at 30°C in the presence of 5 µg/ml polybrene.
953
954 **CRISPR-Cas9/RNP nucleofection:** Naïve CD8+ T cells were enriched from spleen using the EasySep
955 Mouse CD8+ T cell Isolation Kit (STEMCELL Technologies). sgRNAs targeting *ZSCAN20, JDP2,*
956 *Etv5, Prdm1, Hic1* genes or the mouse or human genome nontargeting scramble (control) were obtained
957 from Synthego, IDT, and GeneScript (Extended Data Table 5). Cas9 RNP was prepared immediately
958 before experiments by incubating 1 µl sgRNA (stock: 3 nmol in 10 µl water), 0.6 µl Cas9 nuclease (IDT,
959 stock: 62 µM) and 3.4 µl RNase-free water at room temperature for 10min. Nucleofection of naïve
960 CD8+ T cells was performed using Lonza (Allendale, NJ) P3 primary cell kit and program DN100 with
961 4D-Nucleofector (Lonza Bioscience) for mouse and EO115 for human stimulated T cells . Each
962 nucleofection reaction consisted of approximately 5-10 × 10^6 cells in 20 µl of nucleofection reagent and
963 mixed with 5 µl of RNP: Cas9 complex. After electroporation, 100 µl of T cell culture media was added
964 to the well to transfer the cells to 1.5 mL Eppendorf tube. The cells were rested at 37°C for 3 min. For
965 *in vivo* adoptive transfer, cells were resuspended in PBS at desired concentration and adoptively
966 transferred into recipient mice.
967

968 **CRISPR gene editing validation via Sanger sequencing**: The genomic DNA (gDNA) was isolated
969 from both the KO-induced CD8+ T cells and the control cells using the Quick-DNA MicroPrep Kit
970 (Zymo). The gDNA concentrations were quantified using the NanoDrop One spectrophotometer
971 (ThermoFisher Scientific). Following isolation, PCR amplification was performed with 2x Phusion Plus
972 Green PCR Master Mix (ThermoFisher Scientific) and the respective validation primers under the
973 following conditions: (98 °C for 5 mins; 35 x of 98 °C for 10 secs, 69 °C for 20 secs, 72 °C for 20-30
974 secs/kb; 72 °C for 2 mins; hold at 10 °C). The PCR products were run on a 2% agarose gel with SYBR
975 Safe DNA Gel Stain (Invitrogen), and the appropriate bands on the gel were extracted and purified with
976 the Gel DNA Recovery Kit (Zymo). Concentrations of the purified amplicon samples were measured
977 and then sent off for sequencing along with the designated sequencing primers designed from
978 Benchling's Primer3 tool. The samples with the KOs were compared to the wildtype controls using
979 EditCo's Ice Analysis software, providing the indel percentages, KO score, and the indel distributions
980 used to assess the editing efficiency. Indel% ranges from 56% to 97%, and the KO score throughout
981 experiments ranges from 32 to 74.
982
983 **Flow cytometry, cell sorting, and antibodies:** Both single cell suspensions were incubated with Fc
984 receptor-blocking anti-CD16/32 (BioLegend) on ice for 10 min before staining. Cell suspensions were
985 first stained with Red Dead Cell Stain Kit (ThermoFisher) for 10 min on ice. Surface proteins were then
986 stained in FACS buffer (PBS containing 2% FBS and 0.1% sodium azide) for 30 min at 4˚C. To detect
987 cytokine production *ex-vivo*, cell suspensions were re-suspended in RPMI 1640 containing 10% FBS,
988 stimulated by 50 ng/ml PMA and 3 μM Ionomycin in the presence 2.5 μg/ml Brefeldin A (BioLegend
989 #420601) for 4 h at 37°C. Cells were processed for surface marker staining as described above. For
990 intracellular cytokine staining, cells were fixed in BD Cytofix/Cytoperm (BD #554714) for 30 min at
991 4 °C, then washed with 1× Permeabilization buffer (Invitrogen #00-8333-56). For transcription factor
992 staining, cells were fixed in Foxp3 / Transcription Factor Fixation/Permeabilization buffer (Invitrogen
993 #00-5521-00) for 30 min at 4 °C, then washed with 1× Permeabilization buffer. Cells were then stained
994 with intraceulluar antibodies for 30 min at 4 °C. Samples were processed on LSR-II flow cytometer (BD
995 Biosciences) and data were analyzed with FlowJo V10 (TreeStar). Cells were sorted either on
996 FACSAria™ III sorter or Fusion sorter (BD Biosciences). The following antibodies against mouse
997 proteins were used: anti-CD8a (53-6.7), anti-PD1 (29F.1A12), anti-CX3CR1 (SA011F11), anti-
998 SLAMF6 (13G3), anti-CD38 (90), anti-CD39 (24DMS1), anti-CD101 (Moushi101), anti-KRLG1

999 (2F1), anti-CD69 (H1.2F3), anti-CD103 (M290), anti-CD62L (MEL-14), anti-Tim3 (RMT3-23), anti-
1000 Ly5.1 (A20), anti-Ly5.2 (104), anti-IFN-γ (XMG1.2), anti-TNF-α (MP6-XT22). The following
1001 antibodies against human proteins were used: anti-CD8a (RPA-T8), anti-CD4 (SK3), anti-CD45RA
1002 (H100), anti-CD45RO (UCHL1), anti-CCR7 (G043H7), anti-CD62L (DREG-56), anti-CD69 (FN50),
1003 anti-CD103 (Ber-ACT8), anti-CXCR6 (K041E5), anti-PD1 (EH12.2H7), anti-CD38 (HIT2), anti-CD39
1004 (A1), anti-LAG3 (11C3C65), anti-Tim3 (F38-2E2), anti-TIGIT (A15153G), anti-IFN-γ (4S.B3), anti-
1005 TNF-α (MAb11), anti-IL 2 (JES6-5H4), anti-GZMB (QA16A02), anti-G4S Linker (E7O2V). These
1006 antibodies were purchased from Invitrogen, Biolegend, Cell Signaling, or eBiosciences.

1007

1008 ***In vivo* individual TF KO phenotyping:** To assess the functional impact of individual TF KOs in $CD8^+$
1009 T cells, we used Cas9-expressing P14 donor cells (LCMV-specific TCR transgenic mice, CD45.1
1010 congenic) transduced with GFP-expressing retroviral vectors encoding individual gRNAs.
1011 Transductions were performed on the day of adoptive transfer without prior sorting. Without sorting,
1012 transduced donor cells (0.5-1 x 10^5) were immediately transferred into congenically distinct Cas9-
1013 expressing wild-type recipient mice (CD45.2) infected one day prior with either LCMV-Clone 13 or
1014 Armstrong strains. At ≥ day 20 post-infection, spleens from Clone 13 model and spleens and small
1015 intestines from Armstrong model were harvested. Single-cell suspensions were prepared and analyzed
1016 by flow cytometry. Live, single cells were first gated on $CD8^+$ cells, followed by gating on $CD45.1^+$ P14
1017 donor $CD8^+$ T cells. Successfully transduced ($gRNA^+$) cells were identified by GFP expression, which
1018 ranged from 10% to 70% of P14 $CD8^+$ T cells across experiments. Due to variability in the number of
1019 GFP^+ donor P14 $CD8^+$ T cells obtained from different experiments, all phenotypic analyses were
1020 performed within the GFP^+ ($gRNA^+$) $CD45.1^+$ $CD8^+$ population. PD-1 positive and negative cells,
1021 exhaustion subsets (TEX_{term}: PD-1^+ $SLAMF6^-$ $CX3CR1^-$, TEX_{prog} :PD-1^+ $SLAMF6^+$ $CX3CR1^-$, TEX_{eff}:
1022 PD-1^+ $CX3CR1^+$) or expression of phenotypic markers was reported as a percentage within the $gRNA^+$
1023 (GFP^+) P14 $CD8^+$ T cell population to ensure consistency across samples.

1024

1025 **Co-transfers of control and TF KO/overexpression P14 $CD8^+$ T cells in infection or tumor models:**
1026 Naive $CD8^+$ T cells were isolated from the spleens and lymph nodes of Cas9-expressing LCMV TCR
1027 transgenic (Cas9 P14) or P14 mice using the EasySep™ Mouse $CD8^+$ T Cell Isolation Kit (STEMCELL
1028 Technologies). Purified P14 cells were activated for ~24 hours on plates coated with goat anti-hamster

1029 IgG (ThermoFisher), followed by 1 μg/mL hamster anti-mouse CD3 and 1 μg/mL hamster anti-mouse
1030 CD28 antibodies (ThermoFisher), in complete T cell medium (RPMI-1640 supplemented with 10% FBS
1031 (HyClone), 55 μM 2-mercaptoethanol, 100 IU/mL penicillin-streptomycin, and 1% HEPES). After
1032 activation, cells were transduced with retroviruses encoding Klf6 overexpression or gRNAs targeting
1033 Hic1 or Zscan20, and cultured with 20 IU/mL IL-2, 2.5 ng/mL IL-7, and 2.5 ng/mL IL-15 (PeproTech).
1034 At 48 hours post-transduction, reporter expression was confirmed by flow cytometry. Donor cell mixes
1035 were prepared using control vs. Klf6-overexpressing cells (Fig. 4), or gScramble vs. gHic1/gZscan20
1036 cells (Fig. 5). For LCMV infection studies, 1.5×10^5 transduced P14 $CD8^+$ T cells were transferred into
1037 recipient mice, followed by infection with either 2×10^5 PFU LCMV-Armstrong (acute infection,
1038 intraperitoneal) or 2×10^6 PFU LCMV-Clone 13 (persistent infection, intravenous). For tumor studies,
1039 5×10^5 to 1×10^6 transduced T cells (gScramble vs. gTF) were transferred on day 7 after B16-GP33
1040 tumor implantation. All experiments were conducted according to the University of North Carolina at
1041 Chapel Hill Animal Care and Use Committee.
1042
1043 **Perturb seq screening using the retroviral transcriptional factor library:**
1044 D*ual-guide direct-capture retroviral sgRNA vector.* To generate dual-guide sgRNA vector (MSCV-
1045 hU6-mU6-SV40-EGFP), we replaced the hU6 RNA scaffold region of the previously described
1046 retroviral sgRNA vector MG-guide[79] with an additional scaffold[80] and murine U6 promoter.
1047

1048 *Dual-guide direct-capture retroviral library construction.* For the curated gene list containing 21 TFs,
1049 a total of four gRNA sequences distributed on two individual constructs were designed for each gene.
1050 To construct the library, a customized double-strand DNA fragment pool containing 80 oligonucleotides
1051 targeting those 19 TFs and 4 scramble gRNAs (each oligonucleotide contains two guides targeting the
1052 same gene) (Extended Data Table 5) was ordered from IDT. The dual-guide library was generated using
1053 an In-Fusion (Takara) reaction. In brief, the gRNA containing DNA fragment pool was combined in
1054 MG-guide vector linearized with BpiI (Thermo Fisher). The construct was then transformed into Stellar
1055 competent cells (Takara) and amplified, and the resulting intermediate, individual construct was
1056 assessed for quality using Sanger sequencing. Then, individual dual-gRNA vectors were combined. For
1057 quality control, sgRNA skewing was measured using the MAGeCKFlute[81] to monitor how closely
1058 sgRNAs are represented in a library.

