HHS Public Access

Author manuscript

Nat Immunol. Author manuscript; available in PMC 2019 December 17.

Published in final edited form as:

Nat Immunol. 2019 July ; 20(7): 890–901. doi:10.1038/s41590-019-0403-4.

Single-Cell RNA-Seq Reveals TOX as a Key Regulator of CD8+ T cell Persistence in Chronic Infection

Chen Yao[1], **Hong-Wei Sun**[1], **Neal E. Lacey**[2], **Yun Ji**[2,7], **E. Ashley Moseman**[3,8], **Han-Yu Shih**[1], **Elisabeth F. Heuston**[4], **Martha Kirby**[4], **Stacie Anderson**[4], **Jun Cheng**[4], **Omar Khan**[5], **Robin Handon**[4], **Julie Reilley**[4,6], **Jessica Fioravanti**[2], **Jinhui Hu**[2], **Selamawit Gossa**[3], **E. John Wherry**[5], **Luca Gattinoni**[2], **Dorian B. McGavern**[3], **John J. O'Shea**[1,*], **Pamela L. Schwartzberg**[4,6,*], and **Tuoqi Wu**[4,6,9,*]

[1]National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD, USA.

[2]National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

[3]National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA.

[4]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA.

[5]Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania Perelman School Medicine, Philadelphia, PA, USA.

[6]National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA.

[7]Present address: Cellular Biomedicine Group, Gaithersburg, MD, USA.

[8]Present address: Department of Immunology, Duke University School of Medicine, Durham, NC, USA.

[9]Present address: Department of Immunology and Microbiology, University of Colorado School of Medicine, Aurora, CO, USA.

Abstract

Progenitor-like CD8+ T cells mediate long-term immunity to chronic infection and cancer and respond potently to immune checkpoint blockade. These cells share transcriptional regulators with memory precursor cells, including TCF1, but it is unclear whether they adopt distinct programs to

*Corresponding authors: NAME: John O'Shea, Pamela Schwartzberg, and Tuoqi Wu, osheaj@arb.niams.nih.gov; pams@nih.gov; tuoqi.wu@ucdenver.edu.

Author Contributions

T.W. conceived the study. T.W., C.Y., P.L.S., and J.J.O. designed the experiments. T.W., C.Y., Y.J., E.A.M., H.-Y.S., N.E.L., M.K., S.A., J.C., J.F., J.H., R.H., J.R., and S.G. performed the experiments. O.K. and E.J.W. provided P14 *Tox*$^{f/f}$ CD4-Cre. T.W., C.Y., H.-W.S., P.L.S., J.J.O., D.B.M., L.G., and E.F.H. analyzed and interpreted the results. T.W., C.Y., and P.L.S. drafted the manuscript.

Competing Interests Statement

E.J.W. has consulting agreements with and/or is on the scientific advisory board for Merck, Roche, Pieris, Elstar, and Surface Oncology. E.J.W. has a patent licensing agreement on the PD-1 pathway with Roche/Genentech. Y. J. has stock in Cellular Biomedicine Group.

adapt to the immunosuppressive environment. By comparing single-cell transcriptomes and epigenetic profiles of $CD8^+$ T cells responding to acute and chronic viral infections, we found that progenitor-like $CD8^+$ T cells became distinct from memory precursors before the peak of the T-cell response. We discovered a co-expression gene module containing *Tox* that exhibited higher transcriptional activity associated with more abundant active histone marks in progenitor-like cells than memory precursors. Moreover, TOX promoted persistence of antiviral $CD8^+$ T cells and was required for the programming of progenitor-like $CD8^+$ T cells. Thus, long-term $CD8^+$ T-cell immunity to chronic viral infection requires unique transcriptional and epigenetic programs associated with the transcription factor TOX.

Introduction

Upon acute infection or vaccination, naïve T cells first differentiate into functional effector cells, a subset of which develop into memory cells and mediate immune protection[1]. In contrast, during chronic viral infection and cancer, T cells become exhausted, characterized by progressive loss of T-cell function and memory potential, upregulation of inhibitor receptors such as PD-1 and CTLA-4, and reduced proliferation[2]. In the past decade, checkpoint-blockade immunotherapies directed against inhibitory receptors have achieved remarkable successes in treating cancers. Recently, the hallmarks of T cell subsets with higher potential to respond to immunotherapies have become the focus of intensive study[3].

Effector $CD8^+$ T cells in acute infection are heterogeneous, comprising short-lived effector cells and memory precursor cells[4]. However, the heterogeneity of $CD8^+$ T cells responding to chronic infection has only recently been explored. In mice chronically infected by lymphocytic choriomeningitis virus (LCMV) strain clone 13, PD-1^{int} $CD8^+$ T cells were selectively expanded after PD-1 blockade relative to the PD-1^{hi} subset[5]. More recently, we and others identified a $CD8^+$ subset during chronic LCMV infection and cancer that expresses the transcription factor TCF1 (encoded by *Tcf7*) and is required for long-term $CD8^+$ T-cell immunity and responses to checkpoint blockade[6–11]. This progenitor-like $CD8^+$ subset exhibits characteristics similar to stem cells, being capable of self-renewal and repopulating the terminally exhausted TCF1^{lo} $CD8^+$ subset. Differentiation of these progenitor-like $CD8^+$ T cells is positively regulated by transcription factors TCF1, Bcl6, and E2A, and suppressed by type I interferon, Blimp-1, and IRF4[6–10,12].

Several known pathways that regulate progenitor-like $CD8^+$ differentiation during chronic infection are also involved in memory T-cell differentiation during acute infection[6–10]. However, progenitor-like $CD8^+$ T cells likely require additional transcriptional and epigenetic programs to specifically adapt and persist in the immunosuppressive environment of chronic infection. Population-level mRNA profiling of virus-specific $CD8^+$ T cells in acute versus chronic viral infections indicates that their transcriptomes gradually diverge after the first week post-infection[13]. However, given the heterogeneity within exhausted $CD8^+$ T cells, it is important to determine whether these differences reflect changes in the frequencies of subpopulations or distinct transcriptional programs at the single-cell level. A deeper understanding of these questions can help elucidate the establishment of the T-cell exhaustion program and optimal treatment windows for immunotherapies.

Single-cell RNA sequencing (scRNA-seq) is a powerful tool to identify new cell subsets and uncover transcriptional differences masked by averaging gene expression of pooled cells[14]. Here, we used scRNA-seq to determine the heterogeneity of virus-specific $CD8^+$ T cells during acute and chronic LCMV infections. Our data revealed that the transcriptional programs of virus-specific $CD8^+$ T cells during the two infections diverged before the peak of the $CD8^+$ T-cell response. Although both subsets expressed a *Tcf7*-associated gene module, progenitor-like $CD8^+$ T cells in chronic infection distinguished themselves from memory precursor cells by their enrichment of a gene module containing *Tox*, encoding TOX, a member of HMG (high mobility group) transcription factors[15]. Progenitor-like $CD8^+$ T cells also showed distinct epigenomic features, exhibiting more abundant active histone markers at genes co-expressed with *Tox*. Moreover, TOX promoted the long-term persistence of virus-specific $CD8^+$ T cells during chronic LCMV infection. Conversely, TOX deficiency led to loss of progenitor-like $CD8^+$ T cells and impaired persistence of antiviral $CD8^+$ T cells. Our results suggest that TOX endows $CD8^+$ T cells with longevity that facilitates long-term antiviral $CD8^+$ immunity during chronic infection.

Results

Heterogeneity of antiviral $CD8^+$ T cells in chronic infection

Here, we used scRNA-seq to perform an unbiased analysis of antiviral $CD8^+$ T cells during chronic infection. We adoptively transferred naïve P14 $CD8^+$ T cells, which express a transgene encoding a T cell antigen receptor (TCR) that recognizes H-$2D^b$ presenting the GP33–41 epitope[16], into C57BL/6 recipients. Mice were then infected with LCMV clone 13, which causes chronic infection[17]. On day 7 post-infection, splenic P14 $CD8^+$ T cells were isolated (Supplementary Fig. 1a) and analyzed by scRNA-seq (10x Genomics). T-distributed stochastic neighbor embedding (t-SNE) and unsupervised graph-based clustering partitioned cells into four clusters based on their transcriptomes[18] (Fig. 1a and Supplementary Fig. 1b). Among the top 126 genes upregulated in cluster 3 (Supplementary Table 1) were *Tcf7*, *Id3*, and *Slamf6* (encoding Ly108), known markers of progenitor-like $CD8^+$ T cells[6] (Fig. 1b). In addition, cells in cluster 3 exhibited high expression of *Tox, Ikzf2*, and *Hif1a*, and low expression of *Id2*, *Gzma*, *Gzmb*, *Cxcr6*, and *Lgals3* (Fig. 1c, d and Supplementary Fig. 1c, d). Based on its transcriptional signature, cluster 3 most likely represents the progenitor-like $CD8^+$ population. To determine how cells in cluster 3 overlap with progenitor-like cells at a single-cell transcriptomic level, we performed a single-cell gene enrichment analysis using 207 progenitor-like signature genes previously identified (Supplementary Table 2)[6]. Almost all cells in cluster 3 showed significant enrichment of progenitor-like signature genes, whereas few cells from other clusters showed significant enrichment (Fig. 1e). This conclusion was independently confirmed by using a published method (AUCell)[19] (Supplementary Fig. 1e).

To model the gradual changes in transcriptomes during cell differentiation, we used pseudotime analysis to plot gene expression profiles along two principle components of variance[20] (Fig. 1f, left). The value of principle component 2 positively correlated with *Tcf7* expression (Supplementary Fig. 1f) and enrichment of a progenitor-like gene signature (Fig. 1f, right). This finding suggested the association of principle component 2 with the transition

between progenitor-like CD8$^+$ T cells and their terminally exhausted counterparts. Principle component 1 likely reflects cell cycle progression through G1-S-G2M phases (Supplementary Fig. 1g). Whereas cells in cluster 0 and 3, both with high component 1 values, were mostly in G1 phase, cells in cluster 1 were mostly in G2M or S phase (Supplementary Fig. 1h). Thus, our results suggested that progenitor-like CD8$^+$ T cells (cluster 3) might first undergo terminal differentiation and then commence proliferation.

CD8$^+$ T-cell transcriptomes in acute and chronic infections diverge early

Given the heterogeneity of antiviral CD8$^+$ T cells responding to acute and chronic infections, we used scRNA-seq to compare their transcriptional programs and the timing when the bifurcation occurs. We collected P14 CD8$^+$ T cells from day 4.5 and day 7 after infection by an acute (Armstrong) or chronic (clone 13) LCMV strain (hereafter referred to as D4.5 Arm, D4.5 Cl13, D7 Arm and D7 Cl13) for both scRNA-seq and bulk RNA-seq. Day 4.5 and Day 7 represent the antiviral CD8$^+$ responses at the mid-phase or near the peak of initial expansion, respectively. We performed scRNA-seq experiments twice and observed strong correlations between replicates (Supplementary Fig. 2a, b). In addition, scRNA-seq data and bulk RNA-seq data from the same condition were highly correlated (Supplementary Fig. 2b). To avoid batch effects, scRNA-seq data from one experiment were pooled for downstream analyses. P14 cells were transcriptionally more active with higher numbers of detected mRNA molecules and genes on day 4.5 than day 7 after infection. (Supplementary Fig. 2c).

