Article

Divergent clonal differentiation trajectories of T cell exhaustion

Bence Daniel [1,2,3,11], Kathryn E. Yost[2,11], Sunnie Hsiung [4,11], Katalin Sandor [1,3], Yu Xia[4], Yanyan Qi[1], Kamir J. Hiam-Galvez [1,3], Mollie Black[1,3], Colin J. Raposo [1,3], Quanming Shi[1,2], Stefanie L. Meier[1,3,5], Julia A. Belk [1,3], Josephine R. Giles[6,7,8], E. John Wherry [6,7,8], Howard Y. Chang [2,9,12], Takeshi Egawa [4,12] and Ansuman T. Satpathy [1,3,5,10,12] ✉

Chronic antigen exposure during viral infection or cancer promotes an exhausted T cell (Tex) state with reduced effector function. However, whether all antigen-specific T cell clones follow the same Tex differentiation trajectory remains unclear. Here, we generate a single-cell multiomic atlas of T cell exhaustion in murine chronic viral infection that redefines Tex phenotypic diversity, including two late-stage Tex subsets with either a terminal exhaustion (Tex[term]) or a killer cell lectin-like receptor-expressing cytotoxic (Tex[KLR]) phenotype. We use paired single-cell RNA and T cell receptor sequencing to uncover clonal differentiation trajectories of Tex[term]-biased, Tex[KLR]-biased or divergent clones that acquire both phenotypes. We show that high T cell receptor signaling avidity correlates with Tex[term], whereas low avidity correlates with effector-like Tex[KLR] fate. Finally, we identify similar clonal differentiation trajectories in human tumor-infiltrating lymphocytes. These findings reveal clonal heterogeneity in the T cell response to chronic antigen that influences Tex fates and persistence.

Chronic antigen exposure during viral infections and cancer leads to impaired CD8[+] T cell responses characterized by reduced effector function, diminished proliferative capacity and high expression of inhibitory receptors, including PD-1, LAG-3 and TIM3, termed T cell exhaustion[1,2]. However, exhausted T cells (Tex) maintain some effector functions and persist long-term, suggesting that Tex may control pathogen burden while maintaining immune homeostasis[3,4]. Recent studies have identified heterogeneity in Tex phenotypes characterized by distinct surface receptors, functionality, proliferative capacity and tissue localization[5–12]. Some studies support a linear differentiation model, whereby TCF1[+]CXCR5[+]PD-1[+] progenitor Tex (Tex[prog]) self-renew

and maintain downstream Tex subsets, including CX3CR1[+]PD-1[+] intermediate Tex (Tex[int]) and PD-1[+]TIM3[+] terminal Tex (Tex[term])[5–9,13–15]. These subpopulations exhibit distinct epigenetic states with TCF1 and BACH2 driving Tex[prog] formation, whereas the high mobility group transcription factor (TF), TOX, orchestrates the exhaustion program in all Tex states[4,7,16–21]. Finally, Tex subsets are further distinguished by their ability to respond to immune checkpoint blockade (ICB); Tex[term] possess a stable epigenetic program and cannot be reinvigorated by ICB, whereas Tex[prog] can proliferate in response to ICB and may be important for therapeutic response[3,6,22]. Despite these advances, we lack a comprehensive view of Tex states, their clonal relationships and the molecular

[1]Department of Pathology, Stanford University, Stanford, CA, USA. [2]Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA, USA. [3]Gladstone-UCSF Institute of Genomic Immunology, San Francisco, CA, USA. [4]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO, USA. [5]Parker Institute for Cancer Immunotherapy, San Francisco, CA, USA. [6]Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania, Philadelphia, PA, USA. [7]Institute for Immunology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [8]Parker Institute for Cancer Immunotherapy at University of Pennsylvania, Philadelphia, PA, USA. [9]Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA. [10]Stanford Cancer Institute, Stanford University, Stanford, CA, USA. [11]These authors contributed equally: Bence Daniel, Kathryn E. Yost, Sunnie Hsiung. [12]These authors jointly supervised this work: Howard Y. Chang, Takeshi Egawa, Ansuman T. Satpathy. ✉e-mail: satpathy@stanford.edu

programs underlying their differentiation, particularly in polyclonal T cell responses.

Here, we generate a Tex differentiation atlas using single-cell chromatin accessibility, transcriptome and T cell receptor (TCR) sequencing of antigen-specific CD8[+] T cells during chronic lymphocytic choriomeningitis virus (LCMV) infection. We discover new Tex subsets, including an early effector exhausted subset (Tex[eeff]) that initiates the molecular exhaustion program and the Tex[KLR] subset, which emerges as a late-stage phenotype concurrent with Tex[term]. T cell clone tracing with paired single-cell RNA (scRNA)/TCR sequencing (scRNA/TCR-seq) nominates diverse Tex differentiation trajectories, including Tex[term]-biased, Tex[KLR]-biased or divergent fates, comprising both cell types. Tex clones traffic to multiple organ sites where their differentiation trajectories are conserved; however, Tex[KLR]-biased clones are depleted in the liver, suggesting that Tex[term] may be phenotypically adapted for specific tissue microenvironments. Finally, we show that clone trajectories correlate with TCR signaling avidity; high-avidity TCR clones are biased towards Tex[term], whereas low-avidity TCR clones are biased towards Tex[KLR]. Overall, these results provide an in-depth view of the gene regulatory programs and clonal dynamics of Tex states during chronic infection.

A multiomic atlas of T cell exhaustion

To profile CD8[+] T cell exhaustion, we used mouse models of acute (LCMV Armstrong; Arm) or chronic (LCMV clone 13; Cl13) viral infection. These two viral strains share immunodominant epitopes, enabling direct comparison of antigen-specific T cell responses[23]. We generated paired scRNA/TCR-seq and single-cell assay for transposase accessible chromatin with sequencing (scATAC-seq) data from LCMV glycoprotein 33–41 tetramer-positive (gp33[+]) and tetramer-negative (gp33[−]) splenic CD8[+] T cells at day 8 (D8) and day 21 (D21) postinfection (Fig. 1a–d). At D21 of Cl13 infection, we also generated scRNA/TCR-seq of gp33[+] and gp33[−] populations from two additional organs (lung and liver; Fig. 1a,c and Extended Data Fig. 1a). Finally, we sorted D21 Cl13 splenic T cells using previously defined surface markers that identify Tex[prog] (PD-1[+]SLAMF6[+]CX3CR1[−]), Tex[int] (PD-1[+]CX3CR1[+]SLAMF6[−]) and Tex[term] (PD-1[+]SLAMF6[−]CX3CR1[−]) phenotypes and performed scRNA/TCR-seq and scATAC-seq (Extended Data Fig. 1b)[7–9,13]. In total, we obtained 96,750 scRNA-seq profiles that passed quality control and detected TCR alpha and beta sequences in 88,696 T cells (91.7%), comprising 5,197 expanded T cell clones (clones >1 cell; Fig. 1c and Extended Data Fig. 1c). In addition, we obtained 62,731 scATAC-seq profiles that passed quality control (Fig. 1d and Extended Data Fig. 1d,e).

We performed uniform manifold approximation and projection (UMAP) for dimensionality reduction and identified 11 scRNA-seq clusters, which were annotated based on differentially expressed genes (DEGs). In Arm infection, we observed naive T cells (T[naive]; *Ccr7, Sell, Lef1*), effector T cells (T[eff]; *Klrg1, Ly6c2*), effector memory T cells (T[em]; *Klrb1c, Klrd1, S1pr1*) and memory T cells (T[mem]; *Il7r, Sidt1, Gpr183*; Fig. 1e and Supplementary Table 1). In Cl13 infection, we observed Tex[prog] (*Slamf6, Id3*), Tex[int] (*Lgals3, S100a4, Mki67*) and Tex[term] (*Gzma, Cd101, Entpd1*), as expected (Fig. 1c,e). In addition, we observed Tex[eeff] (*Xcl1, Top2a, Mif*; a predominant population at D8 of C13 infection), Tex[KLR]

(*S1pr5, Cx3cr1, Klrc1*; emerging late in C13 infection), lung terminal exhausted cells (Tex[lung]; *Lag3, Ifng, Ccl4*) and interferon signature gene Tex (*Isg15, Ifit1, Isg20*; Fig. 1c,e). We observed eight analogous T cell populations in the scATAC-seq data based on integration with scRNA-seq data (Fig. 1d and Extended Data Fig. 1f). As our goal was to analyze Tex epigenetic states, we did not perform scATAC-seq at D21 in Arm infection or in lung or liver T cells in Cl13 infection; thus, scATAC-seq clusters did not include T[em], T[mem] or Tex[lung] subsets. However, scATAC-seq clusters did reveal additional heterogeneity among effector populations mainly derived from the Arm condition (T[eff] and T[eff2]; Fig. 1d,f).

scATAC-seq profiles were analyzed at the level of: (1) *cis*-regulatory element chromatin accessibility (open chromatin regions; OCRs); (2) gene activity scores, computed from OCR accessibility, weighted by distance; and (3) TF activity, computed from TF binding site enrichment in OCRs or genome-wide[24,25]. Cell-type-specific OCR accessibility was correlated with gene expression at marker gene loci that define Tex subsets, including *Tcf7, Pdcd1* and *Tox* (Fig. 1g,h and Supplementary Table 2). T[naive]-specific OCRs were enriched for the TCF/LEF motifs, which were also enriched in Tex[prog], along with other known Tex[prog] TFs (for example, BATF, AP-1 and BACH)[14,18,26,27]. Tex[eeff] showed NFAT motif enrichment, whereas KLF motifs were specifically enriched in the Tex[int], T[eff] and Tex[KLR] populations. Finally, Tex[term]-specific OCRs exhibited strong enrichment for NR4A, RUNX and NFAT TF motifs (Extended Data Fig. 1g and Supplementary Table 3)[28–31].

CX3CR1[+] Tex comprise two functionally distinct subsets

We next examined heterogeneity within CX3CR1[+] Tex cells, which have recently been described as a functional intermediate cell state between Tex[prog] and Tex[term][8,9,13]. scRNA-seq of sorted CX3CR1[+] T cells from D21 of Cl13 infection revealed heterogeneity that primarily spanned two distinct phenotypes (Tex[int] and Tex[KLR]; Fig. 2a). To understand the transcriptional programs that distinguish Tex[int] from Tex[KLR], we performed DEG analysis and found 206 Tex[KLR]-biased genes and 384 Tex[int]-biased genes (Fig. 2b and Supplementary Table 4). Pathway analysis revealed enrichment for T cell-exhaustion-related biological terms in Tex[int], cell-cycle-related terms in both populations and T cell activation-related and/or motility-related terms in Tex[KLR] (Extended Data Fig. 2a). Notably, many markers of terminal effector and effector memory T cells, including KLR family members (for example, *Klrd1, Klrk1, Klrc1, Klre1* and *Klrg1*) and the TF *Zeb2* and its target gene *S1pr5* (a marker of tissue emigrating T cells), showed highly specific expression patterns in Tex[KLR][32,33]. By contrast, Tex[int] expressed canonical exhaustion markers (for example, *Ctla4, Pdcd1, Lag3*) and TCR signaling genes (for example, *Coro1a*) (Fig. 2b). Similar gene expression signatures were detected when comparing Tex[KLR] and Tex[term] with a more pronounced exhaustion signature in the latter (for example, *Cxcr6, Cd101, Tigit*) (Fig. 2b and Supplementary Table 5). To support an intermediate exhaustion state in Tex[int], we analyzed CXCR6 surface expression in Tex[prog] (CX3CR1[−]SLAMF6[+]KLRG1[−]), Tex[int] (CX3CR1[+]SLAMF6[−]KLRG1[−]) and Tex[term] (CX3CR1[−]SLAMF6[−]KLRG1[−]) by flow cytometry; the results showed intermediate CXCR6 expression in Tex[int] compared with Tex[term] and Tex[prog] (Extended Data Fig. 2b).

Fig. 1 | Single-cell genomic atlas of T cell exhaustion during LCMV infection. a, scRNA-seq UMAPs colored by sample. **b**, scATAC-seq UMAPs colored by sample. **c**, scRNA-seq UMAP colored by annotated T cell subsets (left) or T cell clone size assigned by scTCR-seq (right). **d**, scATAC-seq UMAP colored by annotated T cell subsets. **e**, Heat map of subset-specific marker genes determined by scRNA-seq (left, log₂ FC > 0.25, Bonferroni-adjusted *P* < 0.01, two-sided Wilcoxon rank-sum test). Feature plots of T cell subset-specific gene markers (right). **f**, Heat map of gene score values (weighted accessibility at gene locus) determined by scATAC-seq (left, log₂ FC > 0.5, false discovery rate (FDR) < 0.01, two-sided Wilcoxon rank-sum test adjusted using the Benjamini

and Hochberg procedure). Feature plots of T cell subset-specific gene score values (right). **g**, Heat map of peak score values at unique OCRs of T cell subsets determined by scATAC-seq (left, log₂ FC > 1, FDR < 0.05, two-sided Wilcoxon rank-sum test adjusted using the Benjamini and Hochberg procedure). Feature plots of chromVAR deviation (dev.) scores for T cell subset-specific motifs (right). **h**, Genome accessibility tracks of indicated gene loci in Tex subsets. Violin plots of gene expression (exp.) determined by scRNA-seq (*n*, number of scRNA-seq profiles; box center line, median; limits, upper and lower quartiles; whiskers, 1.5× interquartile range). Min., minimum; max., maximum.

Next, we focused on Tex^{KLR} and assessed the gene signature of terminal effector memory T cells (T-T^{em}), a recently described T^{em} subset identified during acute LCMV infection that expresses T^{eff} markers, including KLRs[34]. The T-T^{em} gene signature was highly expressed in

Tex^{KLR}, suggesting that this subset may represent a parallel differentiation path to T-T^{em} (Fig. 2c,d). Finally, we assessed the functionality and proliferative capacity of gp33-specific Tex^{term}, Tex^{int} and Tex^{KLR} (CX3CR1$^+$KLRG1$^+$SLAMF6$^-$) by measuring IFNG, LAMP1 (cytotoxic
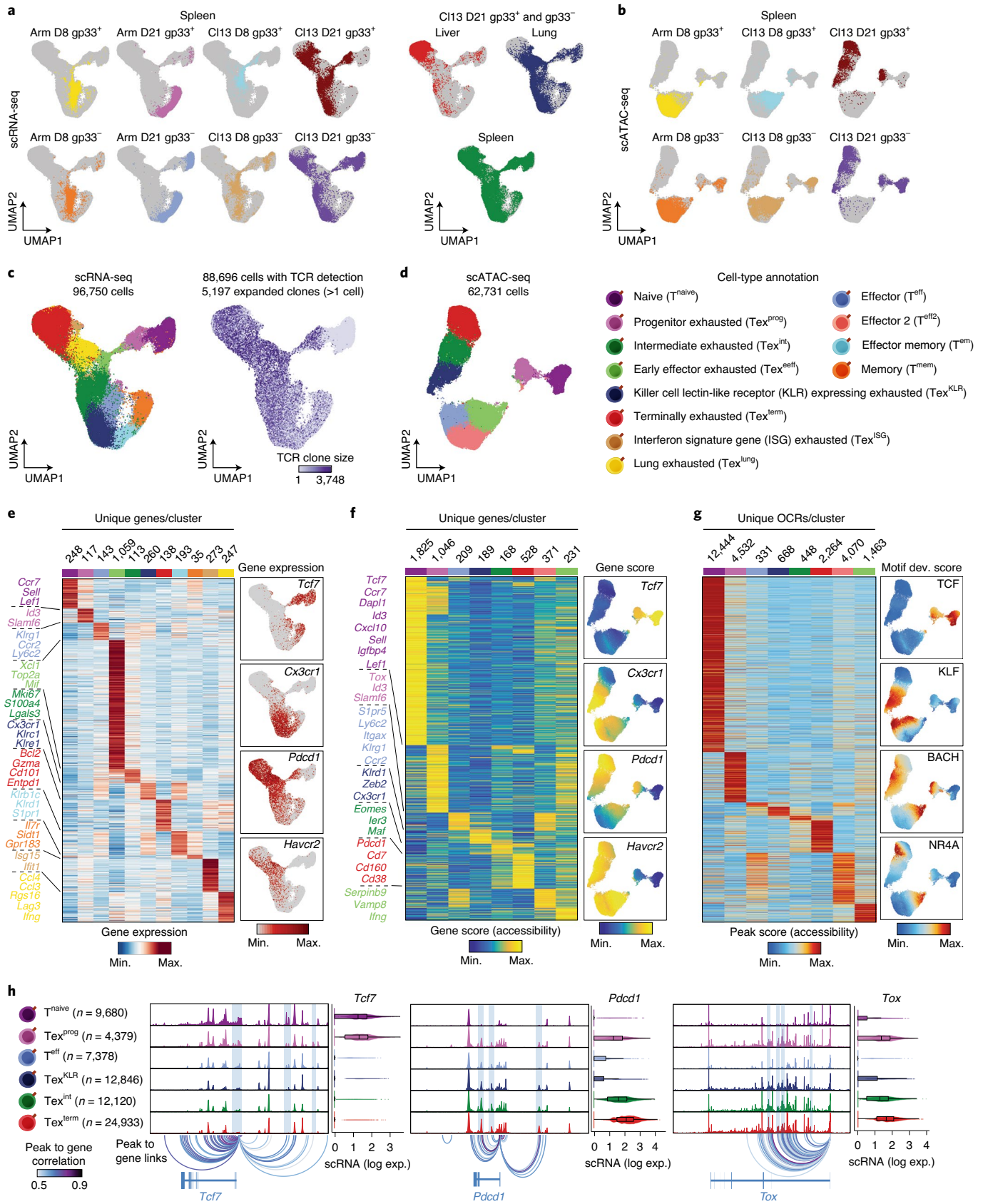
degranulation marker) and MKI67 using flow cytometry. These analyses demonstrated superior functional and proliferative capacity of Tex[KLR] compared with Tex[term] and Tex[int] (Fig. 2e and Extended Data Fig. 2c). In summary, the CX3CR1[+] T cell pool contains two Tex subsets with distinct surface and functional phenotypes.

Tex acquire organ-specific terminal exhaustion signatures

We reclustered scRNA-seq profiles from gp33[+] and gp33[-] CD8[+] T cells from the spleen, lung and liver at D21 of Cl13 infection and examined Tex subset distribution across organs (Fig. 2f). Relative to splenic T cells, cells in the lung exhibited an alternative terminal exhausted phenotype (Tex[lung]) and a reduced Tex[prog] population, with similar Tex[KLR] and Tex[int] proportions (Fig. 2f,g). Strikingly, T cells in the liver almost exclusively adopted the Tex[term] phenotype, as previously described (95.5% Tex[term]; Fig. 2f,g)[15]. We further examined tissue-specific differences in the exhaustion signature by pairwise DEG analyses. Compared with splenic Tex[term], liver-derived Tex[term] possessed a strong tissue-resident T[mem] signature (for example, *Cd69*, *Cxcr6*, *Ccl3*) and a gene program linked to TNF- and glucocorticoid-induced signaling (Extended Data Fig. 2d and Supplementary Table 6). Similarly, lung-derived Tex[term] also exhibited typical markers of lung-resident T[mem] (for example, *Cxcr6*, *Cd44*) and several integrin genes (for example, *Itga4*, *Itgb7*, *Itgb1*; Extended Data Fig. 2d and Supplementary Table 7). Furthermore, both liver-derived and lung-derived Tex[term] expressed higher levels of prosurvival genes (for example, *Bcl2*, *Bcl2a1b*, *Bcl2a1d*) than splenic Tex[term] (Extended Data Fig. 2d and Supplementary Tables 6–8). These results suggest that Tex[term] can obtain tissue residency programs and persist in tissues in the setting of chronic antigen.

