Title: Atlas of initial lineage trajectories and chromatin regulator functions that establish memory CD8+ T cell ontogeny.

**Authors:** Huitian Diao1, Runqiang Chen1, Megan A. Frederick1†, Dapeng Wang1,Thomas Venables1,Natalie Joe1, Adam J. Getzler1, Walter Rogal1, Justin Milner2, Bingfei Yu2, Shane Crotty3, Ananda Goldrath2 and Matthew E. Pipkin1\*

**Affiliations:**

1 Department of Immunology and Microbial Sciences, The Scripps Research Institute, Jupiter, FL.

2 Division of Biological Sciences, University of California, San Diego, La Jolla, California, USA.

3 Division of Vaccine Discovery, The La Jolla Institute of Allergy and Immunology, La Jolla, CA.

\*Correspondence to: [mpipkin@scripps.edu](mailto:mpipkin@scripps.edu)

†Additional author notes should be indicated with symbols (for example, for current addresses).

**Abstract**: Naïve CD8 T cells develop into diverse effector (TEFF), memory (TMEM) and exhausted (TEX) cell subsets that promote adaptive immunity, but the transcriptional control that initiates their developmental divergence is unknown. Unsupervised inferences after single cell RNA sequencing revealed that naïve cells differentiate linearly into effector/memory progenitor cells (EMPs). Distinct branches that radiate from EMPs initiate stem cell-like (TSTEM), circulating (TCIRC), and tissue resident (TRM) memory cell trajectories very early during acute infection; and alternative TSTEM, progenitor and exhausted T cell trajectories during chronic infection. Disparate RNA velocities in EMPs instigate divergence, and arise antecedent to differential expression of lineage-determining transcription factors. An RNA interference screen of all CRFs in vivo identified that BAF- nucleosome remodelers and interleukin-2 receptor signals cooperatively establish uEMP cells. Subsquently, availability of IL-2R signals and differential requirements for specific BAF- and Chd7-remodeler subunits stabilize divergent RNA velocities that crystalize TCIRC, TRM and TSTEM paths. Thus, early TMEM cell ontogeny conforms to defined developmental paths, but stochastic transcriptional fluctuations encourages individual cells to chart unpredictable courses to categorical fates.

**One Sentence Summary:**

**Main Text:** An essential mechanism of mammalian adaptive immuninty to intracellular pathogens and tumors is the differentiation of effector (TEFF) and memory (TMEM) cytotoxic T lymphocytes (CTLs) that directly lyse infected host cells {Kagi, 1996 #348;Pipkin, 2007 #463;Voskoboinik, 2006 #561}. NaïveCD8 T cells are virgin antigen-specific precursors that can differentiate into a compendium of distinct TEFF, TMEM or exhausted (TEX) cell progenies, after being activated in the context of acute or chronic infections {He, 2016 #307;Pauken, 2016 #13;Scharer, 2017 #497;Scott-Browne, 2016 #507;Sen, 2016 #14;Yu, 2017 #592}. These cell types are distinguishable at the level of chromatin structure, gene expression programs and functional states, implying they are phylogenetically distinct. However, it is unresolved the manner in which activated naïve cells diversify into the array of TEFF, TMEM and TEX cell subsets. In addition, although gene-disruption experiments have demonstrated multiple transcription factors and chromatin regulatory factors that are required for generation of these differentiated CD8 T cell subsets, it is still unknown how regulation of chromatin structure and transcription initiates and stabilizes the diversification process.

**An unsuperivised approach to define early developmental paths responding CD8 T cells take during infection**

Lineage tracing studies have provided evidence that following naïve CD8 T cell activation, but prior to geometric population expansion, key developmental decisions commit cells in the accumulating progenies to short-lived TEFF and long-lived TMEM cell, as well as TRM cell differentiaition outcomes {Buchholz, 2016 #189;Kok, 2020 #622}. In addition, previous scRNA-seq studies demonstrated extensive heterogeneity between responding CD8 T cells after intracellular infections, but the potential developmental connections accounting for generation of this diversity was not identified {Kakaradov, 2017 #350;Kok, 2020 #622;Kurd, 2020 #620;Milner, 2020 #621}. In order to define these trajectories, an unsupervised analysis of scRNA-seq data was used to infer connections between naïve cells and their activated progenies early in the response to either an acute or chronic viral infection induced by different strains of *Lymphocytic choriomeningitis virus* (LCMV). Systemic exposure of mice with LCMV-Armstrong (LCMVArm) induces an acute infection in the spleen that is cleared rapidly. Viremia peaks around day 4 post infection (p.i.) and is mainly cleared by day 8. In contrast, by day 8 p.i. with LCMV-Clone 13 (LCMVCl13) the virus disseminates from the spleen and persists chronically in multiple tissues. By day 8 p.i. distinct subsets of activated TEFF, TMEM precursor and TEX CD8 T cells become evident in the spleen in each infection setting [REF]. To generate these subsets, naïve, P14 TCR transgenic (Thy1.1+) CD8 T cells were adoptively transferred to naïve wildytype C57BL/6J (Thy1.2+) hosts, and separate groups were infected with either LCMVArm or LCMVCl13. On days 5 and 8 p.i. donor P14 TCR transgenic CD8 T cells (Thy1.1+), and host-derived endogenous tetramer reactive CD8 T cells (GP33+, Thy1.2+), which both recognize the LCMV epitope GP33-41 presented in H2-Db MHC-I, were purified from the spleens of the same hosts and prepared into droplet-based single-cell RNA-sequencing (scRNA-seq) libraries. To avoid batch effects, isolation of each cell population from the different time points was performed on the same day in parallel with freshly isolated naïve P14 CD8 T cells, by staggering the infections of mice in these groups and using using cell hashing (**fig. S1A**, [schematic showing all transfer and infection groups and timing to coordinate isolation of all cells on the same day]). This approach faciliated sampling a spectrum of developmental stages in the responding ensemble by including analysis of multiple time points, and because naive CD8 T cells are recruited into the response asynchronously over the first ~ 3 days after infection{D'Souza, 2006 #615}.

**Activated naïve CD8 T cells initially differentiate along a linear path into uncommitted progenitor cells.**

Identical numbers of CD8 T cells from each experimental group were sampled randomly and all samples were considered simultaneously to extract Louvain clusters based on gene expression (**Fig. 1A** [UMAP projection, colored/numbered clusters]and S1B [algorithmic strategy flow chart]). Partition-based graph abstraction (PAGA) was used to infer single cell trajectories between Louvain clusters based on correlations, and the clusters were numbered according to pseudotemporal analysis (P0-P8) (**Fig. 1B,**- course PAGA projection – string plots, make nodes color matched with UMAP clusters]). Cluster P0 was designated as the root of pseudotime analysis to create a putative developmental order because it uniformly comprised naïve P14 cells, the *de facto* developmental origin {Wolf, 2019 #471}(**fig S1C,** [UMAP projection of pseudotime]). As expected, naive cells in cluster P0 were delineated from all activated cells, which separated into multiple distinct clusters (P1-P8) (**Fig. 1A-B**). The distribution of P14 and endogenous GP33-tetramer-specific CD8 T cells among the clusters correlated positively in LCMVArm-infected mice (**fig. S2A,** [heatmap of correlation values (all P14 vs Tetramer+ in LCMVArm);and 3 UMAP plots (P14, Tetramer+, P14-Cl13) showing naïve (color 1), day 5 (color 2) and day 8 (color 3) and LCMVCl13 cells (grayed-out in the LCMVArm-plots)]?), confimring previous studies that demonstrated high similarity in gene expression between adoptively transferred TCR transgenic and endogenous CD8 T cells responding to infection after bulk mRNA analysis {Best, 2013 #31}. Cells isolated on day 5 and day 8 p.i., and those from LCMVArm and LCMVCl13 infected hosts each contributed differentially to the clusters (fig. S1B) (**Fig. 1C** colored UMAP projection, and **S2B**[bar chart showing frequencies of P14 cells in each cluster in each infection. Make two plots: One for LCMVArm, the other for LCMVCl13, and use two bars for each day in each cluster (color 1-day 5) and (color 2 – day 8)] – the idea is to show how cells populate the clusters over time]). The putative lineage identities and pseudotemporal arrangement of cells in the unsupervised clusters were interpreted using gene set enrichment analysis (GSEA){Subramanian, 2005 #459} based on ##[how many do you have total now]?? gene expression signatures extracted from previously published bulk genome mRNA expression data from defined CD8 T cell subsets {Best, 2013 #31;Scott-Browne, 2016 #351;Wang, 2018 #406} (**Table S1** – or spread sheets of each signature – or perhaps GitHub link is best). The signatures of virtually all major TEFF, TMEM, TEX-associated subsets were significantly enriched in one or more activated cell clusters (**Fig. 1D**, [bubble plots of GSEA results using only the CD8 T cell TEFF [Naïve, day 5 KLRG1hi KLRG1lo, day 8 EE, DP, TE, MP], TMEM [JJM-TCIRC, TMEM-Rao, TCM, TEM and core TRM] during LCMVArm infection; and TEX subsets [TEXprog1, TEXprog2, TEXint, TEXterm, Tim3hi Blimp1lo, and th general TEX signature from Scott-Browne/Rao; please divide these into separate panels]). Thus, antigen-specific CD8 T cells undergo extensive diversification prior to antigen clearance during acute infection, or extensive viral persistence following chronic infection.