Using t-SNE analysis, we found that whereas D4.5 Arm and D4.5 Cl13 CD8$^+$ T cells commingled and generally occupied the same area, D7 Arm and D7 Cl13 CD8$^+$ T cells were partitioned into distinct areas (Fig. 2a). Thus, transcriptional programs between CD8$^+$ T cells responding to acute infection versus chronic infection diverged before the peak of initial expansion. Furthermore, the differences between D7 Arm and D7 Cl13 CD8$^+$ T cells reflected distinct transcriptomes on a single-cell level, not merely changes in the frequencies of subpopulations. Expression patterns of top differentially expressed genes confirmed that differential programming of antiviral CD8$^+$ T cells responding to acute and chronic viral infections began between day 4.5 and day 7 post-infection (Fig. 2b). Gene ontology analysis revealed that genes involved in translation were upregulated on D7 Cl13 CD8$^+$ T cells (Supplementary Fig. 2d). As expected, D7 Arm CD8$^+$ T cells upregulated genes related to T-cell activation and killer cell cytotoxicity (Supplementary Fig. 2e).

We then used an unsupervised approach to assign virus-specific CD8$^+$ T cells to different clusters based on their transcriptional profiles. We identified eleven clusters, including six (1, 3, 4, 5, 7, 9) from both day 4.5 samples, four (2, 6, 7, 8) from D7 Arm cells, and two (0, 10) unique to D7 Cl13 cells (Fig. 2c). Cluster 7 included cells from D4.5 Arm, D4.5 Cl13, and D7 Arm. Cells in this cluster exhibited elevated expression of type I and type II interferon gene signatures and a transcriptional signature of memory precursor cells, including upregulation of *Bcl2*, *Ccr7*, *Id3*, *Sell*, and *Tcf7*, and downregulation of *Ccr2*, *Gzmb*, *Id2*, *Klrg1*, and *Zeb2* (Fig. 2d–f and Supplementary Fig. 2f). These data suggest that the memory precursor fate was established in a subset of antiviral CD8$^+$ T cells by day 4.5 post-infection. Upregulation of *Bcl2*, *Id3*, and *Tcf7* was also observed in cluster 5 (D4.5

Arm and D4.5 Cl13) and cluster 10 (D7 Cl13). In addition, high expression of *Mt1* encoding metallothionein 1, which promotes T cell dysfunction during cancer[21], was found in cluster 4 (D4.5 Arm and D4.5 Cl13) and cluster 0 (D7 Cl13). It was not present in cluster 10 (D7 Cl13), which is enriched for hypoxia response genes (Fig. 2e, f and Supplementary Fig. 2f). Of note, clusters 10 and 0 correspond to TCF1^{hi} progenitor-like and TCF1^{lo} terminally exhausted CD8^{+} T cells during chronic viral infection, respectively. Thus, single-cell transcriptomes of progenitor-like and memory precursor cells diverged by day 7 post-infection despite similar expression of signature genes such as *Tcf7*.

TOX distinguishes progenitor-like from memory precursor cells

Given their similarities, it is unclear whether the transcriptional program of progenitor-like CD8^{+} T cells is distinct from that of memory precursors. Although the progenitor-like signature previously identified[6] was enriched in most progenitor-like cells (cluster 10), it was also enriched in ~ 60% of cluster 7 cells, a subset with memory precursor characteristics, and in ~30% cluster 5 cells (Fig. 3a and Supplementary Fig. 3a). The memory precursor gene signature (Supplementary Table 3) also failed to distinguish progenitor-like cells from memory precursor cells (Supplementary Fig. 3b). Thus, we sought to identify genes that were differentially expressed between progenitor-like CD8^{+} T cells in chronic infection (cluster 10) and other CD8^{+} populations that expressed *Tcf7* (clusters 5 and 7). Among the top genes upregulated in cluster 10 were ribosomal genes, *Cd200*, *Pdcd4*, and *Tox* (Fig. 3b–d). *Tox* encodes a transcription factor that regulates the development of CD4^{+} T cells and innate lymphocytes[22,23].

Our results that *Tox* and *Tcf7* were co-expressed in progenitor-like CD8^{+} T cells but not memory precursors indicate that these two genes might be involved in two separate gene regulatory circuits. Thus, we used co-expression network analysis of our scRNA-seq data to identify modules of co-expressed genes. Forty-nine modules were identified and color-coded, including a module containing genes such as *Tcf7*, *Tnfsf8*, and *Id3* (module 29, saddlebrown) and a separate module containing *Tox* (module 12, tan) (Fig. 3e, Supplementary Fig. 3c, and Supplementary Tables 4, 5). Based on a previous study that identified TOX binding sites in effector CD8^{+} T cells using chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq)[24], we identified 90 of the 200 genes in the *Tox* module as potential direct targets of TOX. Next, we determined the enrichment of each gene module in individual cells. Module 29 exhibited strong enrichment in cells from cluster 5, 7 and 10, similar to the *Tcf7* expression pattern (Fig. 3f and Supplementary Fig. 3d), whereas most cells with enrichment of module 12 fell in cluster 10 (Fig. 3g, Supplementary Fig. 3e). To confirm these findings, we isolated D7 Cl13 progenitor-like (Tim3^{lo}Blimp-1^{lo}), D7 Cl13 terminally exhausted (Tim3^{hi}Blimp-1^{hi}), D7 Arm memory precursor (KLRG1^{lo}), D7 Arm short-lived effector (KLRG1^{hi}), D4.5 Arm, and D4.5 Cl13 P14 cells for bulk RNA-seq (Supplementary Fig. 3f, g). Consistent with the scRNA-seq data, *Tox* was most abundantly expressed in D7 Cl13 progenitor-like cells (Supplementary Fig. 3h). Terminally exhausted T cells also expressed *Tox*, albeit at lower levels than progenitor-like cells. Differential expression of TOX protein was confirmed by flow cytometric analysis (Supplementary Fig. 3i,j). Thus, high expression of *Tox* and its associated module

distinguishes progenitor-like CD8$^+$ T cells responding to chronic viral infection from memory precursors.

Progenitor-like CD8$^+$ T cells display a distinct epigenome

Epigenomic programming is critical for the differentiation of T cells[25–27]. Here, we used ChIP-seq to characterize the genome-wide profiles of histone H3 lysine 27 acetylation (H3K27ac), a marker of active enhancers[28], in D7 Cl13 progenitor-like, D7 Cl13 terminally exhausted, D7 Arm memory precursor, and D7 Arm short-lived effector P14 cells. Progenitor-like CD8$^+$ T cells had the highest number of unique peaks among all four cell types (Fig. 4a and Supplementary Fig. 4a,b). Compared to common peaks, peaks not shared by all cell types (i.e. variable peaks) were enriched in intergenic regions or introns and reduced in promoters (Fig. 4b and Supplementary Fig. 4c). This finding suggests that the differentiation of virus-specific CD8$^+$ T cells is more likely to be impacted by differential activity of enhancers versus promoters. Hierarchical clustering analysis and heatmap of these H3K27ac profiles showed that progenitor-like CD8$^+$ T cells were most distant from the other three cell types (Fig. 4c,d). Hierarchical clustering of scRNA-seq data also placed progenitor-like CD8$^+$ T cells (cluster 10) on the most distant branch (Supplementary Fig. 4d).

Next, we measured H3K27ac modification of genes co-expressed with *Tox* and found greater H3K27ac modification in progenitor-like CD8$^+$ T cells than in memory precursors (Fig. 5a, left). In addition, genes with TOX binding sites[24] exhibited greater H3K27ac in progenitor-like CD8$^+$ T cells than in memory precursor cells (Fig. 5a, right). Moreover, progenitor-like and terminally exhausted signature genes (Supplementary Tables 2,6) exhibited higher H3K27ac abundance in their respective CD8$^+$ populations (Supplementary Fig. 4e, f). We also compared H3K27ac content and mRNA levels between progenitor-like and terminally exhausted cells or between progenitor-like and memory precursor cells (Fig. 5b). In each case, genes upregulated in progenitor-like CD8$^+$ T cells were often associated with one or more differentially modified peaks with increased H3K27ac (fold change>1.5, FDR<0.1), whereas genes downregulated in progenitor-like cells often contained peaks with significantly less H3K27ac (Fig. 5b).

Then, we examined the H3K27ac profiles at individual gene loci that encode markers of the four cell types. H3K27ac peaks were found at the transcription start site (TSS) and −30-kb enhancer[29] of *Tcf7* in memory precursors and progenitor-like CD8$^+$ T cells, but not in short-lived effectors or terminally exhausted CD8$^+$ T cells (Fig. 5c). *Klrg1* was associated with more H3K27ac peaks in short-lived effectors and terminally exhausted CD8$^+$ T cells than in memory precursors and progenitor-like CD8$^+$ T cells (Fig. 5d). Compared to other CD8$^+$ subsets, progenitor-like CD8$^+$ T cells displayed less H3K27ac peaks at upstream regulatory elements of *Gzmb* (Fig. 5e) and lacked an H3K27ac peak 3′ of *Id2* (Supplementary Fig. 4g). In contrast, there were multiple intronic H3K27ac peaks at *Tox* locus that were unique and/or significantly enriched (e.g. +50 kb, +122 kb, and +133 kb) in progenitor-like CD8$^+$ T cells (Fig. 5f).

Based on published ChIP-seq data, several potential transcriptional regulators of CD8$^+$ T cells during chronic infection, including NFAT1, IRF4, RUNX3, FOXO1, STAT3, and

STAT5B[12,30–33], bind to *Tox* locus (Supplementary Fig. 4h). To determine whether calcineurin-NFAT pathway upregulates *Tox* in virus-specific CD8$^+$ T cells during chronic infection, we treated mice that were infected with LCMV clone 13 with calcineurin inhibitor FK506 starting from day 4 post-infection. On day 7 post-infection, blocking calcineurin-NFAT pathway resulted in reduced TOX protein in virus-specific CD8$^+$ T cells, confirming that calcineurin-NFAT positively regulates TOX expression (Supplementary Fig. 4i).