Despite tissue-specific differences, we observed a common Tex[term] gene signature across all organs. This signature (*n* = 35 genes) contained previously described exhaustion-related genes, including immune checkpoint inhibitory receptors *Pdcd1* and *Tox* (Fig. 2h). Next, we ranked the severity of exhaustion among Tex[term] from each organ using a previously defined exhaustion gene signature[6]. We observed that liver-derived Tex[term] scored the highest, followed by splenic and lung-derived Tex[term] (Fig. 2i). In line with these findings, flow cytometry analysis demonstrated increased frequencies of Tex[term] in the liver (88% of PD-1[+]CD8[+] T cells) compared with the spleen (47%) and lung (51%; Fig. 2j,k,l). By contrast, we observed the highest frequency of Tex[prog] in the spleen (18.3% of PD-1[+]CD8[+] T cells). Tex[int] and Tex[KLR] were increased in the spleen (Tex[int]: 32.33%; Tex[KLR]: 6.59% of PD-1[+]CD8[+] T cells) and lung (Tex[int]: 44.77%; Tex[KLR]: 10.01%) compared with the liver (Tex[int]: 10.72%; Tex[KLR]: 2.58%; Fig. 2j,k,l). We also scored Tex[term] and Tex[int] profiled by scRNA-seq based on cell cycle activity, which ranked liver-derived cells as the least proliferative, followed by those derived from the lung and spleen, inversely correlating with the severity of exhaustion, and we confirmed these observations with flow cytometry (Extended Data Fig. 2e,f). These results demonstrate that T cell exhaustion develops across multiple organs with a common gene expression signature but microenvironment-specific effects are also detectable and shape tissue-specific Tex subset frequencies.
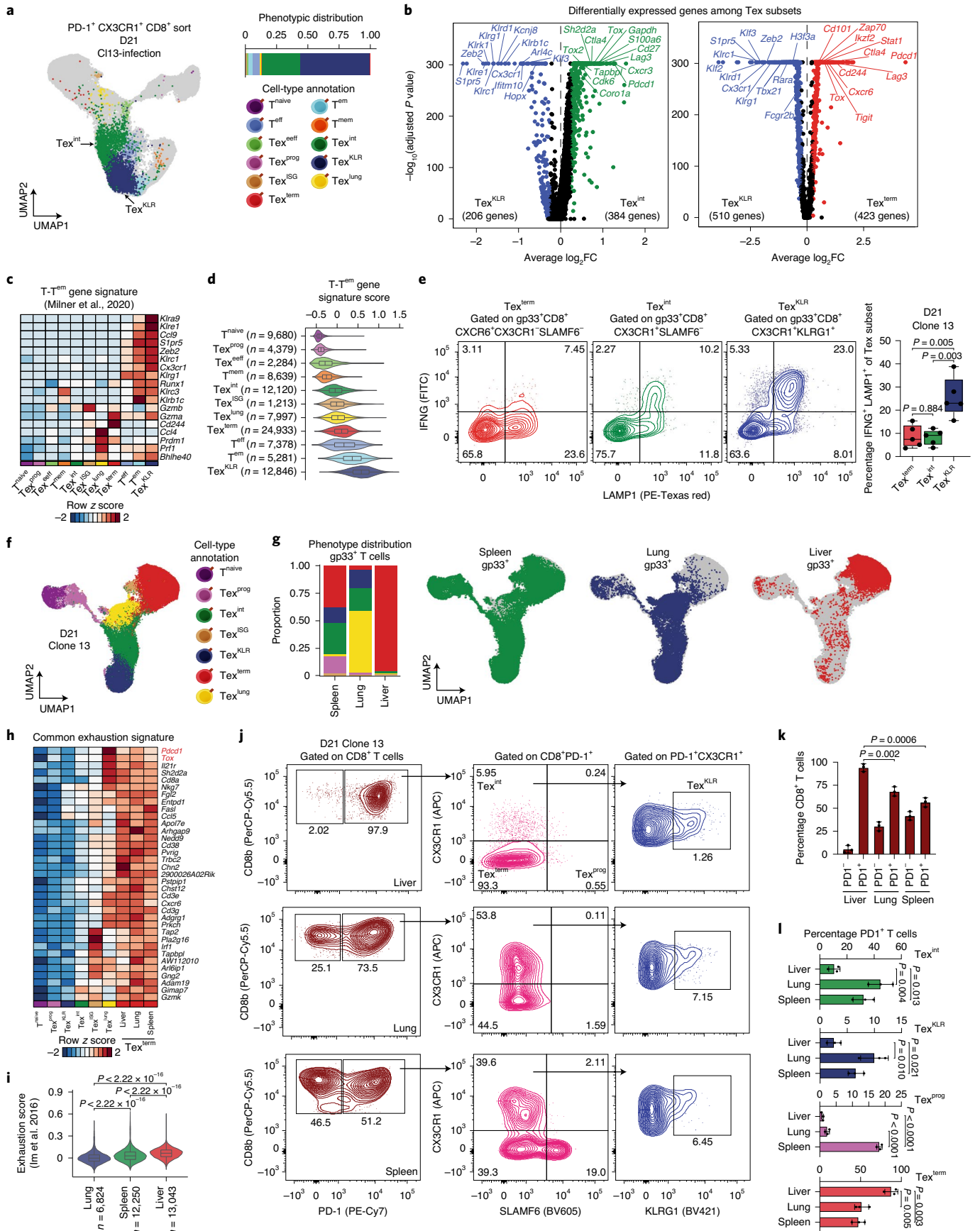
Clone tracing reveals divergent Tex differentiation fates

We next leveraged paired scRNA/TCR-seq data to analyze clonal trajectories of gp33[+] and gp33[-] (other reactive clones and nonreactive T[naive] cells) T cells during Arm and Cl13 (D8 and D21) infection in the spleen (Fig. 3a). We identified 212 and 280 expanded gp33[+] T cell clones (>1 cell; 53 and 40 expanded clones per 1,000 cells sequenced) at D8 and D21 of Arm infection, respectively, and 134 and 338 expanded clones (95 and 36 expanded clones per 1,000 cells sequenced) at D8 and D21 of Cl13 infection, respectively. At D8 of Arm infection, clonally expanded gp33[+] T cells were largely restricted to the T[eff] pool, whereas at D21, clonally expanded gp33[+] T cells exhibited a balanced distribution between T[em] and T[mem] (Fig. 3b and Extended Data Fig. 3a). By contrast, clonal expansion in gp33[+] T cells at D8 in Cl13 infection occurred almost exclusively in Tex[eeff] (Fig. 3c). Importantly, the Tex[prog] population showed weak clonal expansion at D8. However, at D21, gp33[+] specific expanded clones adopted all Tex phenotypes (Fig. 3c,d and Extended Data Fig. 3b).

We visualized the distribution of phenotypes for the top 10 expanded gp33[+] T cell clones at D8 and D21 in each infection. At D8 of Arm infection, almost all cells acquired the T[eff] phenotype. At D21, top expanded clones acquired both T[em] and T[mem] phenotypes, with clone sizes ranging from 77 to 321 (mean 153 cells, 3.8% of 4,030 total cells; Fig. 3e and Extended Data Fig. 3a). By contrast, the top expanded clones in Cl13 infection primarily acquired the Tex[eeff] phenotype at D8, with clone sizes ranging from 26–95 cells (mean 49, 3.5% of 1,414 total cells; Fig. 3e). Analysis of D21 of Cl13 infection identified substantially larger clones, ranging from 146–2,026 cells (mean 525, 5.6% of 9,343 total cells; Fig. 3e). Strikingly, these clones contained cells with multiple Tex phenotypes, although the frequency of each phenotype varied between individual clones. Namely, individual clones either preferentially acquired the Tex[term] or Tex[KLR] phenotypes or developed into both phenotypes (Fig. 3e and Extended Data Fig. 3b). We performed an analysis of the top seven phenotypic distribution of clones (>3 cells), which revealed three main clonal differentiation patterns (clone behaviors): (1) Tex[term]-biased clones, consisting of cells that predominantly acquired the Tex[term] (46% of clones); (2) Tex[KLR]-biased clones, consisting of T cells that predominantly acquired the Tex[KLR] phenotype (18% of clones); and (3) divergent clones, consisting of cells that acquired Tex[term] and Tex[KLR] phenotypes (36% of clones; Fig. 3f). Divergent clones were the most clonally expanded and ranged from 4 to 2,026 cells (mean 145) per clone, whereas Tex[term]-biased clones ranged from 4 to 111 cells (mean 17) per clone. Tex[KLR]-biased clones were relatively small and ranged from 4 to 21 cells (mean 8) per clone (Fig. 3g,h). We also noted several larger clones (4–233 cells, mean 54) that skewed to the Tex[KLR] phenotype (>50% of cells acquired the Tex[KLR] phenotype) but had a small percentage of Tex[term] (Extended Data Fig. 3c). To account for sampling bias, we randomized T cell phenotype

---

Fig. 2 | Identification of intermediate, KLR-expressing and organ-specific Tex subsets. a, scRNA-seq UMAP colored by T cell subset from sorted PD-1[+]CX3CR1[+]CD8[+] T cells (left). Stacked bar plot of the sorted population phenotypic distribution (right) **b**, Volcano plots of DEGs between indicated Tex populations (log$_2$ FC > 0.25, Bonferroni-adjusted *P* < 0.05, two-sided Wilcoxon rank-sum test). **c**, Heat map of scaled marker gene expression of T-T[em] cells (Milner et al. 2020)[34] in indicated T cell subsets. **d**, Violin plot of T-T[em] gene signature scores in indicated T cell subsets (*n*, number of scRNA-seq profiles; box center line, median; limits, upper and lower quartiles; whiskers, 1.5× interquartile range). **e**, Representative IFNG/LAMP1 flow cytometry plots of indicated Tex subsets (left). Box plot quantifying IFNG[+]LAMP1[+] percentages for indicated Tex subsets. Significance determined by two-tailed, unpaired *t*-test (*n* = 5 biologically independent animals). Box center line, mean; limits, upper and lower quartiles; whiskers, minimum and maximum values. **f**, scRNA-seq UMAP from three organs at D21 following Cl13 infection colored by annotated T cell subsets. **g**, Stacked bar plot of gp33[+] phenotypic distribution of annotated T cell subsets in three organs (left). UMAPs colored by organ (right). **h**, Heat map of scaled gene expression values of common exhaustion gene signature among Tex[term] from three organs. **i**, Violin plot of Tex[term] exhaustion scores from the three organs (Im et al. 2016)[6]; *n*, number of scRNA-seq profiles; box center line, median; limits, upper and lower quartiles; whiskers, 1.5× interquartile range. *P* values determined by two-sided Wilcoxon rank-sum test. **j**, Representative flow cytometry plots of Tex subsets detected in organs. **k**, Quantification of the PD-1[+] and PD-1[-] fractions of CD8[+] T cells across organs. Significance determined by two-tailed, unpaired *t*-test (*n* = 3 biologically independent animals). Means with s.d. are shown. **l**, Quantification of fractions of indicated Tex subsets across organs. Significance determined by two-tailed, unpaired *t*-test (*n* = 3 biologically independent animals). Means with s.d. are shown.

Figure showing panels a–l including UMAP plots, volcano plots of differentially expressed genes among Tex subsets, heatmaps of gene signatures, violin plots, and flow cytometry analyses of T cell exhaustion subsets from D21 Clone 13 infection.

and TCR clone assignment to generate a null distribution of clone patterns; this revealed a striking enrichment of Tex^{KLR}-biased and Tex^{term}-biased clone behavior over chance, whereas divergent clonal behavior was twice as likely to be detected by chance than observed in our data (Extended Data Fig. 3d). These results reveal new divergent clonal Tex differentiation trajectories during chronic infection (Fig. 3i).

Directions of clonal Tex differentiation fates

To understand cell state transitions in Tex differentiation paths, we used dynamo to perform RNA velocity analyses[35,36]. This analysis demonstrated that Tex^{eeff} at D8 transitioned to either Tex^{int} or Tex^{prog} fates, and at D21, Tex^{int} showed a bifurcation to either Tex^{KLR} or Tex^{term} phenotypes (Fig. 4a). Furthermore, we applied dynamo to scRNA-seq profiles from individual TCR clones; divergent clones recapitulated bifurcating Tex differentiation trajectories from Tex^{int} to either Tex^{term} or Tex^{KLR}, accompanied by increased transcription of Tex^{term} or Tex^{KLR} genes (Fig. 4b,c). In addition, RNA velocity pseudotime of a divergent clone revealed a gene expression map of Tex differentiation from Tex^{int} to either terminal Tex state (Extended Data Fig. 4a). By contrast, Tex^{term}-biased clones showed a linear RNA velocity path from Tex^{int} to Tex^{term}, whereas Tex^{KLR}-biased clones primarily comprised Tex^{KLR} (Fig. 4b,c). It is important to note that this analysis summarizes the direction and magnitude of dominant local trajectories in UMAP space, rather than nominating terminal differentiation states[35,36]. For comparison, we performed velocity analyses in CD8$^+$ T cells from Arm infection (Extended Data Fig. 4b).

We used scATAC-seq data to define changes in chromatin accessibility that accompany these fate transitions (Extended Data Fig. 4c,d). Comparisons of OCRs and TF motif enrichments between Tex subsets demonstrated a high degree of similarity between Tex^{eeff} and Tex^{prog}. Similarly, Tex^{int} exhibited an intermediate chromatin state between Tex^{term} and Tex^{KLR} states, with very few unique OCRs (Extended Data Fig. 4d). We analyzed 15,809 variable OCRs for TF motif enrichments, as well as gene integration scores, across three differentiation trajectories nominated by RNA velocity and clonal trajectories: (1) Tex^{prog} trajectory (Tex^{eeff} → Tex^{prec} (precursor exhausted Tex, an early Tex^{prog} population present at D8 in Cl13 infection) → Tex^{prog}); (2) Tex^{term} trajectory (Tex^{prog} → Tex^{int} → Tex^{term}); and (3) Tex^{KLR} trajectory (Tex^{prog} → Tex^{int} → Tex^{KLR}; Fig. 4d). The Tex^{prog} trajectory showed a loss of HOMEOBOX TF motifs and enrichment of BATF, AP-1, BACH, NFKB, TCF and CTCF motifs. By contrast, in both Tex^{term} and Tex^{KLR} trajectories, we observed a loss of Tex^{prog}-specific TF motifs in Tex^{int}, followed by enrichment of TF motifs that may guide the differentiation program of Tex^{KLR} (for example, ZEB, KLF, ETS, TBX, RUNX) and Tex^{term} (for example, RUNX, IRF, STAT, NR4A; Fig. 4d).

Finally, we assessed the *Tox* locus, and found a gradual increase in accessibility during the Tex^{eeff} to Tex^{prog} transition. The Tex^{term} trajectory demonstrated a decrease in *Tox* accessibility during the Tex^{prog} to Tex^{int} transition and a subsequent increase in the Tex^{term} state (Fig. 4d). We annotated differentially accessible OCRs (compared with T^{naive}) in a ±250 kb window around the transcription start site (TSS) and identified 88 OCRs. Sixteen and eight OCRs were differentially accessible in Tex^{eeff} and Tex^{prog}, respectively; these results were supported by high

Tox expression in these subsets (relative to T^{naive}), indicating that TOX may initiate the molecular programming of Tex differentiation in these subsets (Extended Data Fig. 4e and Supplementary Table 9).

Early molecular programs of Tex differentiation

Next, we focused on this early stage of exhaustion by comparing D8 scATAC-seq phenotypes in Arm and Cl13 (Extended Data Fig. 5a). Memory precursor cells (T^{mp}) were present at D8 in Arm infection and clustered with an early Tex^{prog} population present at D8 in Cl13 infection that expresses *Tox* and *Tcf7* (Tex^{prec}), but these subsets were relatively infrequent compared with the effector populations in both infection models and at D5 of Cl13 infection (Extended Data Fig. 5b)[14,37]. We compared the gene expression and chromatin state of Tex^{prec} and T^{mp} subsets at D8, which revealed strong exhaustion-induced and interferon-induced programs in Tex^{prec} (Extended Data Fig. 5c and Supplementary Table 10). Similarly, DEGs of effector populations revealed a strong Tex signature in the Tex^{eeff} subset compared with T^{eff}; T^{eff} showed a bona fide effector program (for example, *Gzma*, *Klrd1*, *Ccr2*), whereas Tex^{eeff} expressed higher levels of exhaustion marker genes (for example, *Tnfrsf9*, *Lag3*, *Pdcd1*, *Havcr2*; Extended Data Fig. 5d and Supplementary Table 11). These observations were also supported by the chromatin state programs of these subsets (T^{eff}: 7,066 OCRs versus Tex^{eeff}: 5,211 OCRs) that were associated with T^{eff}-specific (ETS, RUNX) and Tex^{eeff}-specific (NFAT, BATF) TF motifs (Extended Data Fig. 5d). These results support studies demonstrating the formation of Tex^{prec} early during chronic infection that exhibit molecular signatures of exhaustion, distinct from T^{mp}[14,37]. Moreover, we found that the exhaustion program, including *Tox* expression, was present in the Tex^{eeff} stage, which might represent an early, heterogeneous effector pool with plasticity to seed Tex subsets. We tested this concept by adoptively transferring Tex^{eeff}, which specifically expressed *Tnfrsf9* (encoding 4-1BB) (Extended Data Fig. 5e). We transferred ~400,000 CD45.2$^+$TNFRSF9$^+$CX3CR1$^-$SLAMF6$^-$CD8$^+$ T cells from D5 of Cl13 infection to CD45.1$^+$ infection-matched hosts and analyzed the phenotypic distribution of the transferred cells. At D21, CD45.2$^+$ cells comprised ~3.9% Tex^{prog}, ~2.5% Tex^{KLR}, ~17% Tex^{int} and ~76% Tex^{term}, supporting the concept that Tex^{eeff} may represent an early stage of Tex commitment (Extended Data Fig. 5f).