Multiple models have been used to explain TMEM cell development following acute infection, but do not provide a unifying view. Linear models suggest naïve cells develop progressively, but different models place TEFF and TMEM cells in different developmental orders {Buchholz, 2016 #189;Kaech, 2007 #345}. In contrast, separate lineage models suggest deterministic trajectories establish alternative short-lived TEFF and long-lived TMEM cell fates early in the response {Reiner, 2014 #617}, but are not reconciled with lineage tracing studies {Buchholz, 2016 #189}. The PAGA-defined connections resolved unequivocally that naïve cells in P0 were strongly connected to activated cells in cluster P1, but not others (**Fig. 1B**) which implies the initial differentiation path of activated cells is linear.

Cluster P1 was entirely composed of cells from day 5 p.i., which had decreased naive cell gene expression and induced the signatures found in cells two days after TCR stimulation (**Fig. 1E**, Show P0 and P1 NES in 3 bubble plots: **1.** Best et al. clusters, naïve and 48hrs TCR stim; **2.** day 8 TEFF (TE, EE, MP, DP) and; **3.** TMEM (TCM, TRM, TEM, TSTEM) – seprate the bubble plots), confirming the early positioning of P1 cells in the trajectory analysis. P1 cells progressively acquire this gene expression pattern, because multiple genes encoding regulatory factors typical of naïve, TEFF and TMEM cells were expressed in a visually graded fashion within the streak of cells between clusters P0 and P1 in the UMAP projection (**Fig 1F**, 4 UMAP plots of Il7r, Il2ra, Tbx21, Tcf7 showing mature mRNA expression intensity (maybe the ones from scVelo)). Consequently, P1 cells simultaneously expressed intermediate amounts of genes that are characteristic of naïve, TEFF and TMEM cells (**Fig. 1G**, need to show “co-expression” of factors like Tcf7+Tbx21+Prdm1+Runx3 – maybe violin plots for these genes among the different clusters), and they had proliferated extensively by day 5 p.i., because all P14 cells have undergone > 8 cell divisions by this time (**fig S2C**, need to get FACS-plots from Shanel showing CTV dilution [x-axis]vs KLRG1 expression [y-axis], naïve [one color] and day 5 [2nd color]). P1 cell gene expression was not positively enriched with signatures of mature TEFF or TMEM CD8 T cells, whereas these signatures were strongly enriched in other clusters arranged later in pseudotime (**Fig. 1D and 1H** [GSEA bubble plots showing Best clusters enrichments]). This suggests that P1 cells are uncommitted precurors with multilineage differentiation potential. Consistent with this, P1 cell gene expression was partially correlated with the signature of KLRG1lo CD127hi MP cells, which was extracted from cells responding to LCMVArm infection, and strongly correlated with a signature CD69hi Ly108hi (TEXprog2 ) cells, extracted from cells isolated on day 15 after LCMVCl13 infection (**Fig. 1D**). TEXprog2 cells exhibit multilineage potential during LCMVCl13-infection, giving rise to multiple exhausted T cell subsets in adoptive transfer experiments [REFs]. Flow cytometry confirmed the presence of cells with a TEXprog2 cell surface phenotype on day 5 p.i. with LCMVArm and LCMVCl13 (**fig. S2E**, data from Adam – we don’t have LCMVArm data with these markers yet – need to check with Shanel, she might have them). Collectively, these results support the conclusion that the first step of differentiation following naïve cell activation during LCMV infection is a linear differentiation path that results in formation uncommitted progenitor cells, which we refer to as effector/memory progenitors (EMPs) in LCMVArm-infection and TEXprog2 cells in LCMVCl13-infection.

**Stopped here at 9/11/20.**

**Four developmental trajectories initialize TMEM and TEX cell diversification**

Strong connections between cells in cluster P1 with those in clusters P2, P3, P5, and P7 implied activated CD8 T cells initially diversify along four separate trajectories (**Fig. 1E-G**). Trajectory 1 (T1) was a path to cells in cluster P2, which were enriched with the gene expression signatures of TSTEM cells and TEXprog1 cells, and comprised cells from hosts infected with either LCMVArm or LCMVCl13 [REF]. Trajectory 2 (T2) was a path to cluster P5 that connected to cells in cluster P6, whose gene expression was strongly enriched with the signature of TEXterm cells, and was dominated by cells from LCMVCl13-infected hosts. Trajectory 3 (T3) was the path to cluster P7, which led to cells in P8 that were enriched with signatures of KLRG1hi CD127hi double positive (DP) TEFF cells in LCMVArm-infected mice, and TEXprog2 cells in LCMVCl13-infected mice. Trajectory 4 (T4) was a path to cluster P3 which led to cells in cluster P4. This branch was composed entirely of cells from LCMVArm-infected hosts and was a linear continuum that first gave rise to cells with mature TMEM signatures (P3), and then those enriched with both TCIRC and terminally differentiated TE cell signatures (P4). These results demonstrate that multiple separate developmental trajectories are established well before antigen clearance or persistenc and maximal CD8 T cell accumulation, and likely accounts for the initial diversification TEFF, TMEM and TEX cell lineages that arise during acute and chronic LCMV infection.

**Early diversification of TSTEM, TCIRC and TRM cell lineages during acute infection**

Divergence into Trajectories 1 (TSTEM), T3 (TRM) and T4 (TCIRC) in LCMVArm-infected mice was established by day 5 p.i., suggesting that initiation of these lineages is rapid and virtually contemporaneous. To begin defining how divergence is regulated, we examined differential gene expression between cells in clusters at the onset of each trajectory relative to all others, and relative to those in cluster P1. Compared to all other clusters, cells in P1 cells simultaneously expressed intermediate amounts of genes that encode many TFs whose expression is selectively increased in mature CD8 T cell subsets. TE (Tbx21, Zeb2, Id2, Bhlhe40), TCIRC and TSTEM (Tcf7, Id3, Zeb1, KLF2), and TRM (Hobit, Prdm1, Runx3) cell subsets, and are functionally required for development of these cell subsets were expressed significantly more or less in cells from clusters P2, P3, or P7 compared to all other clusters, and relative to P1 cells specifically. These results confirmed that cells in P1 were uncommitted multilineage progenitors, and that selective and reciprocal expression of multiple TFs is evident very early in the onset of early trajectories leading to TSTEM, TRM and TCIRC and TE cells.