1059 *In vivo screening.* Retrovirus was generated by co-transfecting HEK293 cells with the dual-guide,
1060 direct-capture retroviral TF library and the packaging plasmid pCL-Eco. Supernatants were collected at
1061 48 and 72 hours post-transfection, then stored at −80 °C. Cas9-expressing P14 $CD8^+$ T cells were
1062 transduced with the viral supernatant to achieve a transduction efficiency of 20–30%. To ensure
1063 sufficient representation of control cells in downstream analysis, 50% of the viral mixture consisted of
1064 retrovirus encoding a non-targeting control gRNA vector. For in vivo experiments, 5×10^4 transduced
1065 P14 cells were intravenously transferred into Cas9-expressing, puromycin-resistant C57BL/6 recipient
1066 mice infected one day prior with either LCMV-Clone 13 or LCMV-Armstrong strain. A total of 25
1067 LCMV-Clone 13 infected mice were used for five biological replicates, and 10 LCMV-Armstrong-
1068 infected mice were used for three biological replicates. Each biological replicate was labeled using
1069 hashtag antibodies (BioLegend, TotalSeq-C) to enable sample demultiplexing and statistical analysis.
1070 At ≥18 days post-infection, donor-derived P14 $CD8^+$ T cells were sorted and pooled for Perturb-seq
1071 analysis. Preliminary tests indicated that T cells expressing gRNA *in vivo* exhibit a greater tendency for
1072 gRNA silencing over extended periods compared to *ex vivo* cultured cells, despite initial successful KOs.
1073 To mitigate gRNA barcode silencing, we harvested tissue between days 18 and 23. Sorted $EGFP^+$ P14
1074 $CD8^+$ T cells were resuspended and diluted in 10% FBS RPMI at a concentration of 1×10^6 cells per
1075 ml. Both the gene expression library and the CRISPR screening library were prepared using a Chromium
1076 Next GEM Single Cell 5′ kit with Feature Barcode technology for CRISPR Screening (10x Genomics).
1077 In brief, the single-cell suspensions were loaded onto the Chromium Controller according to their
1078 respective cell counts to generate 10,000 single-cell gel beads in emulsion per sample. Each sample was
1079 loaded into four separate channels. Chromium Next GEM Single Cell 5' Kit v2, PN-1000263, Chromium
1080 5' Feature Barcode Kit, PN-1000541, 5' CRISPR Kit, PN-1000451, Chromium Next GEM Chip K
1081 Single Cell Kit, PN-1000287, Dual Index Kit TT Set A, PN-1000215, Dual Index Kit TN Set A, PN-
1082 1000250 (10x Genomics) in total were used for each reaction. The resulting libraries were quantified
1083 and quality checked using TapeStation (Agilent). Samples were diluted and loaded onto a NovaSeq
1084 (Illumina) using 100 cycle kit to obtain a minimum of 20,000 paired-end reads (26x91 bp) per cell for
1085 the gene expression library and 5,000 paired-end reads per cell for the CRISPR screening library,
1086 yielding an average of 42639, 36739, and 53413 reads aligned from cells from *in vivo* LCMV-Clone 13,
1087 *in vivo* LCMV-Armstrong infection, and *in vitro* donor respectively.

1088 *Data analysis.* Alignments and count aggregation of gene expression and sgRNA reads were completed
1089 using Cell Ranger (v.7.0.1). Gene expression and sgRNA reads were aligned using the cell Ranger multi
1090 count command with default settings. Gene expression reads were aligned to the mouse genome (mm10
1091 from ENSEMBL GRCm38 loaded from 10x Genomics). The median average of 4, 2, and 33 unique
1092 molecular identifiers (UMIs) were detected from cells from *in vivo* LCMV-Clone 13 and LCMV-
1093 Armstrong infection, and *in vitro* donor respectively. Droplets with sgRNA UMI passing of default Cell
1094 Ranger CRISPR analysis Protospacer UMI threshold were used in further analysis. The filtered feature
1095 matrices were imported into Seurat (v.4.3.0)[82] to create assays for a Seurat object containing both gene
1096 expression and CRISPR guide capture matrices. Cells detected with sgRNAs targeting two or more
1097 genes were then removed to avoid interference from multi-sgRNA-transduced cells. Low-quality cells
1098 with <200 detected genes, >10% mitochondrial reads, and <5% ribosomal reads were discarded. A total
1099 of 17,257 cells (Clone 13) and 15,211 cells (Armstrong) were passed quality filtering and were used for
1100 downstream analysis. Count data was normalized by a global-scaling normalization method and linear
1101 transformed.[83] Cluster-specific genes were identified using the FindAllMarkers function of Seurat. We
1102 utilized Nebulosa[84] to recover signals from sparse features in single-cell data and made gRNA density
1103 plots with scCustomize[85] based on kernel density estimation. In each biological replicate (Clone 13 n=5
1104 and Armstrong =3), the percent cluster distribution of cells with each TF gRNA vector was calculated.
1105 Among two gRNA vectors per target TF, the gRNA vector with higher TEX_{term} reduction was shown in
1106 Fig. 3d and utilized for Perturb-seq in LCMV-Armstrong infection (Extended Data Table 5). Two-way
1107 ANOVA with Fisher's LSD test was performed to determine statistical significance. DEG was
1108 calculated using the MAST model[86], the results were then used as inputs for GSEA to evaluate the effect
1109 on selected pathways. Genes with p-value < 0.05 were considered as DEG.

1110 UMAP plots were generated by calculating UMAP embeddings using Seurat and then plotting them as
1111 scatter plots using ggplot2. Kernel density calculations for each gRNA were performed on UMAP
1112 embeddings using the MASS package via the kde2d function. The kernel density results were plotted as
1113 a raster layer with ggplot2 over the UMAP scatter plots. Finally, density contour lines were added using
1114 ggplot2's built-in 2D kernel density contour geom (geom_density_2d).

1115 **ATAC-Seq library preparation and sequencing:** ATAC-seq was performed as previously
1116 described[87]. Briefly, 5,000-50,000 viable cells were washed with cold PBS, collected by centrifugation,
1117 then lysed in resuspension buffer (RSB) (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2)

1118 supplemented with 0.1% NP40, 0.1% Tween-20, and 0.01% digitonin. Samples were incubated on ice
1119 for 3 min, then washed out with 1 ml RSB containing 0.1% Tween-20. Nuclei were pelleted by
1120 centrifugation at 500g for 10 min at 4°C then resuspended in 50 ul transposition mix (25ul 2x TD buffer,
1121 2.5 ul transposase (100 nM final), 16.5 ul PBS, 0.5 ul 1% digitonin, 0.5 ul 10% Tween-20, 5 ul H2O)
1122 and incubated at 37°C for 30 min in a thermomixer with 1000 RPM mixing. DNA was purified using a
1123 Qiagen MinElute PCR cleanup kit, then PCR amplified using indexed oligos. The optimal number of
1124 amplification cycles for each sample was determined by qPCR. Libraries were size selected using
1125 AmpureXP beads and sequenced using an Illumina NextSeq500 for 75bp paired-end reads.

1126 **ATAC-Seq analysis:** Paired-end 42-bp, or paired-end 75-bp reads were aligned to the M. musculus
1127 mm10 genome using BWA[88,89]with parameters "bwa mem -M -k 32". ATAC-seq peaks were called
1128 using MACS2[90] program using parameters "callpeaks -qvalue 5.0e-2 –shift -100 –extsize 200".
1129 Differentially accessible regions were identified using DESeq2[91] . Batch effect was removed using
1130 limma[92]. Heatmap visualization of ATAC-seq data was performed using pheatmap.
1131

1132 **Single-cell RNA sequencing metadata analysis:** Analysis was primarily performed in R (v 3.6.1) using
1133 the package Seurat (v 3.1)[82,93], with the package tidyverse (v 1.2.1)[94] used to organize data and the
1134 package ggplot2 (v 3.2.1) used to generate figures. scRNA-seq data from GSE10898, GSE99254,
1135 GSE98638, GSE199565, and GSE181785 were filtered to keep cells with a low percentage of
1136 mitochondrial genes in the transcriptome (< 5%) and between 200 and 3000 unique genes to exclude
1137 poor quality reads and doublets. Cell cycle scores were regressed when scaling gene expression values
1138 and T cell receptor genes were regressed during the clustering process, which was performed with the
1139 Louvain algorithm within Seurat and visualized with UMAP. To quantify the gene expression patterns,
1140 we used Seurat's module score feature to score each cluster based on its per cell expression of TFs.
1141

1142 To obtain Extended Data Fig. 6a, raw single-cell count data and cell annotation data was downloaded
1143 from NCBI GEO (GSE99254)[47]. Count data was normalized and transformed by derivation of the
1144 residuals from a regularized negative binomial regression model for each gene (SCT normalization
1145 method in Seurat[87], v4.1.1), with 5000 variable features retained for downstream dimensionality
1146 reduction techniques. Integration of data was performed on the patient level with Canonical Correlation
1147 Analysis as the dimension reduction technique[95]. Principal component analysis and UMAP dimension

reduction were performed, with the first 50 PCs utilized in UMAP generation. Cells were clustered utilizing the Louvain algorithm with multi-level refinement. The data was subset to CD8+ T cells, which were identified utilizing the labels provided by Guo et al[66]. Cell type labels were confirmed via 1) SingleR[96] (v1.8.1) annotation utilizing the ImmGen[97] database obtained via celda(1.10), 2) cluster marker identification, and 3) cell type annotation with the ProjecTILs T cell atlas[7] (v2.2.1). After sub-setting to $CD8^+$ T cells, cells were again normalized utilizing SCT normalization, with 3000 variable features retained for dimension reduction. Due to the low number of cells on the per-patient level, HArmstrongony[98](v1.0) was utilized to integrate the data at the patient level, rather than Seurat. PCA and UMAP dimensionality reduction were performed as above.

Statistical analyses: Statistical tests for flow cytometry data were performed using Graphpad Prism 10. *p-values* were calculated using either two-tailed unpaired Student's t-tests, one-way ANOVA or two-way ANOVA as indicated in each figure. Linear regressions were performed using the ordinary least squares method in R (v 3.6.1). All data were presented as the mean ± s.e.m. The P values were represented as follows: ****$P < 0.0001$, ***$P < 0.001$, **$P < 0.01$, and *$P < 0.05$.

Data availability

ATAC-sequencing data from this paper will be deposited in the GEO database (GSE279498). Taiji v2.0 output of this study (TF activity atlas, TF-TF interaction maps, and TF activity on genome browser view) will be available at our $CD8^+$ T cell TF atlas portal (https://wangweilab.shinyapps.io/Tcellstates/) and interactive interface for TF atlas exploration (https://huggingface.co/spaces/taijichat/test-taijichat).