TOX promotes persistence of virus-specific CD8$^+$ T cells

The pathways that endow progenitor-like CD8$^+$ T cells with longevity in the face of immunosuppressive environments, such as chronic infection, remain unclear. To test whether TOX enhances the persistence of virus-specific CD8$^+$ T cells during chronic infection, we overexpressed TOX in P14 CD8$^+$ T cells by retroviral transduction (Supplementary Fig. 5a,b) and transferred transduced cells into C57BL/6 mice, which were immediately infected with LCMV clone 13. TOX overexpression greatly increased the number and frequency of splenic P14 cells on day 14 and day 28 post-infection (Fig. 6a,b) but did not affect the initial expansion of P14 cells (Fig. 6a,b) or the frequency of TCF1^{hi} cells (Fig. 6c and Supplementary Fig. 5c). Thus, TOX overexpression enhanced the persistence of virus-specific CD8$^+$ T cells without skewing their differentiation towards progenitor-like cells. Next, we sorted control and TOX-overexpressing Tim3^{hi} (TCF1^{lo} terminally exhausted) P14 CD8$^+$ T cells on day 7 post-infection (Supplementary Fig. 5d) and transferred them into infection-matched mice. On day 5 post-transfer, significantly more TOX-overexpressing P14 cells were present than control cells, although both groups remained committed to the TCF1^{lo} lineage (Fig. 6d and Supplementary Fig. 5e). These data suggest that ectopic expression of TOX is sufficient to endow terminally exhausted CD8$^+$ T cells with the ability to persist during chronic viral infection.

We profiled the transcriptomes of Tim3^{lo}Ly108^{hi}(TCF1^{hi} progenitor-like) and Tim3^{hi}Ly108^{lo}(TCF1^{lo} terminally exhausted) P14 cells (Supplementary Fig. 5f) from control and TOX-overexpression samples. TOX overexpression led to upregulation of 266 genes and downregulation of 358 genes in progenitor-like P14 cells, and upregulation of 48 genes and downregulation of 57 genes in terminally exhausted P14 cells (fold change>1.5, P<0.05, FDR<0.1) (Supplementary Tables 7,8). Among the differentially expressed genes were transcription factors (e.g. *Ikzf2*), inhibitory receptors (e.g. *Pdcd1*), and effector molecules (e.g. *Tnf*) (Fig. 6e). We confirmed that TOX overexpression moderately increased PD1 and TIGIT protein on P14 cells on day 7 post-infection (Supplementary Fig. 5g,h). However, a slight decrease in LILRB4 level was observed in TOX-overexpressing P14 cells (Supplementary Fig. 5i). In addition, we found reduced production of interferon-γ (IFN-γ), tumor necrosis factor (TNF), and interleukin 2 (IL-2) by TOX-overexpressing P14 cells upon re-stimulation on day 7 post-infection (Supplementary Fig. 5j–m). However, *in vitro* cytotoxicity of CD8$^+$ T cells was unaffected by TOX overexpression (Supplementary Fig. 5n).

Using Gene Set Enrichment Analysis (GSEA) to determine whether TOX governs the activity of *Tox* gene module, we found that TOX overexpression upregulated the gene signature associated with this module in both progenitor-like and terminally exhausted P14

CD8$^+$ T cells (Fig. 6f,g). In addition, in both progenitor-like and terminally exhausted CD8$^+$ T cells, TOX overexpression led to upregulation of genes in the hypoxia pathway and downregulation of signatures such as oxidative phosphorylation, mTOR signaling, IFN-α response, and DNA repair (Fig. 6h and Supplementary Fig. 5o–s). To determine the effects of TOX overexpression on activity of the phosphatidylinositol-3-OH kinase (PI(3)K)–AKT–mTOR pathway, we quantified phospho-AKT(S473) and phospho-ribosome protein S6 (S235/S236) in control and TOX-overexpressing P14 cells on day 7 post-infection. Consistent with GSEA results, TOX-overexpressing P14 cells exhibited lower PI(3)K–AKT–mTOR signaling than control cells (Fig. 6i,j). Taken together, TOX overexpression promotes long-term persistence of virus-specific CD8$^+$ T cells during chronic infection but reduces PI(3)K–AKT–mTOR signaling and cytokine production.

Progenitor-like cells and persistent CD8$^+$ immunity require TOX

To determine the cell-intrinsic role of TOX in the antiviral CD8 T cell response during chronic infection, we generated mixed-bone marrow chimeras. Reconstitution of lethally irradiated wild-type CD45.2 mice with mixed-bone marrow either from wild-type CD45.1 and wild-type CD45.2 donors or from *Tox*$^{-/-}$CD45.1 and wild-type CD45.2 donors enabled comparison of *Tox*$^{-/-}$ and wild-type CD8$^+$ T cells exposed to the same environment. Despite comparable frequencies of LCMV-specific cells between the wild-type and *Tox*$^{-/-}$ CD8$^+$ compartments on day 7 after LCMV clone 13 infection (Supplementary Fig. 6a), frequencies of TCF1^{hi}Tim3^{lo} cells were significantly lower in *Tox*$^{-/-}$ LCMV-specific CD8$^+$ T cells than in their wild-type counterparts in the same mice (Fig. 7a,b). This finding suggests that progenitor-like CD8$^+$ T cells require cell-intrinsic TOX activity. To confirm these findings, we adoptively transferred CD45.1$^-$ *Tox*$^{f/f}$ CD4-Cre (*Tox* cKO) and *Tox*$^{f/f}$ (wild-type) P14 cells into wild-type CD45.1$^+$ recipients and analyzed their number and phenotype on day 7 post-infection. Consistently, *Tox* deficiency led to a substantial reduction in the frequency of TCF1^{hi}Tim3^{lo} virus-specific CD8$^+$ T cells without impacting the magnitude of initial expansion (Supplementary Fig. 6b,c).

Next, we analyzed pooled scRNA-seq data from 4,409 wild-type and 5,296 *Tox*$^{-/-}$ LCMV-specific CD8$^+$ T cells from mixed-bone marrow chimeras on day 7 after LCMV clone 13 infection. Significant portions of wild-type and *Tox*$^{-/-}$ cells occupied distinct spaces in the t-SNE plots, suggesting that *Tox* deficiency led to a substantial change in the differentiation program of virus-specific CD8$^+$ T cells (Fig. 7c). Unsupervised clustering assigned cells into four subsets. Fewer cluster 2 cells, which expressed high levels of *Tcf7, Id3*, and *Slamf6* and represent the progenitor-like population, were present amongst *Tox*$^{-/-}$ cells (Fig. 7d,e). In addition, more cluster 0 cells, which expressed signature genes typical of terminally exhausted cells, were present in *Tox*$^{-/-}$ cells. However, *Tox*$^{-/-}$ cells in cluster 0 did not fully comingle with wild-type cells in the same cluster, suggesting that *Tox* deficiency altered the differentiation of terminally exhausted cells (Fig. 7c,d). Indeed, *Tox*$^{-/-}$ cells upregulated *Klrg1*, typically associated with short-lived effectors during acute infection (Fig. 7f). In addition, *Tox*$^{-/-}$ cells upregulated *Gzma* and *Gzmb*, and downregulated inhibitor receptor *Pdcd1*, indicating a potential increase in the effector function. However, *Tox*$^{-/-}$ cells also downregulated *Hif1a* and *Batf*, both encoding transcription factors required for sustaining

antiviral CD8$^+$ response during chronic infection[34,35]. Moreover, the increased *Mt1* expression in *Tox*$^{-/-}$ cells also suggests that these cells are more dysfunctional[21].

To determine whether cell-intrinsic *Tox* deficiency affects the maintenance of virus-specific CD8$^+$ T cells, we tracked CD8$^+$ responses in these chimeras four weeks after infection. Notably, there were significantly lower frequencies of LCMV-specific cells within *Tox*$^{-/-}$ CD8$^+$ T cells than within wild-type CD8$^+$ T cells in the same mice (Fig. 7g,h). Therefore, cell-intrinsic TOX activity is required both for the programming of progenitor-like CD8$^+$ T cells and long-term CD8$^+$ immunity during chronic viral infection.

Discussion

Progenitor-like CD8$^+$ T cells found in chronic viral infection and cancer have attracted great attention as a potential target for adoptive cell therapies and checkpoint blockade therapies[6–11,36–40]. However, it is unclear whether progenitor-like CD8$^+$ T cells require distinct transcriptional and epigenetic programs to persist in the immunosuppressive environment induced by chronic viral infection and cancer. Here, we revealed distinct transcriptional and epigenetic programs of progenitor-like CD8$^+$ T cells in chronic infection using scRNA-seq and H3K27ac ChIP-seq. Compared to memory precursors in acute infection, progenitor-like cells exhibited a greater transcriptional activity and higher H3K27ac in a gene module containing *Tox*, identified through our scRNA-seq analyses. We further demonstrated the critical role of TOX in the persistence of antiviral CD8$^+$ T cells and differentiation of progenitor-like CD8$^+$ T cells during chronic viral infection.

Progenitor-like CD8$^+$ T cells are exposed to continuous antigen stimulation and immunosuppression during chronic infection[41]. Therefore, compared to memory precursors, progenitor-like CD8$^+$ T cells likely exhibit differential activities in pathways required for the adaptation to chronic viral infection. Previous studies mostly relied on bulk transcriptomic analysis of progenitor-like CD8$^+$ T cells, which reflect the populational average and are affected by the bias introduced by the markers selected to purify this population. Using single-cell transcriptomes of virus-specific CD8$^+$ T cells from mice after acute and chronic viral infections, we found that *Tcf7*$^+$ CD8$^+$ T cells from day 7 acute and chronic LCMV infections partitioned in separate areas in the t-SNE plot. This finding suggests that memory precursors and progenitor-like cells have distinct transcriptional programs despite sharing common markers. For example, we found pronounced upregulation of gene signatures, such as translation and hypoxia, in progenitor-like CD8$^+$ T cells relative to memory precursors. Progenitor-like CD8$^+$ T cells and memory precursors also differed in their H3K27ac profiles, suggesting that their differentiation might be regulated by activation of different sets of enhancers. Nonetheless, while intriguing, the epigenomic data in this study suffer from similar limitations of population-level profiles as discussed above. Thus, future studies that employ single-cell epigenomics might help further elucidate the differential regulation of progenitor-like CD8$^+$ T cells and memory precursor cells.

Long-term persistence is a cardinal hallmark of T cell stemness and is critical for the CD8$^+$ T cell response against chronic infection and cancer[42,43]. Moreover, the longevity of therapeutic T cells directly affects the efficacy of adoptive cell therapies[44]. Thus, it is

important to determine molecular mechanisms underlying the longevity of progenitor-like CD8$^+$ T cells. We found that *Tox* was among the most differentially expressed genes between progenitor-like CD8$^+$ T cells and memory precursor cells and is in a gene module most highly expressed by progenitor-like CD8$^+$ T cells. Consistent with our results, a previous study that used microarrays to compare the transcriptomes between CD8$^+$ T cells responding to acute and chronic LCMV infections identified *Tox* as a hub gene in the difference network between acute and chronic infections[13]. In this study, we have shown that *Tox* deficiency led to defects in long-term persistence of antiviral CD8$^+$ T cells and a loss of progenitor-like CD8$^+$ T cells, and that TOX overexpression enhanced the persistence of both progenitor-like and terminally exhausted virus-specific CD8$^+$ T cells. In addition, our comparison of single-cell transcriptomes between wild-type and *Tox* deficient antiviral CD8$^+$ T cells underlines a critical role of TOX in dictating the transcriptional program of virus-specific CD8 T cells responding to chronic infection. These findings suggest that TOX plays an essential role in promoting persistence of T cells in chronic infection and might promote the persistence of therapeutic T cells in immunotherapies. Interestingly, a recent study has shown that TOX is required for the pathogenicity of autoreactive CD8$^+$ T cells in the brain[24]. Similar to our observation in chronic viral infection, *Tox*-deficient autoreactive CD8$^+$ T cells downregulated *Tcf7* and upregulated short-lived effector markers such as *Gzmb* and *Klrg1*. These data suggest a potential overlap of the molecular circuit involving TOX between autoreactive and antiviral CD8$^+$ T cells. Of note, although less prevalent, there were terminally exhausted CD8$^+$ T cells that expressed *Tox* and/or were enriched with *Tox* module. Moreover, *Tox* deficiency also altered the differentiation program of terminally exhausted CD8$^+$ T cells, suggesting a role of TOX in these cells. Whether progenitor-like cells directly give rise to TOX-expressing terminally exhausted cells remains to be investigated.