The initial clusters of each trajectory were present in the spleen on day 5 p.i., suggesting these lineage are established contemporaneously.

Cells of cluster P2 were the first branch of T1, and expressed the most *Tcf7*, *Tox*, *Cxcr5*, *Id*3, *Bcl6* and *Slamf6,* and the least *Runx3, Prdm1*, *Id2*, *Il2ra*, *Cx3cr1*, *Prf1*, and *Gzmb*, compared to all other activated cells in the analysis, including P1 cells, confirming their desgination as TSTEM cells. In addition, these gene expression in these cells was positively enriched with genes whose expression is promoted by the TFs Klf2, Tcf1 and Tox, implying this constellation of TFs could promote the early activation of TSTEM gene expression.

These cells were not enriched with MP cell signatures, implying they are distinct precursors. However, they were positively enriched with signatures of naïve, TCIRC and TRM signatures, and expressed the TEFF-associated genes *Prdm1, Tbx21* and *Runx2* compared to naïve cells. The genes encoding Tcf1 and Tox were upregulated in P2 cells relative to P1 cells indicating these programs developed after commitment to development into P2 cells.

Consistent with this, Gene expression in P2 cells from LCMVArm-infected hosts

The origins of distinct TMEM subsets are still unclear. Following an acute infection, the TMEM cells that develop are broadly delineated into circulating (TCIRC) or tissue-resident memory (TRM) subsets. TCIRC cells include effector memory (TEM) and long-lived effector (TLLE) cells that reside in the vasculature; peripheral memory (TPM) cells that recirculate through peripheral tissues; and both central memory (TCM) cells, and cells that manifest stem cell like qualities (TSTEM), which both localize within T cell areas of secondary lymphoid organs. In contrast, TRM cells depart the circulation and mature in non-lymphoid tissues outside of the circulation, where they remain resident *in situ* without re-circulating. KRLG1lo IL-7Rahi MP cells in the spleen efficiently form TCM cells, but might not be the major source of TRM cells. In addition, the exact relationship to either subset

but differ at the level of chromatin structure and transcription from the earliest precursors of TRM cells found in tissues. Consistent with this, cells that migrate to non-lymphoid tissues to become TRM cells are present in the spleen ~ 5 days following LCMV infection, are much less prevalent by day 8 p.i., when MP cells have formed. activated cells near the peak of the response can be delineated into terminally differentiated effector (TE) cells which are mainly short-lived, and memory precursor (MP) cells which preferentially form TMEM cells.

Organization of trajectoris in chronic infection

Compared to all other activated cell clusters, P2 cells expressed the most *Tcf7*, *Tox*, *Cxcr5*, *Id*3, *Bcl6* and *Slamf6,* and the least *Runx3, Prdm1*, *Id2*, *Il2ra*, *Cx3cr1*, *Prf1*, and *Gzmb*. Gene expression in P2 cells from LCMVArm-infected hosts was not enriched with MP cell signatures, but was positively enriched with signatures of both naïve and TMEM cells, and had upregulated expression of the TEFF-associated genes *Prdm1, Tbx21* and *Runx2*, distinguishing them from naïve cells. Cluster P2 cells developed by day 5 p.i. after either LCMVArm and LCMVCl13 infection, and cells consistent with gene expression in P2 cells were readily identified by flow cytometry at this time after either LCMVArm orLCMVCl13 infection, indicating they develop rapidly in both settings. However, these TSTEM-like cells in LCMVCl13-infected hosts were more prevalent, and exhibited distinct enrichments of multiple gene expression signatures, compared to those from LCMVArm-infected hosts. These results imply TSTEM cells develop in a dedicated lineage at early times after either acute or chronic viral infection, but are distinct in each setting.

Trajectory 2 (T2) defined a distinct pathway that gives rise to TEXterm cells. T2 was a path from cluster P1 into two closely related clusters P5 and P7 and then to cluster P6. Cells in cluster P1 from LCMVCl13-infected mice were strongly enriched with the TEXprog2 signature. Cluster P5 was comprised of cells from day 5 p.i., their gene expression was partially enriched with the signatures of TEXint and TEXterm cells and strongly enriched with the signature of EE TEFF cells. Cluster P6 was comprised exclusively of cells from LCMVCl13-infected hosts on day 8 p.i., and their gene expression was enriched with the signatures of both intermediately (TEXint)and terminally (TEXterm) exhausted cells (**Fig. 1F**). These cells also expressed the most *Pdcd1* (PD-1), *Lag3* (LAG3) and *Havcr2* (TIM3) compared to all other clusters in the analysis. Cells in cluster P6 and those in P2 (TSTEM/TEXprog1) each expressed the most *Tox* compared to all other clusters in the analysis, none of the clusters in T2 were enriched with either TSTEM-like or TEXprog1 signatures, nor were they strongly connected to cells in cluster P2. These results indicate that during EMPs that form during chronic LCMV infection, TEXprog2 cells give rise to TEXint and TEXterm cells in a linear continuum (T2), which initially develops separately from TEXprog1/TSTEM-like cells.

Trajectory 3 (T3) led to cells in cluster P8 which comprised DP TEFF cells in LCMVArm-infected mice, and an alternative form of TEXprog2 cells in LCMVCl13-infected mice. T3 was defined by the connection of P1 cells to those in cluster P7 and then P8, and were most prevalent on day 5 p.i., indicating these cells are transient populations in the spleen (**Fig. 1F**).

P5 cells were strongly connected to cluster P7, which were also comprised of cells from day 5 p.i., and whose gene expression was positively enriched with the signature of TEXprog2 cells

A second branch of T2 was a path to cluster P8, which comprised cells from day 8 p.i.. Their gene expression was positively enriched with genes that are upregulated in TEXprog2 cells, but was not positively enriched with the signatures specific to either TEXint or TEXterm cells. Divergence toward either P6 or P8 originated in clusters P5 or P7, which were prevalent on day 5 p.i. but not on day 8 p.i., suggesting they are transient intermediates. Consistent with this, gene expression in Cluster P5 cells was not enriched with any signatures of TEX cell subsets, whereas in Cluster P7 cells it was partially positively enriched with genes that are upregulated in TEXprog2 cells.

correlated with sustained positive enrichment of genes that are upregulated during antigen exposure (Best clusters 1-3 and ), whereas divergence toward cells in cluster P6 was inversely enriched with these genes. This result suggests sustained antigen exposure drives development of TEXprog2 cells, whereas development of TEXterm cells follows interruption of gene expression promoted by these signals.

Cluster P5 also comprised cells from LCMVArm-infected mice on day 5 p.i., but cells from LCMVArm-infected hosts did not contribute to Cluster 6. Instead, cells from LCMVArm-infected hosts were enriched with gene expression signatures extracted from EE, DP and TE, effector cell subsets that develop during acute infection. Thus, cells

In addition, interconnection between clusters P5 and P7 suggests potential interconvergence between cells in T2 and T3 at this point, consistent with each cells in each cluster representing distinct intermediates. Cells from both LCMVArm- and LCMVCl13-infected hosts were found in cluster P7 and were connected to those in P8, and were enriched with gene expression induced by TCR stimulation, suggesting P7 cells, and to less extent P8 cells, recently received antigen stimulation in vivo. Cells from LCVMArm and LCMVCl13 infected hosts in cluster P8 occupied distinct space within the UMAP projection, indicating their gene expression programs diverged appreciably after cluster P7. Notably, cells from LCMVArm-infected hosts were most evident in clusters P7 and P8 on day 5 p.i. compared to day 8 p.i., implying these cells are transient in the spleen, whereas those from LCMVCl13-infected hosts were present at both times. Thus, cells in both infections initially develop along similar paths, but contribute differentially to this trajectory over time.