1. Login: Hugging Face account at https://huggingface.co/login using ID: taijichat-reviewers, password: HelloTaij1Chat#3425
2. Access TaijiChat: After logging in, navigate to the TaijiChat testing page (https://huggingface.co/spaces/taijichat/test-taijichat)
3. Initial Loading: Upon your first visit, TaijiChat will automatically load necessary packages and build the system from scratch. This initial setup may take a few minutes. You will know TaijiChat is ready when a green "running" indicator appears in the top-left corner of the webpage.
4. Testing Chat Functionality: Once TaijiChat is running, select the "Chat" tab on the website. You may then begin entering questions to test the system. Note that the first query might respond slowly due to system initialization.
5. Accuracy and Limitations: TaijiChat is powered by state-of-the-art Large Language Models (LLMs), which can still occasionally make mistakes. Always verify critical scientific information against original research papers and authoritative sources.

All other raw data are available from the corresponding author upon request.

Code availability

All scripts and Taiji v2.0 package will be uploaded at GitHub

(https://github.com/Wang-lab-UCSD/Taiji2).

Methods references

74. Zheng, C. *et al.* Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. *Cell* **169**, 1342–1356.e16 (2017).
75. Weirauch, M. T. *et al.* Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell* vol. 158 1431–1443 Preprint at https://doi.org/10.1016/j.cell.2014.08.009 (2014).
76. Zhao, T., Liu, H., Roeder, K., Lafferty, J. & Wasserman, L. The huge Package for High-dimensional Undirected Graph Estimation in R. (2020) doi:10.48550/arXiv.2006.14781.
77. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
78. Schönfeld, M. & Pfeffer, J. Fruchterman/Reingold (1991): Graph Drawing by Force-Directed Placement. *Schlüsselwerke der Netzwerkforschung* 217–220 Preprint at https://doi.org/10.1007/978-3-658-21742-6_49 (2019).
79. Bailis, W. *et al.* Author Correction: Distinct modes of mitochondrial metabolism uncouple T cell differentiation and function. *Nature* **573**, E2 (2019).
80. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
81. Wang, B. *et al.* Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nat. Protoc.* **14**, 756–780 (2019).
82. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* vol. 177 1888–1902.e21 Preprint at https://doi.org/10.1016/j.cell.2019.05.031 (2019).
83. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
84. Alquicira-Hernandez, J. & Powell, J. E. Nebulosa recovers single-cell gene expression signals by kernel density estimation. *Bioinformatics* **37**, 2485–2487 (2021).
85. Marsh, S. E. scCustomize: custom visualizations & functions for streamlined analyses of single cell sequencing. *Preprint at https://doi. org/10.5281/zenodo*.
86. Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
87. McDonald, B. *et al.* Canonical BAF complex activity shapes the enhancer landscape that licenses CD8 T cell effector and memory fates. *Immunity* **56**, 1303–1319.e5 (2023).
88. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, (2009).
89. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, (2010).
90. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, 1–9 (2008).
91. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
92. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
93. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
94. Wickham, H. *et al.* Welcome to the Tidyverse. *Journal of Open Source Software* vol. 4 1686 Preprint at https://doi.org/10.21105/joss.01686 (2019).
95. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
96. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
97. Heng, T. S. P., Painter, M. W. & Immunological Genome Project Consortium. The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008).
98. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
99. Zhang, L. *et al.* Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. *Cell* **181**, 442–459.e29 (2020).

Acknowledgments

We thank members of the Chung, Kaech and Wang laboratory for their advice and assistance with experiments. We appreciate UNC Lineberger Preclinical Research Unit at the University of North Carolina at Chapel Hill assisted animal experiments which is supported in part by an NCI Center Core Support Grant (CA16086) to the UNC Lineberger Comprehensive Cancer Center. This work was supported by NIH grants to S.M.K. R37AI066232, R01AI123864, R21AI151986, and R01CA240909, to W.W. R01AI150282, R01HG009626 (in part) to H.K.C. K01EB034321, to J.J.M. R01AI177864 to J.E.T. R01CA248359 and R01CA244361, to D.C.H. AI151123, to Y.W. EB029122 and GM140929, and to L.L. K01EB035649. H.K.C. is a Damon Runyon Fellow supported by Damon Runyon Cancer Research Foundation grant [DRG-2374-19].

Author contributions

H.K.C. designed and performed experiments. A.B., B.M.R., F.X., Y.L., K.Y., D.J., M.S., J.G., S.M., E.C., B.M. D.C., B.C., Q.Y., F.H., T.H.M., S.K.V., V.T., U.H.C., G.D., B.S., Y.W., L.L., P.H., and J.H. assist with experiments. H.K.C., C.L., A.J., N.J.B., M.S. analyzed the sequencing and bioinformatics data. J.W., B.M., M.S., Y.H., L.H., U.H.C., J.M., contribute to bioinformatics data analysis. Z.W., A.W. contributed to TaijiChat and portal generation. J.H. and D.H. provided resources and training for ATAC-sequencing. B.P.R. and J.E.T. performed proteasome activity experiments. H.K.C., S.M.K., W.W., Y.W., L.L., B.S., G.P., J.E.T., J.J.M., D.C.H, provided scientific input and acquired funding. H.K.C, and C.L. wrote the original manuscript. H.K.C., W.W., and S.M.K. reviewed and edited the paper. H.K.C., S.M.K., and W.W. conceptualized, designed, and supervised the study.

Competing interests

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

Additional information

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

1268 **Extended Data Figure Legends**

1269 **Extended Data Fig. 1 | Parallel differentiation of T_{RM} and TEX_{term} and their transactional and**
1270 **epigenetic similarity. a,** UMAP of scRNA-seq data of T cells from blood, tumor, and adjacent normal
1271 tissues of CRC[99], NSCLC[47], and HCC[74] patients. Unbiased clustering identified multiple T cell states
1272 consistent with those observed murine LCMV infection and tumors. **b,** T_{RM} marker genes show higher
1273 expression in TEX_{term} cluster from Pan-cancer scRNA-seq in **a**. **c,** Both T_{RM} and TEX_{term} clusters
1274 upregulate exhaustion-[17] and T_{RM}[39]-associated gene signatures. **d,** Pearson correlation matrix of batch-
1275 corrected ATAC-seq datasets[3,9,37,38]. Color and size are both proportional to correlation strength.

1276

1277 **Extended Data Figure 2 | Composition of multi-omic atlas**. **a,** A total of 121 experiments across
1278 multiple data sets[3,9,20,27,36–40] were utilized to generate an epigenetic and transcriptional atlas of $CD8^+$ T
1279 cells under chronic and acute antigen exposure. **b,** Batch-effect corrected ATAC-seq signal at the *Ifng*
1280 locus across nine $CD8^+$ T cell states.

1281

1282 **Extended Data Fig. 3 | Cataloging key TFs across $CD8^+$ T cell states. a,** Logic flow of the unbiased
1283 PageRank comparison to classify single-state/multi-state TFs. **b,** Number of TFs catalogued in each cell
1284 state. **c,** Venn Diagrams showing overlap of TFs with the T_{RM} cell state. **d,** TF activity score (normalized
1285 PageRank) of previously reported TEX_{term}-preventing TFs (TFAP4, JUN, BATF3, and BATF), newly
1286 identified TEX_{term} single-state TFs (JDP2, ZFP324, ZBTB49, ZFP143, ZSCAN20), and NFATC1, a
1287 known TEX_{term}-associated TF.

1288

1289 **Extended Data Fig. 4 | TF wave analysis. a,** Schematic of the analysis pipeline. **b,** Selection of
1290 algorithms and parameters for TF wave analysis. The Pearson correlation was chosen for the distance
1291 metric, with k=7 chosen as the optimal cluster number. **c,** Seven TF waves. Circles represent specific
1292 cell states. Red color indicates normalized PageRank scores. **d,** List of TF members in each wave. **e**,
1293 Heatmap of biological pathways enriched in each TF wave. Red-blue color scale indicates the p-value.

1294

1295 **Extended Data Fig. 5 | TF–TF association network in TEX_{term} and T_{RM} cell states. a–c,** TF–TF
1296 associations of (**a**) TEX_{term} single-state TFs (JDP2, ZFP324, and IRF8); (**b**) T_{RM} single-state TFs (FOSB,
1297 SNAI1, and KLF6); and (**c**) multi-state TFs shared by TEX_{term} and T_{RM} (PRDM1, FLI1, and GFI1). Line

1298 color indicates state specificity: T_{RM} (green) or TEX_{term} (brown) state. Line thickness represents
1299 interaction intensity.
1300
1301 **Extended Data Fig. 6 | TF network analysis reveals proteasome pathway enrichment in TEX_{term}**
1302 **state with diminished tumor control function. a,** Pseudotime analysis of $CD8^+$ TIL scRNA-seq data
1303 from NSCLC patients (n = 14), showing a positive correlation between proteasome gene scores (KEGG:
1304 M10680) and T cell exhaustion. **b,** Gene set enrichment analysis (GO:0043161, Proteasome-mediated
1305 ubiquitin-dependent protein catabolic process) of RNA-seq from $CD8^+$ splenocytes (black) and TILs
1306 (purple), n = 3 each. **c,** Tumor growth of $Rag2^{-/-}$ mice bearing MCA-205 sarcomas infused with
1307 $proteasome^{high}$ or $proteasome^{low}$ $CD8^+$ TILs isolated from C57BL/6 mice bearing MCA-205 tumors. n =
1308 5 per group. Two-way ANOVA Tukey's multiple comparison test were performed. ****P<0.0001,
1309 ***P<0.001, **P<0.01, *P<0.05.
1310
1311 **Extended Data Fig. 7 | *in vivo* Perturb-seq with dual-guide RNA in LCMV chronic infection. a,**
1312 Retroviral vector design for dual-gRNA delivery. **b,** Feature plot of differentiation markers (Pan-
1313 exhaustion: red, TEX_{prog}: pink, TEX_{eff}: blue, Cell cycle: purple, TEX_{term}: brown). **c,** Heatmap of marker
1314 gene expression across cell state clusters identified by Seurat's Find Markers() function.
1315
1316 **Extended Data Fig. 8 | Individual and co-transfer of TF KO P14 $CD8^+$ T cells identifies novel TFs**
1317 **involved in TEX_{term} differentiation during chronic viral infection. a–b,** Quantification of $Cas9^+$ P14
1318 $CD8^+$ T cells transduced with TF-targeting gRNAs isolated from LCMV-Clone 13-infected mice.
1319 TEX_{term} single-state TFs are in bold. **a,** Frequency of $PD1^+$ $SLAMF6^-$ $CD101^+$ cells. **b,** Frequency of
1320 $CD38^+$ $CD39^+$ double-positive cells. Statistical analysis: Ordinary one-way ANOVA with Dunnett's
1321 multiple comparisons test versus gScramble (n ≥ 5 from ≥2 biological replicates). Data are presented as
1322 mean ± s.e.m. ****P< 0.0001; ***P < 0.001; **P < 0.01; *P < 0.05.
1323
1324 **Extended Data Fig. 9 | Single-cell transcriptomic profiling of $CD8^+$ T cells from Perturb-seq in**
1325 **acute LCMV infection. a,** Feature plots of differentiation marker genes (T_{CM}: yellow, T_{EM}: blue, T_{RM}-
1326 *Itgae*^{low}: dark blue, T_{RM}: green). **b,** Heatmap of differentially expressed genes between T_{RM}-*Itgae*^{low} and
1327 T_{RM} clusters. **c,** Volcano plots of differentially expressed genes in $Cas9^+$ P14 $CD8^+$ T cells expressing
1328 *gEtv5*, *gArid3a*, g*Hic1*, or *gGfi1*.