In summary, through scRNA-seq and H3K27ac profiling, we found that progenitor-like CD8$^+$ T cells responding to chronic viral infection are regulated by distinct transcriptional and epigenetic programs despite their similarities with memory precursor cells. Importantly, this includes up-regulation of *Tox*, which promotes persistent antiviral CD8$^+$ T cell response and is necessary for progenitor-like CD8$^+$ T-cell differentiation. Our study adds to our understanding of T-cell stemness and may shed new light on the development of more effective immunotherapies.

Methods

Mice, infection, adoptive transfer, mixed-bone marrow chimeras, and FK506 treatment

C57BL/6 and wild-type CD45.1 (B6.SJL-Ptprca Pepcb/BoyJ) were purchased from the Jackson Laboratory. P14 mice carry a transgenic TCR that recognizes H-2D^b GP33–41 epitope on LCMV[16]. Blimp-1-YFP reporter mice[45], purchased from the Jackson Laboratory, were bred to P14 to generate P14 Blimp-1-YFP mice. *Tox*-deficient mice were generated by deleting the first exon and promoter region using CRISPR/Cas9. *Tox*$^{f/f}$ mice have been described previously[22] and were crossed to CD4-Cre and P14. All mice were kept on a C57BL/6 background. Mice used in the experiments were age- and sex- matched. All animal husbandry and experiments were approved by the National Human Genome Research

Institute (NHGRI) or the National Institute of Neurological Disorders and Stroke (NINDS) Animal Care and Use Committees (protocols G98–3 and 1295–12).

For acute viral infections, mice were intravenously (i.v.) injected with 2×10^5 plaque-forming units (PFU) LCMV Armstrong. For chronic viral infections, mice were intravenously (i.v.) injected with 2×10^6 PFU LCMV clone 13. For adoptive transfer experiments, 5,000 naïve splenic P14 cells were adoptively transferred into mice before infection. To generate mixed-bone marrow chimeras, lethally irradiated (950 rads) wild-type mice were reconstituted with 1:1 mixed-bone marrows from a CD45.1 and a CD45.2 donors that are either wild-type or *Tox*-deficient. After at least two months from the date of reconstitution, chimeras were used for experiments. For FK506 treatment, 10 mg/kg FK506 were subcutaneously injected to mice daily from day 4 through day 7 post-infection, as previously described[46].

Retroviral transduction

TOX (UniProt ID: Q66JW3–1) overexpression construct was made with an MSCV-IRES-GFP backbone (pMIG) as described previously[29]. Activated P14 $CD8^+$ T cells were spinoculated with retroviruses carrying a control (pMIG) or a TOX overexpression construct for 90 min, cultured overnight with 2ng/mL IL-7, and sorted by flow cytometry for GFP^+-transduced P14 cells. 5,000 GFP^+ P14 cells were injected into each C57BL/6 recipient, which was immediately infected with LCMV clone 13.

Antibodies, dyes, flow cytometry and cell sorting

Anti-CD8a (53–6.7), anti-CD45.1 (A20), anti-B220 (RA3–6B2), anti-Tim3 (RMT3–23), anti-TOX (TXRX10), anti-TIGIT (GIGD7), anti-IFN-γ (XMG1.2), and LIVE/DEAD™ Fixable Aqua were from Thermo Fisher Scientific; anti-CD45.2 (104), anti-KLRG1 (2F1), anti-PD1 (RMP1–30), anti-LILRB4 (H1.1), and anti-CD44 (IM7) were from BioLegend; anti-p-AKT(S473) (M89–61), anti-TNF (MP6-XT22), anti-IL-2 (JES6–5H4), anti-Ly6G (1A8), and anti-Ly108 (13G3) were from BD Biosciences; anti-p-S6(S235/236) (D57.2.2E) was from Cell Signaling Technology. H-$2D^b$ GP33–41 tetramer was obtained from NIAID Tetramer Core Facility. TCF1 staining has been described in our previous study[29]. BD™ LSR II and BD FACSAria™ II were used for flow cytometry analysis and FACS sorting, respectively. Data analysis was performed with FlowJo 9.9.

Single-cell RNA-sequencing (scRNA-seq)

The scRNA-seq libraries were generated using Chromium Single Cell 3′ Library & Gel Bead Kit v2 (10X Genomics) according to manufacturer's protocol. Briefly, $1–1.5 \times 10^4$ live cells were FACS-sorted and used to generate single-cell gel-bead in emulsion (GEM). After reverse transcription, GEMs were disrupted. Barcoded cDNA was isolated and amplified by PCR (12 cycles). Following fragmentation, end repair, and A-tailing, sample indexes were added during index PCR (8 cycles). The purified libraries were sequenced on a HiSeq 3000 (Illumina) with 26 cycles of Read 1, 8 cycles of i7 Index, and 98 cycles of Read2.

Single-cell RNA-sequencing analysis

Alignment, filtering, barcode counting, and unique molecular identifier (UMI) counting were performed using Cell Ranger v2.1.0. Data were further analyzed using Seurat 2[47]. Briefly, cells with percentage of mitochondrial genes below 0.05% were included. Cells with highest numbers of detected genes (top 0.2%) or lowest numbers of detected genes (bottom 0.2%) were considered as outliers and excluded from downstream analyses. Raw UMI counts were normalized to UMI count per million total counts and log-transformed. Variable genes were selected based on average expression and dispersion. PCA analysis was performed using variable genes. Clusters and t-SNE plots were generated based on selected PCA dimensions. Marker genes were identified by Seurat function FindAllMarkers. Scaled expression data of these marker genes were used for making Heatmaps. Normalized data were shown in the form of feature plots or violin plots. Gene set enrichment was performed on each cell by calculating *P*-values of Fisher's exact test to determine the enrichment of a gene set in each cell. Alternatively, cells with active gene sets were identified using AUCell, as previously described[19]. Trajectory analysis was performed using Monocle 2[20]. Gene ontology analysis was performed using maker genes in Metascape[48] (http://metascape.org), and plotted with R ggplot2. Topological overlap matrix, hierarchical clustering, and gene modules were generated by Weighted Correlation Network Analysis (WGCNA)[49].

Chromatin immunoprecipitation sequencing (ChIP-seq)

1×10^6 FACS-sorted stem-like (Blimp-1-YFP^{lo}Tim3^{lo}), terminally exhausted (Blimp-1-YFP^{hi}Tim3^{hi}), memory precursor (KLRG1^{lo}), and short-lived effector (KLRG1^{hi}) P14 cells were used for ChIP analyses. More than ten mice from each infection were pooled for each sort. Two biological replicates of each CD8$^+$ T cell population were collected. DNA-protein crosslinking, nuclei isolation, and chromatin sonication were performed using truChIP Chromatin Shearing Kit (Covaris) according to manufacturer's instructions. After immunoprecipitation by anti-H3K27Ac (ab4729, AbCam), ChIPmentation was performed according to the published protocol[50]. The libraries were sequenced for 50 cycles (single read) on a HiSeq 3000 (Illumina).

ChIP-seq analysis

ChIP-seq reads were mapped to the mouse genome (build mm10) with Bowtie 1.1.1[51]. Peaks were identified using MACS (v 1.4.2; default *P*-value threshold of 1E-5)[52]. Only peaks identified in both biological replicates were selected for downstream analysis. Peak annotation was performed with the Hypergeometric Optimization of Motif EnRichment program (HOMER) version 4.9[53]. Peaks were visualized in Integrative Genomics Viewer (IGV). To determine differentially modified H3K27ac peaks between different CD8$^+$ subsets, reads per peak were calculated by HOMER. Statistically differentially modified peaks were then determined using edgeR 3.20.9, defined as those with FDR (Benjamini-Hochberg) less than 0.1 and fold-change greater than 1.5.

RNA Sequencing (RNA-seq)

For each biological replicate, cells from at least three mice were pooled. 5×10^4 cells were sorted and resuspended in Qiazol (Qiagen). Total RNA was extracted with miRNeasy Mini

kit (Qiagen). mRNA was selected using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB). RNAseq libraries were generated using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB) and NEBNext Multiplex Oligos for (NEB) according to manufacturer's instruction. The libraries were sequenced on HiSeq 3000 (Illumina) with 50 cycles of single reads.

RNA-seq analysis

RNA-seq reads were mapped to the mouse genome (mm10) using TopHat 2.1.0[54]. Normalization to reads per kb exon per million mapped reads (RPKM), and ANOVA statistical analyses were performed with Partek Genomics Suite 6.6. Differentially expressed genes were selected with following criteria: *P*-value < 0.05, FDR< 0.1, fold change > 1.5, and expression > 4 RPKM in at least one condition. To compare bulk RNA-seq data with scRNA-seq data, bulk RNA-seq TPM were calculated by pseudo-aligning RNA-seq reads to ENSEMBL transcript sequences using kallisto (0.44.0). ClusterProfiler 3.8.1 was used for GSEA analysis[55].

FACS-based *in vitro* killing assay

EL4 cells were labeled with CellTraceViolet (Thermo Fisher Scientific), pulsed with GP33 peptide at 1μg/mL for 1 h at 37 °C, and used as target cells. *In vitro* activated control or TOX-overexpressing $CD8^+$ T cells were collected and incubated with peptide-pulsed EL4 cells at different ratios for 4 h at 37 °C. The percentages of dead EL4 cells were measured by LIVE/DEAD™ Fixable Aqua (Thermo Fisher Scientific).

Statistical analysis

Statistical analyses were conducted with R3.4.0 and GraphPad Prism 6. Pearson's R was calculated by R function Cor (). Unless otherwise indicated, two-tailed paired or unpaired Student's t-test was used to determine statistical significance. $*P< 0.05$, $**P< 0.01$, $***P< 0.001$, $****P< 0.0001$. The mean and standard deviation (SD) are presented in the figures. Error bars represent SD.

Data Availability Statement

All data generated during this study are available within the paper. The RNA-seq, ChIP-seq, and scRNA-seq data have been deposited in the Gene Expression Omnibus (GEO accession number: GSE119943). TOX, NFAT1, IRF4, RUNX3, FOXO1, STAT3, and STAT5B ChIP-seq data are publicly available (GEO accession numbers: GSE93953, GSE64407, GSE49930, GSE46943, GSE50128, and GSE102317).