Specifically, cells from LCMVArm-infected hosts in cluster P7 repressed gene expression induced by Klf2, a TF that promotes lymphocyte recirculation, and which is downregulated in TRM cells. Those in P8 were most similar to DP TEFF cells overall and had upuregulated genes expressed in multiple signatures of late effector and memory cells. However, their gene expression was not positively or negatively enriched with any specific TMEM signature, suggesting they retained potential to differerentate into alternative TMEM subsets. These LCMVArm derived cells expressed genes upregulated in the TEXprog2 and TEXint signatures, although they were unlikely predecessors of TEX cells because they were not enriched with the TEX-specific signature and they strongly repressed genes promoted by both TCF1 and Tox TFs, which both promote features of TEX cells. In contrast, cells in cluster P7 from LCMVCl13-infected hosts were clearly precursors of TEXprog2cells. Genes expressed in these cells were positively enriched with those upregulated in the TEXprog2 and TEX cell specific signatures, and concomittantly had downregulated genes expressed in the core TCIRC and TMEM-specific signatures. Subsequently, cells in cluster P8 more strongly expressed genes upregulated in the TEXprog2 and TEX-specific signatures compared to those in P7, but they did not induce gene expression characteristic of TSTEM-like, TEXprog1, TEXint or TEXterm signatures.

Trajectory 4 (T4) led to cannonical TCIRC and TE populations and was entirely composed of cells from LCMVArm-infected hosts. Cells from P1 directly connected to those in cluster P3, which were connected to those in cluster P4. P3 cells were derived from both days 5 and 8 p.i., and enriched with signatures of both TCIRC and TRM cells, as well as, EE (KLRG1lo CD127lo) and TE (KLRG1hi CD127lo) cells. In contrast, P4 cells were exclusively derived from day 8 p.i., had lost enrichment of the TRM signature, but had increased enrichment of both the TCIRC and the TE signatures. Thus, T4 unambiguously reveals that cells with TMEM gene expression develop first and are the progenitors of TE cells, but that both cell types are highly related.

A clear subdivision between the trajectory framework was enrichment with gene expression that is dependent upon antigen stimulation.

Instead, these cells could be a source of TRM precrusors because they were most similar to DP effector cells, which have TRM potential [REF], and lacked gene expression to promote recirculation.

Cells from LCMVArm-infected hosts were derived from day 5 p.i., and were enriched with genes highly expressed in double positive effector cells (DP, KLRG1hiCD127hi), and late effector and memory cells (Fig. XX, Best DEG, 3,6,8-9). These cells were also negatively enriched with expression of genes upregulated by TFs encoded by *Klf2,* *Tcf7* and *Tox*, suggesting they could be primed for emigrating the spleen and seeding non-lypmhoid tissues. Consistent with this, both *Runx3* and *Prdm1* (Blimp1), which both encode TFs that promote TRM cell development were most highly expressed in clusters P7 and P8 relative to all others, and cells in these clusters were most prevalent on day 5 p.i., when enrichment TRM precursors in the spleen has been shown to peak.

Notably, gene expression driven by TCR stimulation was not enriched in cells from clusters P5 and P6, indicating that formation of TEXterm cells was not correlated with sustained, immunogenic antigen-receptor stimulation, which implies that earlier transcriptional events establish this trajectory in LCMVCl13-infected hosts.

TMEM cells that develop after resolution of a prototypical acute infection are broadly delineated into circulating (TCIRC) cells, which comprise effector memory (TEM), long lived effector (TLLE) and central memory (TCM) subsets, or tissue-resident memory (TRM) cells. In addition, within the TCM population, cells that manifest stem cell like qualities (TSTEM) have been identified functionally. However, the initial derivation of each of these subsets remain unclear. CD8 T cells responding to acute infection can be subdivided into TE cells and classical memory precursor (cMP) cells based on differential surface expression of Killer like lectin receptor G-1 (Klrg1) and interleukin-7 receptor alpha (IL-7Ra/CD127). However, these subsets only become distinguishable in the spleen near the peak in T cell accumulation, when re-expression of CD127 (*Il7ra*) occurs on some cells. KRLG1lo CD127hi cMP cells preferentially develop into TCM and TEM cells, but are distinct from TRM cell precursors found in non-lymphoid tissues at the same time in terms of both chromatin accessibility and gene expression. Consistent with this, cells that most efficiently form TRM in gut non-lymphoid tissue are most prevalent in the spleen several days prior to when cMPs become evident.

**Heterogeneous RNA velocities in uEMPs initiate lineage specification**

To elucidate how transcriptional heterogeneity initiates lineage choice, we analyzed single cell RNA velocity vectors, which reflect nascent transcription of genes that precede future mature mRNA expression regimes in single cells. To identify putative developmental roots and endpoints in an unsupervised fashion, transition probalities were calculated based on divergence in RNA velocity vectors. Cells in clusters P1, P2, P5, P7 and P8 harbored cells with strong transition probabilities. Cells from LCMVArm and LCMVCl13-infected hosts harbored distinct fractions of cells with root probabilities within each cluster (display the % of cells in each cluster with > 50% transition probability for each infection type; grouped bar chart). In contrast, cells in clusters P0 and P4 harbored the weakest transition probabilities, suggesting they were developmental endpoints. Consistent with this, cluster P0 comprised naïve cells prior to stimulation, whereas cluster P4 comprised cells from day 8 p.i. in LCMVArm-infected hosts that were most similar to TE cells, that have limited potential to form TMEM cells. These results indicated cells throughout trajectories 1, 2 and 3 experience divergent RNA velocities, and reveals multiple decision points in each trajectory exist that could alter the trajectory cells in these developmental paths.

To characterize the RNA velocities within the trajcetories, we examined the both the grid average and individual single cell velocities to To clarify the gene expression changes that distinguished cells at the initial divergence of trajectories during LCMVArm infection we defined differential expression between cells in each of the intial clusters of trajectories 1, 3 and 4, and cells in cluster P1. The initial clusters of each trajectory were present in the spleen on day 5 p.i., suggesting these lineage are established contemporaneously.

The GSEA analysis of these clusters demonstrated that cells in clusters P1 and P2 were enriched with

In contrast, cells from cluster P1, P2, P7 and P8 demosntrated in this space of the of UMAP projection manifested strong transition probabilities and highly divergent RNA velocities.

P1 cells were heterogenous. Those located proximal to naïve cells in the projection of PAGA clusters were likely those most recently recruited into the antiviral response (**Fig. XX**), and exhibited RNA velocities vectored away from naïve cells, and toward cells in cluster P1 located proximal to other activated cell clusters in the projection. Individual This implied these P1 cells were the earliest developmental roots of T1-T4, confirming their designation as uEMP cells. In addition, cells in P2, P7 and P8 each also exhibited highly divergent RNA velocity vectors and strong transition probabilities, implying that multiple differentiation roots developed downstream of the initial manifold that originated within cluster P1. Consistent with this, the scVelo transition stream revealed underlying conduits within the PAGA-defined connections that defined alternative paths individual cells can take to the same lineage endpoints. These results imply that after the P1 cell stage, differentiating CD8 T cells might encounter multiple opportunities to engage differential transcriptional programs that lead to divergent developmental paths.