1329

1330 **Extended Data Fig. 10 | T_{RM} single-state TF, *Klf6* overexpression does not accelerate T cell**
1331 **exhaustion. a,** Experimental setup: *Klf6*-RV or control-RV transduced P14 $CD8^+$ T cells co-transferred
1332 into mice infected with chronic LCMV-Clone 13. **b-e,** Quantification of the frequency of (**b**) $PD1^+$, (**c**)
1333 TEX_{prog} ($PD1^+$ $SLAMF6^+$ $CX3CR1^-$), (**d**) TEX_{eff} ($PD1^+$ $SLAMF6^-$ $CX3CR1^+$) and (**e**) TEX_{term} ($PD1^+$
1334 $SLAMF6^-$ $CX3CR1^-$) populations. Paired t-tests ($n \geq 6$ from ≥ 2 biological replicates). Data are presented
1335 as mean ± s.e.m. ****P <0.0001, ***P < 0.001, **P < 0.01, *P < 0.05.

1336

1337 **Extended Data Fig. 11 | Depletion of TEX_{term}-single state TF, *Zscan20* and TEX_{term} and T_{RM} multi-**
1338 **state TF, *Hic1* reduces T cell exhaustion. a,** Co-transfer of $Cas9^+$ P14 $CD8^+$ T cells transduced with
1339 RV-g*Zscan20* or RV-g*Hic1*, mixed with gRNA control RV transduced cells into B16-GP33 tumor-
1340 bearing mice. **b,** Representative flow plots of $PD1^+$ P14 $CD8^+$ T cells stained for SLAMF6 and CX3CR1.
1341 **c-e,** Quantification of (**c**) TEX_{prog} ($PD1^+$ $SLAMF6^+$ $TIM3^-$), (**d**) TEX_{term} ($PD1^+$ $SLAMF6^-$ $CX3CR1^-$),
1342 and (**e**) non-progenitor, exhausted ($PD1^+$ $SLAMF6^-$ $TIM3^+$) popultation. Paired t-tests ($n \geq 6$ from ≥ 2
1343 biological replicates). Data are presented as mean ± s.e.m. ****P < 0.0001, ***P < 0.001, **P < 0.01,
1344 *P < 0.05.

1345

1346 **Extended Data Fig. 12 | Conservation of single- and multi-state TFs in human pan-cancer TEX_{term}**
1347 **and T_{RM} cell states. a,** Human pan-cancer datasets utilized in this study[52–59]. **b,** Cluster-specific marker
1348 mRNA expression across integrated pan-cancer single-cell multi-omics datasets. **c,** TF activity analysis
1349 using Taiji on matched matched scRNA-seq and scATAC-seq datasets from various human cancers,
1350 including ccRCC, GBM, BCC, HNSCC, HCC, and RCC. $CD8^+$ T cell states were annotated using
1351 canonical marker gene expression. PageRank scores derived from Taiji were log-transformed, averaged
1352 per state, and z-score normalized. Results were visualized with a focus on TF activity T_{RM} and TEX_{term}
1353 cell states. **d,** mRNA expression of TFs were compared across human $CD8^+$ T cell states. The dataset
1354 encompassed T cells from 15 tumor types. Cell type annotations provided by the original study were
1355 retained and mapped to the following broad categories of clusters: T_{RM}, TEX_{term}, T_{EM}, Memory, Naive,
1356 and cell cycle. Seurat's AverageExpression function was used, followed by z-score normalization. Data
1357 visualization emphasized comparisons between T_{RM} and TEX_{term} cell states. Cell-state-selectivity
1358 conserved TFs in humans and mice are highlighted in bold (**c, d**).

1359

1360 **Extended Data Fig. 13 | CRISPR KO validation of *ZSCAN20* and *JDP2* in human PBMCs.** Indel
1361 frequencies, representative Sanger sequencing traces, and indel distributions for (**a**) *ZSCAN20* and (**b**)
1362 *JDP2* KOs.
1363
1364 **Extended Data Fig. 14 | Tumor control of TEX_{term} single-state TF deficient T cells. a,** Survival of
1365 B16-GP33 bearing mice receiving *Zscan20* KO or control P14 $CD8^+$ T cells, followed by anti-PD1 or
1366 isotype IgG2a treatment. **b,** Survival of tumor-bearing mice with *Jdp2* KO P14 $CD8^+$ T cell transfer.
1367
1368 **Extended Data Tables**
1369 Extended Data Table 1 | Composition of multi-omic atlas
1370 Extended Data Table 2 | TF PageRank scores of multi-state and single-state TFs
1371 Extended Data Table 3 | TF PageRank scores of T_{RM}- and TEX_{term} single-state TFs and multi-state TFs
1372 Extended Data Table 4 | Catalog of cell-state important TFs
1373 Extended Data Table 5 | Catalog of TF communities in T_{RM} and TEX_{term}
1374 Extended Data Table 6 | gRNA sequences for TF gRNA library and RNP gRNA
1375 Extended Data Table 7 | Perturb seq cluster statistics (Fig. 3 and Fig. 4)
1376 Extended Data Table 8 | Gene signatures
1377

Extended Data Note

Development of TaijiChat: An Integrated Conversational Interface for Multi-Omics Data Exploration

Overview of the Multi-Agent Architecture and Design Principles of TaijiChat: To facilitate intuitive and interactive exploration of our Taiji multi-omics atlas, we developed TaijiChat, a web-based chatbot platform enabling natural language querying of both the study's datasets and associated literature. Unlike generic large language models (LLMs), TaijiChat provides domain-specific responses grounded in the content of our manuscript, supplementary materials, and underlying transcriptomic and epigenomic datasets. This tight integration with structured outputs—such as TF activity scores, cell clustering results, and motif enrichment profiles—enables users to dynamically engage in context-aware dialogues with the data. Users can directly query TF activity (e.g., "Which TFs are enriched in TRM cells in GBM?"), request visualizations (e.g., UMAP plots or violin plots of selected genes), or retrieve ranked TF lists by specific cell states or tumor contexts.

In addition to data interrogation, TaijiChat incorporates a literature-aware component. Each generated response is cross-referenced with relevant supporting or contrasting studies from databases such as PubMed and GEO, allowing users to validate findings, assess consistency with prior work, and develop novel hypotheses. For example, querying the role of KLF6 in TRM differentiation yields not only a summary of its activity across tumor-infiltrating lymphocytes (TILs) and associated figures from our atlas, but also citations to independent literature corroborating its specific function in TRM cells.

To systematically enable this advanced functionality, we designed TaijiChat as a sophisticated multi-agent artificial intelligence system, integrating natural language processing, safety-verified code generation, and domain-specific expertise. The system employs a manager-orchestrated architecture consisting of specialized agents responsible for planning, validation, and secure code execution. Specifically, this architecture embodies four core design principles critical for scientific computing applications:

1. Safety-first architecture: Ensures all generated analytical code undergoes mandatory validation through multiple security layers prior to execution.

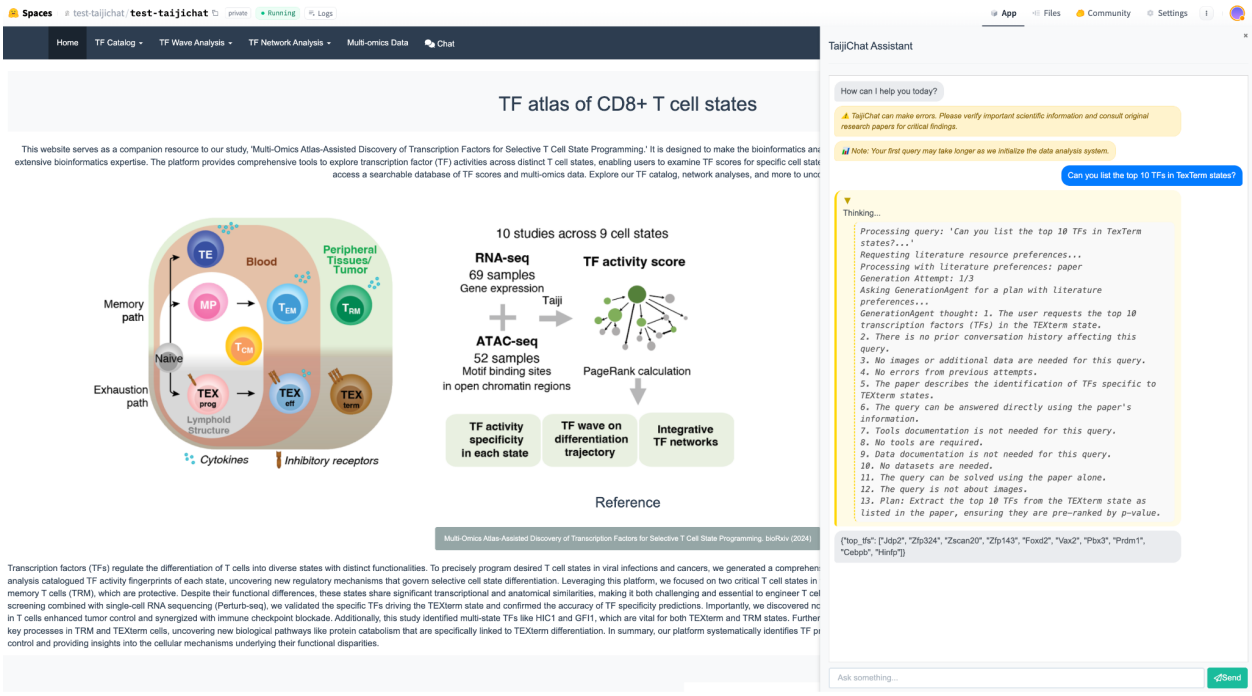
2. Separation of concerns: Assigns distinct, specialized roles to individual agents, facilitating modularity, scalability, and maintainability.

3. Intelligent orchestration: Provides centralized coordination of complex multi-step workflows, ensuring coherent interactions across system components.

4. Domain expertise integration: Directly incorporates scientific methodologies and domain-specific knowledge into the reasoning and analysis processes, enabling contextually appropriate and scientifically rigorous responses.