Differentiation of Tex^{int} to Tex^{KLR} and Tex^{term}

Tex differentiation downstream of Tex^{prog} has been suggested to follow bifurcating differentiation paths, and the RNA velocity analysis suggested that Tex^{int} may represent the bifurcation point between Tex^{KLR} and Tex^{term}[3,9]. We nominated TFs specific for each Tex phenotype based on differential RNA expression (Extended Data Fig. 5g) and Tex^{term}-specific and Tex^{KLR}-specific TF motif enrichments (for example, Tex^{term}: NFAT, STAT, NR4A, IRF; and Tex^{KLR}: TBX, KLF; Extended Data Fig. 5h). Notably, differential analysis of Tex^{term} and Tex^{KLR} relative to Tex^{int} identified only four shared OCRs, suggesting that these two cell states are epigenetically divergent. To test whether Tex^{int} could represent a bifurcation point between Tex^{KLR} and Tex^{term}, we adoptively transferred ~200,000 CD45.2$^+$ Tex^{int} cells (CX3CR1$^+$SLAMF6$^-$KLRG1$^-$) from Cl13 infection at D21 into CD45.1$^+$ infection-matched hosts. After 10 days, transferred Tex^{int} cells gave rise to both Tex^{term} and Tex^{KLR}, although

---

Fig. 3 | TCR-based lineage tracing reveals divergent Tex clonal trajectories. a, UMAP of scRNA-seq results from the gp33$^+$ and gp33$^-$ T cell fractions in the spleen at D8 and D21 of Arm and Cl13 infection. UMAP is colored by the annotated T cell subsets. **b**, UMAP of scRNA-seq results of gp33$^+$ T cells colored by the size of the detected TCR clones at D8 and D21 in the Arm infection model. **c**, UMAP of scRNA-seq results of gp33$^+$ T cells colored by the size of the detected TCR clones at D8 in the Cl13 infection model (left). Same UMAP of gp33$^+$ T cells colored by the TCR clone size at the D21 time point in the Cl13 infection model (right). **d**, UMAPs colored by the expression of the indicated gene transcripts by scRNA-seq. **e**, Stacked bar plot of the phenotypic distribution of the top 10 expanded clones in the gp33$^+$ fraction at D8 and D21 of Arm infection (left) and
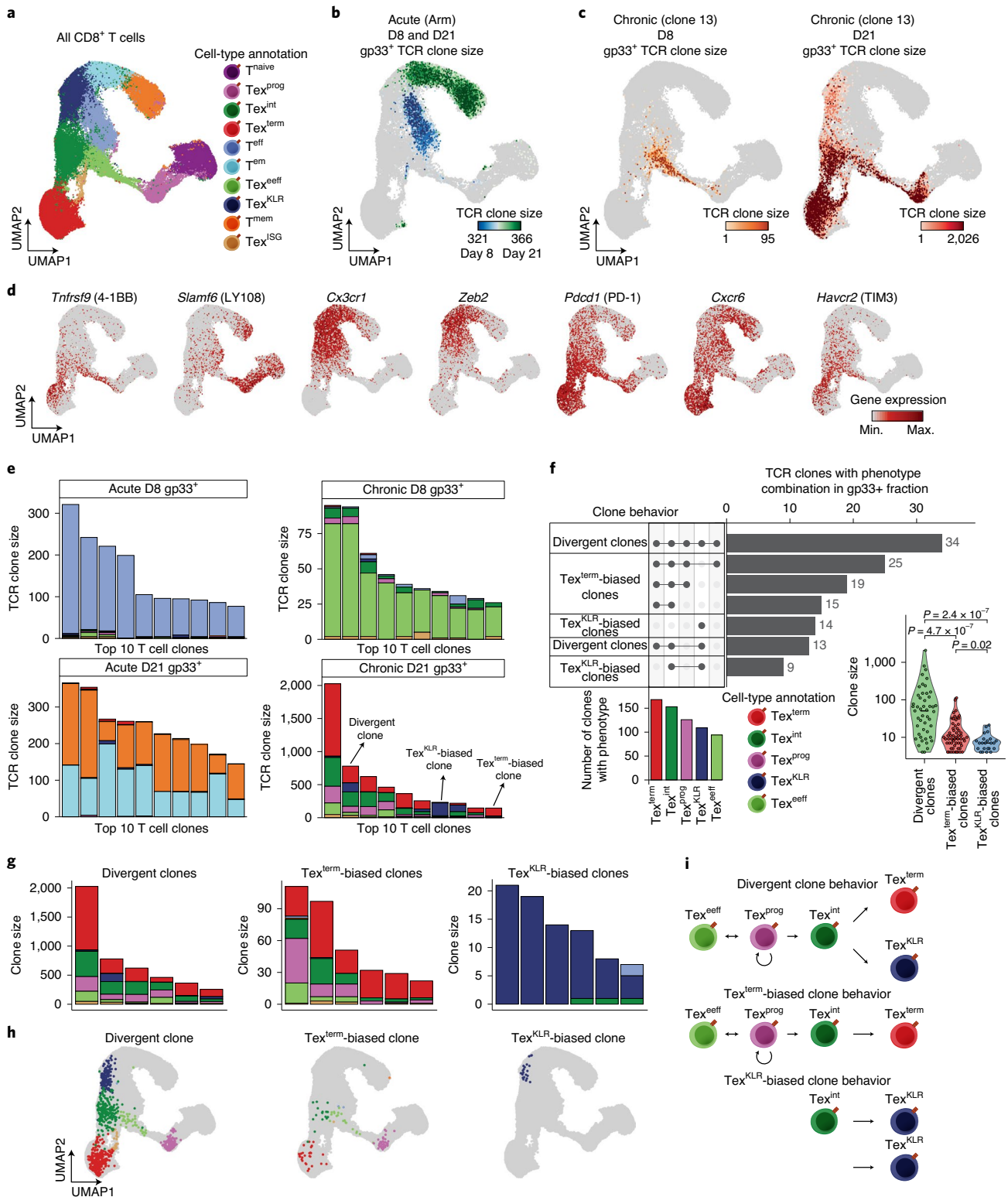
Cl13 infection (right). **f**, Upset plot depicting the number of expanded clones with specific phenotype combinations (clone behaviors). For clarity, the top seven most common clone behaviors are shown. Bar plot of the number of clones with the indicated phenotypes (bottom left). Violin plot of the clone size distribution of the detected clone behaviors (bottom right, *P* values determined by two-sided Wilcoxon rank-sum test). **g**, Stacked bar plots show the top six expanded clones with the indicated clone behaviors. **h**, UMAPs show representative examples of T cell clones with the detected clone behaviors. **i**, Schematic of the phenotypic composition and the potential differentiation trajectories of the identified clone behaviors.

there was relatively more Tex[term] (Fig. 4e). To further test the potential for endogenous Tex[int] to give rise to Tex[term], we performed genetic fate mapping using *Cx3cr1*[CreER]*Rosa26*[LSL-tdT] mice that can trace the fate of *Cx3cr1*-expressing cells. Short-term fate mapping of *Cx3cr1*-expressing cells (which includes both Tex[KLR] and Tex[int] cells) at D22 in Cl13 demonstrated that by D25, 6.5% of Tex[term] cells were labeled with tdTomato, supporting the concept that Tex[int] cells can give rise to Tex[term] (Extended Data Fig. 5i). These results support the concept that the Tex[int] stage is a bifurcation point of Tex differentiation; however, it is important to

note two caveats: (1) the genetic fate mapping results do not inform the overall contribution of Tex[int] to Tex[KLR], and it is possible that Tex[KLR] also arise from a Tex[int]-independent path, which is also supported by our TCR clone analysis; and (2) we cannot rule out the possibility that Tex[term] are generated directly from Tex[prog] in some settings.

Tex clone behaviors are shared across tissues

We next asked whether clonal differentiation patterns are intrinsically programmed. We first determined whether expanded Tex clones

a All CD8+ T cells

Cell-type annotation
T[naive]
Tex[prog]
Tex[int]
Tex[term]
T[eff]
T[em]
Tex[eeff]
Tex[KLR]
T[mem]
Tex[ISG]

b Acute (Arm) D8 and D21 gp33+ TCR clone size

TCR clone size
321 — 366
Day 8 Day 21

c Chronic (clone 13) D8 gp33+ TCR clone size

TCR clone size
1 — 95

Chronic (clone 13) D21 gp33+ TCR clone size

TCR clone size
1 — 2,026

d *Tnfrsf9* (4-1BB) *Slamf6* (LY108) *Cx3cr1* *Zeb2* *Pdcd1* (PD-1) *Cxcr6* *Havcr2* (TIM3)

Gene expression
Min. — Max.

e Acute D8 gp33+ / Chronic D8 gp33+ / Acute D21 gp33+ / Chronic D21 gp33+

TCR clone size — Top 10 T cell clones

Divergent clone
Tex[KLR]-biased clone
Tex[term]-biased clone

f Clone behavior

TCR clones with phenotype combination in gp33+ fraction

Divergent clones — 34
Tex[term]-biased clones — 25 / 19 / 15
Tex[KLR]-biased clones — 14
Divergent clones — 13
Tex[KLR]-biased clones — 9

Number of clones with phenotype

Cell-type annotation
Tex[term]
Tex[int]
Tex[prog]
Tex[KLR]
Tex[eeff]

Clone size
$P = 2.4 \times 10^{-7}$
$P = 4.7 \times 10^{-7}$
$P = 0.02$

Divergent clones / Tex[term]-biased clones / Tex[KLR]-biased clones

g Divergent clones / Tex[term]-biased clones / Tex[KLR]-biased clones

Clone size

h Divergent clone / Tex[term]-biased clone / Tex[KLR]-biased clone

i Divergent clone behavior
Tex[eeff] ↔ Tex[prog] → Tex[int] ↗ Tex[term] ↘ Tex[KLR]

Tex[term]-biased clone behavior
Tex[eeff] ↔ Tex[prog] → Tex[int] → Tex[term]

Tex[KLR]-biased clone behavior
Tex[int] → Tex[KLR] → Tex[KLR]

could be found across organs (animal-matched) in Cl13 at D21 (Fig. 5a). In spleen-derived, liver-derived, and lung-derived scRNA/TCR-seq datasets, we detected expanded T cell clones across all three tissues and found that the gp33[+] and gp33[−] fractions showed minimal TCR overlap (Fig. 5b and Extended Data Fig. 6a). Importantly, there was substantial TCR sharing across the different organs within both gp33[+] and gp33[−] fractions (Extended Data Fig. 6a). We identified expanded organ-shared T cell clones that had at least five T cells, which consisted of at least one cell from each organ. This analysis identified 100 shared T cell clones among all organs, 37 clones shared between the lung and spleen, and 22 spleen-specific clones (Fig. 5c,d).

Next, we examined the phenotypic distribution of organ-shared clones (Fig. 5e,f and Extended Data Fig. 6b,c). First, we focused on comparisons between the spleen and lung and found highly conserved clone behaviors. Divergent clones in the spleen also maintained Tex[term] and Tex[KLR] phenotypes in the lung (although instead exhibiting the Tex[lung] phenotype; 35 of 48 divergent clones detected in both organs; Fig. 5e,f and Extended Data Fig. 6b,c). Similarly, the majority of splenic Tex[KLR]-biased clones (four of seven clones) and Tex[term]-biased clones (11 of 12 clones) maintained their behavior in the lung (Fig. 5e and Extended Data Fig. 6c). In particular, we did not observe appreciable interconversion between Tex[KLR]-biased and Tex[term]-biased clones between these two organs (0 of 19 shared clones). Accordingly, quantification of Tex[KLR] and Tex[term] frequencies within individual clones showed a high concordance across organs (Fig. 5g). These results demonstrate that Tex clones are shared across organs and that clonal differentiation behavior is primarily intrinsically programmed.

Depletion of Tex[KLR] clones in the liver microenvironment

We next analyzed clonal behavior in the liver, which showed an overall enrichment of Tex[term] compared with other organs (94% Tex[term]). Thus, we expected an enrichment in clonal Tex[term] frequency; however, this could be driven by a depletion of Tex[KLR]-biased clones and/or interconversion of Tex[KLR]-biased or divergent clones to Tex[term]-biased clones. We first analyzed the Tex[KLR]-biased clones from the spleen and found that only one of these clones was present in the liver (one of seven clones), suggesting that Tex[KLR]-biased clones are depleted in the liver niche (Extended Data Fig. 6c). Similarly, although divergent clones were largely detectable in the liver (42 of 48 clones shared between spleen and liver), we again observed a depletion of Tex[KLR] cells, resulting in Tex[term]-biased behavior in the majority of the cases (32 of 42 shared clones) (Extended Data Fig. 6c). By contrast, the majority of Tex[term]-biased clones remained Tex[term]-biased in the liver, although they were heavily skewed towards Tex[term], with relative loss of Tex[prog] and Tex[int] (nine of nine clones; Fig. 5e,f and Extended Data Fig. 6c). Quantification of frequencies of Tex[KLR] and Tex[term] phenotypes of shared clones in the spleen and liver confirmed the depletion of Tex[KLR] in the liver and a skewing of Tex[term]-biased clones to the Tex[term] fate (Fig. 5g,h).

To investigate differences in viral antigen levels in the three organ niches, we measured LCMV glycoprotein viral transcripts (*LCMV-gp*) and found that *LCMV-gp* transcripts were higher in the lung and liver compared with the spleen (Extended Data Fig. 6d). However, *LCMV-gp*

transcripts were the highest in the lung, suggesting that additional microenvironment differences may underlie the depletion of Tex[KLR] from the liver. As IL-21 has been shown to be a critical cytokine supporting the maintenance of CX3CR1[+] Tex cells, we also measured *Il21* transcript levels in each organ[9,13]. We found substantially lower *Il21* levels in the liver compared with the spleen and lung, suggesting that a lack of IL-21 may contribute to Tex[KLR] depletion (Extended Data Fig. 6d).

TCR signaling avidity correlates with clone behavior

The difference in clonal expansion between Tex[term]-biased clones and Tex[KLR]-biased clones led us to examine whether Tex differentiation trajectories correlated with differences in TCR signaling avidity. We first used tetramer binding level as a proxy for TCR avidity and sorted gp33[−] ($n = 8,914$), gp33-intermediate (gp33[int]; $n = 5,875$) and gp33-high (gp33[high]; $n = 8,194$) splenic CD8[+] T cells from Cl13-infected mice at D21 and performed scRNA/TCR-seq (Fig. 6a and Extended Data Fig. 7a). Analysis of TCR sequences identified 313 TCR clonotypes in gp33[high] cells, 1,576 in gp33[int] cells and 3,803 in gp33[−] cells (Extended Data Fig. 7b). The TCR repertoire showed a small overlap between gp33[high] and gp33[−] cells (13 shared TCRs) compared with the overlap between gp33[high] and gp33[int] cells (158 shared TCRs) and between gp33[int] and gp33[−] cells (306 shared TCRs). TCR repertoire similarity analysis demonstrated that gp33[int] sorting captured a unique repertoire (Extended Data Fig. 7b,c).

Next, we evaluated the clone size distribution of the sorted populations, which revealed an increase in the percentage of large clones (clones with 50–200 or >200 cells) as a function of higher tetramer binding (fluorescence), with an accompanying decrease in clonal diversity (Extended Data Fig. 7d). To link unique TCR clones to each gp33[−] tetramer fraction, we compared the overlap of clones between gp33[−] fractions and identified 592 unique gp33[−] clones, 114 unique gp33[int] clones and 88 unique gp33[high] clones (Fig. 6b,c). Importantly, we found phenotypic skewing in the unique clones (Fig. 6c and Extended Data Fig. 7e,f). Namely, gp33[high] cells contained ~3.3 times more cells with Tex[term] and Tex[int] phenotypes compared with either gp33[−] or gp33[int] cells (39% Tex[term], 19% Tex[int] in gp33[high]; 11% Tex[term], 6.7% Tex[int] in gp33[int]; 11% Tex[term], 7.0% Tex[int] in gp33[−]). By contrast, gp33[int] cells exhibited phenotypic skewing towards the Tex[KLR] phenotype compared with gp33[high] and gp33[−] cells (27% Tex[KLR] in gp33[int]; 7.9% Tex[KLR] in gp33[high]; 13% Tex[KLR] in gp33[−]; Fig. 6d and Extended Data Fig. 7e). We validated differences in Tex phenotype distribution and function across gp33 fractions using flow cytometry (Extended Data Fig. 7g,h).

To analyze differentiation trajectories, we visualized the top 10 unique expanded clones in each gp33 tetramer fraction and assessed their phenotypic composition. We found that the top clones in the gp33[−] and gp33[high] fractions were biased towards Tex[term] or divergent phenotypes (10 of 10 gp33[−] clones and 10 of 10 gp33[high] clones; expanded gp33[−] clones probably represent other antigen specificities), whereas the largest clones in the gp33[int] pool exhibited skewing towards the Tex[KLR] phenotype (five of 10 gp33[int] clones; Fig. 6d). We analyzed the top seven clone behaviors of the three gp33 fractions and found that clones from the gp33[−] fraction primarily exhibited Tex[term]-biased

Fig. 4 | Tex[int] represent a bifurcation point of Tex fate differentiation.
a, UMAPs of scRNA-seq results of D8 and D21 gp33[+] T cells from the Cl13 infection model. UMAPs are colored by the annotated T cell subsets and arrows represent Tex differentiation fate directions predicted by RNA velocity analysis. **b**, UMAPs of individual Tex clones with the indicated clone behavior. Color gradient (RNA velocity pseudotime order) indicates directions of Tex differentiation fates determined by RNA velocity analysis. **c**, Scatter plots of the expression level and spliced to unspliced ratio of the indicated gene transcripts in T cells of a divergent clone over pseudotime as determined by RNA velocity analysis. Expression of spliced and unspliced transcripts was derived from the mean expression (first moment) of each gene calculated using *k*-nearest neighbors to alleviate