Although gene disruption has demonstrated that multiple TFs which are expressed in a lineage-specific fashion are required for the formation of specific TEFF, TMEM and TEX cell subsets, the underlying transcriptional dynamics that drive formation of these lineages are still unresolved and it is still unclear whether these TFs establish or stabilize these lineages. To define how transcription establishes lineage-bias that leads to distinct CD8 T cell subsets, differential RNA velocities were computed between all clusters to identify when single cells activated transcription of specific gene expression programs by using GSEA (Fig. XX). The TSTEM signature was enriched within RNA velocities in cells from clusters P1, P7 (LCMVArm MP and DP; LCMVCl13 EE, TEX and T) and P8 (DP and ) that were strongly oriented toward cells in cluster P2. Consistent with this, transcription of the gene expression signatures promoted by Tcf1 and Tox and repressed the TF Runx3 were strongly enriched within these RNA velocities. Thus, some activated effector-like cells engage TSTEM like transcription and develop via ‘retrograde’ differentiation (**Fig. XX**).

T2 indicates TEXterm cells, which are found in cluster P6, develop rapidly in the context of LCMVCl13 infection and derive from cells in cluster P5 (enriched with TEXprog2 and TEXint signatures), confiming the PAGA-inferred T2 (**Fig. XX**).

Previous studies have implied that spontaneous transcriptional fluctuations promote early lineage bias independent of particular lineage-determining factors. In every CD8 T cell trajectory, origination of lineage specific RNA velocities was not correlated with the highest expression of cognate TFs that are normally definitive of the specific lineage, implying that spontaneous fluctuations in protein expression of these factors in conjuction with other lineage-determining factors are critical for driving the specific RNA velocities.

In that Downstream of uEMP cells in P1, addition, other cells from cluster P5 developed vectors aimed at cells in P7. Cells in P7 then either developed velocities aimed toward cells in P2 (TSTEM), or toward cells in cluster P8 (TEXprog2,Ly608+CD69-, LCMVCl13 hosts), confirming the PAGA-inferred T3 and heterogeneity in the origins of TSTEM cells. Finally, RNA velocities in some cells from clusters P1, P2 and P5 were vectored toward cells in cluster P3 (TCIRC/TRM), suggesting that cells which develop TCIRC gene expression derive from at least 3 developmental paths. The velocity vectors in P3 cells either became weakly retrograde, or skewed weakly forward toward cells in P4 which lacked strong vectors toward any other gene expression space within this analysis. These results suggest early TMEM ontogeny is organized by defined developmental paths, but that stochastic transcriptional fluctuations drive individual cells to undertake variable courses through these trajectories to one of several endopoints.

TCR stimulation appeared to play an important role in promoting root developmental potential within the trajectories, because clusters P1, P2, P7 and P8 were all enriched with gene expression Multiple gene expression signatures associated with recent TCR stimulation were significantly enrichecd

**Transcriptional programs that establish TEFF and TMEM trajectories during acute viral infection**

The PAGA-inferred trajectories and underlying RNA velocities indicated that specific TMEM lineages are established from uEMPs via distinct early trajectories. However, a dedicated trajectory enriched with TRM gene expression was not obvious, perhaps because full TRM differentiation proceeds after entering non-lymphoid sites. Both TSTEM (cluser P2) and TCIRC (cluster P3) precursor cells were partially enriched with the core TRM gene expression signature, suggesting TRM cells could originate from these lineages. Cells with the capacity to seed non-lymphoid sites are most highly enriched within the spleen ~ 5 days after infection with LCMV, but relatively few TSTEM and TCIRC precursors were evident at this time, arguing that the other effector subsets harbored TRM potential. The TF Runx3 drives the TRM transcriptional program, in part by repressing expression of the TF KLF2, which normally enforces expression of surface receptors that promote homing and recirculation of T cells in lymphoid tissue. The gene expression signatures promoted by Runx3 and repressed by KLF2 were positively enriched within cEMPs and DP (clusters P1 and P7), whereas the signatures promoted by KLF2 and repressed by Runx3 were positively enriched within TSTEM, TCIRC and TE cells (clusters P2, P3 and P4), which both suggested that cells in P1 and P7, which were prolific on day 5 p.i., expressed mature gene expression signatures that were likely to be inducing the TRM program. To determine this directly, we computed differential RNA velocities between all clusters and determined the ernichement of the diffferntial velocities with the cell type specific gene expression sigantures, which revealed that RNA velocities in P1 cells were strongly polarized toward the TRM gene expression program.

extracted from cells isolated from non-lymphoid tissues. Although TRM differentiation is thought occur in situ, some TRM are localized and secondary lymphoid organs, or cells that activate a significant portion of the core TRM program prior to departing the spleen. Thus, these cells could The differential origin of effector cells that give rise to TCIRC and TRM

The developmental process by which TCIRC cells emerge from the large burst of responding TEFF cells following acute viral infections is still unresolved. Trajectory T4 unequivocally demonstrated that activated naïve cells differentiate into TMEM cells (enriched with both TCIRC and TRM gene signatures in cluster P3), and then terminally differentiate into TE cells in cluster P4, which was a terminal node in the analysis. Enhanced expression of multiple definitive TE markers (e.g., *Tbx21*, *Klrg1*, *Zeb2*, *Cx3cr1*) in cluster P4 relative to P3 confimed this transition. The real time-series confirmed the pseudotime estimation because cells in P3 were identified on day 5 p.i., whereas those in P4 had not formed at this time. Thus, TMEM-like cells are the precursors of TE cells, as suggested by lineage-tracing studies [REF]. However, the close connection between clusters P3 and P4, and the fact that P4 cells were enrinched with both TE and TCIRC signatures, both indicate that P4 cells retain the potential for TCIRC memory formation, consistent with some effector cells from day 8 p.i. reverting to a TCM phenotype [REF].

**Transcriptional programs that establish TSTEM and TEX cell subsets during chronic viral infection**

The development of TEX cells during LCMVCl13 infection is attributed to chronic antigen stimulation [REF]. However, very early gene expression alterations correlated with initiation of TEX cell development and defective formation of protective TMEM cell lineages in mice responding to LCMVCl13. LCMVCl13-infected mice were severely depleted of cEMPs and completely lacked cells in P3 on day 5 p.i., implying that very early events after LCMVCl13 infection prevented TMEM cell formation via the T4 trajectory (**Fig. 1A-C**). Instead, cells from LCMVCl13-infected mice dominated Trajectory 2 (TSTEM-like) and branches of Trajectory 3 leading to TEX cells of cluster P6 (**Fig. 1F**, top), and TEXprog2 cells in cluster P8 (**Fig. 1F**, bottom). In hosts infected with either LCMVArm or LCMVCl13, P1 cells were strongly enriched with multiple gene signatures indicative of recent or ongoing TCR stimulation (fig. S1X). However, P1 cells from LCMVCl13-infected mice compared to those from LCMVArm-infected mice expressed much less *Jun, Junb, Fos* and *Fosb*, which encode bZIP proteins that cooperatively bind DNA with NFAT-family TFs to drive immunogenic T cell activation [REF], and conversely, much more *Batf* which encodes the bZIP protein that drives TEX cell devevelopment [REF]. In addition, P1 cells from LCMVCl13-infected mice were enriched with the gene expression signatures promoted by both Tcf1 and Tox TFs, whereas those from LCMVArm-infected mice were not (Fig. 1XX). Although the signatures of mature TPROG and TEX cells were not yet evident P1 cells from LCMVCl13-infected mice, they nevertheless strongly overexpressed *Tox* itself as well as *Pdcd1* and *Lag3* (**Fig. 1X**). Thus, early differential mRNA expression establishes lineage-bias that promotes T cell exhaustion during LCMVCl13 infection and functional memory cells during LCMVArm infection. Furthermore, these results also suggest that TCR stimulation of cEMPs in the face of reduced expression of multiple AP-1-family TFs during LCMVCl13 infection initiates the TEX cell transcriptional program.