To evaluate TaijiChat's effectiveness, we curated a benchmark set of 21 expert-driven immunology questions, comparing responses from TaijiChat to those from a general-purpose LLM lacking direct data integration. TaijiChat demonstrated accurate, evidence-supported responses in all tested scenarios, highlighting its utility as a domain-specific scientific assistant.

Collectively, TaijiChat represents a platform for hypothesis generation and data interpretation, effectively combining the intuitive accessibility of conversational AI with rigorous, integrated multi-omics data analysis and automated literature synthesis.



1.1 Agent Architecture and Coordination: The system consists of four specialized agents operating under centralized coordination: the Manager Agent, Generation Agent, Supervisor Agent, and Executor Agent.

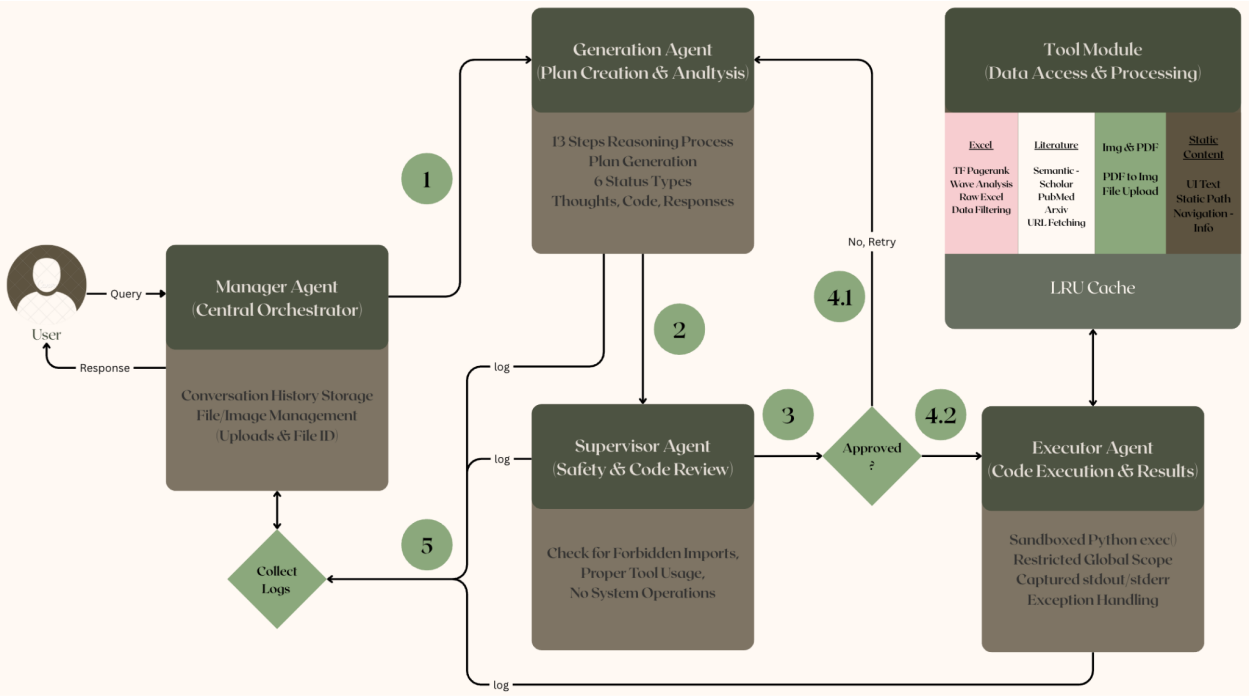
The Manager Agent serves as the central orchestrator in a multi-agent architecture designed for scientific computing and code generation tasks. The system employs a hierarchical workflow where the Manager coordinates three specialized sub-agents: a GenerationAgent that transforms natural language queries into executable Python code and structured execution plans, a SupervisorAgent that performs safety validation and code review to prevent potentially harmful operations, and an ExecutorAgent that executes approved code in a controlled environment. The Manager implements a sophisticated query processing pipeline with built-in retry logic (up to three attempts) to handle generation failures, maintains conversation history for context-aware responses, and supports multi-step computational workflows where intermediate results inform subsequent processing steps. This design ensures robust, safe, and efficient processing of complex scientific queries while maintaining clear separation of concerns across the agent hierarchy.RetryClaude can make mistakes.

The Generation Agent functions as the intelligent planner, employing a structured 13-step reasoning process for query understanding and solution development. This agent integrates domain-specific knowledge of transcription factor biology and experimental methodologies, enabling scientifically appropriate interpretations of natural language queries. The reasoning framework encompasses query decomposition, conversation history integration, multi-modal content evaluation, scientific domain knowledge activation, tool capability assessment, and comprehensive execution plan synthesis.

The Supervisor Agent implements a critical safety validation layer through multi-stage code review processes. This agent performs static code analysis, security pattern detection, and tool usage verification before any code execution. The validation process examines code structure, prevents dangerous system operations, validates tool usage patterns, and provides detailed feedback for rejected code with specific remediation suggestions.

The Executor Agent provides secure code execution within a sandboxed Python environment. This agent implements restricted global scope access, limiting available modules to approved scientific computing libraries and custom tools. All code execution occurs in isolation with comprehensive output capture

and sanitized error reporting, preventing system exposure while maintaining informative debugging capabilities.



1.2 Safety and Security Architecture: The system implements a three-layer security model designed to prevent unsafe operations while maintaining scientific functionality. Layer 1 provides generation-time safety through security-aware code templates and built-in constraint knowledge within the Generation Agent. Layer 2 implements mandatory pre-execution review through the Supervisor Agent's pattern-based security analysis and tool compliance verification. Layer 3 enforces execution isolation through the Executor Agent's sandboxed environment with restricted resource access and output sanitization. Error recovery mechanisms include intelligent retry strategies with supervisor feedback integration, progressive query simplification for complex requests, and maintenance of conversation context across retry attempts. The system limits retry attempts to three per query while incorporating learned patterns from previous failures to improve subsequent attempts.

1.3 Tools Integration and Performance Optimization: The system interfaces with a comprehensive tools module providing specialized functions for scientific data analysis. Data processing capabilities include Excel file processing with proper formatting, temporal pattern analysis, correlation matrix access, and intelligent data subsetting. Literature integration tools provide multi-source academic database searches

across PubMed, ArXiv, and Semantic Scholar, with automated content extraction and deduplication algorithms. Multi-modal analysis capabilities support AI-powered image analysis, PDF document processing, and file format detection. Performance optimization employs a multi-level caching strategy including function-level LRU caching, data-level timeout caching, schema caching for persistent structures, and session-based query-response caching. Resource efficiency measures include lazy loading of datasets, batch processing for large operations, automatic memory management, and intelligent API rate limiting for external services

1.4 Technical Implementation and Performance: The system utilizes OpenAI GPT-4o models for both text and vision processing, operating within a Python 3.x runtime environment with scientific computing libraries including pandas and openpyxl. External API integration provides access to major academic databases while maintaining security through input validation and sandboxed execution.

1.5 Architectural Advantages for Scientific Applications: This multi-agent architecture provides several key advantages for scientific computing applications. The multi-layer validation system ensures research integrity through multiple checkpoints, preventing unsafe operations while maintaining detailed audit trails of all agent interactions. The modular design enables easy maintenance and extensibility, allowing individual agent updates without system-wide modifications and seamless integration of new analytical capabilities. Scientific rigor is emphasized through reproducible operations with consistent outputs, transparent reasoning traces for scientific validation, comprehensive academic literature integration, and built-in domain expertise for methodologically appropriate analyses. The system's fail-safe design defaults to safe states during error conditions while providing robust error recovery with intelligent retry mechanisms.