dropout. **d**, Pseudotime trajectory analyses of three potential Tex differentiation paths in scATAC-seq space (top). Heat maps of TF motif deviation scores and gene integration scores (integrated scRNA-seq expression values) over the pseudotime trajectories (middle). Gene integration scores for *Tox* expression in the three pseudotime trajectories (bottom). **e**, Schematic of adoptive transfer experiments for Tex[int] subset (top). Pretransfer enrichment strategy for Tex[int] cells (FSC-A, forward scatter area; middle). Stacked bar plot of the phenotypic composition of transferred T cells with all Tex subsets shown or only the CX3CR1[+] subsets (bottom, *n* = 5 biologically independent animals). The mean percentage of each subset is indicated.
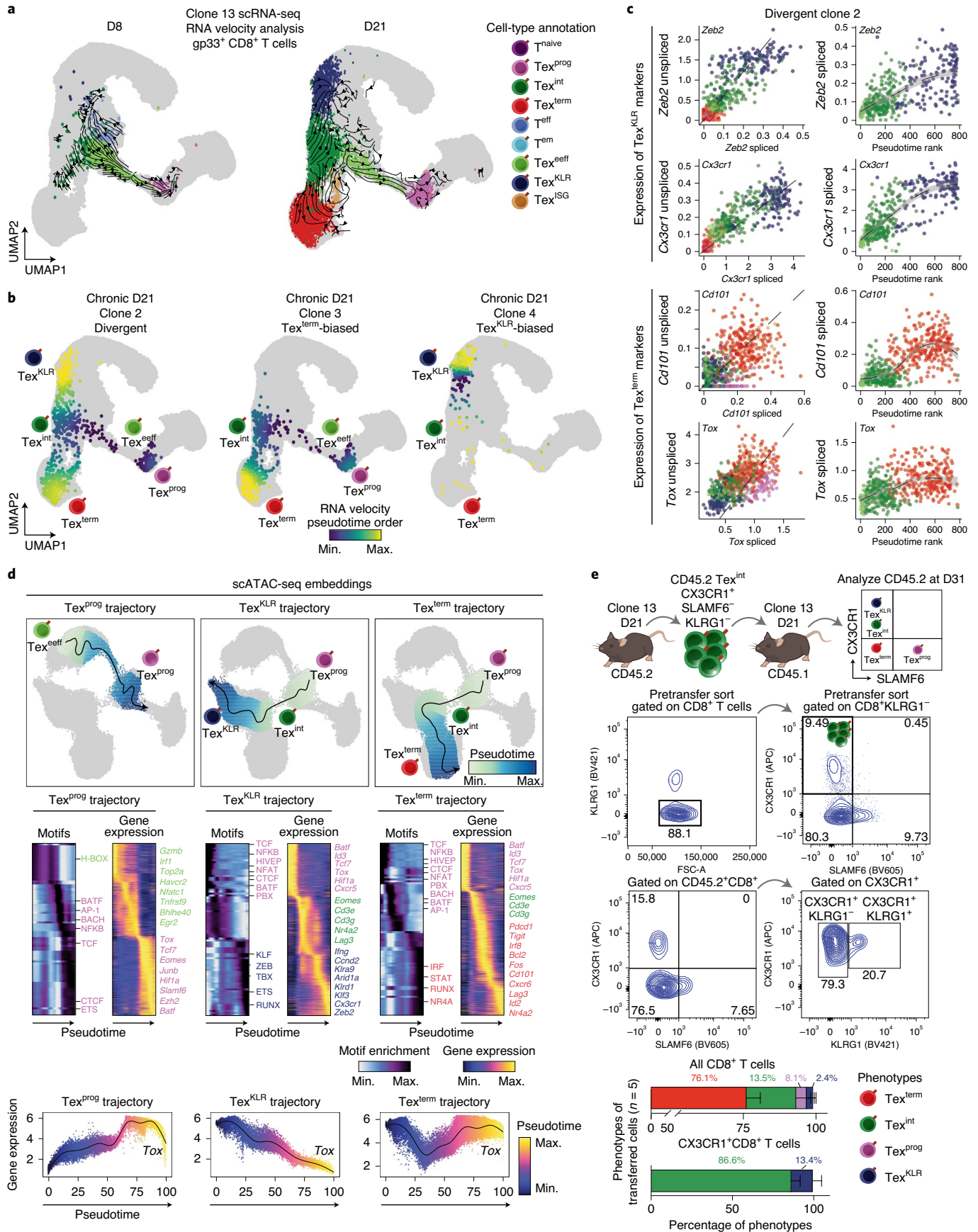
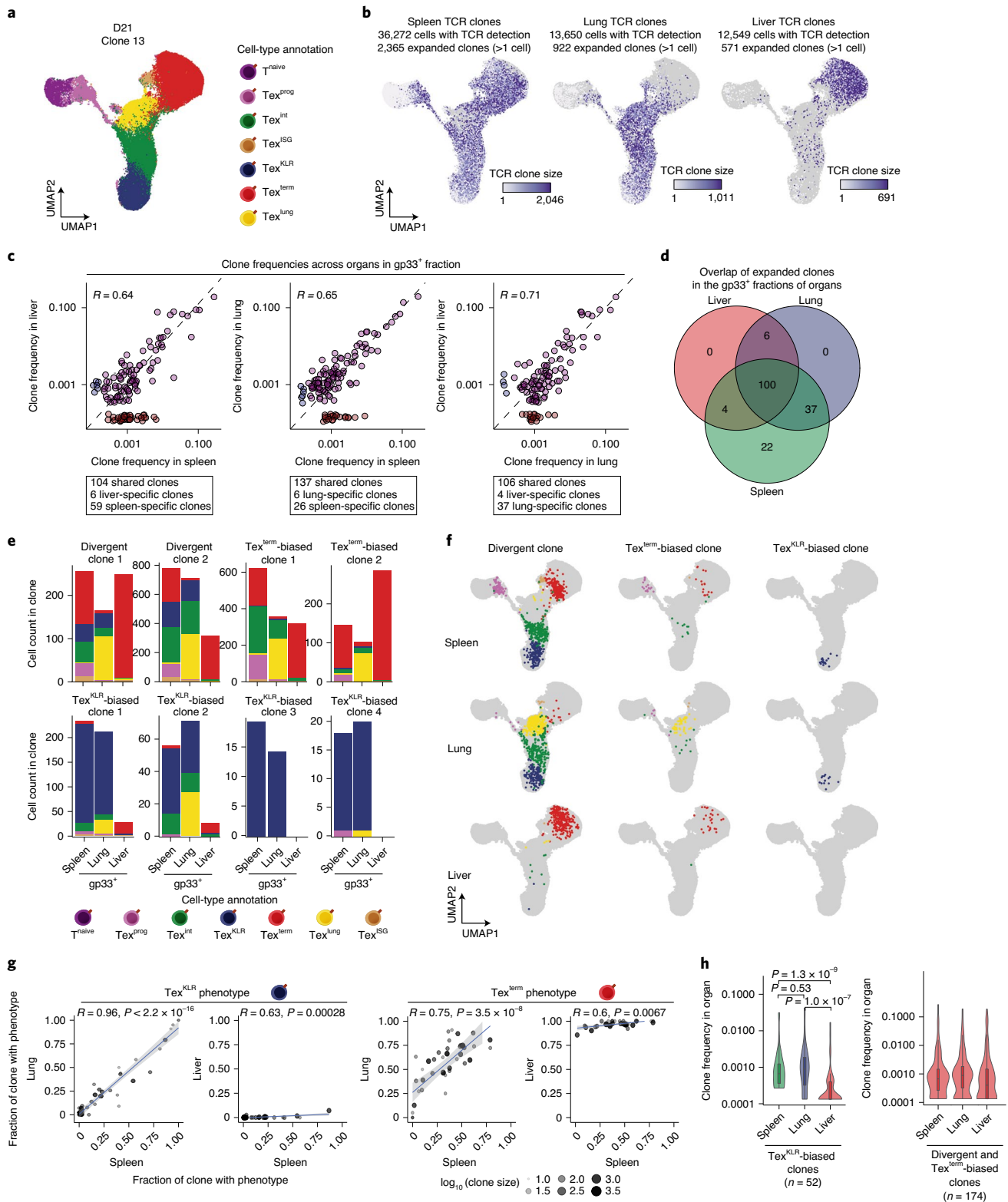**Fig. 5 | Conserved clonal T cell trajectories across organs and depletion of Tex^KLR in the liver microenvironment. a**, UMAP of organ-derived T cells at D21 in Cl13 infection colored by the annotated T cell subsets. **b**, UMAPs colored by the detected TCR clone sizes in the different organs. **c**, Scatter plots comparing the frequencies of expanded T cell clones from the indicated organs. The correlation coefficient (Pearson's R) and specific and shared clone numbers are indicated for each comparison. **d**, Venn diagram depicting the overlap of expanded T cell clones in the gp33+ fraction of the indicated organs. **e**, Stacked bar plot of the phenotypic composition of individual clones across organs. **f**, UMAPs of individual clones with specific clone behaviors across organs. **g**, Scatter plots comparing the fraction of cells in individual clones with Tex^KLR and Tex^term phenotypes between the indicated organs. Correlation coefficient calculated using Pearson's R and P value determined using a two-sided t distribution with n − 2 degrees of freedom. Shaded area represents 95% confidence interval of linear model. **h**, Violin plot of Tex^KLR-biased clone frequencies across the organs, defined as clones with >50% Tex^KLR phenotype in the spleen (left). Violin plot of Tex^term-biased and divergent clone frequencies across the organs (right; n, number of scRNA-seq profiles; box center line, median; box limits, upper and lower quartiles; box whiskers, 1.5× interquartile range). P values determined by two-sided Wilcoxon rank-sum test.
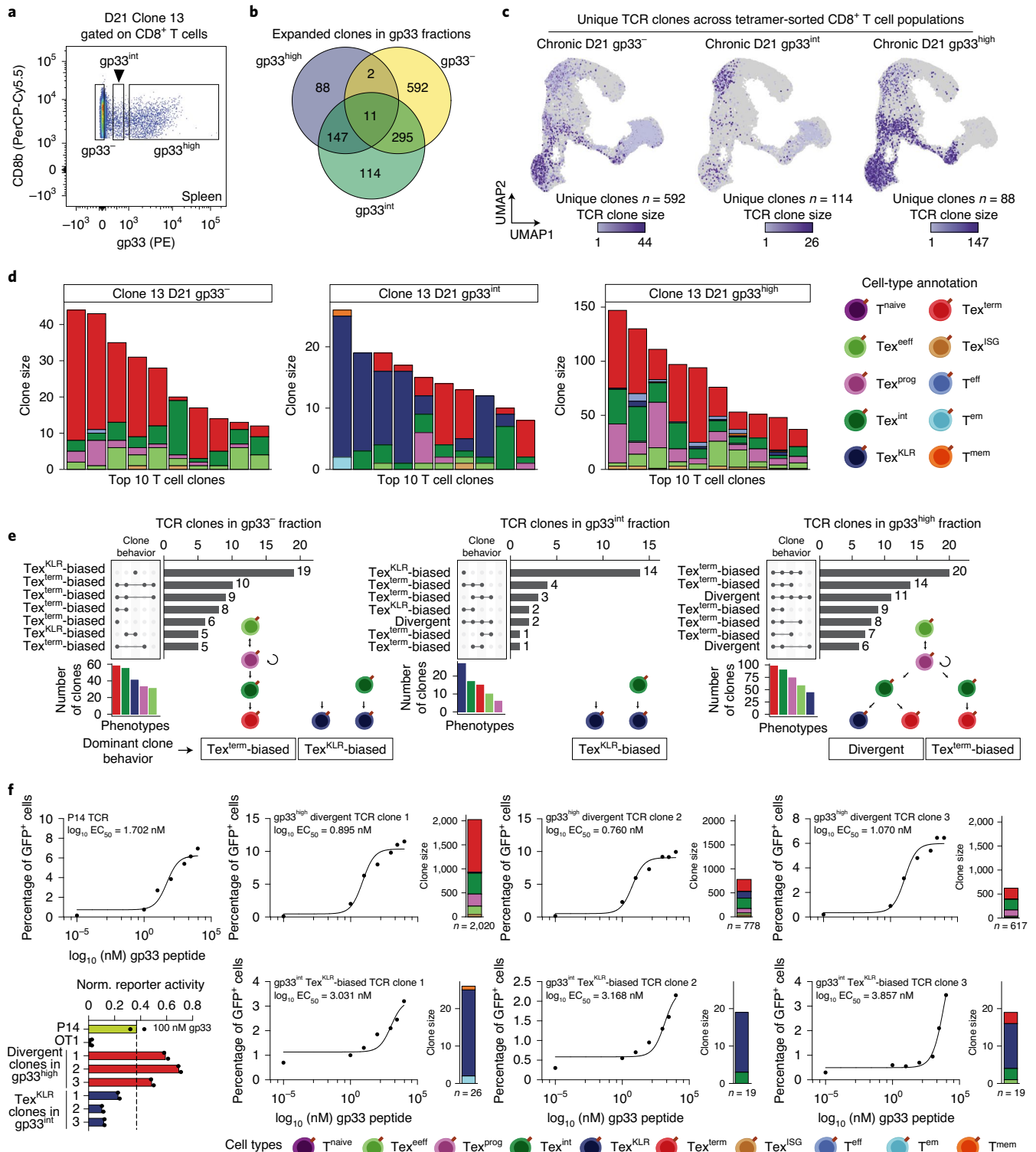
Fig. 6 | Clonal differentiation of Tex^KLR and Tex^term correlates with TCR signaling avidity. a, Sorting strategy to obtain gp33⁻, gp33^int and gp33^high CD8⁺ T cell populations from the spleens of LCMV-Cl13-infected animals 21 days following infection. **b**, Venn diagram of the overlap of expanded clones from the gp33 T cell fractions. **c**, UMAPs colored by size of the unique expanded clones in the three gp33 T cell fractions. **d**, Stacked bar plot of the top 10 uniquely expanded T cell clones from gp33 T cell fractions colored by T cell phenotype. **e**, Upset plots of phenotype combinations (clone behavior) in unique expanded clones from the three gp33 T cell fractions. For clarity, the top seven most common clone behaviors are shown. Bar plots show the number of clones with the indicated phenotypes. Dominant clone behaviors are indicated at the bottom. **f**, Dose response curves of gp33 peptide stimulation of the indicated

TCRs in an NFAT-GFP reporter T cell hybridoma cell line. TCRs were selected from gp33^high (divergent clones) and gp33^int (Tex^KLR-biased clones) fractions. Peptide concentration that triggers half of the maximum signaling avidity is indicated based on percentage of GFP⁺ cells (log₁₀ EC50 values). Clone size and phenotype distribution of each clone is shown as a stacked bar plot. Bar plot shows normalized (to TCRβ expression) percentage of GFP⁺ cells (normalized, Norm. reporter activity) for all TCRs stimulated with a constant peptide concentration (100 nM, bottom left). P14 LCMV-specific TCR was used as a positive control and OT-1 TCR (ovalbumin peptide specific) as a negative control. The mean of two technical replicates are shown of a representative experiment of three independent experiments that all showed the same results.
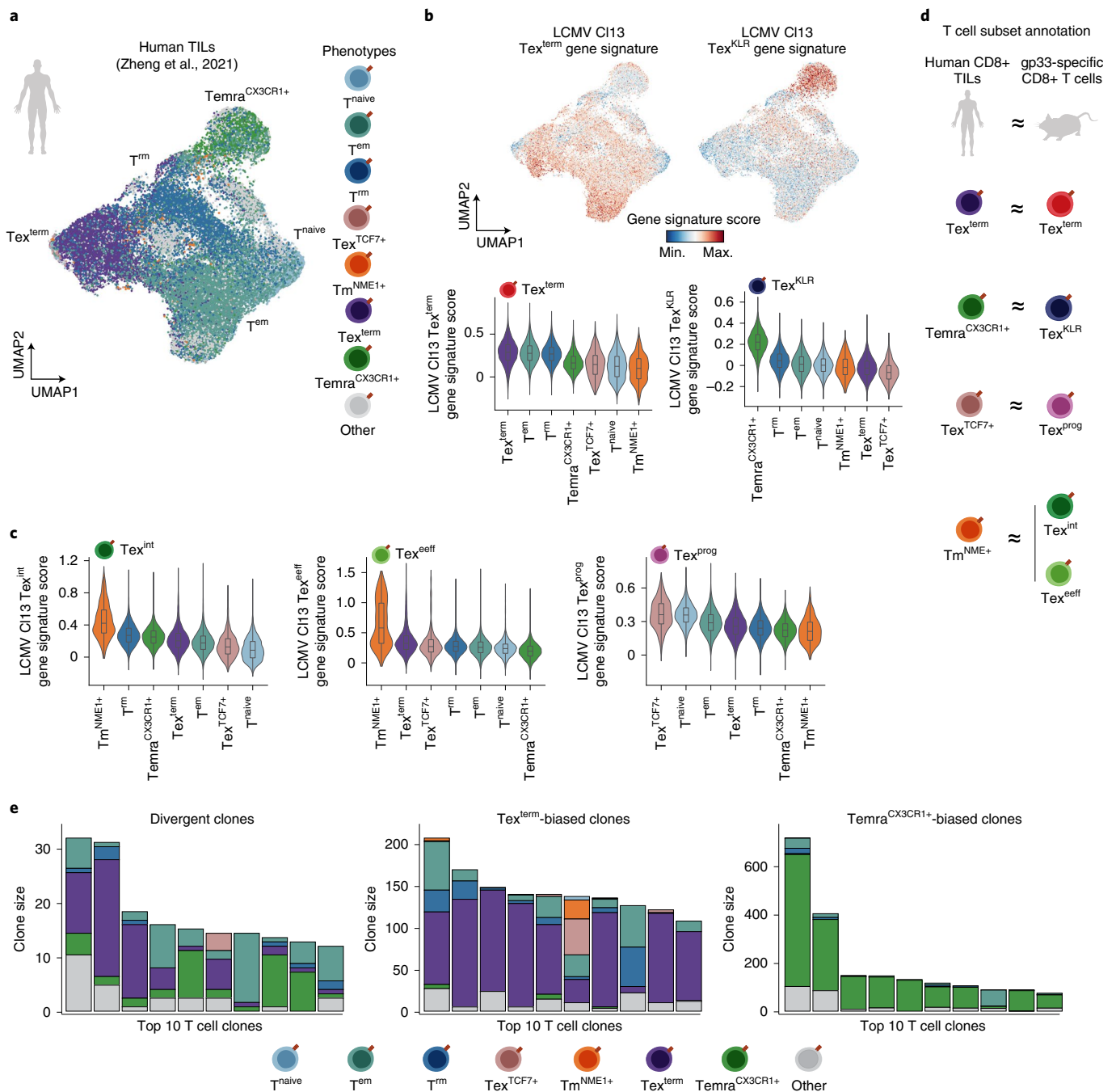
Fig. 7 | Human TILs exhibit shared Tex transcriptional programs and divergent clonal differentiation trajectories. a, UMAP of TILs profiled by scRNA-seq from Zheng et al. (2021)[38], colored by T cell phenotypes. **b**, UMAPs colored by the indicated LCMV Cl13 Tex gene signature scores (top). Violin plots of LCMV Cl13 Tex gene signature scores across human TIL clusters (bottom, box center line, median; box limits, upper and lower quartiles; box whiskers, 1.5× interquartile range). For each violin plot, n is the number of scRNA-seq profiles: T^{naive}, $n = 6,879$; T^{em}, $n = 37,057$; T^{rm}, $n = 18,348$; Tex^{TCF7+}, $n = 757$; Tm^{NME1+}, $n = 593$; Tex^{term}, $n = 13,024$; Temra$^{CX3CR1+}$, $n = 11,702$. **c**, Violin plots of additional

LCMV Cl13 Tex gene signature scores across human TIL clusters (box center line, median; box limits, upper and lower quartiles; box whiskers, 1.5× interquartile range). Human TIL clusters are ranked in decreasing order, with the cluster with highest enrichment of LCMV Cl13 Tex gene signature score on the left. For each violin plot, cell numbers are same as in **b**. **d**, Schematic of the relationships found between human TIL subsets and mouse LCMV-reactive (gp33$^+$) Tex subsets. **e**, Stacked bar plots of the phenotype distribution of the top 10 human expanded TIL clones for the indicated clone behaviors that were detected among LCMV-reactive mouse T cells.

and Tex^{KLR}-biased behaviors, clones from the gp33^{int} fraction were heavily enriched for Tex^{KLR}-biased differentiation, and clones from the gp33^{high} fraction were biased towards Tex^{term}-biased and divergent clone behaviors and lacked Tex^{KLR}-biased clones (Fig. 6e).

Finally, we tested TCR signaling avidity of unique TCR clonotypes from gp33^{high} and gp33^{int} fractions (top three TCRs derived from divergent and Tex^{KLR}-biased clones). We computationally assembled

full-length TCRα and TCRβ sequences from scRNA/TCR-seq data (Extended Data Fig. 7i and Supplementary Table 12), cloned and virally transduced each TCR into an NFAT-GFP reporter cell line. We also cloned the P14 TCR (positive control; high-affinity gp33-specific TCR) and an OT-1 TCR recognizing ovalbumin (negative control). We cocultured each T cell hybridoma TCR line with increasing concentrations of gp33 peptide-pulsed splenocytes and measured GFP expression

as an indicator of downstream of TCR signaling. These studies supported two concepts: (1) TCRs from the gp33[int] and gp33[high] fractions induced GFP expression after coculture with gp33 peptide, confirming gp33-reactivity; and (2) divergent TCRs from the gp33[high] fraction exhibited higher signaling avidities (mean half maximal effective concentration (EC50), 8.4 nM; positive control P14, 50 nM) compared with Tex[KLR] clones from the gp33[int] fraction (mean EC50, 3,200 nM; Fig. 6f). These results indicate that divergent/gp33[high] TCRs exhibit ~400× higher signaling avidity than Tex[KLR]/gp33[int] TCRs, supporting a role for TCR signal strength in driving divergent Tex trajectories.

Human TILs show conserved Tex clone behavior

We analyzed tumor-infiltrating lymphocytes (TIL) to investigate whether Tex clone behaviors are conserved in human cancer. We reanalyzed a dataset of 109,089 CD8[+] T cells profiled by scRNA/TCR-seq from 47 patients spanning 10 cancer types, which had previously been categorized into seven predominant subsets: (1) T[naive]; (2) T[em]; (3) tissue-resident memory (T[rm]); (4) TCF7[+] exhausted (Tex[TCF7+]); (5) NME1[+] memory-like (Tm[NME1+]); (6) Tex[term]; and (7) CX3CR1[+] terminally differentiated effector memory (Temra[CX3CR1+]; Fig. 7a)[38]. We scored each TIL subset for enrichment of gene signatures from Tex subsets in Cl13 infection and found a concordance between murine and human T cell types. Namely, the murine Tex[term] gene signature was highly enriched in human Tex[term], the Tex[KLR] gene signature was highly enriched in human Temra[CX3CR1+], the Tex[int] and Tex[eeff] gene signatures were enriched in Tm[NME1+] and the Tex[prog] gene signature was enriched in Tex[TCF7+] (Fig. 7b–d).

Next, we analyzed the phenotypic distribution of expanded TIL clones (n = 729) and found that most clones exhibited Tex[term]-biased behavior (n = 483), comprising Tex[TCF7+], T[em], T[rm] and Tex[term] (Extended Data Fig. 8a,b). In addition, we detected Temra[CX3CR1+]-biased clones (n = 225), which primarily comprised cells with a Tex[KLR]-like phenotype (Supplementary Fig. 8a,b). Finally, we also observed TIL clones with divergent behavior, comprising Tex[term] and Temra[CX3CR1+] (n = 24 clones; Extended Data Fig. 8a,b). Analyzing the top 10 clones with each behavior demonstrated that apart from a few large Temra[CX3CR1+]-biased clones, clone sizes were similar between Tex[term] and Temra[CX3CR1+] clones and smaller in divergent clones (Fig. 7e). Taken together, these results demonstrate similarities in gene expression signatures of human CD8[+] TIL and LCMV-reactive murine Tex subsets and show that human TIL clones can exhibit divergent, Tex[KLR]-biased and Tex[term]-biased differentiation paths analogous to those observed in Cl13 infection.

Discussion

Here we report a single-cell multiomic atlas of T cell exhaustion during chronic viral infection, which reveals new Tex subsets and their molecular programs, identifies multiple differentiation trajectories of Tex clones and nominates TCR signal strength as a driver of clonal behavior. Previous studies have described multiple Tex subsets with distinct phenotypic and functional traits, including CX3CR1[+] transitory exhausted cells[8,9,13]. Here, we show that this CX3CR1[+] population encompasses two Tex subsets with distinct functionalities: (1) Tex[int], which represent a bifurcation point in Tex differentiation to either Tex[term] or Tex[KLR]; and (2) Tex[KLR] with superior functionality and proliferative capacity compared with Tex[int] and Tex[term]. This heterogeneity may help to explain previous discrepancies in understanding the lineage potential and differentiation path of CX3CR1[+] Tex cells[8,9,15]. Notably, the development of Tex[KLR] occurs via two distinct clonal differentiation paths (Tex[KLR]-biased and divergent clones), and future studies are needed to understand whether clonally unrelated Tex[KLR] cells possess distinct functional characteristics.