In both humans and mice, a Tcf1+ subset of TSTEM-like cells develops after acute or chronic viral infections, and in response to tumors, but their origins are unclear. Trajectory 2 demonstrates that TSTEM cells (P2) develop directly from cEMPs (P1) within 5 days after either LCMVArm and LCMVCl13 infections in a dedicated pathway that is initially unrelated to other trajectories. Many more cells in P2 develop after infection with LCMVCl13 compared to LCMVArm, consistent with the Tcf1 and Tox-dependent lineage-bias observed in P1 cells from LCMVCl13 infected mice. P2 cells from LCMVArm and LCMVCl13 were also distinct, differential expression demonstated

P2 cells from mice responding to LCVMCl13 were much more abundant compared to P2 cells from LCMVArm-infected hosts, and were distinct based on gene expression. Unlike P2 cells from LCMVArm-infected hosts, those from LCVMCl13-infected hosts expressed genes that were positively enriched with genes upregulated in the exhausted TEXprog1 signature, and also repressed genes that are upregulated in multiple signatures from late effector and memory cells during acute infection. In contrast, cells in P2 from LCMVArm-infected hosts significantly upregulated genes expressed in both naïve cells, TMEM cells responding to acute infection, and multiple… . Thus, TSTEM cells develop in a dedicated lineage at early times during both acute and chronic infection, but are distinct at the level of mRNA expression in each context.

TSTEM cells function as progenitor cells that develop into intermediate (TEXint) and terminally exhausted (TEXterm) subsets during LCMVCl13 infection and dysfunctional T cells (TDYS) in tumors. Following acute viral infections, TSTEM cells develop and are enriched within the TCM compartment, and function to sustain the TMEM compartment during homeostasis, and to repopulate multiple TEFF and TMEM subsets during anamnestic responses.

Cells from LCMVCl13-infected mice were not enriched with putative endpoints, confirming their distinction from those that develop in LCMVArm-infected hosts, and could suggest that terminal differentiation results after antigen can be cleared. Consistent with this, there was strong positive correlation between transition probabilities and the enrichment of gene expression from TCR stimulated CD8 T cells.

We initially focused on cells that developed in the context of LCMVArm infection, which ultimately give rise to multiple protective memory subsets including TCIRC and TRM cells.

Cells with TSTEM-like qualities that are enriched within the TCM compartment.

The pseudotemporal order inferred from PAGA analysis was ratified by the enrichment of dynamically expressed genes clustered after unsupervised analysis of an extensive longitdunal time series [Best, et al] ratified (**Fig. 1X**). However, estimations of developmental orders using pseudotime are potentially limited because the analyzed cells were separated in actual time and their kinships were unkown [REF]. Thus, we computed single cell RNA velocities which confirmed the PAGA-initialized connections and improved the resoluation of their inferred developmental orders (**Fig. 1Y**).

The developmental paths that account for Distinct models have been used to described how long-lived TMEM cells emerge from the large number of effector cells, most of which die,. Two linear models have been described, but each place the precursors of

, the majority of responding cells at the peak of the response on day 8 p.i. comprise TE (KLRG1hi CD127lo) cells which inefficiently give rise The developmental paths that account for how classical MP (KLRG1lo CD127hi) cells differentiate from TE (KLRG1hi CD127lo) cells has not been resolved. Linear differentiation models have arranged effector and memory precursor cells in opposite developmental orders, whereas additional bifur TRM cells. which most efficientaly give rise to TCIRC (central memory (TCM) and effector memory (TEM) cells,

the inferred trajectories indicated that TSTEM (P2) and TCIRC (P3) precursor cells develop in different paths. However, TSTEM cells appeared to arise earlier and some could give rise to TCIRC precursors.

TSTEM cells result from clarified

The developmental order in which The pseudotemporal order inferred from PAGA analysis was ratified by the enrichment of dynamically expressed genes clustered after unsupervised analysis of an extensive longitdunal time series [Best, et al] ratified (**Fig. 1X**). However, estimations of developmental orders using gene expression pseudotime is hampered because the kinships of individual cells are unknown and are separated in actual time [REF]. We computed single cell RNA velocities which confirmed the PAGA-initialized connections and improved the resoluation of their inferred developmental orders (**Fig. 1Y**). This facilitated estimating the transition probabilities of individual cells in the trajectories, and using the enrichment of gene expression signatures within the RNA velocities to define when cells establish transcription that leads to future cell fates. Differential RNA velocities and was used to define causal transcriptional regulatory networks that drive lineage formation. The first critical step in TMEM cell formation involves differentiation of cEMPs and results from transcription that leads to gene expression in P1. This was evident by virtue of activated cells that clustered with naïve cells in gene expression space (P0), but that manifested RNA velocities vectored toward gene expression in P1. Notably, some activated cells positioned in an intermediate location in the UMAP projection exhibit retrograde vectors, implying that continuous signals or a threshold drives differentiation forward into P1. Development of cells in P1 involves extensive cell division, because all transferred P14 cells had undergone > 8 cell divisions by day 5 p.i., and had, to varying extents, upregulated KLRG1 expression relative to naive cells.

Responding CD8 T cells isolated from hosts infected with either LCMVArm and LCMVCl13 contributed differentially to each trajectory.

Divergence between T2 and T3 appears to be governed by antigen receptor stimulation of cells in P5 and P7.

Classically defined TCIRC memory cells. linear continuum that ultimately gives rise to classical TE cells as defined using differential KLRG1 and CD127 expression, and only develops during LCMVArm infection after day 5.

**Transcriptional programs that establish TSTEM and TEX cell subsets during chronic viral infection**

which highly expressed genes characteristic of TSTEM cells (*Tcf7, Tox, Bcl6, Id3, Slamf6, Cxcr5*)

This signature was strongly enriched, even when cells from LCMVArm or LCMVCl13 infected hosts were considered independently. However, P2 cells from LMCVArm- and LCMVCl13-infected hosts were distinct. P2 cells from LCMVCl13-infected hosts repressed multiple effector and memory cell gene signatures and were more significantly enriched with the TEX gene signature than P2 cells from LCMVArm-infected hosts. What are the key features that distinguish TSTEM in Cl13 vs Arm? (Differential expression between P2 cells from hosts infected with either LCMVCl13 or LCMVArm demonstrated that…. (what is significantly different between the two)). Conversely, P2 cells from LCMVArm-infected hosts were strongly enriched with signatures of naïve cells, and both mature TCIRC and TRM cells. Differential analyses of P2 cells from LMCVArm and LCMVCl13 demonstrated that….. expressed much hiher levels of. Thus, TSTEM-like cells develop at early times as a distinct pathway from other developing lineages during both acute and chronic infection settings, but they manifest distinct gene expression biases. These results provide an explaination for why responding CD8 T cells isolated at very early times after LCMVCl13 infection can form protective memory CD8 T cells upon transfer into LCMVArm-infected hosts.

Trajectory 2 develops through an intermediate (P5 (L4)) that subsequently diverges into two branches giving rise to cells with TMEM potential in LCMVArm infected hosts, and two alternative forms of TEX cells in LCMVCl13 infected hosts. Cluster P5 cells (L4) are enriched with the gene expression signature of TCR-stimulated CD8 T cells, but do not express signatures of mature TEFF or TMEM subsets. Thus, P5 cells have recently experienced antigen stimulation but appear uncommitted. P5 cells responding to LCMVArm and LCMVCl13 diverge in their gene expression in a manner that correlates differentially with their response to TCR stimulation. P5 cells in LCMVArm hosts upregulate genes associated with early T cell activation (Best #2/3 and 2-day TCR), whereas those from LCMVCl13 infected hosts less strongly do, and instead upregulate genes expressed in EE and TEX cells, which suggests reduced immunogenic TCR stimulation in LCMVCl13-infected hosts leads to TEX gene expression.