Figure 1

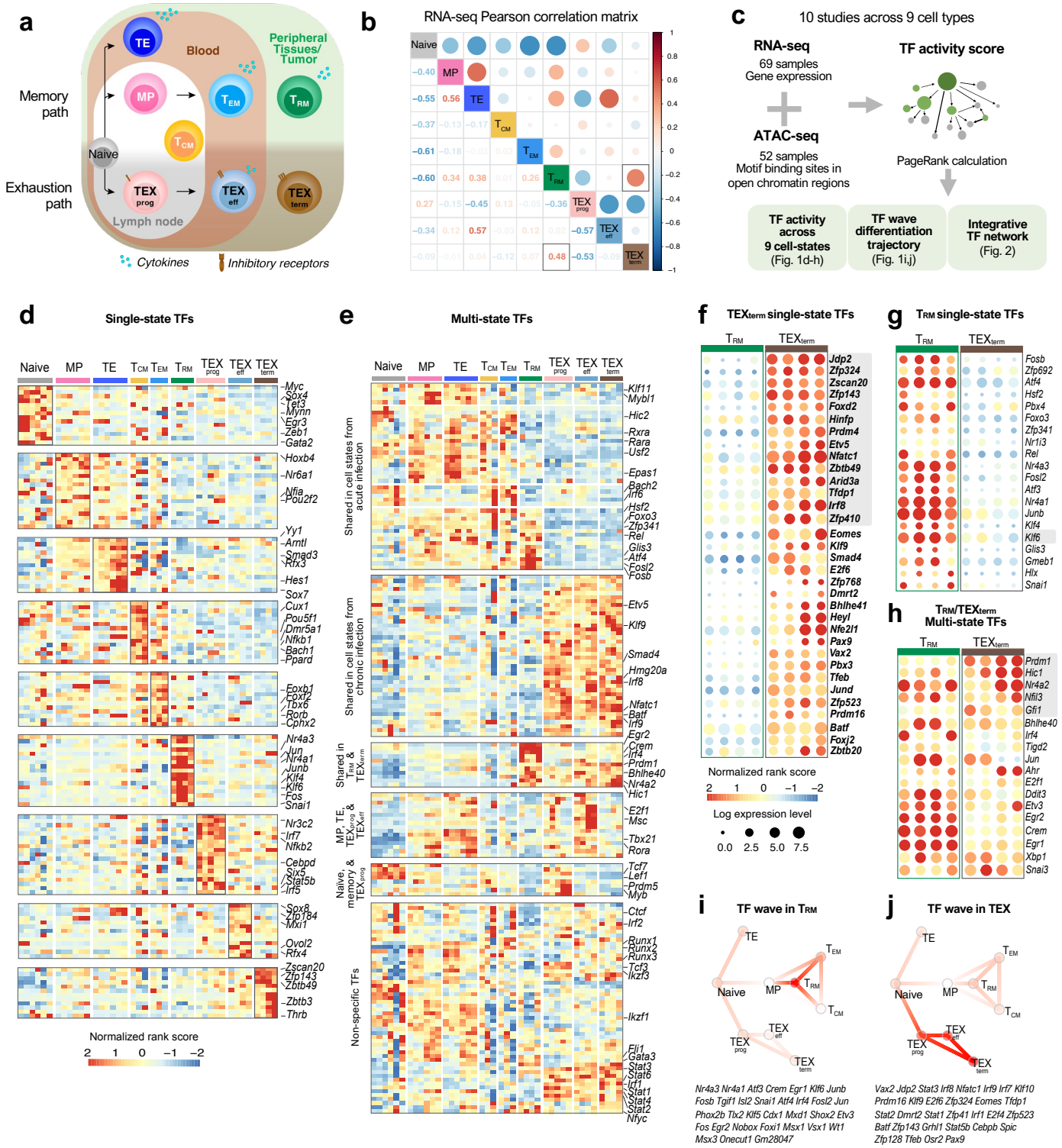


Fig. 1| Transcription and epigenomic atlas of CD8⁺ T cell differentiation states and cell state-selective TF identification pipeline. a, The diagram outlines several CD8⁺ T cell state trajectories in acute and chronic infection/tumor where various memory and exhaustion states are acquired. Parallel differentiation of T_{RM} and TEX_{term} cells and their similarity in anatomical location. **b,** Pearson correlation matrix of batch effect corrected RNA-seq. Color and size are both proportional to correlation values, with red being the highest correlation. **c,** Pipeline of our integrative analysis. The matched RNA-seq and ATAC-seq data[3,9,20,27,36–40] were used as input for Taiji algorithm to construct a regulatory network and output PageRank score matrix representing the activity of TF. The downstream analysis included the identification of single-state TFs, as well as the construction of transcriptional waves and TF networks. **d-h,** TFs in rows (z- normalized), samples in columns, and the colors of the cell in the matrix indicate the normalized PageRank scores. Each column represents a different data set. **d,** PageRank scores of 136 single-state TFs. **e,** 173 multi-state TFs. **f-h,** Bubble plots of PageRank scores and expression of TFs that are selectively active in TEX_{term} state **(f),** T_{RM} state **(g),** and multi-state TFs that are active in both TEX_{term} and T_{RM} cells **(h).** TEX_{term} single-state TFs are in bold. Color represents the normalized PageRank score with red displaying high values, and circle size represents the logarithm of mRNA expression across 5 different data sets. TFs are ordered based on the p-value. TFs validated *in vivo* are highlighted in gray. **i-j,** Transcription factor waves associated with exhaustion **(i)** or T_{RM} state **(j).** Sample numbers for each cell state and the statistics used to curate the TF catalogue are provided in Extended Data Fig. 2a and Extended Data Fig. 3a, respectively.

Figure 2

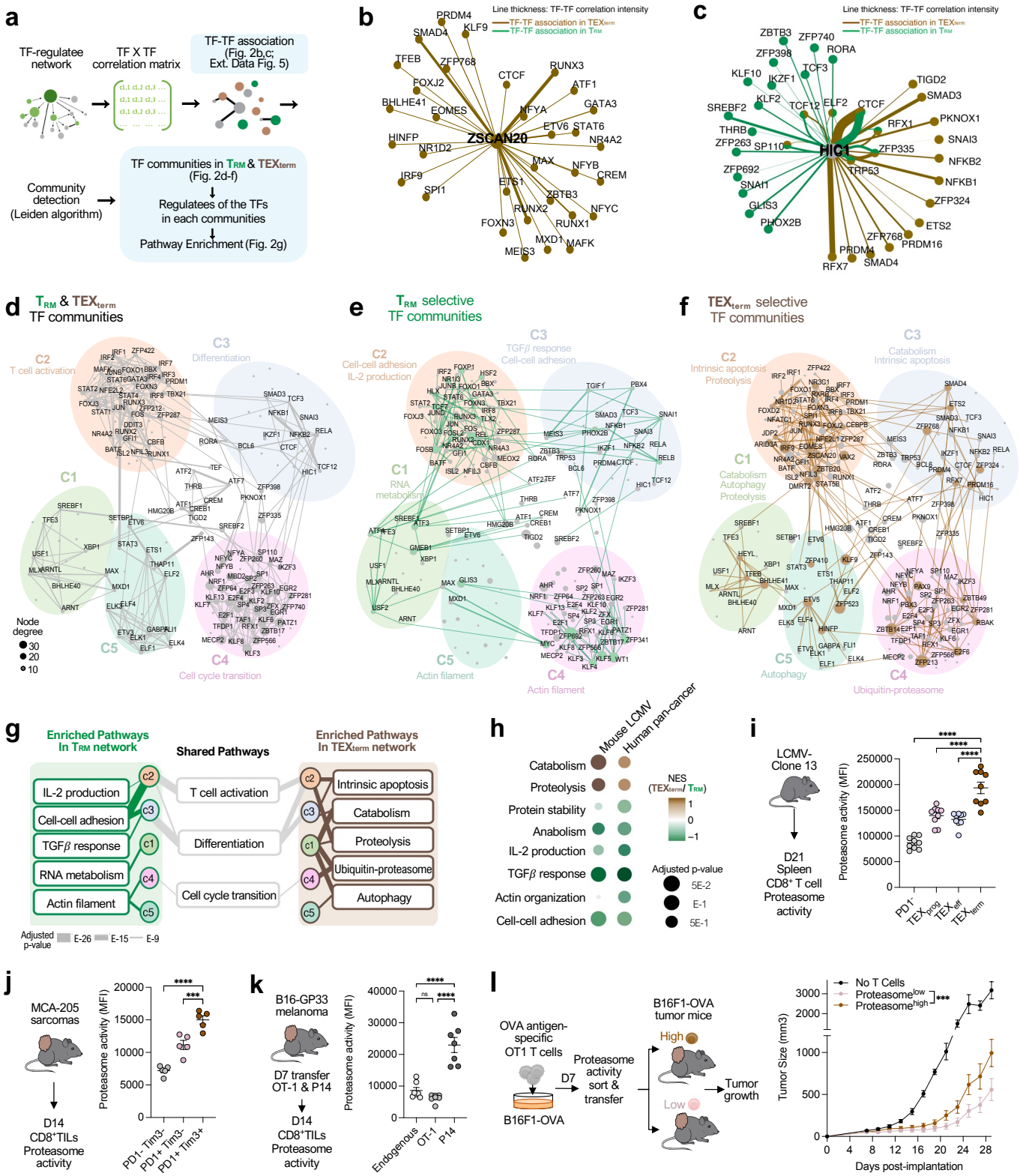


Fig. 2| Global analysis of TF networks in TEX_{term} and T_{RM} cell states. a, TF network analysis in T_{RM} and TEX_{term} cells encompasses both TF-TF association (blue, Fig. 2b-g). **b-c,** TF-TF association network plots focus on TEX_{term} single-state TFs like ZSCAN20 **(b)** and multi-state TFs like HIC1 **(c)** showing predicted TF-TF network in T_{RM} (green) or TEX_{term} (brown) cells. **d-f,** TF-TF association clustering generates five TF communities in T_{RM} and TEX_{term} cells. The overall community topology is shaped by shared TFs (gray) in both T_{RM} and TEX_{term} cells **(d)** while the TF-TF associations exclusive in T_{RM} **(e)** and TEX_{term} **(f)** cells are shown as green and brown lines, respectively. **g,** Summary of the shared and unique pathways enriched in each community in T_{RM} and TEX_{term} cells. The thickness of the lines reflects p-value. (Gene sets: Extended Data Table. 8). **h,** GSEA of indicated pathways in TEX_{term} vs. T_{RM} cells using batch-effect corrected LCMV bulk-RNA-seq[3,9,20,27,36–40] and human pan-cancer scRNA-seq data sets[47,59,74]. **i-k,** Flow cytometry analysis of proteasome activity. TEX_{term} cells show the highest proteasome activity in LCMV-Clone 13 (chronic strain) infected mice **(i),** or MCA-205 tumors **(j). k,** Dual transfer of GP33-41 antigen-specific (P14) $CD8^+$ T cells and bystander (OT-1) $CD8^+$ T cells in B16-GP33 tumor bearing mice. **l,** Evaluation of B16F1-OVA tumor growth by OVA antigen-specific OT-1 cells with differential proteasome activity. C57BL/6 mice bearing tumors were infused with $proteasome^{high}$ or $proteasome^{low}$ sorted OT-1 cells, which were pre-stimulated for 7 days with B16F1-OVA tumor cells. Tumor growth was monitored every two days. Data are presented as the mean ±s.e.m. Ordinary one-way ANOVA **(i-k),** and two-way ANOVA Tukey's multiple comparison test **(l)** were performed. ****P<0.0001, ***P<0.001, **P<0.01, *P<0.05.