Given the stable epigenetic state of Tex[22,39], it is important to understand the stage at which the Tex epigenetic program is initiated. Previous studies have demonstrated that early TCF1[+] Tex[prec] cells possess the epigenetic signature of Tex and can seed additional Tex subsets[14,37]. Here, we find that the Tex program is initiated at an earlier stage in

Tex[eeff]. scATAC-seq analysis and adoptive transfer experiments suggest that this fate decision may be driven by NFAT and BATF, followed by subsequent activation of BACH2 and TCF1 to give rise to Tex[prec]/Tex[prog14,18]. These findings, coupled with early clonal expansion of Tex[eeff], support a model in which the Tex[prec] and Tex[prog] pool originates from Tex[eeff]. However, it is important to note that our results do not exclude the possibility that Tex[prog] can also differentiate to Tex[eeff]. Future lineage tracing experiments will further clarify the relationship between these early Tex cell types.

Downstream of Tex[prog], the differentiation trajectory of Tex has largely been thought to follow a linear path, although some studies have suggested a bifurcating path[3,9]. Our data identify two late-stage Tex types (Tex[KLR] and Tex[term]) and show that individual clones can follow three differentiation trajectories resulting in Tex[term]-biased, Tex[KLR]-biased or divergent fates, comprising both cell types. Furthermore, we find that the differentiation trajectory of Tex clones correlates with TCR signaling avidity and thus seems to be programmed by the TCR and conserved across different tissues. However, additional mechanisms to induce TCR signal strength variation—for example, via inhibitory receptor signaling or access to antigens, cytokines or other factors—may also promote the development of the Tex[KLR] phenotype. Importantly, Tex[KLR]-biased clones were depleted in the liver microenvironment, suggesting that these clones are unable to persist in the liver, whereas Tex[term] were able to persist. Given the high viral load and inflammatory microenvironment of the liver during infection, these results suggest that the Tex[term] phenotype may improve Tex persistence and preserve antiviral effector function in specific organ systems[40].

Finally, these findings may have several implications for cancer. First, several ongoing therapeutic strategies aim to reverse exhaustion; however, our results suggest that Tex[term] may be specifically adapted to survive in high-antigen niches[4,16,17,19–21,41]. Whether the pro-survival aspects of T cell exhaustion can be maintained while reinvigorating effector function will require further study. Second, our findings reinforce the notion that TCR signal strength can direct the fate of T cells[42,43]. Thus, the generation of TCR-based cellular therapies should assess the phenotypic outcomes of TCR activation in addition to antigen recognition. Finally, the observation that a polyclonal T cell response to chronic antigen balances persistence, effector and potential memory functions via the development of two Tex states suggests that future cellular therapies may wish to establish divergent phenotypes[44]. Indeed, a recent study identified a natural killer (NK)-cell-like signature in exhausted human chimeric antigen receptor (CAR)-T cells, which resembles the Tex[KLR] signature, suggesting that this cell type may be present in adoptive cell therapy settings as well[45]. Manipulation of these features and gene regulatory programs of the Tex state may provide important avenues for improved T cell-based immunotherapies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41590-022-01337-5.

References

1. Hashimoto, M. et al. CD8 T cell exhaustion in chronic infection and cancer: opportunities for interventions. Annu. Rev. Med **69**, 301–318 (2018).
2. McLane, L. M., Abdel-Hakeem, M. S. & Wherry, E. J. CD8 T cell exhaustion during chronic viral infection and cancer. Annu. Rev. Immunol. **37**, 457–495 (2019).
3. Blank, C. U. et al. Defining 'T cell exhaustion'. Nat. Rev. Immunol. **19**, 665–674 (2019).
4. Scott, A. C. et al. TOX is a critical regulator of tumour-specific T cell differentiation. Nature **571**, 270–274 (2019).

5. Paley, M. A. et al. Progenitor and terminal subsets of CD8+ T cells cooperate to contain chronic viral infection. *Science* **338**, 1220–1225 (2012).

6. Im, S. J. et al. Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. *Nature* **537**, 417–421 (2016).

7. Utzschneider, D. T. et al. T cell factor 1-expressing memory-like CD8+ T cells sustain the immune response to chronic viral infections. *Immunity* **45**, 415–427 (2016).

8. Hudson, W. H. et al. Proliferating transitory T cells with an effector-like transcriptional signature emerge from PD-1+ stem-like CD8+ T cells during chronic infection. *Immunity* **51**, 1043–1058. e1044 (2019).

9. Zander, R. et al. CD4+ T cell help is required for the formation of a cytolytic CD8+ T cell subset that protects against chronic infection and cancer. *Immunity* **51**, 1028–1042.e1024 (2019).

10. Beltra, J. C. et al. Developmental relationships of four exhausted CD8+ T cell subsets reveals underlying transcriptional and epigenetic landscape control mechanisms. *Immunity* **52**, 825–841.e828 (2020).

11. He, R. et al. Follicular CXCR5-expressing CD8+ T cells curtail chronic viral infection. *Nature* **537**, 412–428 (2016).

12. Wu, T. et al. The TCF1-Bcl6 axis counteracts type I interferon to repress exhaustion and maintain T cell stemness. *Sci. Immunol.* **1**, eaai8593 (2016).

13. Raju, S. et al. Identification of a T-bet^hi quiescent exhausted CD8 T cell subpopulation that can differentiate into TIM3+CX3CR1+ effectors and memory-like Cells. *J. Immunol.* **206**, 2924–2936 (2021).

14. Utzschneider, D. T. et al. Early precursor T cells establish and propagate T cell exhaustion in chronic infection. *Nat. Immunol.* **21**, 1256–1266 (2020).

15. Sandu, I. et al. Landscape of exhausted virus-specific CD8 T cells in chronic LCMV infection. *Cell Rep.* **32**, 108078 (2020).

16. Khan, O. et al. TOX transcriptionally and epigenetically programs CD8+ T cell exhaustion. *Nature* **571**, 211–218 (2019).

17. Alfei, F. et al. TOX reinforces the phenotype and longevity of exhausted T cells in chronic viral infection. *Nature* **571**, 265–269 (2019).

18. Yao, C. et al. BACH2 enforces the transcriptional and epigenetic programs of stem-like CD8+ T cells. *Nat. Immunol.* **22**, 370–380 (2021).

19. Yao, C. et al. Single-cell RNA-seq reveals TOX as a key regulator of CD8+ T cell persistence in chronic infection. *Nat. Immunol.* **20**, 890–901 (2019).

20. Seo, H. et al. TOX and TOX2 transcription factors cooperate with NR4A transcription factors to impose CD8+ T cell exhaustion. *Proc. Natl Acad. Sci. USA* **116**, 12410–12415 (2019).

21. Belk, J. A., Daniel, B. & Satpathy, A. T. Epigenetic regulation of T cell exhaustion. *Nat. Immunol.* **23**, 848–860 (2022).

22. Pauken, K. E. et al. Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science* **354**, 1160–1165 (2016).

23. Matloubian, M., Kolhekar, S. R., Somasundaram, T. & Ahmed, R. Molecular determinants of macrophage tropism and viral persistence: importance of single amino acid changes in the polymerase and glycoprotein of lymphocytic choriomeningitis virus. *J. Virol.* **67**, 7340–7349 (1993).

24. Pliner, H. A. et al. Cicero predicts *cis*-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell* **71**, 858–871.e858 (2018).

25. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).

26. Chen, Y. et al. BATF regulates progenitor to cytolytic effector CD8+ T cell transition during chronic viral infection. *Nat. Immunol.* **22**, 996–1007 (2021).

27. Stelekati, E. et al. Long-term persistence of exhausted CD8 T cells in chronic infection is regulated by microRNA-155. *Cell Rep.* **23**, 2142–2156 (2018).

28. Chen, J. et al. NR4A transcription factors limit CAR T cell function in solid tumours. *Nature* **567**, 530–534 (2019).

29. Liu, X. et al. Genome-wide analysis identifies NR4A1 as a key mediator of T cell dysfunction. *Nature* **567**, 525–529 (2019).

30. Shan, Q. et al. The transcription factor Runx3 guards cytotoxic CD8+ effector T cells against deviation towards follicular helper T cell lineage. *Nat. Immunol.* **18**, 931–939 (2017).

31. Martinez, G. J. et al. The transcription factor NFAT promotes exhaustion of activated CD8+ T cells. *Immunity* **42**, 265–278 (2015).

32. Chung, H. K., McDonald, B. & Kaech, S. M. The architectural design of CD8+ T cell responses in acute and chronic infection: parallel structures with divergent fates. *J. Exp. Med.* **218**, e20201730 (2021).

33. Evrard, M. et al. Sphingosine 1-phosphate receptor 5 (S1PR5) regulates the peripheral retention of tissue-resident lymphocytes. *J. Exp. Med.* **219**, e20210116 (2022).

34. Milner, J. J. et al. Delineation of a molecularly distinct terminally differentiated memory CD8 T cell population. *Proc. Natl Acad. Sci. USA* **117**, 25667–25678 (2020).

35. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

36. Qiu, X. et al. Mapping transcriptomic vector fields of single cells. *Cell* **185**, 690–711.e645 (2022).

37. Pritykin, Y. et al. A unified atlas of CD8 T cell dysfunctional states in cancer and infection. *Mol. Cell* **81**, 2477–2493.e2410 (2021).

38. Zheng, L. et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science* **374**, abe6474 (2021).

39. Sen, D. R. et al. The epigenetic landscape of T cell exhaustion. *Science* **354**, 1165–1169 (2016).

40. Philip, M. & Schietinger, A. Heterogeneity and fate choice: T cell exhaustion in cancer and chronic infections. *Curr. Opin. Immunol.* **58**, 98–103 (2019).

41. Stadtmauer, E. A. et al. CRISPR-engineered T cells in patients with refractory cancer. *Science* **367**, eaba7365 (2020).

42. Chang, J. T., Wherry, E. J. & Goldrath, A. W. Molecular regulation of effector and memory T cell differentiation. *Nat. Immunol.* **15**, 1104–1115 (2014).

43. Daniels, M. A. & Teixeiro, E. TCR signaling in T cell memory. *Front. Immunol.* **6**, 617 (2015).

44. Mold, J. E. et al. Divergent clonal differentiation trajectories establish CD8+ memory T cell heterogeneity during acute viral infections in humans. *Cell Rep.* **35**, 109174 (2021).

45. Good, C. R. et al. An NK-like CAR T cell transition in CAR T cell dysfunction. *Cell* **184**, 6081–6100.e6026 (2021).

Methods

Mice and infection

Male C57BL/6J (CD45.2) and B6.SJL-*Ptprc*[a] *Pepc*[b]/BoyJ (CD45.1) mice were purchased from Charles River Laboratories or The Jackson Laboratories. All mice were housed in a specific-pathogen-free facility and were used for infection at 8–12 weeks of age. Mice were housed with a 12 h light/12 h dark cycle, and the temperature was kept between 18 and 24 °C with 50% humidity. LCMV infection was performed essentially as described previously[46]. All experiments were performed according to protocols approved by the Institutional Animal Care and Use Committees of Stanford University (protocol number: 33814) and Washington University (protocol number: 21–0244).

Tissue preparation

Single-cell suspensions of the different organs were prepared by manual dissociation. Organs were minced and gently pushed through a 40-μm strainer. Spleen single-cell suspensions were spun, and red blood cells were lysed with ACK-lysis buffer by resuspending the cell pellet followed by 2 min incubation. Cells were then washed with ice-cold phosphate-buffered saline (PBS) and stained for sorting in FACS buffer (PBS, 0.1% bovine serum albumin (BSA), 2 mM EDTA, 5% fetal bovine serum). For the lung and liver single-cell suspensions, organs were cut into small pieces and gently pushed through a 40-μm-diameter strainer. Single-cell suspensions were then layered on top of Ficoll-Paque Plus (Cytiva) and centrifuged according to the manufacturer's recommendations. The lymphocyte fraction was collected and washed with ice-cold PBS and then stained for sorting.

RNA isolation and viral transcript determination

Approximately 10 mg of spleen, liver and lung pieces were cut and tissues were flash frozen on dry ice, followed by the addition of 750 μl TRIzol, and smashed with a plastic pestle, before being further homogenized with a 25-gauge needle. From this tissue homogenate, 375 μl was used and further diluted with 125 μl TRIzol. This mixture was mixed with 100 μl chloroform and spun at maximum speed at 4 °C. RNA was precipitated from the aqueous phase with isopropanol, followed by centrifugation and 70% ethanol washing of the RNA pellet. Then, 500 ng RNA was used for reverse transcription (Qscript), and quantitative real-time PCR (SYBR Green) was performed with the following primers. *LCMV-gp* forward primer: CATTCACCTGGACTTTGTCAGACTC; *LCMV-gp* reverse primer: GCAACTGCTGTGTTCCCGAAAC; *Il21* forward primer: TCATCATTGACCTCGTGGCCC; *Il21* reverse primer: ATCG-TACTTCTCCACTTGCAATCCC. Gene expression was normalized to the expression of the *Hprt* housekeeping gene with the following primers. *Hprt* forward primer: AGGTTGCAAGCTTGCTGGT; *Hprt* reverse primer: TGAAGTACTCATTATAGTCAAGGGCA.

Staining T cells for sorting and analysis

Single-cell suspensions were stained with the following antibodies: CD8b (BioLegend catalog number 126610; conjugated with PerCP-Cy5.5, dilution 1:800), CD4 (BioLegend catalog number 100414; conjugated with APC-Cy7, dilution 1:400), PD-1 (BioLegend catalog number 135216; conjugated with PE-Cy7, dilution 1:200), CX3CR1 (BioLegend catalog number 149008; conjugated with APC, dilution 1:400), SLAMF6 (BD Biosciences catalog number 745250; conjugated with BV605, dilution 1:100), CXCR6 (BioLegend catalog number 151119; conjugated with PE-Cy7 or APC-Cy7, dilution 1:200), TNFRSF9 (BioLegend catalog number 106105; conjugated with PE, dilution 1:200), KLRG1 (BioLegend catalog number 138413; conjugated with BV421, dilution 1:200) and the class I tetramer, H-2Db LCMV gp33–41 (KAVY-NFATC) (PE conjugated, dilution 1:100). Cells were stained with the tetramer for 20 min at 4 °C followed by staining with the combination of the other antibodies for 20 min. Cells were washed in FACS buffer and stained with LIVE/DEAD Fixable Aqua dead cell stain for 20 min in PBS. Intracellular staining was performed with the Foxp3 intracellular

staining kit (eBioscience) according to the manufacturer's protocol with the following antibodies: MKI67 (BioLegend catalog number 652410; conjugated with FITC or PerCP-Cy5.5, dilution 1:200), IFNG (BioLegend catalog number 505806; conjugated with FITC, dilution 1:200) and LAMP-1 (BioLegend catalog number 121624; conjugated with PE-Dazzle 594, dilution 1:200). Flow cytometry analysis and cell sorting were performed with a BD FACSAria III (BD Biosciences) using BD FACSDiva Software 6.0 (BD Biosciences). Flow cytometry data were analyzed with FlowJo v.10 (FlowJo; BD Biosciences).

Adoptive transfers

Adoptive transfer experiments were performed by transferring Tex[eeff] (CD8[+] TNFRSF9[+], SLAMF6[−], CX3CR1[−]) into an infection-matched CD45.1 host. Approximately 4×10^5 cells were transferred from D5 following Cl13 infection into infection-matched hosts ($n = 3$). Transferred cells (CD45.2[+]) were analyzed at D21 following Cl13 infection. Owing to this stringent sorting strategy that excludes Tex[int], Tex[KLR] and Tex[prog] from the transferred cell pool, the Tex[eeff] population comprised 8.7% of the total CD8[+] CX3CR1[−] SLAMF6[−] fraction at D5, which represented a smaller fraction of total Tex[eeff] than would be expected from the scRNA-seq results (52.36% of the gp33[+] fraction was Tex[eeff]). This may have also been in part due to inaccuracies in scRNA-seq cluster-based cell classification. Tex[int] (CD8[+], CX3CR1[+], KLRG1[−], SLAMF6[−]) were isolated from D21 following Cl13 infection and transferred into infection-matched CD45.1 hosts ($n = 5$). Approximately 2×10^5 cells were transferred from D21 and transferred; CD45.2[+] Tex subsets were analyzed at D31. Tex subsets were identified by the following markers. Tex[int]: CX3CR1[+], SLAMF6[−], KLRG1[−]; Tex[prog]: SLAMF6[+], CX3CR1[−]; Tex[KLR]: CX3CR1[+], KLRG1[+], SLAMF6[−], Tex[term]: CX3CR1[−], SLAMF6[−]. We emphasize that whereas both Tex[eeff] and Tex[term] were defined as CX3CR1[−] SLAMF6[−], the double-negative fraction at D21 after Cl13 infection contained ~1% Tex[eeff] cells based on scRNA-seq (Extended Data Fig. 1b, scRNA-seq UMAP) 1; therefore, no additional gating for Tex[term] cells was performed in the Tex[int] transfer experiment. Flow cytometry analysis and cell sorting were performed with a BD FACSAria III (BD Biosciences) using BD FACSDiva Software 6.0 (BD Biosciences). Flow cytometry data were analyzed with FlowJo v.10 (FlowJo; BD Biosciences).

scATAC-seq sample and library generation

scATAC-seq experiments were performed on the 10x Chromium platform as described previously[47]. Briefly, after sorting, T cells were washed with PBS with 0.04% BSA and then subjected to nuclei isolation according to the protocol of the manufacturer. Nuclei were counted, and on average ~10,000 nuclei were submitted for tagmentation. After tagmentation, nuclei were loaded for capture using the 10x Chromium controller. After gel emulsion generation, linear amplification was performed, followed by DNA purification according to the manufacturer's protocol. The resulting DNA was used for library construction as described on the website of the manufacturer. Libraries were quantified with an Agilent Bioanalyzer and sequenced on an Illumina NovaSeq S4 sequencer, using the following setup: 50 bp read 1N, 8 bp i7 index, 16 bp i5 index and 50 bp read 2N. In this reaction, 1N and 2N refer to the DNA insert sequencing, whereas i5 and i7 sequencing identify the individual barcodes of single cells.

scRNA-seq library preparation

scRNA-seq libraries were prepared using a 10× 5' Single Cell Immune Profiling Solution Kit (v.1.1 Chemistry) according to the manufacturer's instructions. Briefly, FACS-sorted cells were washed once with PBS and 0.04% BSA, and on average 10,000 cells were submitted for capture using the 10× Chromium controller. Following reverse transcription and cell barcoding in droplets, emulsions were broken, and complementary DNA (cDNA) was purified using Dynabeads MyOne SILANE followed by PCR amplification (98 °C for 45 s; 14 cycles of 98 °C for 20 s, 67 °C for 30 s, 72 °C for 1 min; 72 °C for 1 min). For gene expression library

construction, 50 ng of amplified cDNA was fragmented, end-repaired and double-sided size-selected with SPRIselect beads. Purified DNA was subjected to PCR amplification with sample indexing primers (98 °C for 45 s; 14 cycles of 98 °C for 20 s, 54 °C for 30 s, 72 °C for 20 s; 72 °C for 1 min). Amplified DNA was double-sided size-selected with SPRIselect beads and quantified using an Agilent Bioanalyzer. scRNA-seq libraries were sequenced on an Illumina NovaSeq S4 sequencer using the following read configuration: 26 bp Read1, 8 bp i7 Index, 91 bp Read2.

scTCR library generation

Single-cell TCR (scTCR) libraries were prepared with a 10× Chromium Single Cell V(D)J Enrichment Kit for mouse T cells (v.1.1 Chemistry) following the manufacturer's protocol. Briefly, after cDNA amplification and clean-up, 2 µl of cDNA was used for target enrichment. The first target enrichment was performed with specific primers, followed by a SPRIselect bead clean-up. The second target enrichment was performed with specific primers, followed by double-sided size-selection with SPRIselect beads. After the two target enrichment steps, the quality of the product was assessed with an Agilent Bioanalyzer. Amplified product was then subjected for fragmentation, followed by end-repair and A-tailing. End-repaired product was then subjected to adapter ligation, followed by SPRIselect bead purification. Product was amplified and barcoded with adapter-specific primers, and the quality of the resulting libraries was determined with an Agilent Bioanalyzer. scTCR-seq libraries were sequenced on an Illumina NovaSeq S4 sequencer using the following read configuration: 26 bp Read1, 8 bp i7 Index, 91 bp Read2.