Life Sciences Reporting Summary

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

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank L. Garrett (NHGRI), E. Escobar (NHGRI), C. Rivas (NHGRI), I. Ginty (NHGRI), W. Pridgen (NHGRI), G. Gutierrez-Cruz (NIAMS), and S. Dell'Orso (NIAMS) for their excellent technical support and J. Kaye for the *Tox*$^{f/f}$ mice. This work was supported in part by the intramural programs of the NHGRI, NIAID, NIAMS, NINDS, and NCI, NIH, DDIR Innovation Awards (to P.L.S., J.J.O., and L.G.), and NIH grant AG056524 (to T.W.). E.J.W. is a member of the Parker Institute for Cancer Immunotherapy which supported the University of Pennsylvania cancer immunotherapy program.

References

1. Ahmed R & Gray D Immunological memory and protective immunity: understanding their relation. Science 272, 54–60 (1996). [PubMed: 8600537]
2. Hashimoto M et al. CD8 T Cell Exhaustion in Chronic Infection and Cancer: Opportunities for Interventions. Annu Rev Med 69, 301–318 (2018). [PubMed: 29414259]
3. Thommen DS & Schumacher TN T Cell Dysfunction in Cancer. Cancer cell 33, 547–562 (2018). [PubMed: 29634943]
4. Kaech SM & Wherry EJ Heterogeneity and cell-fate decisions in effector and memory CD8+ T cell differentiation during viral infection. Immunity 27, 393–405 (2007). [PubMed: 17892848]
5. Blackburn SD, Shin H, Freeman GJ & Wherry EJ Selective expansion of a subset of exhausted CD8 T cells by alphaPD-L1 blockade. Proc Natl Acad Sci U S A 105, 15016–15021 (2008). [PubMed: 18809920]
6. 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). [PubMed: 28018990]
7. Im SJ et al. Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. Nature 537, 417–421 (2016). [PubMed: 27501248]
8. He R et al. Follicular CXCR5- expressing CD8(+) T cells curtail chronic viral infection. Nature 537, 412–428 (2016). [PubMed: 27501245]
9. Utzschneider DT 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). [PubMed: 27533016]
10. Leong YA et al. CXCR5(+) follicular cytotoxic T cells control viral infection in B cell follicles. Nat Immunol 17, 1187–1196 (2016). [PubMed: 27487330]
11. Siddiqui I et al. Intratumoral Tcf1(+)PD-1(+)CD8(+) T Cells with Stem-like Properties Promote Tumor Control in Response to Vaccination and Checkpoint Blockade Immunotherapy. Immunity 50, 195–211 e110 (2019). [PubMed: 30635237]
12. Man K et al. Transcription Factor IRF4 Promotes CD8(+) T Cell Exhaustion and Limits the Development of Memory-like T Cells during Chronic Infection. Immunity 47, 1129–1141 e1125 (2017). [PubMed: 29246443]
13. Doering TA et al. Network analysis reveals centrally connected genes and pathways involved in CD8+ T cell exhaustion versus memory. Immunity 37, 1130–1144 (2012). [PubMed: 23159438]
14. Papalexi E & Satija R Single-cell RNA sequencing to explore immune cell heterogeneity. Nat Rev Immunol 18, 35–45 (2018). [PubMed: 28787399]
15. O'Flaherty E & Kaye J TOX defines a conserved subfamily of HMG-box proteins. BMC Genomics 4, 13 (2003). [PubMed: 12697058]
16. Pircher H, Burki K, Lang R, Hengartner H & Zinkernagel RM Tolerance induction in double specific T-cell receptor transgenic mice varies with antigen. Nature 342, 559–561 (1989). [PubMed: 2573841]
17. Ahmed R, Salmi A, Butler LD, Chiller JM & Oldstone MB Selection of genetic variants of lymphocytic choriomeningitis virus in spleens of persistently infected mice. Role in suppression of cytotoxic T lymphocyte response and viral persistence. J Exp Med 160, 521–540 (1984). [PubMed: 6332167]
18. Butler A, Hoffman P, Smibert P, Papalexi E & Satija R Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 36, 411–420 (2018). [PubMed: 29608179]

19. Aibar S et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods 14, 1083–1086 (2017). [PubMed: 28991892]
20. Trapnell C et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 32, 381–386 (2014). [PubMed: 24658644]
21. Singer M et al. A Distinct Gene Module for Dysfunction Uncoupled from Activation in Tumor-Infiltrating T Cells. Cell 171, 1221–1223 (2017). [PubMed: 29149608]
22. Aliahmad P & Kaye J Development of all CD4 T lineages requires nuclear factor TOX. J Exp Med 205, 245–256 (2008). [PubMed: 18195075]
23. Seehus CR et al. The development of innate lymphoid cells requires TOX-dependent generation of a common innate lymphoid cell progenitor. Nat Immunol 16, 599–608 (2015). [PubMed: 25915732]
24. Page N et al. Expression of the DNA-Binding Factor TOX Promotes the Encephalitogenic Potential of Microbe-Induced Autoreactive CD8(+) T Cells. Immunity 48, 937–950 e938 (2018). [PubMed: 29768177]
25. Youngblood B et al. Effector CD8 T cells dedifferentiate into long-lived memory cells. Nature 552, 404–409 (2017). [PubMed: 29236683]
26. Gray SM, Amezquita RA, Guan T, Kleinstein SH & Kaech SM Polycomb Repressive Complex 2-Mediated Chromatin Repression Guides Effector CD8(+) T Cell Terminal Differentiation and Loss of Multipotency. Immunity 46, 596–608 (2017). [PubMed: 28410989]
27. He B et al. CD8(+) T Cells Utilize Highly Dynamic Enhancer Repertoires and Regulatory Circuitry in Response to Infections. Immunity 45, 1341–1354 (2016). [PubMed: 27986453]
28. Creyghton MP et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A 107, 21931–21936 (2010). [PubMed: 21106759]
29. Wu T et al. TCF1 Is Required for the T Follicular Helper Cell Response to Viral Infection. Cell reports 12, 2099–2110 (2015). [PubMed: 26365183]
30. Martinez GJ et al. The transcription factor NFAT promotes exhaustion of activated CD8(+) T cells. Immunity 42, 265–278 (2015). [PubMed: 25680272]
31. 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). [PubMed: 28604718]
32. Staron MM et al. The transcription factor FoxO1 sustains expression of the inhibitory receptor PD-1 and survival of antiviral CD8(+) T cells during chronic infection. Immunity 41, 802–814 (2014). [PubMed: 25464856]
33. Cox MA, Kahan SM & Zajac AJ Anti-viral CD8 T cells and the cytokines that they love. Virology 435, 157–169 (2013). [PubMed: 23217625]
34. Xin G et al. A Critical Role of IL-21-Induced BATF in Sustaining CD8-T-Cell-Mediated Chronic Viral Control. Cell reports 13, 1118–1124 (2015). [PubMed: 26527008]
35. Doedens AL et al. Hypoxia-inducible factors enhance the effector responses of CD8(+) T cells to persistent antigen. Nat Immunol 14, 1173–1182 (2013). [PubMed: 24076634]
36. Gattinoni L et al. Wnt signaling arrests effector T cell differentiation and generates CD8+ memory stem cells. Nat Med 15, 808–813 (2009). [PubMed: 19525962]
37. Miller BC et al. Subsets of exhausted CD8(+) T cells differentially mediate tumor control and respond to checkpoint blockade. Nat Immunol 20, 326–336 (2019). [PubMed: 30778252]
38. Sade-Feldman M et al. Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma. Cell 175, 998–1013 e1020 (2018). [PubMed: 30388456]
39. Kurtulus S et al. Checkpoint Blockade Immunotherapy Induces Dynamic Changes in PD-1(−)CD8(+) Tumor-Infiltrating T Cells. Immunity 50, 181–194 e186 (2019). [PubMed: 30635236]
40. Snell LM et al. CD8(+) T Cell Priming in Established Chronic Viral Infection Preferentially Directs Differentiation of Memory-like Cells for Sustained Immunity. Immunity 49, 678–694 e675 (2018). [PubMed: 30314757]
41. Wherry EJ & Kurachi M Molecular and cellular insights into T cell exhaustion. Nat Rev Immunol 15, 486–499 (2015). [PubMed: 26205583]

42. Gattinoni L, Speiser DE, Lichterfeld M & Bonini C T memory stem cells in health and disease. Nat Med 23, 18–27 (2017). [PubMed: 28060797]
43. Gautam S et al. The transcription factor c-Myb regulates CD8(+) T cell stemness and antitumor immunity. Nat Immunol 20, 337–349 (2019). [PubMed: 30778251]
44. Rosenberg SA & Restifo NP Adoptive cell transfer as personalized immunotherapy for human cancer. Science 348, 62–68 (2015). [PubMed: 25838374]
45. Fooksman DR, Nussenzweig MC & Dustin ML Myeloid cells limit production of antibody-secreting cells after immunization in the lymph node. J Immunol 192, 1004–1012 (2014). [PubMed: 24376270]
46. Araki K et al. Pathogenic virus-specific T cells cause disease during treatment with the calcineurin inhibitor FK506: implications for transplantation. J Exp Med 207, 2355–2367 (2010). [PubMed: 20921283]
47. Macosko EZ et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 161, 1202–1214 (2015). [PubMed: 26000488]
48. Zhou Y et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 10, 1523 (2019). [PubMed: 30944313]
49. Langfelder P & Horvath S WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559 (2008). [PubMed: 19114008]
50. Schmidl C, Rendeiro AF, Sheffield NC & Bock C ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat Methods 12, 963–965 (2015). [PubMed: 26280331]
51. Langmead B, Trapnell C, Pop M & Salzberg SL Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10, R25 (2009). [PubMed: 19261174]
52. Zhang Y et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol 9, R137 (2008). [PubMed: 18798982]
53. Heinz S et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38, 576–589 (2010). [PubMed: 20513432]
54. Trapnell C & Salzberg SL How to map billions of short reads onto genomes. Nat Biotechnol 27, 455–457 (2009). [PubMed: 19430453]
55. Yu G, Wang LG, Han Y & He QY clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16, 284–287 (2012). [PubMed: 22455463]