Figure 3

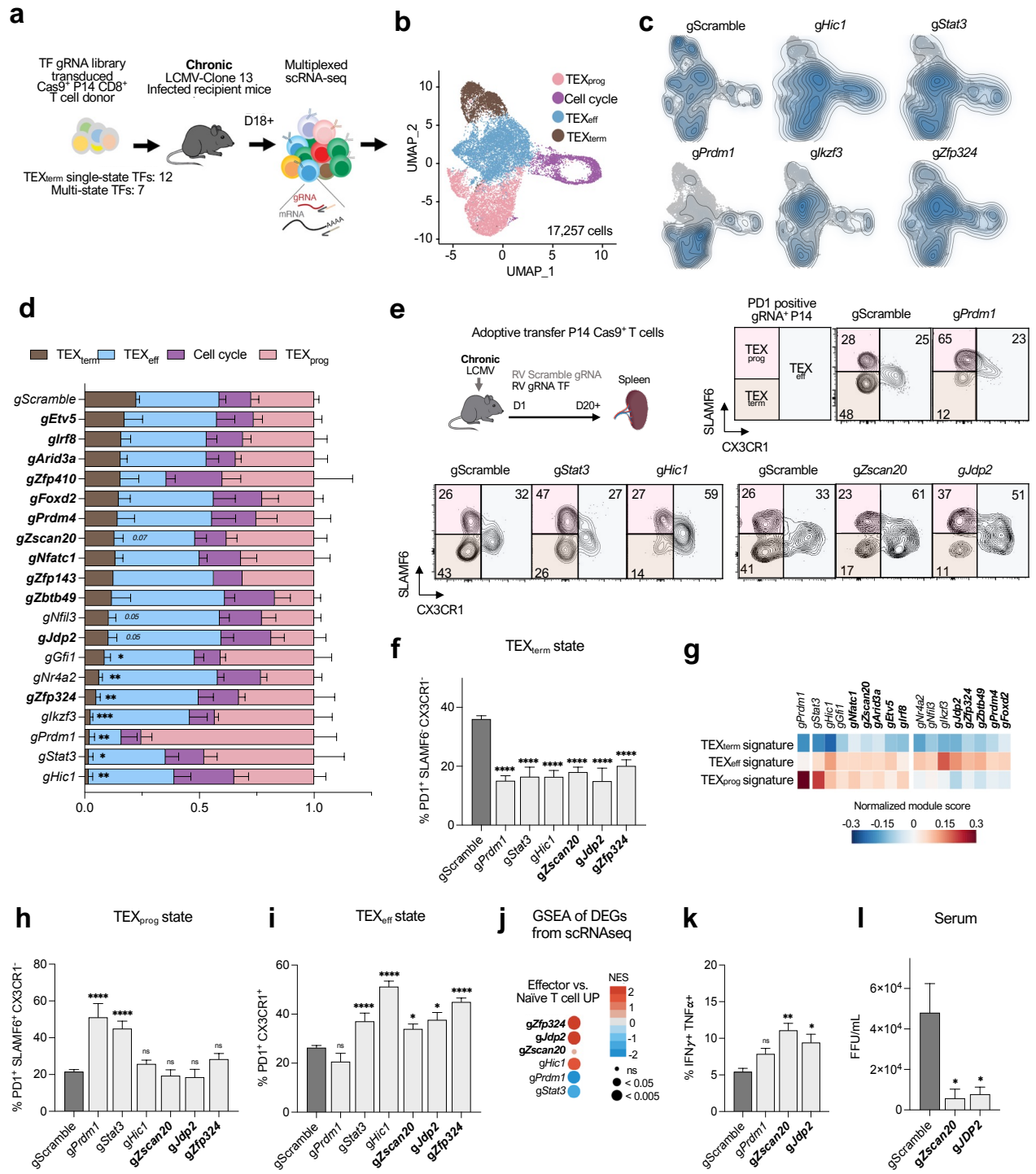


Fig. 3| *In vivo* Perturb-seq validation of TEX$_{term}$-driving TFs. a, Schematic of the *in vivo* Perturb-seq strategy. Dual Cas9$^+$ and P14$^+$ TCR transgenic CD8$^+$ T cells that recognize LCMV epitope GP33-41 were transduced with retroviruses (RVs) expressing the gRNA library, adoptively transferred into mice previously infected one day earlier with LCMV-Clone 13, and analyzed 18-23 days later by scRNA-seq. **b**, UMAP showing TEX$_{prog}$, TEX$_{eff}$, TEX$_{term}$, and cell cycle clusters. Feature plots and heatmap of the cluster marker mRNA expression are in Extended Data Fig. 7b,c. **c**, Kernel density of each TF gRNA-expressing cells. **d**, Cluster distribution of cells for each TF gRNA$^+$ cells. TEX$_{term}$ single-state TFs are indicated in bold. Data represent five pooled replicates from three independent experiments. Values are shown as mean ± s.e.m. Statistical analysis was performed using two-way ANOVA followed by Fisher's LSD test, with comparisons made relative to the gScramble control. Statistical results for TEX$_{term}$ and TEX$_{prog}$ populations are shown; additional statistical comparisons are provided in Extended Data Table 7. ****P < 0.0001, ***P < 0.001, **P < 0.01, *P < 0.05. **e**, Individual phenotyping of the TF KOs in LCMV-Clone 13 infected mice based on flow cytometry and representative flow plots. **f**, Quantification of the frequency of TEX$_{term}$ (PD1$^+$ SLAMF6$^-$ CX3CR1$^-$) subsets within P14 CD8$^+$ T cells. **g**, TEX$_{term}$, TEX$_{prog}$, and TEX$_{eff}$ gene signature analysis in each gRNA KO compared to gScramble. Gene signatures were generated from ref. 49 (Extended Data Table 8). **h**, Quantification of the frequency of TEX$_{prog}$ (PD1$^+$ SLAMF6$^+$ CX3CR1$^-$) cells. **i**, Quantification of the frequency of TEX$_{eff}$ (PD1$^+$ CX3CR1$^+$) cells. **j**, GSEA of an effector-associated gene set between each TF KOs *vs.* control. **k-l**, P14 CD8$^+$ T cells lacking the indicated TFs were compared to gScramble controls for IFNγ and TNFα production **(k)** or virus titers **(l)** in LCMV-Clone 13 infected mice. Statistic analysis for **f, h, i, k, l**: Ordinary one-way ANOVA with Dunnett's multiple comparison test versus gScramble (n ≥ 6 from ≥ 3 biological replicates). Data are presented as the mean ± s.e.m. ****P<0.0001, ***P<0.001, **P<0.01, *P<0.05.

Figure 4

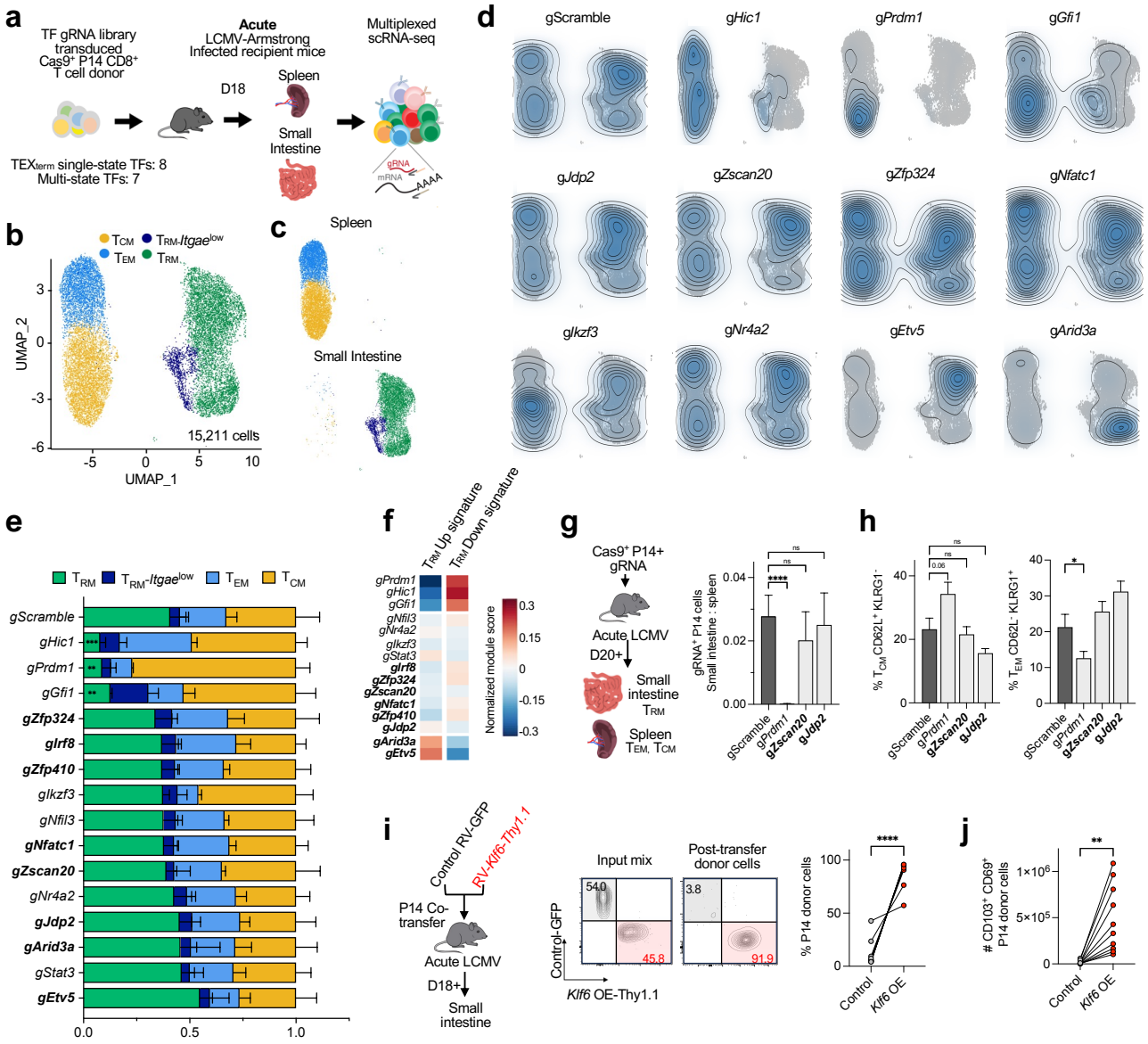


Fig. 4| Functional validation of TFs with distinct roles in TEX_{term} and T_{RM} differentiation. a, Schematic of *in vivo* Perturb-seq screening in memory $CD8^+$ T cells during acute LCMV-Armstrong infection. The donor cells were then analyzed for T_{RM} cells in the small intestine (SI) and T_{CM} and T_{EM} cells in the spleen. **b,** UMAP showing T_{CM}, T_{EM}, T_{RM} and T_{RM}- *Itgae*^{low} clusters. **c,** Differential clustering of cells from different tissue origins. **d,** Kernel density plot of $gRNA^+$ cells on UMAP. **e,** Cluster distribution of cells with each gRNA. TEX_{term} single-state TFs are in bold letters. Data represent three replicates. Values are shown as mean ± s.e.m. Statistical analysis was performed using two-way ANOVA followed by Fisher's LSD test, with comparisons made relative to the gScramble control. Statistical results for T_{RM} populations are shown; additional statistical comparisons are provided in Extended Data Table 5. ****$P < 0.0001$, ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$. **f,** Normalized expression score of T_{RM} "up and down" gene-signatures[9,39] in cells with each gRNA KO compared to gScramble cells. **g-h,** Individual TF KO phenotyping. **g,** The ratio of gRNA KO cells in SI:spleen. **h,** The frequency of splenic T_{CM} ($CD62L^+$ $KLRG1^-$) and T_{EM} ($CD62L^-$ $KLRG1^+$) cells. **i-j,** P14 $CD8^+$ T cells were transduced with RVs expressing the T_{RM} single-state TF *Klf6* or empty-control vectors and co-transferred at ~1:1 ratios into mice infected with acute LCMV-Armstrong one day earlier. **i,** Representative plots of donor P14 cells pre- and post-transfer, with the latter isolated from the SI. **j,** Quantification of donor P14 cells expressing T_{RM} markers ($CD69^+$ $CD103^+$) in the SI. Data are presented as the mean ± s.e.m. ***$P<0.001$, ** $P<0.01$, * $P<0.05$. Ordinary one-way ANOVA with Dunnett's multiple comparison test versus gScramble ($n \geq 5$ from ≥ 2 biological replicates, **g-h**). Co-transfer experiments (**i–j**) were analyzed using paired t-tests ($n \geq 6$ from ≥ 2 biological replicates).