In vitro assessment of TCR signaling avidity

Sequences of CDR3 variable regions and VDJ gene annotations of paired *Tcra*/*Tcrb* expressed in gp33^high and gp33^int cells were extracted from the single-cell sequencing data[48]. Full-length TCR sequences were assembled at www.imgt.org and confirmed using TRUST4. The P14 TCR sequence was obtained from GenBank (accession numbers: X06772 and X06771). Each Vb-TRBC2-(GSG linker)-T2A-Va sequence was synthesized with 5′ linker (5′-AGGCGCCGGAATTAGATCTCTCGAGCCACC -3′) and 3′ linker (5′-GCAGAGGGTGCTGTCCTGAGACCGAGGATC-3′) sequences (Twist Biosciences) and cloned into XhoI and BamHI sites of the MSCV-based retroviral plasmid MigcaRCh (a gift from C.-S. Hsieh, Washington University), which contains a TRAC *IRES* mCherry sequence downstream of the cloning sites, using an NEB HiFi assembly kit. Each paired *Tcra*/*Tcrb* was retrovirally expressed in 58a⁻b⁻ NFAT-GFP reporter cells[49,50] (Thy1.1⁺, a gift from K. Murphy, Washington University) that had been engineered to express *Cd8a/b* and sorted based on expression of CD8a/b, mCherry and TCRb. Then, 2.5 × 10⁴ cells of each hybridoma were stimulated with 5 × 10⁵ irradiated C57BL/6 splenocytes (3,000 rad) in the presence of varying concentrations of LCMV-gp33 peptide (Genscript) for 40 h in RPMI 1640 media supplemented with 10% fetal bovine serum and 50 µM 2-mercaptoethanol, and frequencies of GFP⁺ cells in Thy1.1⁺ cells were measured by flow cytometry using an Attune flow cytometer.

Cx3cr1-creER-mediated fate mapping

Lineage tracing of *Cx3cr1*-expressing cells was performed as described previously[13]. Briefly, *Cx3cr1*-creER^T2 mice were crossed to *Rosa26*-CAG-loxP-stop-loxP-tdTomato reporter mice and infected with LCMV Cl13. Infected mice were treated with 1 mg tamoxifen (Sigma) dissolved in corn oil (Sigma) by oral gavage 22 days postinfection, followed by analysis of tdTomato expression in LCMV-gp33-specific CD8 T cells 25 days postinfection.

scATAC-seq data processing and analysis

scATAC-seq datasets were processed as described previously[51]. Briefly, reads were filtered, trimmed and aligned to the mm10 reference genome using the 10x Genomics cellranger-atac count pipeline (v.1.2.0).

Processed fragment files were loaded into ArchR (v.1.0.1) for additional processing and analysis. All functions used default parameters unless otherwise specified. Cells were filtered during Arrow file generation using the ArchR createArrowFiles function to remove cells with an enrichment of Tn5 insertions in TSS (TSS enrichment) of less than 4 or less than 1,000 unique fragments. Doublets were identified using the ArchR addDoubletScores function, and predicted doublets were removed using the filterDoublets function. Dimensionality reduction was performed using iterative latent semantic indexing using the ArchR addIterativeLSI function. After initial clustering and UMAP projection, we excluded a small cluster of non-T cells. Cell clustering was performed using the ArchR addClusters function on IterativeLSI reduced dimensions 1:10 and a resolution of 0.4 (reducedDims = 'IterativeLSI', dimsToUse = 1:10, resolution = 0.4). The same dimensions were used for single-cell embedding by UMAP using the ArchR addUMAP function using IterativeLSI reduced dimensions 1:10 and a minimum distance of 0.1 (reducedDims = 'IterativeLSI', dimsToUse = 1:10, minDist = 0.1). Cell clustering and UMAP projection for Chronic LCMV (D8 and D21; Fig. 3) and D8 (chronic and acute; Extended Data Fig. 5a) subsets were performed as described above with the following modifications: dimsToUse = NULL, resolution = 0.2 and minDist = 0.4.

GeneScore matrices were computed by summing Tn5 insertions in the gene promoter and gene body during Arrow file generation using the ArchR createArrowFiles function[51]. Gene score imputation was performed with Magic using the ArchR addImputeWeights function[52]. After clustering of the cells, peaks were called by MACS2 on pseudoreplicates sampled from each cluster to obtain a reproducible peak set retaining cell-type-specific peaks using the ArchR addReproduciblePeakSet function. Peak coaccessibility and Peak2Gene linkages were computed using the ArchR addCoAccessibility and addPeak2GeneLinks functions. TF motif deviations were computed with chromVar using the ArchR addDeviationsMatrix function[25]. Pseudobulk tracks for indicated groups of cells were plotted using the ArchR plotBrowserTrack function with default normalization method based on reads in TSS ('ReadsInTSS'). Differential peak testing was performed using the ArchR getMarkerFeatures function with testMethod = 'wilcoxon' and bias = c('TSSEnrichment', 'log10(nFrags)'). TF motif enrichment in differential peaks was performed using the ArchR peakAnnoEnrichment function. Trajectory analysis was performed using the ArchR addTrajectory and plotTrajectory functions.

scRNA-seq and TCR-seq computational methods

scRNA-seq reads were aligned to the mm10 reference genome and quantified using cellranger count (10x Genomics, v.3.1.0). Filtered gene-barcode matrices that contained only barcodes with unique molecular identifier counts that passed the threshold for cell detection were used for further analysis. scTCR reads were aligned to the mm10 reference genome, and consensus TCR annotation was performed using cellranger vdj (10x Genomics, v.3.1.0). TCR annotation was performed using the 10x cellranger vdj pipeline as described.

Additional analysis was performed in R (v.4.0.3) using Seurat (v.4.0.1) with default function parameters unless otherwise noted[53]. Doublets were predicted using DoubletFinder (v.2.0.3)[54]. Cell types were predicted using SingleR (v.1.4.1) based on mouse bulk RNA-seq reference data (MouseRNAseqData) from celldex (v.1.0.0)[55]. Cells with fewer than 200 genes detected, greater than 5% mitochondrial RNA content or predicted doublets from DoubletFinder and cells annotated as non-T and non-NK cells by SingleR were excluded from analysis. We predicted cell cycle phase based on previously defined gene sets using the CellCycleScoring function[56]. We then split cells by experimental batch and cell cycle (noncycling or G1 versus cycling or G2M/S) into four datasets using Seurat SplitObject and performed batch correction using the Seurat reciprocal principal component analysis (PCA) workflow. First, we normalized and identified variable features for each dataset independently using the Seurat NormalizeData and

FindVariableFeatures. Then, we selected variable features across datasets using Seurat SelectIntegrationFeatures. We excluded variable TCR (^Tr.v) genes, variable Ig (^Ig.v) genes, cell cycle genes (used for cell cycle scoring) and mitochondrial genes (^mt-) from integration features used for downstream analysis. We then scaled data and ran PCA on each dataset independently using these features with Seurat ScaleData and RunPCA. We identified integration anchors using Seurat FindIntegrationAnchors using noncycling datasets as reference datasets and rpca for dimensionality reduction. We integrated all datasets using Seurat IntegrateData using dims = 1:50. Integrated data were used for data scaling with ScaleData and PCA dimensionality reduction with RunPCA. After initial clustering, we noted three small clusters representing 7% of total cells that had low numbers of genes detected and high mitochondrial RNA content; these were excluded from further analysis. Clusters were identified using shared-nearest-neighbor-based clustering based on the first 10 PCs with resolution = 0.45. The same principal components were used to generate the UMAP projections, which were generated with a minimum distance of 0.1. Cell clustering and UMAP projection for chronic D21 T cells (all tissues; Figs. 2 and 5), spleen-derived T cells (chronic and acute, D8 and D21; Figs. 3, 4 and 6), D8 T cells (spleen, chronic and acute, Extended Data Fig. 5), chronic including D5 (spleen, chronic, D5, D8 and D21, Extended Data Fig. 5) were performed as described above with the following modifications:

chronic D21 T cells: dims = 1:10, resolution = 0.2, min.dist = 0.1;

spleen-derived T cells: dims = 1:8, k.param = 50, resolution = 0.5, min.dist = 0.1;

D8 T cells: dims = 1:12, k.param = 40, resolution = 0.28, min. dist = 0.2;

chronic including D5: dims = 1:6, resolution = 0.22, min.dist = 0.01, spread = 1.5.

Expression of selected genes was plotted using log-normalized gene expression values based on original RNA count data before data integration. Marker genes were identified using Seurat FindAllMarkers using a cutoff of p_val_adj < 0.01. Differential gene expression analysis was performed using Seurat FindMarkers using cutoffs of p_val_adj < 0.05 and abs(avg_log2 fold change (FC)) > 0.25. Gene module scoring was performed using Seurat AddModuleScore. TCR clone behaviors were visualized using UpSetR (v.1.4.0). Null distribution of TCR clone behaviors was determined by randomly shuffling TCR clonotype and scRNA phenotype and generating a distribution of TCR clone phenotype combinations (n = 50 iterations). The Morisita–Horn index for quantifying TCR overlap was calculated using the mh function from the R package divo (v.1.0.1).

scRNA velocity analysis

Spliced and unspliced transcript counts were computed using velocyto run10x (v.0.17.17) on scRNA-seq cellranger outputs[35]. The resulting loom files were used for subsequent RNA velocity analysis and visualization with dynamo (v.1.0.0). Preprocessing was performed using dynamo. pp.recipe_monocle. RNA velocity was computed using dynamo. tl.dynamics with model = 'stochastic'. Cell transition probabilities were computed using dynamo.tl.cell_velocities with method = 'pearson' and other_kernels_dict = {'transform': 'sqrt'}. RNA velocity was visualized using dynamo.pl.streamline_plot. Pseudotime was estimated using single-cell potential (umap_ddhodge_potential) based on the vector field topology computed using dynamo.vf.VectorField with basis = 'umap' and dynamo.ext.ddhodge with basis = 'umap'. A caveat of these analyses is that highly distinct transcriptional cell states (for example, Tex^{term} or Tex^{prog}; similarly, in Arm infection, T^{eff}, T^{em} or T^{naive}), as well as local minima and/or maxima neighboring distinct cell states, can drive the velocity signal, influencing the 'start' or 'end' points of the analysis. Further, RNA velocity pseudotime is calculated from the potential of the RNA velocity field and summarizes the direction and magnitude of the dominant local trajectories in UMAP space, rather than nominating terminal differentiation states.

Human scRNA-seq and TCR-seq analysis

Processed scRNA-seq and TCR-seq data from human tumor-infiltrating T cells were downloaded from zenodo[38]. For further analysis, we used scRNA-seq expression data from the R object data/expression/CD8/integration/CD8.thisStudy_10X.seu.rds and scTCR-seq data from the R object data/tcr/byCell/tcr.zhangLab.comb.flt.rds. Additional analysis was performed using Seurat (v.4.0.1). Phenotypic clusters were defined using the provided meta.cluster.simplified assignment. Enrichment of mouse LCMV gene signatures was calculated using the Seurat AddModuleScore. TCR clones were defined using the provided cloneID annotation. TCR clone behaviors for tumor-infiltrating T cells (loc = 'T') were visualized using UpSetR (v.1.4.0). For UMAP visualization of single cells, provided normalized expression data were batch-corrected using RunHarmony in harmony based on the 'patient' variable (v.1.0)[57]. The first 10 harmonized embeddings were used to generate the UMAP projections, which were generated with a minimum distance of 0.1.

Statistics and reproducibility

Statistical analyses were performed in R or GraphPad Prism. Gene expression measurements by quantitative PCR were presented as mean ± s.d. In the bar graphs, significant changes were determined by two-tailed, unpaired t-test at $P < 0.05$. The exact replicate numbers are indicated in the figure legends for each experiment. Differential gene expression analysis was performed using a cutoff of adjusted $P < 0.05$ and absolute (average log$_2$ FC) > 0.25. Statistical parameters are also reported in the figure legends. Data distribution was assumed to be normal, but this was not formally tested. No statistical test was used to predetermine sample size. No data were excluded from the analysis. The experiments were not randomized. The investigators were not blind to allocation during experiments or outcome assessment.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Source data are provided with this paper. All sequencing data generated in this study are available under GEO accession GSE188670.

Code availability

Custom code used in this study is available at https://github.com/katieyost/LCMV-code-2022.

References

46. Chou, C. et al. The transcription factor AP4 mediates resolution of chronic viral infection through amplification of germinal center B cell responses. *Immunity* **45**, 570–582 (2016).

47. Satpathy, A. T. et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**, 925–936 (2019).

48. Song, L. et al. TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat. Methods* **18**, 627–630 (2021).

49. Ise, W. et al. CTLA-4 suppresses the pathogenicity of self antigen-specific T cells by cell-intrinsic and cell-extrinsic mechanisms. *Nat. Immunol.* **11**, 129–135 (2010).

50. Letourneur, F. & Malissen, B. Derivation of a T cell hybridoma variant deprived of functional T cell receptor alpha and beta chain transcripts reveals a nonfunctional alpha-mRNA of BW5147 origin. *Eur. J. Immunol.* **19**, 2269–2274 (1989).

51. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).

52. van Dijk, D. et al. Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716–729.e727 (2018).

53. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e3529 (2021).
54. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e324 (2019).
55. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
56. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
57. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

Author contributions

B.D., K.E.Y. and A.T.S. conceptualized the study. B.D., K.E.Y. and A.T.S. wrote and edited the manuscript, and all authors reviewed and provided comments on the manuscript. B.D., K.E.Y., S.H., K.S., K.J.H.G., X.Y., M.B., C.J.R., Q.S. and Y.Q. performed experiments. K.E.Y., S.L.M. and J.A.B. analyzed data. B.D., K.E.Y., J.R.G., E.J.W., H.Y.C., T.E. and A.T.S. guided experiments and data analysis.

Competing interests

A.T.S. is a founder of Immunai and Cartography Biosciences and receives research funding from Allogene Therapeutics and Merck Research Laboratories. H.Y.C. is a cofounder of Accent Therapeutics, Boundless Bio and Cartography Biosciences and an advisor to 10x Genomics, Arsenal Biosciences and Spring Discovery. K.E.Y. is a consultant for Cartography Biosciences. J.A.B. is a consultant for Immunai. E.J.W. is an advisor for Danger Bio, Marengo, Janssen, NewLimit Inc., Pluto Immunotherapeutics, Related Sciences, Rubius Therapeutics, Synthekine and Surface Oncology. E.J.W. is a founder of Surface Oncology, Danger Bio, and Arsenal Biosciences. The remaining authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41590-022-01337-5.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41590-022-01337-5.

Correspondence and requests for materials should be addressed to Ansuman T. Satpathy.

Peer review information *Nature Immunology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: L. A. Dempsey, in collaboration with the *Nature Immunology* team.