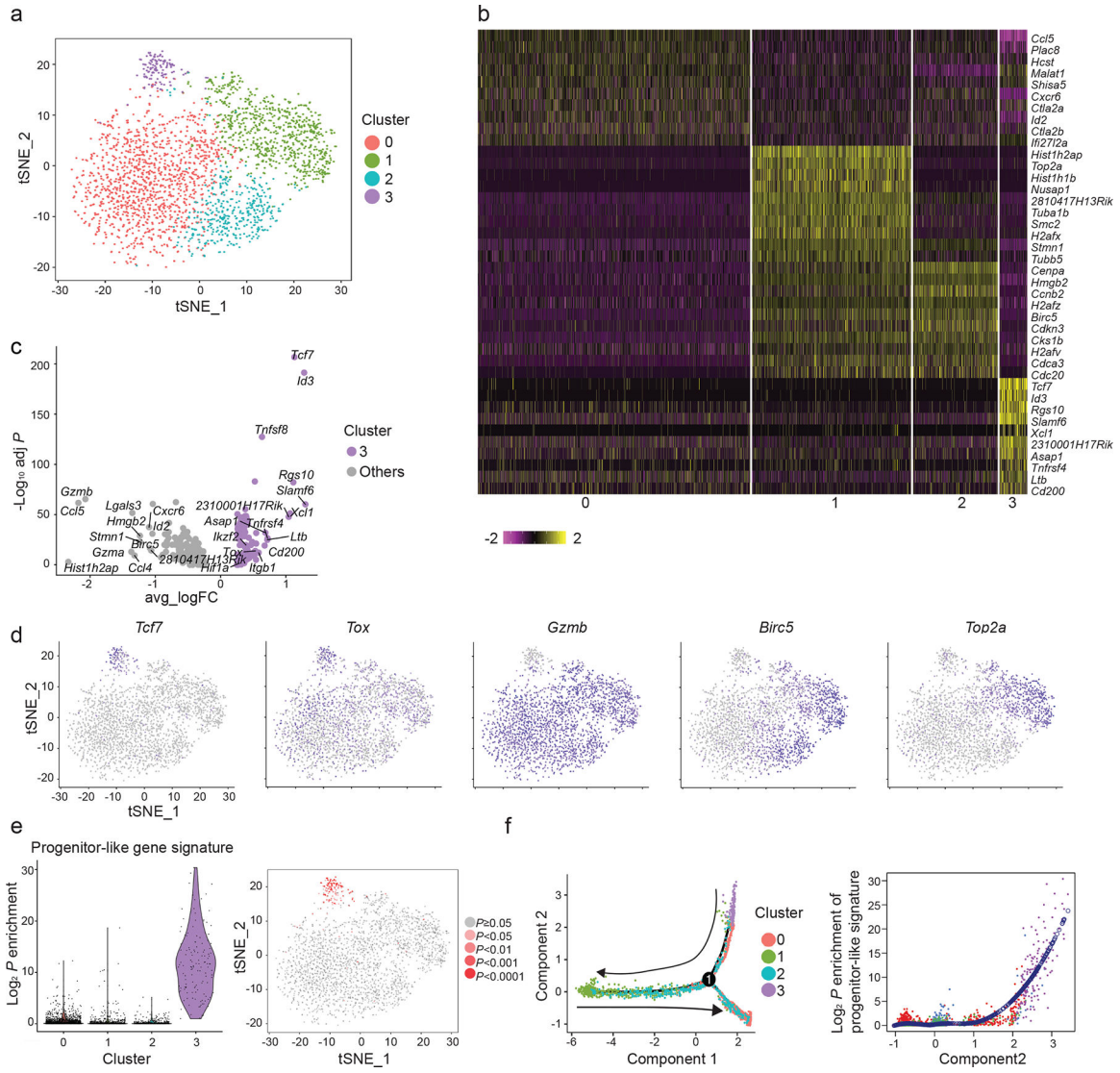

Fig. 1. Heterogeneity of virus-specific CD8$^+$ T cells from chronic LCMV infection delineated by scRNA-seq.

Naïve P14 CD8$^+$ T cell were transferred to C57BL/6 mice that were subsequently infected with LCMV clone 13. P14 cells were isolated on day 7 post-infection. N= 2,597 cells were used for scRNA-seq analyses in (**a-f**). (**a**) The t-SNE projection of P14 cells, determined by Seurat 2. Each dot corresponds to one individual cell. A total of four clusters (cluster 0 through 3) were identified and color-coded. (**b**) A heatmap of top 10 genes expressed in each cluster defined in **Fig. 1a**. Columns correspond to cells; rows correspond to genes. Cells are grouped by clusters. Color scale is based on z-score distribution from −2 (purple) to 2 (yellow). (**c**) Volcano plot showing the differentially expressed genes between cells within cluster 3 and cells outside cluster 3 (purple: upregulated in cluster 3; grey: downregulated in cluster 3). X-axis represents log fold changes; Y-axis presents log10 adjusted *P*-values. Two-sided Wilcoxon rank sum test was used. (**d**) Single-cell transcript levels of *Tcf7*, *Tox*, *Gzmb*, *Birc5*, and *Top2a* illustrated in t-SNE plots. Transcript levels are color-coded: grey, not expressed; purple, expressed. (**e**) Left panel: Enrichment (log2 *P*-values) of progenitor-like

gene signature in each cell, determined by one-sided Fisher's exact test, illustrated in violin plots. Cells are separated into the four clusters defined in **Fig. 1a**. The violin represents the probability density at each value; each dot represents one cell. Right panel: Enrichment (log2 *P*-values) of progenitor-like gene signature in each cell illustrated in t-SNE plots. *P*-values are color-coded. (**f**) Left panel: The trajectory of P14 cells state transition in a two-dimensional state-space determined by Monocle 2. Each dot represents a single cell. Colors represent the clusters to which the cells belong, as defined in **Fig. 1a**. Right panel: X-axis represents values of component 2 determined by Monocle 2; Y-axis represents log2 *P*-values derived from the enrichment of progenitor-like gene signature, as in **Fig. 1e**. Each dot represents a single cell. Dark blue circles represent the spline.

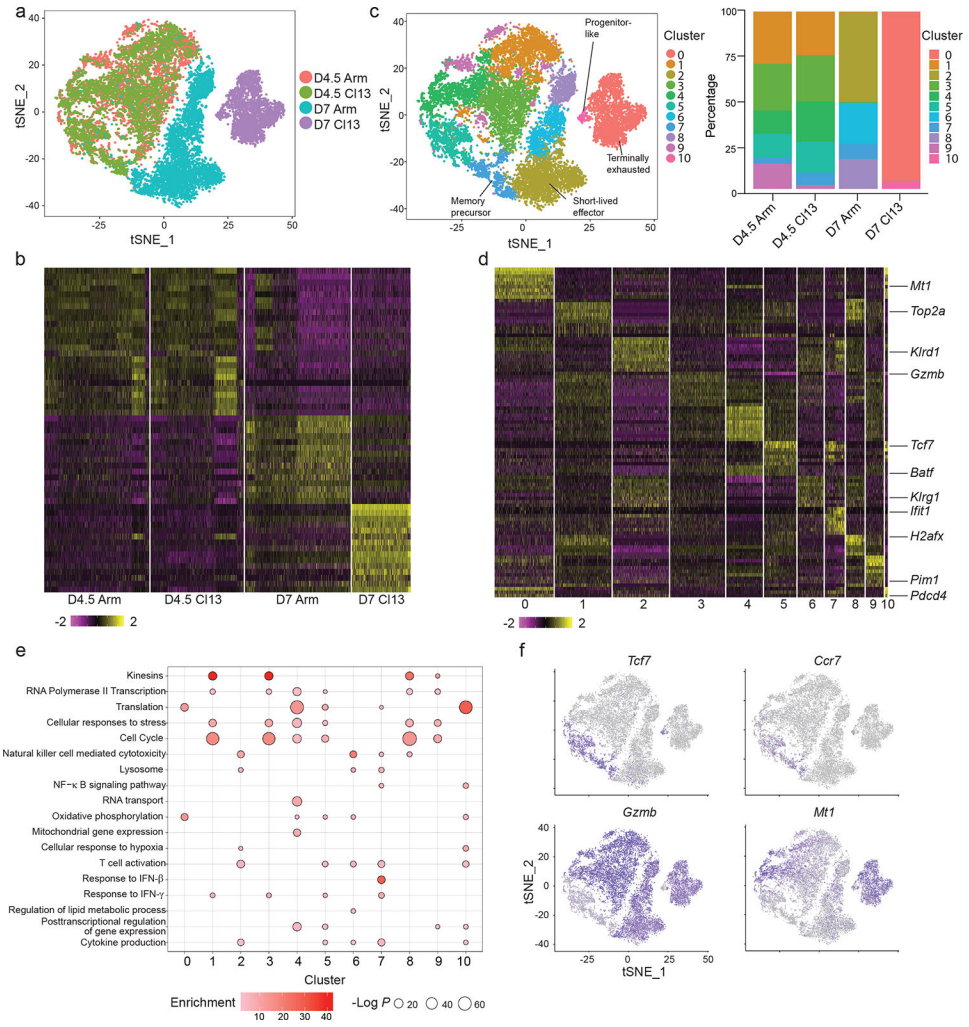

Fig. 2. Single-cell transcriptomes of virus-specific CD8$^+$ T cells responding to acute and chronic viral infections diverged between day 4.5 and day 7 post-infection.

Naïve P14 cells were transferred into C57BL/6 mice that were subsequently infected with either LCMV Armstrong or LCMV clone 13. P14 cells were collected on day 4.5 and day 7 post-infection. N=16,042 cells were used for scRNA-seq analyses in (**a-f**). (**a**) t-SNE plots of 16,042 P14 cells (D4.5 Arm: 4, 651 cells; D4.5 Cl13: 4,116 cells; D7 Arm: 4, 678 cells; D7 Cl13: 2,597 cells), determined by Seurat 2. Each dot represents a single cell. Cells from different samples are color-coded (D4.5 Arm: red; D4.5 Cl13: green; D7 Arm: turquoise; D7 Cl13: purple). (**b**) A heatmap of top 15 genes expressed in each sample. Columns correspond to cells; rows correspond to genes. Cells are grouped by samples. Color scale is based on z-score distribution from −2 (purple) to 2 (yellow). (**c**) Left panel: t-SNE plots of cells from all four samples, determined by Seurat 2. A total of eleven clusters (cluster 0 through 10) were identified and color-coded. Right panel: Percentages of cells from each cluster in each sample. (**d**) A heatmap of top 10 genes expressed in each cluster. Cells are grouped by clusters. Color scale is based on z-score distribution from −2 (purple) to 2 (yellow). **e**, Dot plots of gene ontology, determined by Metascape. Each column represents one cluster; each row represents a pathway. The enrichment scores are color-coded. The

ranges of log10 *P*-values are represented by the diameter of the circles. (**f**) Single-cell transcript levels of *Tcf7*, *Ccr7*, *Gzmb*, and *Mt1* illustrated in t-SNE plots. Transcript levels are color-coded: grey, not expressed; purple, expressed.