Cells in P5 diverged into the unrelated clusters P6 (L6) and P7 (L5). Most cells in P6 are from day 8 p.i., LCMVCl13 infected hosts, and are terminally differentiated TEX cells that highly express *Pdcd1* (PD-1), *Lag3* (LAG3) and *Havcr2* (TIM3), as well as the TF *Tox*. In contrast, cells in P6 responding to LCMVArm express genes typical of TRM cells. Thus, TEX and TRM cells both appear to arise along a common developmental trajectory, which might explain previously observed parallels between gene expression in both subsets. The second branch from P5 leads to clusters P7 and then P8. Trajectory 3 directly connects P1 to the P7-P8 path bypassing P5. The P7 path appears to be driven by continued TCR stimulation, because cells in both LCMVArm and LCMVCl13-infected hosts strongly enrich with genes induced upon TCR stimulation. However, P7 and P8 cells from LCMVCl13-infected hosts progressively increase their enrichment with gene expression of TEX cells, whereas those from LCMVArm-infected hosts do not, and instead increase expression of EE and DP cells. P8 cells from LCMVArm-infected hosts do not express a mature TMEM gene expression signature, but both EE and DP cells exhibit robust TMEM cell developmental potential, indicating they are TMEM cell precursors. Consistent with this, these cells were not enriched with the TE cell gene expression signature, and arise in a distinct trajectory from those that do.

Trajectory 4 branches from cluster P1 into P3 and P4, which forms a linear continuum that ultimately gives rise to classical TE cells as defined using differential KLRG1 and CD127 expression, and only develops during LCMVArm infection after day 5. P1 cells are most strongly enriched with the MP cell signature compared to all other clusters. P3 cells strongly express EE, TE, TCIRC and TRM memory cell signatures, whereas P4 cells downregulate the EE and TRM cell signatures, and upregulate expression of both TCIRC and TE cell signatures. In addition, P3 cells are connected earlier with TSTEM cells in P2, indicating they might give rise to cells in P3 (L0) between days 5 and 8 p.i., during LCMVArm infection. Thus, cells with TMEM gene expression develop prior to and serve as the precurosrs of those that induce TE cell gene expression.

**Distinct RNA velocities in individual cEMPs underlie lineage divergence**

Single cell RNA velocities confirmed the PAGA-initialized connections and their inferred developmental orders, and was used to extract causal transcriptional regulatory networks that drive lineage formation. The first critical step in TMEM cell formation involves differentiation of cEMPs and results from transcription that leads to gene expression in P1. This was evident by virtue of activated cells that clustered with naïve cells in gene expression space (P0), but that manifested RNA velocities vectored toward gene expression in P1. Notably, some activated cells positioned in an intermediate location in the UMAP projection exhibit retrograde vectors, implying that continuous signals or a threshold drives differentiation forward into P1. Development of cells in P1 involves extensive cell division, because all transferred P14 cells had undergone > 8 cell divisions by day 5 p.i., and had, to varying extents, upregulated KLRG1 expression relative to naive cells.

Strong omnidirectional RNA velocities observed in the average vector field of P1 indicated individual P1 cells establish differential RNA velocities that precedes their divergence into to connected PAGA clusters with mature gene expression signatures (**Fig 1X**). Vectors from individual cells demonstrated that the averages represented a differentiation manifold flowing toward mature gene expression signatures in clusters P2, P3, P4 and P5. Thus, cells in P1 are cEMPs that establish lineage bias at the transcriptional level which precedes their development along any one particular lineage.

Although PAGA connections suggest TSTEM cells mainly develop from cells in P1, RNA velocities that develop in

The gene expression signatur of TRM cells. *RNA velocities indicate that both TSTEM-like cells and effector-like branches feed a linear path that initially acquires TMEM cell gene expression that develops into TE cell gene expression and then RNA velocities collapse:*

*RNA velocities point to a common differentiation pathway* that gives rise to TCIRC and TE cells*:*

*RNA velocities indicate TRM cells develop in multiple pathways:*

**Naïve CD8 T cells initially differentiate along a linear path into common effector and memory progenitor cells**

To classify the identities PAGA-clusters and their single cell trajectories, we used gene set enrichment analysis (GSEA){Subramanian, 2005 #459} (**Fig 1D**). Each PAGA-cluster mRNA expression profile was compared to multiple gene sets, using previously defined dynamically expressed genes (DEGs) in CD8 T cells that were clustered because they exhibit similar kinetics throughout the response to intracellular infection {Best, 2013 #31}; gene expression “signatures” from mature TE and MP populations defined at the peak response to LCMV infection (day 8 KLRG1 and CD127 subsets) and TMEM subsets (KLRG1-negative cells from day 45 LCMV p.i.) derived from bulk RNA-seq analyses {Scott-Browne, 2016 #351;Wang, 2018 #406}; and gene expression signatures from single cell RNA-seq analyses that identified progenitor-like CD8 T cells during chronic infection settings.

Compared to naive cells in cluster P0, those in P1 and P2 progressively downregulate genes that are highly expressed in naïve cells (**Fig 1D**, DEG 4 and 5), followed by activation of genes characteristic of proliferating effector cells (**Fig 1D**, DEG 1-3), and was highly enriched with genes that are up and down regulated within the first 48 hours of naive CD8 T cell stimulation (**Fig 1D**, far right), confirming the early pseudotemporal assignemnts of cells in clusters P0-P2. Cells in cluster P1 strongly expressed *Cd44*, demonstrating they were activated and had already induced genes that are highly expressed in fully differentiated TMEM and TE cells (*Prdm1, Id2, Tbx21, Bhlhe40, Klrg1, Zeb2*). They also upregulated *Ifng, Tnf* and *Il2* which encode cytokines that are expressed by both effector and TMEM CD8 T cells*,* butthey had not upregulated *Prf1* or *Gzmb,* which encode effector molecules expressed by mature cytotoxic T lymphocytes (CTLs) (maybe a heatmap of TPMs to show expression in these clusters vs mature TE, MP and N). Moreover, they also strongly expressed many genes that are repressed in mature effector cells that encode regulatory proteins which define naïve, MP and TMEM cells (*Ccr7*, *Sell,* *Tcf7*, *Klf2, Id3*). Cells in cluster P2 downregulated these genes, and more strongly expressed genes that are upregulated in recently activated naïve CD8 T cells (Fig 1D, DEG 1-2, Act 48h). P2 cell gene expression was the most similar to the MP cell gene signatures compared of all other PAGA-clusters, and was also very strongly enriched with signature genes that are upregulated and downregulated in DP cells, which like MP cells, also have TMEM potential. In contrast, P2 gene expression did not positively correlate with signatures of EE cells, TE cells, TPRO-like cells, or mature TMEM subsets, nor that from CD8 T cells in late phases of the response (**Fig 1D**, DEG 7-10). These results demonstrate that naïve CD8 T cells initially differentiate along a common linear path into cells with gene expression indicating that they are early MP cells.

**Common effector and memory progenitor cells develop lineage-bias and then diverge into separate memory stem cell-like and effector-like trajectories**

Cells from cluster P2 diverge along two distinct pathways. The earliest branch leads to cluster P3 which is most highly enriched with signatures of memory stem cell- or progenitor-like (TPRO) CD8 T cells {Yao, 2019 #613} compared to all other activated P14 cell PAGA-clusters (**Fig 1D**, middle row). Cells in P3 reexpressed multiple genes encoding TFs that are expressed in naïve cells but are transiently downregulated in cells from P2 (*Tcf7*, *Tox*, *Id3*, *Myb*, *Ccr7, Slamf6*). P3 cells appear to be immediate predcessors of mature TMEM cells, because their gene expression is enriched with the up and downregulated gene signatures of mature circulating (TCirc) and Trm subsets (**Fig 1D**, …), and because cells from cluster P6 which are arranged immediatedly downstream of cluster P3 are even more enriched with these signatures compared to P3, and are not enriched with the signatures of MP cell or TPRO cells. Thus, CD8 T cells with TSTEM-like gene expression develop after traversing a common differentiation pathway, and then differentiate into those that acquire gene expression of mature TMEM cells.