Reprints and permissions information is available at www.nature.com/reprints.
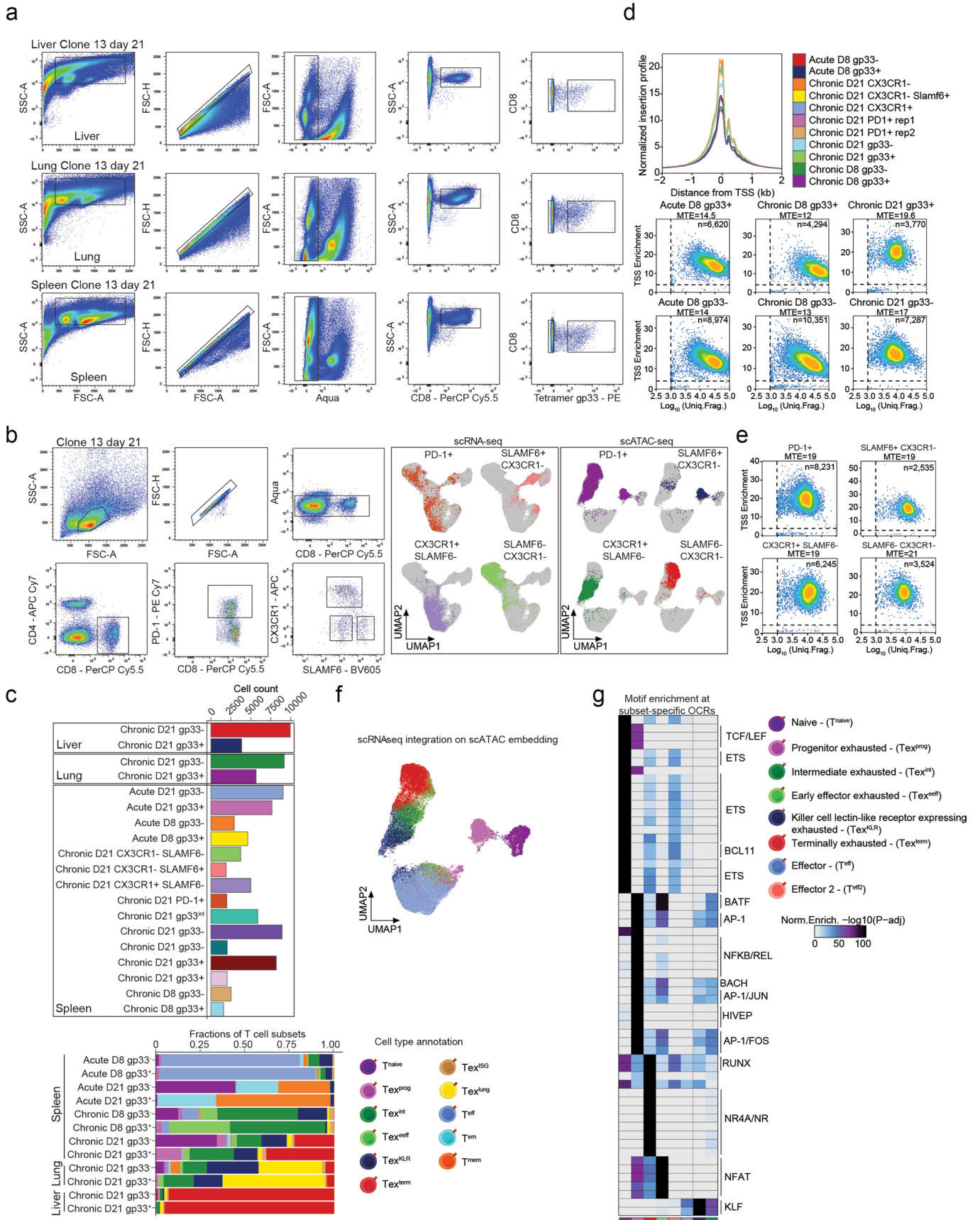
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Sorting strategy and quality controls for scATAC-seq data. (a) Sorting strategy to obtain antigen specific gp33+ and gp33− CD8+ T cells from different organs. **(b)** Sorting strategy to obtain the main Tex subsets (left). UMAPs of scRNA-seq and scATAC-seq results, originating from the indicated Tex subsets. **(c)** Bar plot of cell counts from the scRNA-seq samples (top). Stacked bar plot of the phenotypic composition of the indicated scRNA-seq samples (bottom). **(d)** Quality control of scATAC-seq data. Histogram shows normalized read enrichment on the transcription start sites (TSS) of genes from the indicated samples (top). Density plots depict the cells that passed the TSS enrichment and Log_{10} unique fragment count threshold. Median TSS enrichment (MTE) is also indicated. **(e)** Density plots of scATAC-seq data from the main Tex populations depicting the same quality controls as panel (c). **(f)** UMAP of scATAC-seq data colored by integrated scRNA-seq cluster labels. **(g)** Heat map of TF motif enrichment at the specific open chromatin regions (OCRs) of the annotated T cell populations (p-values determined by hypergeometric enrichment and adjusted using the Bonferroni correction method).
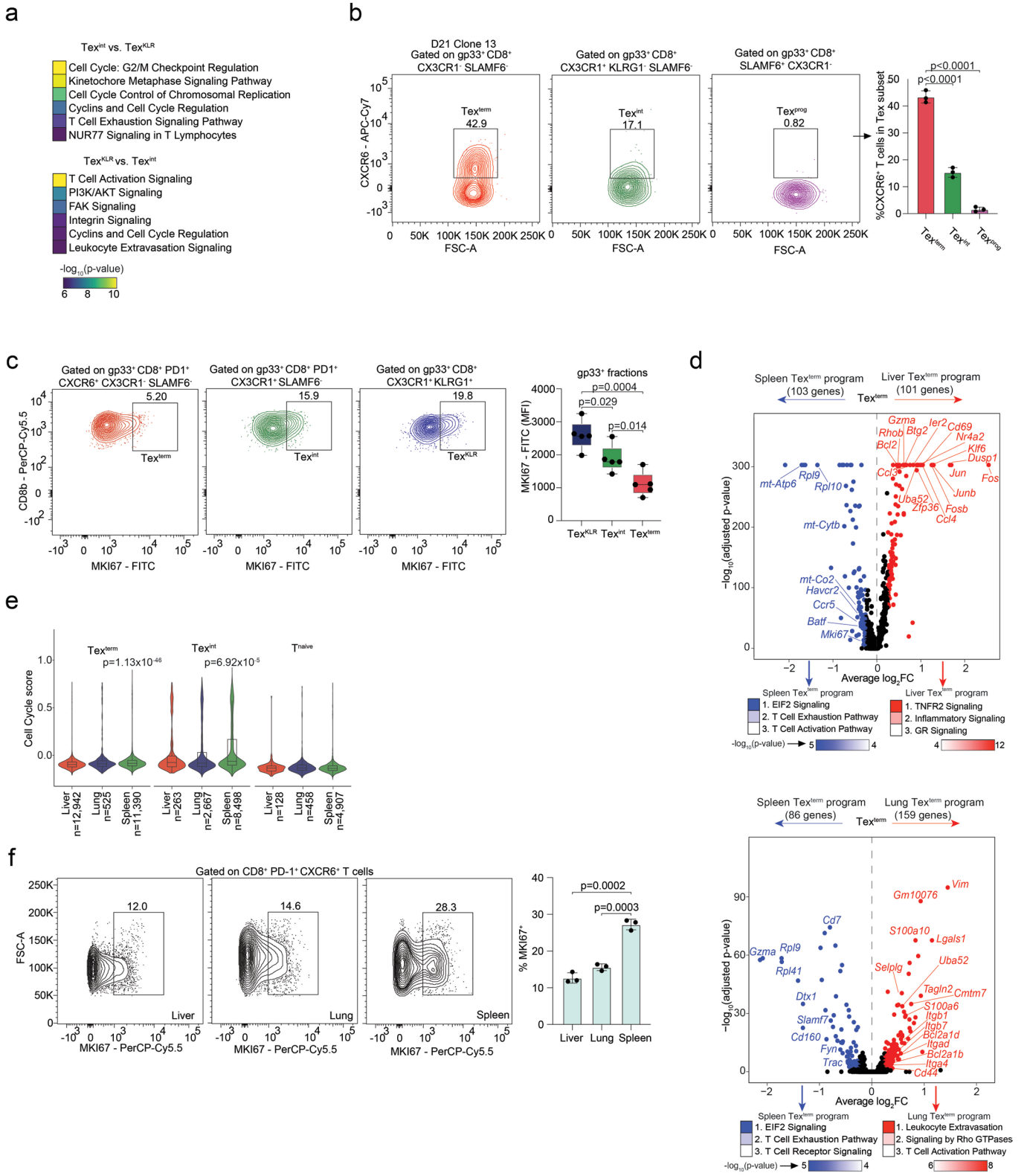
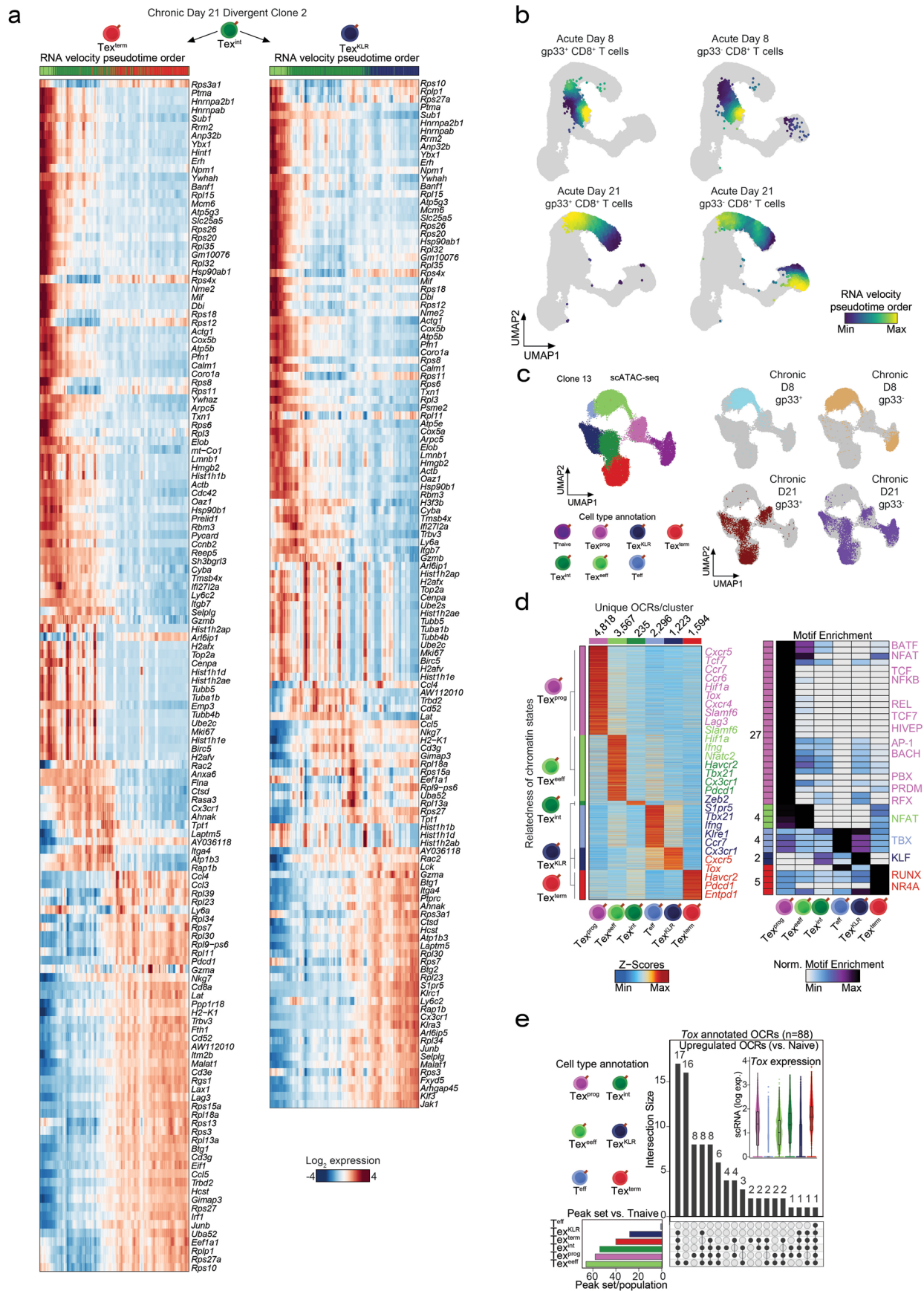Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Characterization of Tex^{int} and Tex^{KLR} subsets and organ-specific exhaustion signatures. (a) Ingenuity pathway analyses of the differentially expressed genes identifying enriched biological pathways between the two subsets (Tex^{int} versus Tex^{KLR}). Top 6 hits are shown. **(b)** Representative flow cytometry plots that quantify CXCR6 expression in Tex^{term}, Tex^{int} and Tex^{prog}. Barplots summarize the quantification across three biological replicates. Significant changes were determined by two tailed, unpaired t-test at $p < 0.05$ (n = 3). Shown are means with SDs. **(c)** Representative flow cytometry plots show the MKI67$^+$ fractions of the indicated Tex subsets. Boxplot depicts the quantification of MKI67$^+$ Tex subsets. Significant changes were determined by two tailed, unpaired t-test at $p < 0.05$ (n = 5 biologically independent animals). Box center line, mean; limits, upper and lower quartiles; whiskers, minimum and maximum values. **(d)** Volcano plots of differentially expressed genes comparing Tex^{term} populations from different organs (log$_2$ FC > 0.25, Bonferroni adjusted p-value < 0.05, p-values determined by two-sided Wilcoxon Rank Sum test). Ingenuity pathway analysis results on the differentially expressed gene groups (bottom). Top 3 hits are shown. **(e)** Violin plots of the Cell Cycle score of the indicated T cell populations across organs (n = number of scRNA-seq profiles, box center line, median; box limits, upper and lower quartiles; box whiskers, 1.5× interquartile range). P-values determined by two-sided Wilcoxon Rank Sum test relative to overall distribution of single cells from the indicated Tex subsets across all organs. **(f)** Representative flow cytometry of the MKI67$^+$ fraction of Tex^{term} subsets in the indicated organs. Bar plot summarizes MKI67$^+$ fractions across organs. Significant changes were determined by two tailed, unpaired t-test at $p < 0.05$ (n = 3 biologically independent animals). Shown are means with SDs.
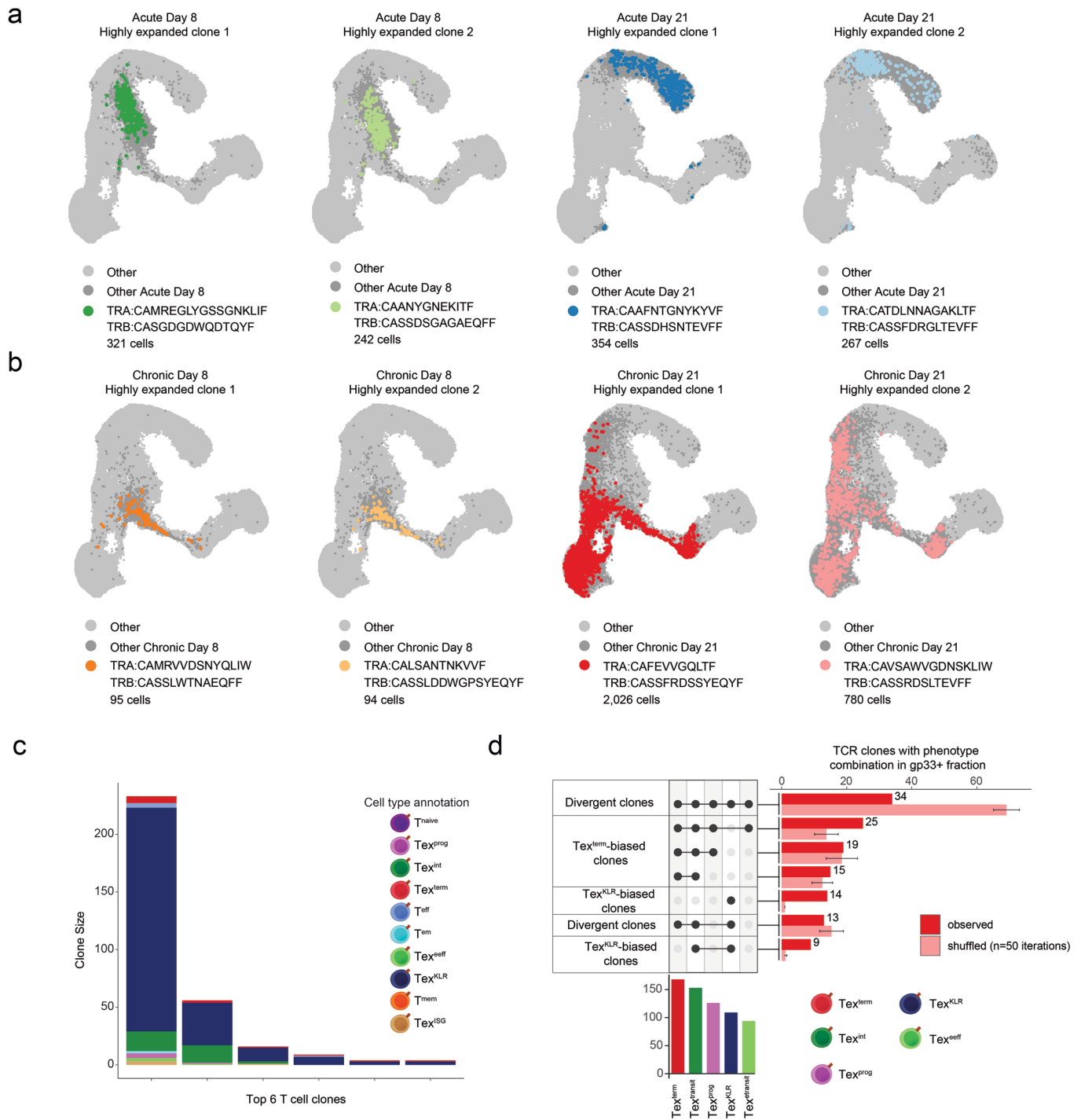
Extended Data Fig. 3 | Analysis of highly-expanded T cell clones in Arm and Cl13 infection. (a) UMAPs of highly expanded clones from the Arm infection model at the indicated time points. **(b)** UMAPs of highly expanded clones of the Cl13 infection model at the indicated time points. **(c)** Stacked bar plot of the phenotypic composition of individual T cell clones with a bias towards the Tex^{KLR} fate that also contain some cells with the Tex^{term} phenotype. Top 6 clones are shown. **(d)** Upset plot of the phenotype combinations of the observed and shuffled TCR clones. Bar represents the mean and error bars represent standard deviation for 50 randomized TCR shuffling iterations performed to obtain the shuffled distribution.

Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | RNA velocity analysis of a divergent clone and regulatory programs of exhaustion. (a) Heat maps depict the gene expression program of a gp33–reactive divergent clone, differentiating from the Tex^{int} stage into either Tex^{term} or Tex^{KLR} fates (left). Pseudotime order (direction of differentiation) was determined by RNA velocity analysis and is presented on a UMAP (right). **(b)** UMAPs of gp33$^+$ CD8$^+$ T cells from Arm infection at D8 and D21 and gp33$^-$ at D21. Color gradient (RNA velocity pseudotime order) indicates directions of T cell differentiation fates determined by RNA velocity analysis. **(c)** UMAP of scATAC-seq results of D8 and D21 gp33$^+$ and gp33$^-$T cells from Cl13 infection. UMAP is colored by the annotated T cell subsets. Small UMAPs (right) show T cells that originate from the indicated gp33 fractions and timepoints. **(d)** Heat map of Peak score values at the unique open chromatin regions (OCRs) of the T cell subsets determined by scATAC-seq with a list of annotated putative target genes based on proximity (left, log$_2$ FC > 1, FDR < 0.05, p-values determined by two-sided Wilcoxon Rank Sum test and adjusted using the Benjamini & Hochberg procedure to obtain FDRs). Heat map of motif enrichment results at the unique OCR sets of the annotated T cell subsets (right, p-values determined by hypergeometric enrichment and adjusted using the Bonferroni correction method). **(e)** Upset plot of differentially accessible OCRs relative to T^{naive} at the *Tox* gene locus and their overlap among the different Tex subsets (log$_2$ FC > 1, FDR < 0.01, p-values determined by two-sided Wilcoxon Rank Sum test and adjusted using the Benjamini & Hochberg procedure to obtain FDRs). Violin plot shows the gene expression level of *Tox* in the identified Tex subsets. Box center line, median; box limits, upper and lower quartiles; box whiskers, 1.5× interquartile range.
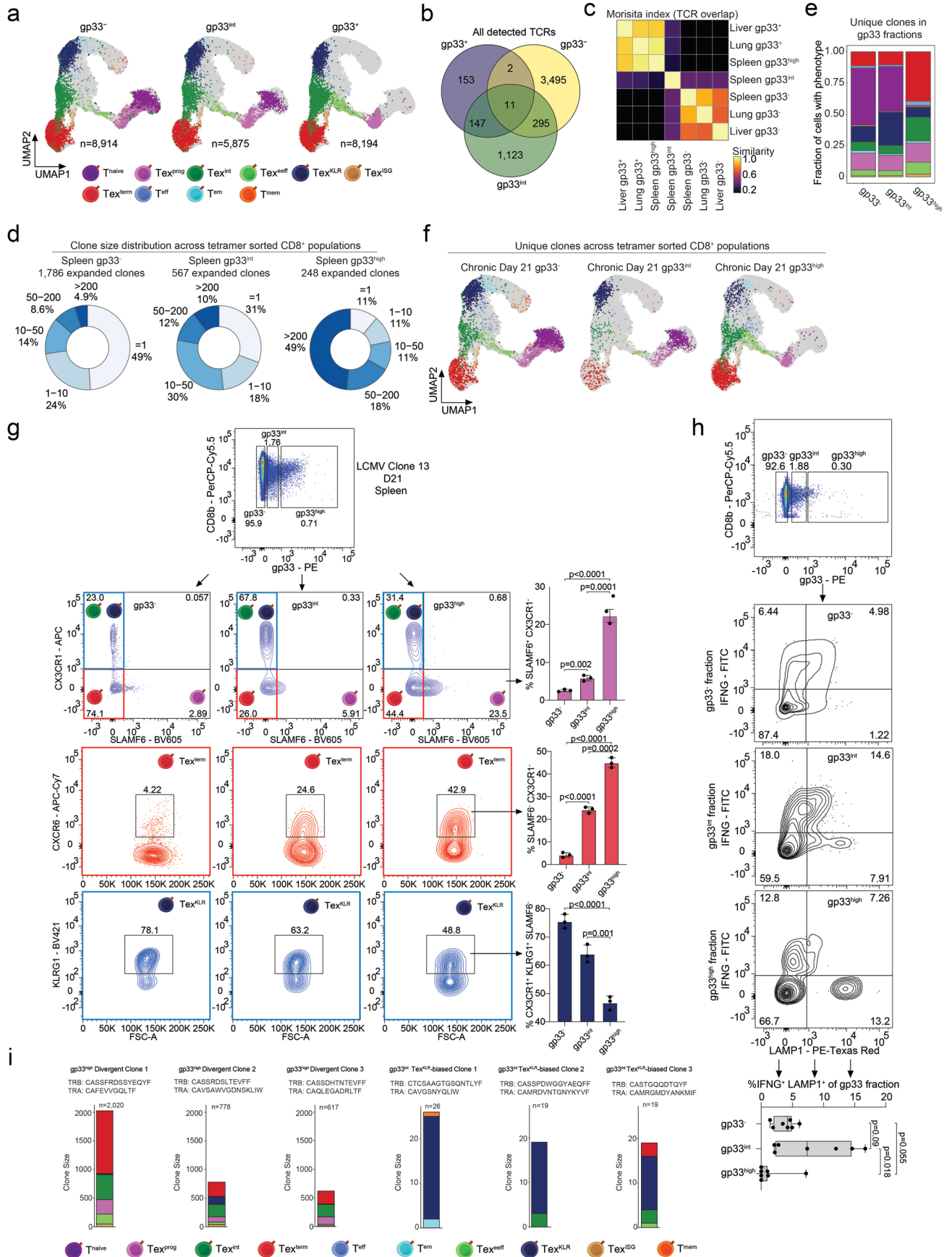
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Molecular programs of early effector- and progenitor-exhausted T cells and fate mapping experiments. (a) UMAPs of scATAC-seq (left) and scRNA-seq (right) results from the infection models. Small UMAPs are colored by sample of origin (bottom). **(b)** Stacked bar plot depicts the phenotypic distribution of gp33$^+$ CD8 T cells from scRNA-seq. **(c)** Volcano plot of differentially expressed genes (DEGs) between memory precursor T cells (T^{mp}) from Arm and the precursor exhausted T cells (Tex^{prec}) from Cl13 infections (top left). Differential gene expression analyses were performed as follows: log$_2$ FC > 0.25, Bonferroni adjusted p-value < 0.05, p-values determined by two-sided Wilcoxon Rank Sum test. Ingenuity pathway analyses of the T^{mp} and Tex^{prec} specific gene sets. Volcano plot of the differential open chromatin regions (OCRs) of the T^{mp} and Tex^{prec} populations (top right). Differential OCR analyses were performed as follows: log$_2$ FC > 1, FDR < 0.1, p-values determined by two-sided Wilcoxon Rank Sum test and adjusted using the Benjamini & Hochberg procedure to obtain FDRs. Enriched transcription factor (TF) motifs in specific OCRs of T^{mp} and Tex^{prec} subsets are shown (p-values determined by hypergeometric enrichment and adjusted using the Bonferroni correction method). **(d)** DEGs between the D8 effector T cells (T^{eff}) from Arm and early effector Tex cells (Tex^{eeff}) from Cl13 infections (top left). Same statistical approach was used as in (c). Ingenuity pathway analyses of the T^{eff}