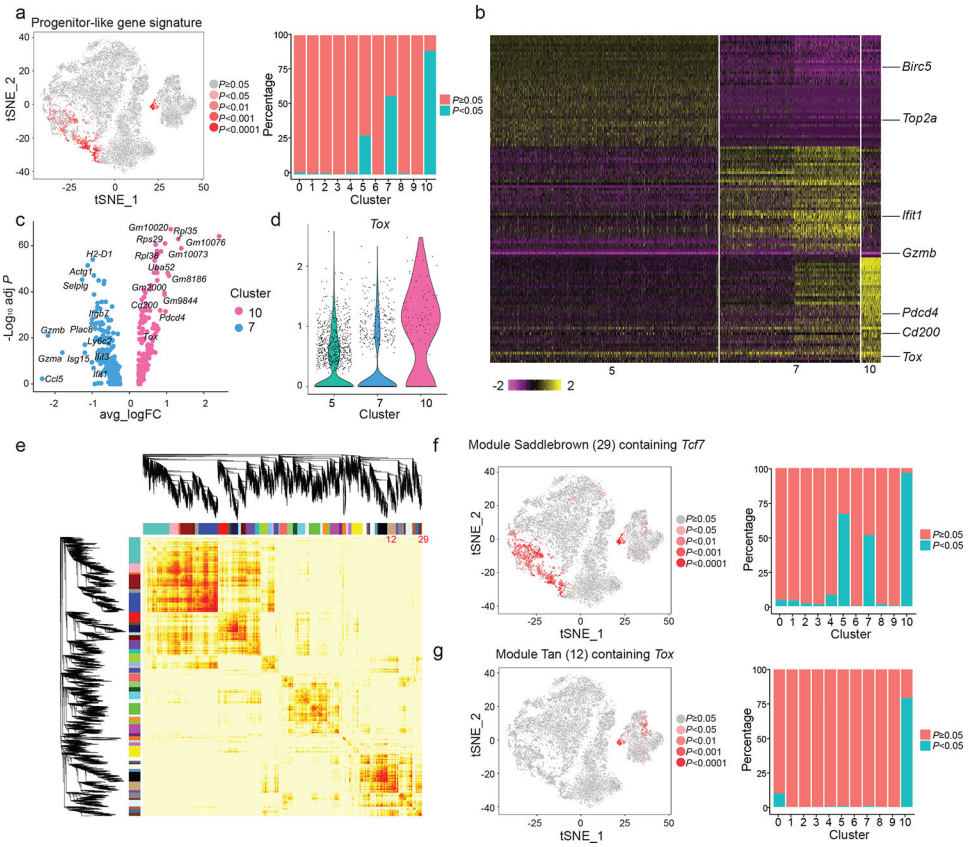

Fig. 3. *Tox* gene module distinguishes progenitor-like CD8$^+$ T cells from memory precursor cells.

Experimental setup has been described in Fig. 2. (**a**) Enrichment (log2 *P*-values) of progenitor-like gene signature in each cell (n=16,042 cells) determined by one-sided Fisher's exact test. Left panel: t-SNE plots with color-coded *P*-values. Right panel: Percentages of cells in each cluster, defined in Fig. 2c, with *P*-values above (red) or below (turquoise) 0.05. (**b**) A heatmap of top 40 genes expressed in each cluster. Cells are from cluster 5 (n=1,357 cells), 7 (n=834 cells), or 10 (n=116 cells), as defined in Fig. 2c. Columns correspond to cells; rows correspond to genes. Cells are grouped by clusters. Color scale is based on z-score distribution from −2 (purple) to 2 (yellow). (**c**) Differentially expressed genes between cells in cluster 10 (n=116 cells) and cells in cluster 7 (n=834 cells), illustrated by volcano plots (pink: upregulated in cluster 10; blue: upregulated in cluster 7). X-axis represents log fold changes; Y-axis presents log10 adjusted *P*-values. Two-sided Wilcoxon rank sum test was used. (**d**) Violin plots of *Tox* expression in cells from clusters 5 (n=1,357 cells), 7 (n=834 cells), or 10 (n=116 cells). The violin represents the probability density at each value; each dot represents one cell. (**e**) A gene co-expression network of the scRNA-seq data described in Fig. 2 (n=16,042 cells) was constructed by Weighted Correlation Network Analysis (WGCNA). The heatmap shows the topological overlap matrix among all genes used in the analysis. Darker color represents higher overlap. Hierarchical clustering and module assignment of genes are shown along the left side and the top. Numbers indicate module saddlebrown (29) and module tan (12). (**f-g**) Left panels: Enrichment (log2 *P*-values) of genes in module 29 (saddlebrown, **f**) or in module 12 (tan, **g**) in each cell (n=16,042 cells) was determined by one-sided Fisher's exact test and illustrated

by t-SNE plots. Right panels: Percentages of cells (n=16,042 cells) in each cluster, defined in Fig. 2c, with *P*-values above (red) or below (turquoise) 0.05.

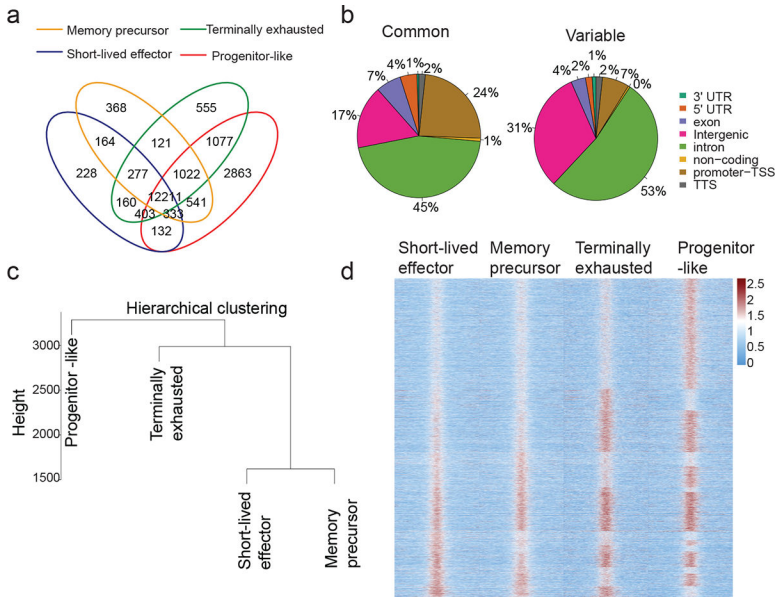

Fig. 4. Progenitor-like CD8$^+$ T cells exhibit a H3K27ac profile distinct from that of memory precursor cells.

H3K27ac ChIP-seq was performed with progenitor-like (Tim3^{lo}Blimp-1^{lo}), terminally exhausted (Tim3^{hi}Blimp-1^{hi}), memory precursor (KLRG1^{lo}), and short-lived effector (KLRG1^{hi}) P14 CD8$^+$ T cells collected from mice seven days after infection with LCMV clone 13 or Armstrong. (**a**) Venn diagram illustrating H3K27ac peaks that are commonly or differentially present in progenitor-like (red), terminally exhausted (green), memory precursor (orange), and short-lived effector (blue) CD8$^+$ T cells. The number of peaks in each category is indicated. (**b**) Pie charts demonstrating the distribution of common (left) or variable (right) H3K27ac peaks across the genome (three prime untranslated region [3′ UTR], five prime untranslated region [5′ UTR], exon, intergenic, intron, non-coding, promoter-transcription start sites [TSS], and transcription termination site [TTS]). (**c**) Hierarchical clustering of progenitor-like, terminally exhausted, memory precursor, and short-lived effector cells based on their H3K27ac profiles. (**d**) Deposition of H3K27ac centered on variable peaks (±4kb) in progenitor-like, terminally exhausted, memory precursor, and short-lived effector CD8$^+$ T cells. Each row represents a peak. Red represents higher signal intensity; blue represents lower signal intensity.

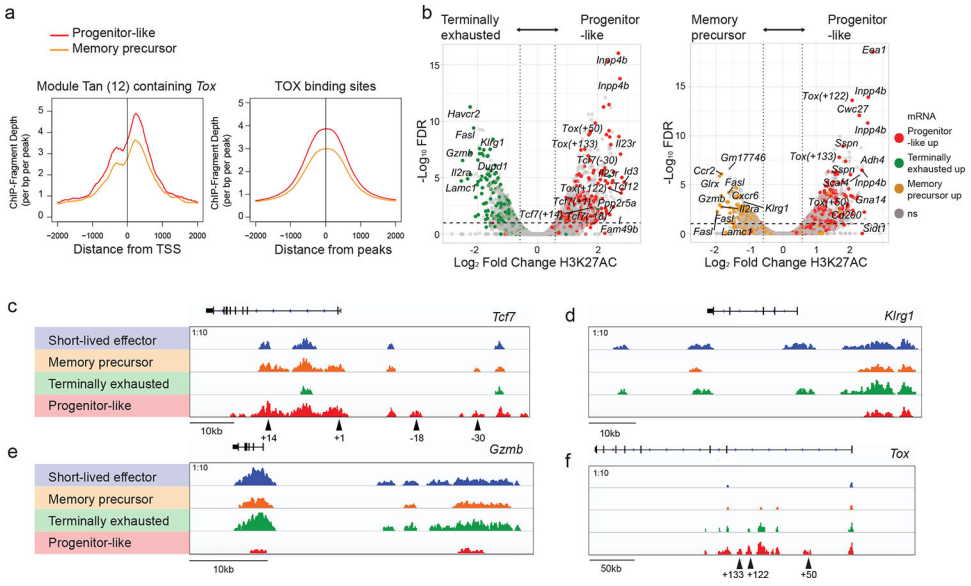

Fig. 5. *Tox* associated genes epigenetically active in progenitor-like CD8$^+$ T cells.

(**a**) Deposition of H3K27ac centered on TSS of genes in module 12 (left) or on TOX binding peaks (right) in progenitor-like (red) and memory precursor (orange) CD8$^+$ T cells. (**b**) Volcano plots of differentially modified H3K27ac peaks between progenitor-like and terminally exhausted CD8$^+$ T cells (left) or between progenitor-like and memory precursor CD8$^+$ T cells (right). H3K27ac data include n=2 biological replicates per group. Differential abundance of H3K27ac is shown as log2 fold change and is plotted against -log10(FDR). Horizontal dash lines denote FDR=0.1, whereas vertical dash lines denote fold change=±1.5. Each dot represents a peak. Peaks associated with significantly upregulated genes (fold change>2, FDR< 0.05) in progenitor-like (red, n=2 biological replicates), terminally exhausted (green, n=2 biological replicates), or memory precursor (orange, n=3 biological replicates) cells are color coded. FDR is determined by edgeR. (**c-f**) Normalized H3K27ac profiles at *Tcf7* (**c**), *Klrg1* (**d**), *Gzmb* (**e**), and *Tox* (**f**) loci in progenitor-like (red), terminally exhausted (green), memory precursor (orange), and short-lived effector (blue) CD8$^+$ T cells. Data is representative of two independent experiments.