Figure 5

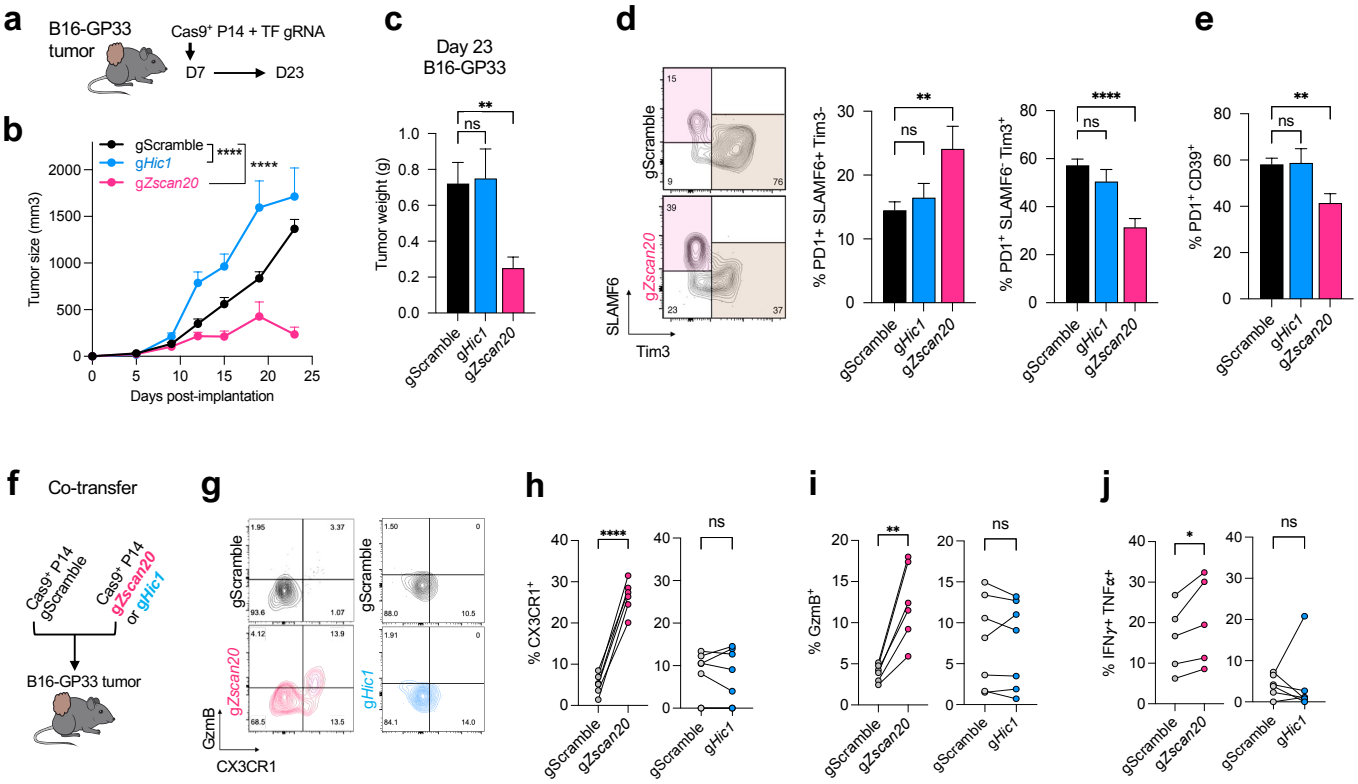


Fig. 5 | Distinct roles of TEX_{term} single-state TF *Zscan20* and multi-state TF *Hic1* in tumor control. a–c, Experimental design (**a**) for adoptive transfer of P14 $CD8^+$ T cells with CRISPR-mediated KO of either the TEX_{term} single-state TF (*Zscan20*) or a multi-state TF (*Hic1*, active in both TEX_{term} and T_{RM}) into B16-GP33 melanoma-bearing mice. Tumor volumes were measured twice per week (**b**) and weights were assessed at day 23 (**c**). **d**, Representative flow cytometry plots of $PD1^+$ P14 $CD8^+$ T cells stained for SLAMF6 and TIM3, with quantification of TEX_{prog} ($PD1^+$ $SLAMF6^+$ $TIM3^-$) and exhausted ($PD1^+$ $SLAMF6^-$ $TIM3^+$) subsets. **e**, Frequency of $PD1^+$ $CD39^+$ double-positive cells. **f-j**, Co-transfer experiments in which $Cas9^+$ P14 $CD8^+$ T cells expressing gRNA targeting *Zscan20* or *Hic1* were mixed with control (scramble gRNA) and transferred into B16-GP33 tumor-bearing mice. Experimental design (**f**). Representative flow cytometry plots of effector marker CX3CR1 and the cytotoxic cytokine Granzyme B (GzmB) in TF KO and control cells (**g**). Quantification of $CX3CR1^+$ (**h**), $GzmB^+$ (**i**), and $IFN\gamma^+$ $TNF\alpha^+$ (**j**) population in *Zscan20* KO vs. control cells (left), and *Hic1* KO vs. control cells (right). Statistical analysis was performed using two-way ANOVA with Tukey's multiple comparisons test (**b**) and one-way ANOVA with Dunnett's test versus gScramble (**c**). Data are from $n \geq 9$ from ≥ 2 biological replicates (**b-e**). Co-transfer experiments (**g–j**) were analyzed using paired t-tests ($n \geq 6$ from ≥ 2 biological replicates). Data are presented as mean ± s.e.m. ****$P < 0.0001$, ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$.

Figure 6

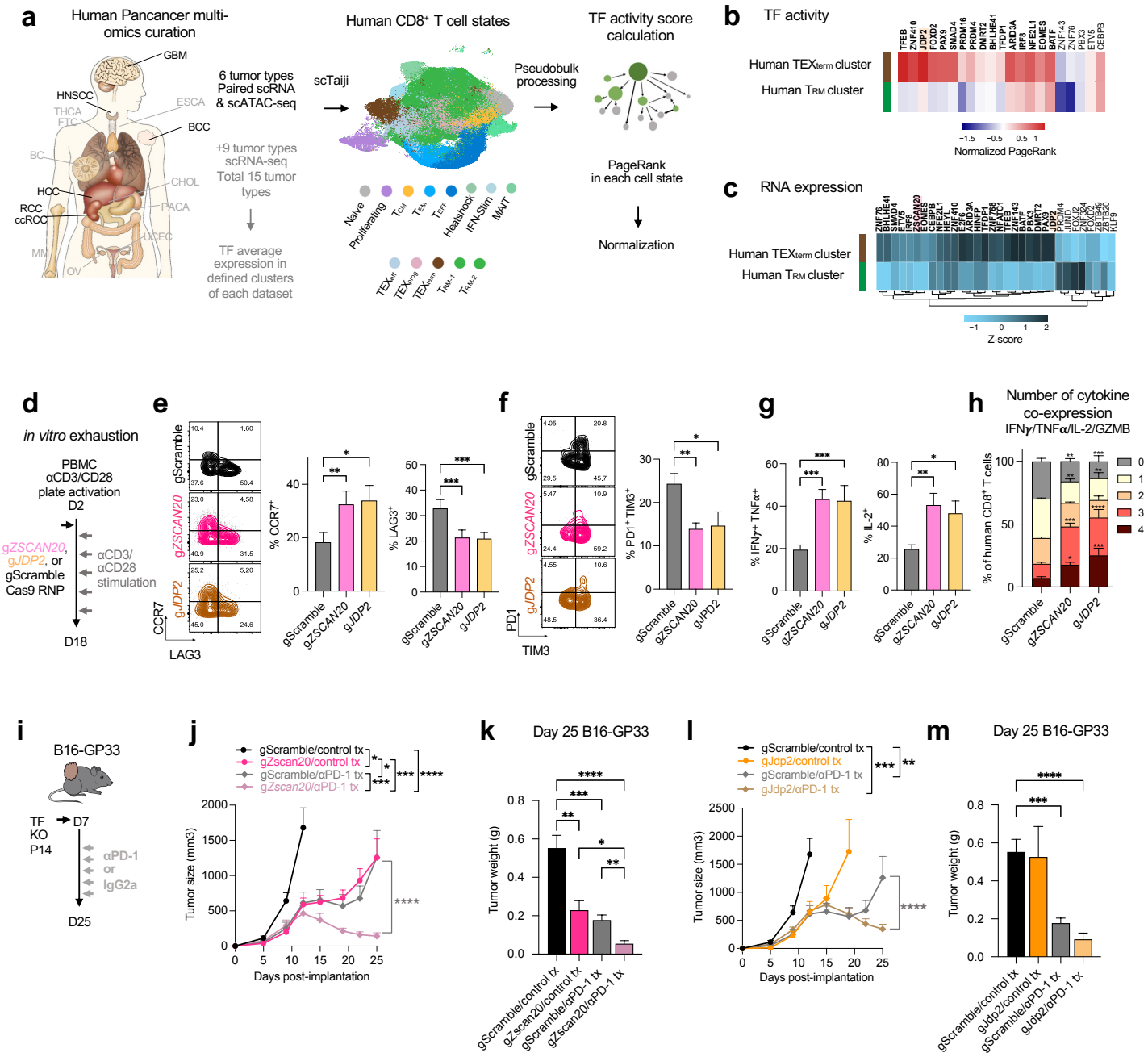


Fig. 6 | Targeting conserved TEX_{term} single-state TFs enhances human T cell effector function and immunotherapy responses

a, Human pan-cancer single-cell multi-omics and scRNA-seq datasets[52–59] were curated to evaluate TF mRNA expression across diverse T cell states and to infer TF activity using the scTaiji algorithm. **b,** For each sample, paired scRNA-seq and scATAC-seq data were used to construct a regulatory network and compute PageRank-based TF activity scores with scTaiji. Shown are normalized PageRank scores of TEX_{term} single-state TFs identified in this study (Fig. 1f) with known binding motifs conserved in the human genome. **c,** mRNA expression of TEX_{term} single-state TFs across TEX_{term} and T_{RM} cell states in human pan-cancer datasets. TFs with conserved cell-state-selectivity in human and mice are highlighted in bold (**b,c**). **d,** Human PBMCs with *ZSCAN20* or *JDP2* KO were stimulated with anti-CD3/CD28 antibodies to induce chronic activation and mimic T cell exhaustion. Cells were expanded every 2–3 days at a density of 1×10^6 cells/mL. **e,** KO and control human $CD8^+$ T cells were analyzed for CCR7(central memory/stem-like marker) and LAG3 (exhaustion marker) by flow cytometry as shown in representative flow plots and cumulative bar graphs. **f,** TF KO and control human $CD8^+$ T cells were analyzed for inhibitory receptors $PD1^+$ $TIM3^+$ double positive cells by flow cytometry as shown in representative flow plots and cumulative bar graphs. **g,** Frequency of $IFN\gamma^+$ $TNF\alpha^+$ double-positive cells and and IL-2 producing cells. **i,** Polyfunctionality analysis based on the number of cytokines produced per cell. **i,** Experimental design for testing the synergy between TEX_{term} single-state TF KO and anti–PD-1 treatment. $Cas9^+$ P14 $CD8^+$ T cells (with or without TF KO) were adoptively transferred into B16-GP33 tumor-bearing mice on day 7 post-implantation, followed by anti–PD-1 or isotype IgG2a treatment. **j–k,** Tumor growth curves (**j**) and tumor weights (**k**) of *Zscan20* KO vs. control. **l–m,** Tumor growth curves (**l**) and tumor weights (**m**) of *Jdp2* KO vs. control. Data are presented as mean ± s.e.m. $n \geq 6$ from ≥2 biological replicates. One-way ANOVA with Dunnett's test versus gScramble (**e-g**); two-way ANOVA with Dunnett's multiple comparisons test versus gScramble (**h**); two-way ANOVA with Tukey's multiple comparisons test (**j,l**); one-way ANOVA with Tukey's test (**k, m**) $****P < 0.0001$, $***P < 0.001$, $**P < 0.01$, $*P < 0.05$.