and Tex^{eeff} specific gene sets (bottom left). Volcano plot depicts the differentially accessible OCRs of T^{eff} and Tex^{eeff} populations (right). Same statistical approach was used as in (c). Enriched TF motifs in T^{eff} and Tex^{eeff} specific OCR sets (p-values determined by hypergeometric enrichment and adjusted using the Bonferroni correction method). **(e)** Expression of the indicated genes profiled by scRNA-seq. Box center line, mean; limits, upper and lower quartiles; whiskers, minimum and maximum values. **(f)** Schematic of the Tex^{eeff} adoptive transfer experiment with pre-transfer sorting strategy (top). Representative flow cytometry plots show the analysis of the phenotypic content of recovered T cells. Stacked bar plots show the phenotypic distribution from 3 biologically independent animals (mean % of each subset is shown with SDs). **(g)** Heat map of differentially expressed TFs in Tex subsets. **(h)** Upset plot of differentially accessible OCRs of Tex^{KLR} and Tex^{term} relative to Tex^{int} (log$_2$ FC > 1, FDR < 0.01, p-values determined by two-sided Wilcoxon Rank Sum test and adjusted using the Benjamini & Hochberg procedure to obtain FDRs) and their overlap. **(i)** Schematic of lineage tracing experiment (left). Gating strategy to analyze tdTomato$^+$ fractions of Tex subsets. Boxplot of the % of tdTomato$^+$ Tex subsets are shown. Box center line, mean; limits, upper and lower quartiles; whiskers, minimum and maximum values (n = 4 biologically independent animals).
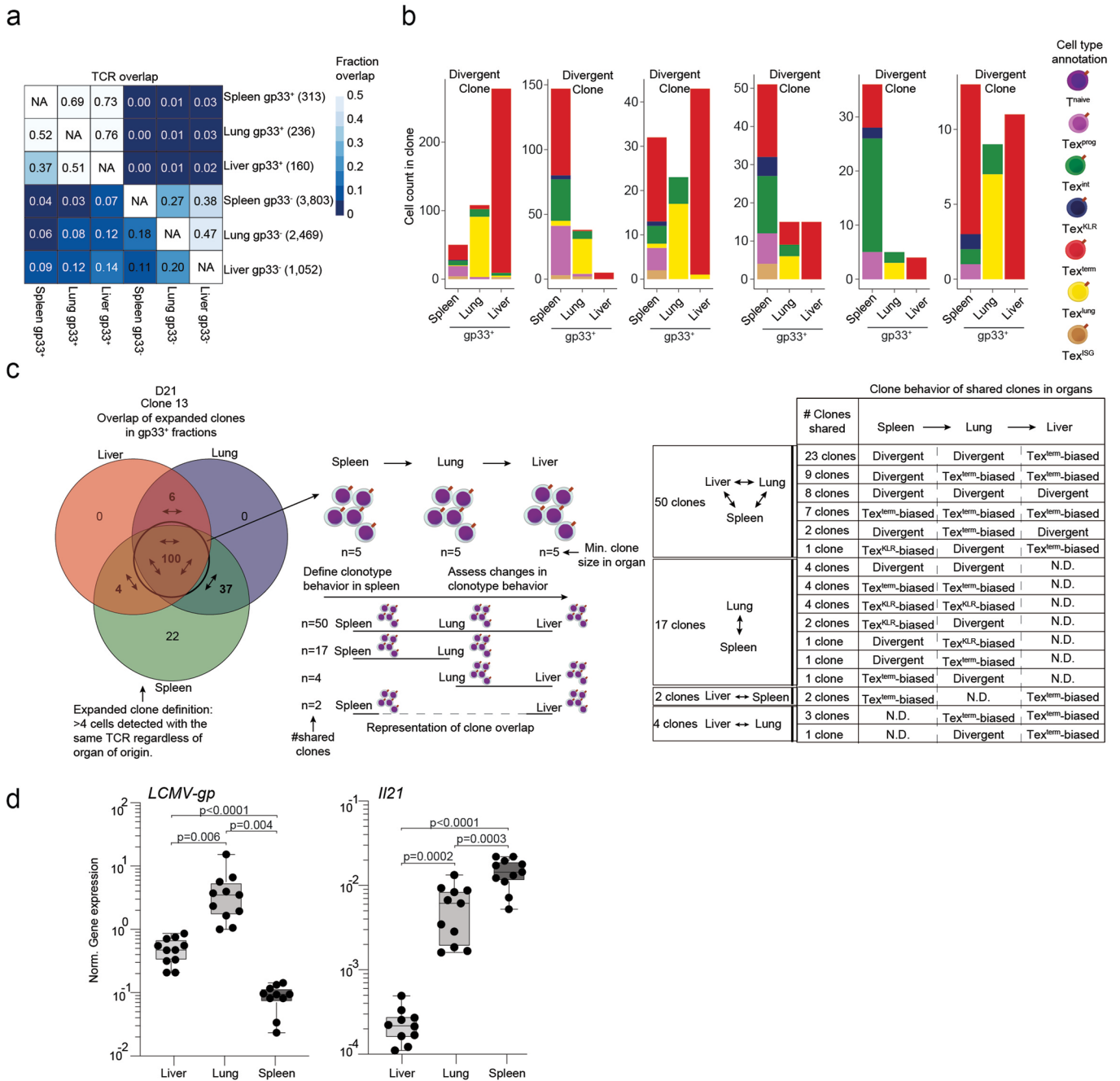
Extended Data Fig. 6 | scRNA/TCR-seq reveals T cell clone behaviors in different organs. (a) Heat map of the fraction overlap between the TCR repertoires of the indicated gp33+ and gp33‾ CD8+ T cell subsets from different organs. **(b)** Stacked bar plot of the phenotypic composition of individual clones with divergent behavior across organs. **(c)** Schematic of the definition of an expanded, organ-shared T cell clone for clone behavior analysis. Only clones that had at least 5 T cells present in each organ were considered. Shared clone numbers across the organs are indicated (left). Table depicting the number of expanded clones that are shared across tissues and their clone behaviors (right). **(d)** Gene expression analysis of *LCMV-gp* transcript (left) and *Il21* transcript (right) in indicated organs at D22 following Cl13 infection. Significant changes were determined by two tailed, unpaired t-test (n = 10 biologically independent animals). Box center line, mean; limits, upper and lower quartiles; whiskers, minimum and maximum values.

Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | scRNA-seq reveals the phenotypic composition of gp33⁻, gp33^int, and gp33^high T cell subsets. (a) UMAPs of cells profiled by scRNA-seq in the three gp33 T cell fractions colored by the annotated T cell subsets. **(b)** Venn diagram shows the overlap of all detected TCR clones among the three gp33 T cell fractions. **(c)** Heat map depicting TCR repertoire overlap (Morisita index) among the different gp33 fractions from the indicated samples. **(d)** Pie chart representation of the fraction of the detected clone sizes in the three gp33 T cell fractions. **(e)** Stacked bar plot of the phenotypic distribution of the unique clones from the three gp33 T cell fractions. **(f)** UMAPs of unique TCR clones determined by scRNA/TCR-seq and colored by the phenotypic distribution of the three gp33 fractions of T cells. **(g)** Representative flow cytometry plots depict the gating strategy to analyze the fractions of Tex subsets in gp33⁻, gp33^int, and gp33^high CD8⁺ T cells. Bar plots quantify the frequencies of the indicated Tex subsets in each gp33 fraction. Significant changes were determined by two tailed, unpaired t-test at $p < 0.05$ (n = 3 biologically independent animals). Shown are means with SDs. **(h)** Representative flow cytometry plots depict the gating strategy to analyze the functionality (IFNG/LAMP1 double positive CD8⁺ T cells) of gp33⁻, gp33^int, and gp33^high CD8⁺ T cells. Boxplot quantifies the double positive fractions of T cells in each gp33 fraction. Significant changes were determined by two tailed, unpaired t-test at $p < 0.05$ (n = 7 biologically independent animals). Box center line, mean; limits, upper and lower quartiles; whiskers, minimum and maximum values. **(i)** Stacked bar plots show the phenotypic content of TCR clones that were tested for TCR signaling avidity. CDR3 amino acid sequences are indicated.

a

b
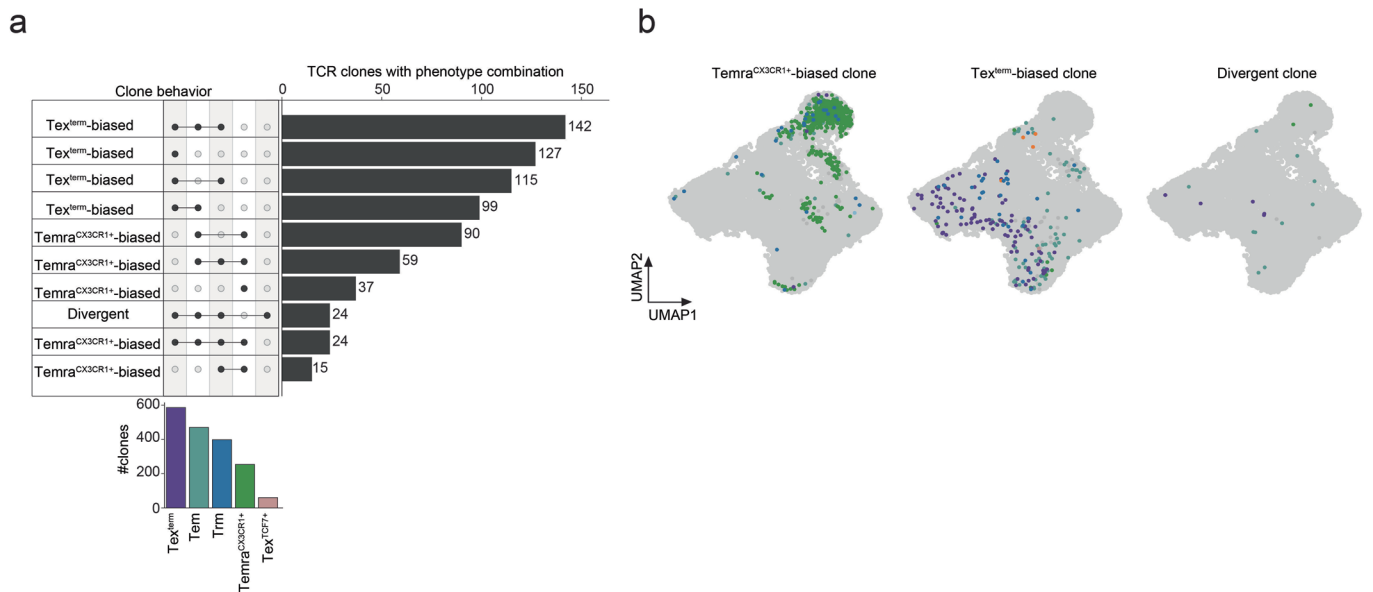
Extended Data Fig. 8 | Human expanded TILs exhibit divergent, Tex^term^- and Tex^KLR^-biased clone behaviors. (a) Upset plot depicting the combination of phenotypes (clone behaviors) for expanded TIL clones. For clarity, the top 10 most common clone behaviors are shown. Bar plot shows the number of clones with the indicated phenotypes. **(b)** UMAPs of representative expanded TIL clones with the indicated clone behaviors.

nature portfolio

Corresponding author(s):   Ansuman T. Satpathy

Last updated by author(s):   Sep 12, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	scATAC-seq datasets were processed as described previously. Briefly, reads were filtered, trimmed, and aligned to the mm10 reference genome using 10X Genomics' cellranger-atac count pipeline (version 1.2.0).
	scRNA-seq reads were aligned to the mm10 reference genome and quantified using cellranger count (10x Genomics, version 3.1.0). Filtered gene-barcode matrices that contained only barcodes with unique molecular identifier (UMI) counts that passed the threshold for cell detection were used for further analysis. scTCR reads were aligned to the mm10 reference genome and consensus TCR annotation was performed using cellranger vdj (10x Genomics, version 3.1.0). TCR annotation was performed using the 10x cellranger vdj pipeline as described.
Data analysis	For scATAC-seq data, processed fragment files were loaded into ArchR (version 1.0.1) for additional processing and analysis.
	For scRNA-seq data, analysis was performed in R (version 4.0.3) using Seurat (version 4.0.1) using default function parameters unless otherwise noted. Doublets were predicted using DoubletFinder (version 2.0.3). Cell types were predicted using SingleR (version 1.4.1) based on mouse bulk RNA-seq reference data (MouseRNAseqData) from celldex (version 1.0.0). TCR clone behaviors were visualized using UpSetR (version 1.4.0). Morisita-Horn index for quantifying TCR overlap was calculated using the mh function from the R package divo (version 1.0.1).
	scRNA velocity analysis Spliced and unspliced transcript counts were computed using velocyto's run10x (version) 0.17.17 on scRNA-seq cellranger outputs [41]. The resulting loom files were used for subsequent RNA velocity analysis and visualization with dynamo (version 1.0.0). Preprocessing was performed using dynamo.pp.recipe_monocle. RNA velocity was computed using dynamo.tl.dynamics with model='stochastic'. Cell transition probabilities were computed using dynamo.tl.cell_velocities with method='pearson' and other_kernels_dict={'transform': 'sqrt'}. RNA velocity was visualized using dynamo.pl.streamline_plot. Pseudotime was estimated using single cell potential (umap_ddhodge_potential) based on the vector field topology computed using dynamo.vf.VectorField with basis='umap' and dynamo.ext.ddhodge with basis='umap'.

Human scRNA-, TCR-seq analysis
Processed scRNA-, TCR-seq data from human tumor-infiltrating T cells were downloaded from zenodo [66]. For further analysis we used scRNA-seq expression data from the R object data/expression/CD8/integration/CD8.thisStudy_10X.seu.rds and scTCR-seq data from the R object data/tcr/byCell/tcr.zhangLab.comb.flt.rds. Additional analysis was performed using Seurat (version 4.0.1). Phenotypic clusters were defined using the provided meta.cluster.simplified assignment. Enrichment of mouse LCMV gene signatures was calculated using Seurat's AddModuleScore. TCR clones were defined using the provided cloneID annotation. TCR clone behaviors for tumor-infiltrating T cells (loc == 'T') were visualized using UpSetR (version 1.4.0). For UMAP visualization of single cells, provided normalized expression data was batch corrected using harmony's RunHarmony based on the 'patient' variable (version 1.0) [67]. The first 10 harmonized embeddings were used to generate the UMAP projections, which were generated with a minimum distance of 0.1.

Custom code used in this study is available at https://github.com/katieyost/LCMV-code-2022.

Flow cytometry analyses and sorting was perfomed by DB FACSAria III using BD FACSDiva Software 6.0 (BD Biosciences). The acquired data were analyzed with FlowJo v10 (BD Biosciences).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All sequencing data generated in this study is available under GEO accession: GSE188670.

Single cell data was aligned to the following pre-built mm10 references:
https://cf.10xgenomics.com/supp/cell-atac/refdata-cellranger-atac-mm10-1.2.0.tar.gz
https://cf.10xgenomics.com/supp/cell-exp/refdata-cellranger-mm10-3.0.0.tar.gz
https://cf.10xgenomics.com/supp/cell-vdj/refdata-cellranger-vdj-GRCm38-alts-ensembl-3.1.0.tar.gz

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. The number of timepoints and conditions included in this study was determined based on previous studies to capture the variability in T cell response over the course of either chronic or acute LCMV infection. For single-cell sample size per sample, we attempted to obtain greater than 1,000 paired scRNA/TCR or scATAC profiles per sample to enable analysis of rare cell populations.
Data exclusions	No data were excluded from analysis.
Replication	Samples in this study were derived from mouse samples and biological replicates are available from CD8+ T cell populations that contain overlapping T cell subsets that we sorted and processed for single cell analysis separately. Time course experiments do not have biological replicates, only sorted gp33+ and gp33- populations from the same animal, which report good technical reproducibility. T cell functional assays and experiments assessing the described T cell phenotypes were successfully reproduced.
Randomization	Animals were randomly assigned to either acute (Armstrong) or chronic (Clone 13) LCMV infection groups.
Blinding	No blinding was used for experimental condition. Because these data were generated using objective quantifications, researchers assessing results were not blinded for the experimental design. Blinding is not relevant to this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	CD8b - BioLegend - 126610 - clone YTS156.7.7 CD4 - BioLegend - 100414 - clone GK1.5 PD1 - BD - 135216 - clone 29F.1A12 SLAMF6 - BD - 745250 - clone 13G3 CX3CR1 - BD - 149007 - clone SA011F11 KLRG1 - BioLegend - clone 2F1/KLRG1 CXCR6 - BioLegend - clone SA051D1 MKI67 - BD - clone B56 LAMP1 - BioLegend - clone 1D4B IFNG - BioLegend - clone XMG1.2 TNFRSF9 (4-1BB) - BioLegend - clone17B5 HAVCR2 (TIM3) - BioLegend - RMT3-23
Validation	All antibodies that have been used in this study are commercially available and tested by the manufacturer.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Mouse T cell hybridoma cell line from mouse thymus. The description of the cell line is referenced in the methods section of the study. The NFAT-GFP reporter cells were a generous gift from Kenneth Murphy, Washington University.
Authentication	These cells induce strong GFP signals in the presence of an artificially introduced T cell receptor (TCR) in the presence of the cognate epitope. The cell lines were tested by introducing positive (P14 - high affinity, LCMV-gp33 specific TCR) and negative (OT-1, ovalbumin peptide specific TCR) control TCRs in the presence of the immunodominant LCMV epitope, gp33.
Mycoplasma contamination	We confirm that the T cell NFAT-GFP hybridoma cell line was tested, and were mycoplasma contamination free.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Male C57BL/6J (CD45.2) and B6.SJL-Ptprca Pepcb/BoyJ (CD45.1) mice were purchased from Charles River Laboratories or The Jackson Laboratories. All mice were housed in a specific pathogen-free facility and were used for infection at 8–12 week of age. For lineage tracing experiments, Cx3cr1-creERT2 mice were crossed to Rosa26-CAG-loxP-stop-loxP-tdTomato reporter mice to generate Cx3cr1CreER Rosa26LSL-tdT mice. These animals were housed in a specific pathogen-free facility and were used for infection at 8–12 week of age. Male mice were used. Mice were housed in a 12 light/12dark cycle, and temperature was kept in between 65-75F with 50% humidity.
Wild animals	No wild animals have been used in the study.
Field-collected samples	No field collected samples were used in this study.
Ethics oversight	All experiments were performed according to protocols approved by Stanford University's (protocol number: 33814) and Washington University's (protocol number: 21-0244) Institutional Animal Care and Use Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Single cell suspension of the different organs was prepared by manual dissociation. Organs were minced and gently pushed through a 40-micron strainer. Spleen single cell suspensions were spun, and red blood cells were lysed with ACK-lysis buffer by resuspending the cell pellet followed by 2 minutes incubation. Cells were then washed with ice-cold PBS and stained for sorting in FACS buffer (PBS, 0.1% BSA, 2mM EDTA, 5% FBS). For the lung and liver single-cell suspension, organs were cut into small pieces and gently pushed through a 40-micron diameter strainer. Single-cell suspensions were then layered on top of Ficoll-Paque Plus (Cytiva) and centrifuged according to the manufacturer's recommendations. The lymphocyte fraction was collected and washed with ice-cold PBS, and then stained for sorting.
Instrument	BD FACSAria III was used for sorting.
Software	FlowJo was used for data analysis.
Cell population abundance	All sorted samples were checked for after-sorting puirty (>99%).
Gating strategy	For all the sorting experiments, cells were gated on single lymphocytes by a forward side scatter gate. After excluding dead cells (Aqua positive cells), LCMV antigen specific T cells were defined as CD8+ and tetramer (gp33+) positive. This gating scheme was used for scATAC- and scRNA/TCR-seq experiments. Experiments that used additional exhausted T cell subsets for single cell experiments utilized CD8+ CD4- cells, and further gated on PD1+ cells. CD8+ PD1+ cells were sorted in a three way sorting scheme based on CX3CR1 and SLAMF6 signal. The following three populations were collected: CD8+ PD1+ CX3CR1+; CD8+ PD1+ SLAMF6+; and CD8+ PD1+ CX3CR1- SLAMF6-. CD8+ PD1+ double positive populations were also sorted and processed for scATAC- and scRNA/TCR-seq.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.