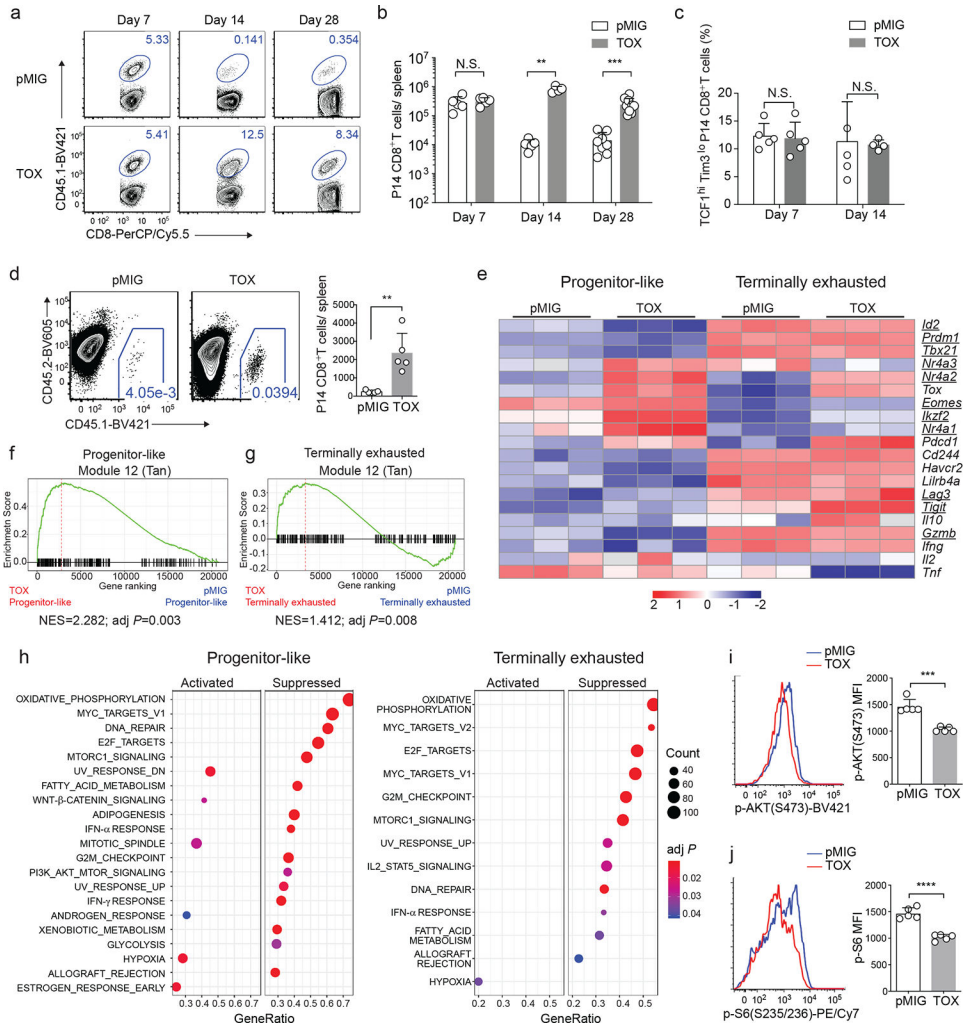

Fig. 6. TOX promotes persistence of virus-specific CD8$^+$ T cells during chronic LCMV infection.

P14 CD8$^+$ T cells transduced with control MSCV-IRES-GFP (pMIG) or TOX overexpression constructs were transferred to C57BL/6 mice that were infected with LCMV clone 13. (**a**) Representative FACS plots of control or TOX-overexpressing P14 cells within CD8$^+$ T cells on day 7, 14, and 28 post-infection. (**b**) Numbers of control or TOX-overexpressing splenic P14 cells on day 7 (pMIG: n=5; TOX: n=5), 14 (pMIG: n=5; TOX: n=4), and 28 (pMIG: n=8; TOX: n=10) post-infection. (**c**) Percentages of TCF1^{hi}Tim3^{lo} control or TOX-overexpressing P14 cells on day 7 (pMIG: n=5; TOX: n=5) and 14 (pMIG: n=5; TOX: n=4) post-infection. (**d**) Day 7 control or TOX overexpression Tim3^{hi} splenic P14 cells were transferred into infection-matched mice (200,000 cells/recipient; n=5 mice per group). Representative FACS plots (left, gated on CD8$^+$ cells) and numbers (right) of splenic donor control or TOX-overexpressing P14 cells on day 5 post-transfer are shown. (**e**) A heatmap of gene expression in day 7 control or TOX-overexpressing progenitor-like (Tim3^{lo}Ly108^{hi}) or terminally exhausted (Tim3^{hi}Ly108^{lo}) P14 cells. Color scale is based on relative fold change. Genes with TOX binding sites are underlined. (**f-g**) GSEA by clusterProfiler illustrating the enrichment of module 12, defined in Fig. 3e, in TOX-overexpressing versus control progenitor-like (**f**) or terminally exhausted (**g**) P14 cells.

Normalized Enrichment Scores (NES) and adjusted *P*-values (adj *P*) are indicated. (**h**) GSEA by clusterProfiler illustrates the activated or suppressed pathways in TOX-overexpressing versus control progenitor-like (left) or terminally exhausted (right) P14 cells. Circle size reflects the count of enriched genes. Adjusted *P*-values are color-coded. (**i-j**) Phospho-AKT(S473) (**i**) and phospho-S6(S235/236) (**j**) staining in splenic control (pMIG: n=5) and TOX-overexpressing (n=5) P14 cells on day 7 post-infection after 30-min re-stimulation with 1μg/mL GP33 peptide. Data in **a-d** and **i-j** are representative of two independent experiments. Three biological replicates per group were used to generate RNA-seq data in **e-h**. In **b-d** and **i-j**, each dot represents one mouse; statistical significance was determined by two-sided Student's t-test; centers and error bars represent the mean and SD. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$.

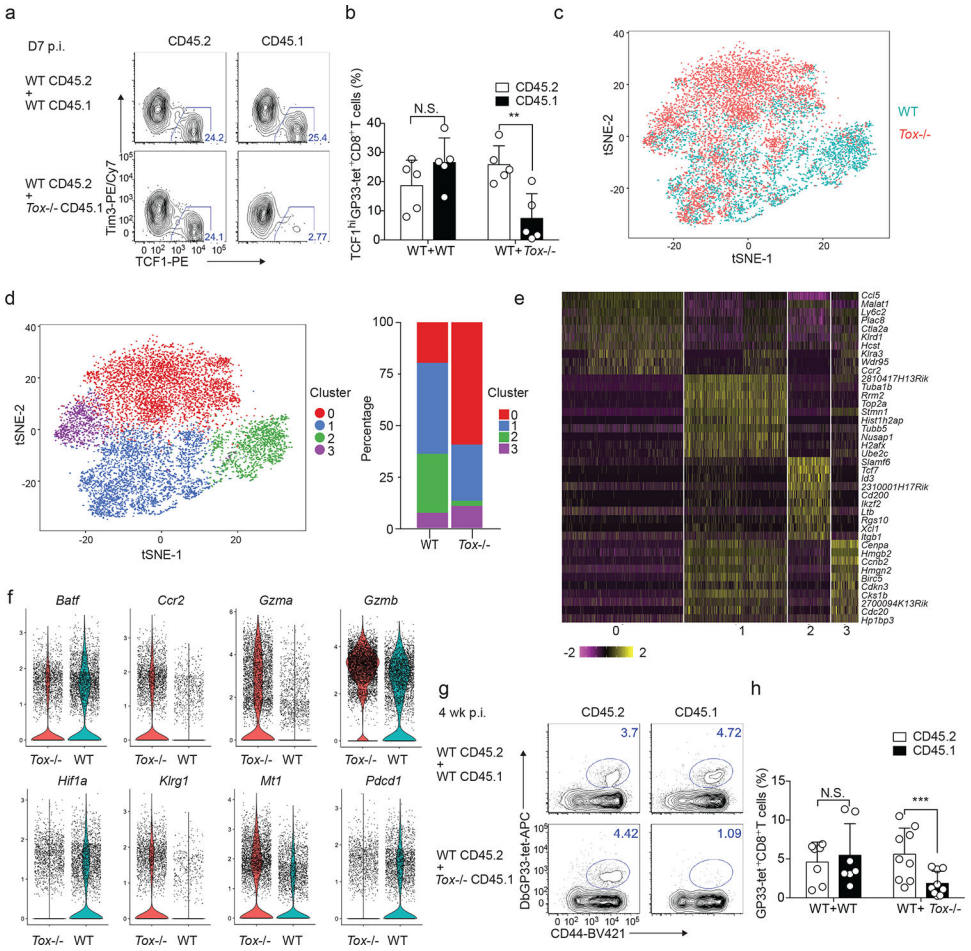

Fig. 7. TOX is required for progenitor-like CD8$^+$ T cell differentiation and persistent antiviral CD8$^+$ T cell responses.

Mixed-bone marrow chimeras that received wild-type CD45.1 and wild-type CD45.2 (WT + WT) or *Tox*−/− CD45.1 and wild-type CD45.2 bone marrows (WT + KO) were infected with LCMV clone 13. (**a**) Representative FACS plots of TCF1 and Tim3 staining in splenic H-2D^b GP33 tetramer$^+$ CD45.1 or CD45.2 CD8$^+$ T cells in chimeras on day 7. (**b**) Frequencies of TCF1^{hi}Tim3^{lo} cells within H-2D^b GP33 tetramer$^+$ CD45.2 wild-type (white) and CD45.1 wild-type or *Tox*$^{-/-}$ (filled) splenic CD8$^+$ T cells on day 7. N=5 mice in each group. (**c-f**) H-2D^b GP33 tetramer$^+$ wild-type and *Tox*$^{-/-}$ CD8$^+$ T cells were collected from mixed-bone marrow chimeras on day 7 post-infection and analyzed by scRNA-seq. (**c**) t-SNE plots of wild-type (n=4,409 cells) and *Tox*$^{-/-}$ (n=5,296 cells) CD8$^+$ T cells, determined by Seurat 2. (**d**) t-SNE plots of wild-type (n=4,409 cells) and *Tox*$^{-/-}$ (n=5,296 cells) cells that were assigned to four clusters and color-coded based on clusters. (**e**) A heatmap of top 10 genes expressed in each cluster defined in **Fig. 7d**. Columns correspond to cells; rows correspond to genes. Cells are grouped by clusters. Color scale is based on z-score distribution from −2 (purple) to 2 (yellow). (**f**) Violin plots of *Batf*, *Ccr2*, *Gzma*, *Gzmb*, *Hif1a*, *Klrg1*, *Mt1*, and *Pdcd1* expression in wild-type (n=4,409 cells) and *Tox*$^{-/-}$ (n=5,296 cells) cells. The violin represents the probability density at each value; each dot represents one cell. (**g**) Representative FACS plots of H-2D^b GP33 tetramer staining on CD45.1 or

CD45.2 $CD8^+$ T cells in the spleen of chimeras 4 weeks after infection. (**h**) Frequencies of H-$2D^b$ GP33 $tetramer^+$ cells within CD45.2 wild-type (white) and CD45.1 wild-type or *Tox*$^{-/-}$ (filled) splenic $CD8^+$ T cells four weeks after infection. N=7 mice in the WT+WT group; n=9 mice in the WT+KO group. Data in **a**, **b**, **g**, **h** are representative of two independent experiments. In **b**, **h**, statistical significance was determined by two-sided paired t-test; centers and error bars represent the mean and SD. *P< 0.05, **P< 0.01, ***P< 0.001.