In contrast, the second and later branch deriving from cluster P2 leads into cells from cluster P4, whose gene expression is enriched with upregulated and downregulated genes in signatures from both TE and DP cells, as well as effector CD8 T cells near the peak response to infection (**Fig 1D**, DEGs 6, 8-9). However, P4 cells still also highly express genes upregulated in CD8 T cells in early phases after infection (DEGs 1-3, and Act 48hours), indicating they are early intermediates in the effector cell pathway. In line with this, cells in P4 undergo further branching, leading to clusters P5 and P7, whose gene expression more strongly enriches within the TE and DP cell signatures, and with those expressed at late times after infection (DEGs 6, 8-10). Compared to all other clusters, cells in P5 and P7 most highly expressed the hallmark effector genes of cytotoxic T lymphocytes, *Prf1* and *Gzmb*, demonstrating they comprise *bona fide* mature effector CTLs.

**IL-2R-dependent transcription establishes common effector memory progenitor cells and promotes effector-like differentiation bias**

In addition, both wildtype (WT) and *Il2ra*-deficient (*Il2ra-/-*) P14 CD8 T cells were co-transferred, and were analyzed simultaneously with ex vivo naïve WT P14 CD8 T cells. We reasoned this approach would facilitate orientating the de novo assembled differentiation trajectories and interpretating identities of cells in the pathways, because naïve CD8 T cells are the initiation point in the trajectories, and *Il2ra-/-* CD8 T cells inefficiently differentiate into effector memory (TEM) and TE progenies{Mitchell, 2010 #424;Pipkin, 2010 #468}, and instead develop into cells that manifest a TCM phenotype but are defective{Williams, 2006 #573}(**Fig. 1A –** schematic of co-transfer).

**Memory stem cell-like and effector-like trajectories each give rise to both memory and terminally differenated effector cells via downstream convergence.**

Although distinct memory (P2-P3-P6) and effector branches (P2-P4-P7) of differentiation develop, at late times cells from both pathways converge. Clusters P6 and P7 are closely correlated, but P7 is arranged later in the pseudotemporal order, suggesting that some cells in P6 give rise to those in P7. Gene expression in both P6 and P7 cells postively correlated with gene expression signatures of EE and TE cells, and genes whose expression is upregulated and sustained at late times after infection (DEGs 6, 8-10), demonstrating that both clusters harbor cells with qualities of maturing effector cells. However, P6 cell gene expression positively correlates with all analyzed TMEM signatures, whereas P7 cell gene expression does not, and instead more strongly correlates with TE and DP cells. Therefore, P6 cells are more memory-like compared to P7 cells. In addition, compared to all other clusters, P6 cell gene expression is the most strongly enriched and positively correlated with the gene expression signature of EE cells, which are known to give rise to both TE and TMEM cells. Thus, P6 cells exhibit both memory and effector cell potential, suggesting that after divergence from a common differentiation pathway at P2, cells differentiate linearly into TMEM, some of which lose TMEM gene expression and gain TE gene expression.

Cells from the effector pathway also appear give rise to TMEM cells. Although cells in P7 expressed a more effector cell like profile compared to those in P6, they most likely retain TMEM potential because they did not repress genes that are upregulated in TMEM cells and they highly expressed genes that were enriched within the EE cell signature. Furthermore, Gene expression in P5 cells differed substantially from that in P7 cells, implying these cells represent distinct effector cell differentiation outcomes (Huitian – do GO enrichment for genes different between P5 and P7). Gene expression in P5 cells was enriched with the TMEM gene expression signature, but correlated inversely, indicating P5 cells are unrelated with mature TMEM cells (We need to use GSEA of Tcm, Tem and Tlle cells). In contrast, there was virtually no enrichment of genes that are upregulated in cells from cluster P7 with those that are upregulated in TMEM cells, but downregulated genes in P7 cells were enriched among those that are downregulated in TMEM  cells, suggesting that P7 cells are partially related to TMEM cells. Consistent with this,

We confirmed that the de novo organization of the developmental pathways presented above and their correlation to effector and memory CD8 T cell susbset identities is robust by using a separate unsupervised clustering of all cells based on expression of cell cycle regulated genes. Previous lineage tracing studies indicate development of termainlly differentiated effector cells positively correlates with cell progenies that have undergone extensive proliferation. Gene expression in naïve cells and activated cells in cluster P1 indicated few were in G1/S or G2/M (clusters P0 [13%], P0.1[10%], P0.2 [16%], P1[5%]). A substantially greater fraction in cluster P2 were in cycle (P2, 33%), and the fraction of cycling cells increased to nearly 100% in all clusters along the effector cell trajectories (P4 96%, P5 98%, P7 99%). In contrast, clusters in the memory trajectory comprised much smaller fractions of cycling cells (P3, 6%). However, a much greater fraction of cells in cluster P6 were cycling (52%), consistent with their close relatedness to effector cells in P7. These results demonstrate that cycling cells positively correlate with TE cell gene expression regardless of previous developmental history, and that some cells which initially develop along a memory cell trajectory while proliferating slowly are closely related to those that increase proliferation and give rise to more effector-like cells. Thus, TE cells are derived both from TMEM cells that initially differentiate into TMEM cells, but then proliferate and induce terminal effector cell gene expression, together with cells that directly differentiate along the effector cell devleopmental pathway [Let’s estimate quantitatively the relative contribution of each lineage to TE cell development].

**Cells in both stem-like memory and effector-like pathways**

The effector genes *Prf1* and Gzmb were highly expressed in cells from P4 compared to those in P3.

MP and TMEM cells (Bcl2, , but cells in P7 have not. P5 are becoming terminally differentiated. They have

In addition, genes that are normally repressed in TMEM cells, were less highly expressed in cells cluster P7 compared to those in P5. Finally, both the core TCIRC and TRM cell signatures were positively enriched with genes expressed in cells from P7, whereas these signatures, especially those in the TCIRC signature, were downregulated in cells from cluster P5. Thus, cells in the effector-like pathway appear to diverge into cells that acquire both effector and TMEM characteristics along the P7 trajectory, and to more strongly repress TMEM gene expression in P5.

which differentially manifest gene expression of more mature effector and memory like CD8 T cell subsets.

At late times the memory-like and effector-like pathways converge, which indicates that each divergent pathway can give rise to both effector and memory CD8 T cells. Gene expression in cells from P6 was substantially enriched with gene signatures of both TMEM and TE cells. In addition, cells in P7, although more strongly enriched with the gene expression signatures of effector cells that some cells that initially differentiate along the memory-like pathway subsequently engage gene expression of TE cells.

Thus, cells in cluster P4 are an early effector-like intermediate that has diverged from an alternative early memory-like pathway comprising cells in cluster P3.

Independent unsuperivised single cell analysis of cell cycle regulated genes confirmed this hypothesis because > 90% of cells from clusters P4, P7 and P1 expressed specifically defined gene signatures of cycling cells (G1/S or G2/M), whereas conversely, 90% of cells in cluster P8, and ~ 40% of cells in cluster P2 expressed a specific gene signature indicating they were quiescent (G0). To define the origins of the branched pathways into PAGA-clusters P4 or P8, we investigated the heterogeneity among cells in cluster P9. Sub-dividing cells in P9 using UMAP followed by clustering with Louvain delinated three partions, P9.L0, P9.L1 and P9.L2. Gene expression profiles in cluster P9.L0 was strongly enriched with gene expression typical of cluster P4, whereas gene expression in cluster P9.L1 was strongly enriched for the gene expression signature of cluster P8. The gene expression profile of cluster P9.L0 was comparatively unbiased (maybe determine if P9.L0 is more similar to naïve cells or cluster P3 signature?). Thus, cluster P9 is heterogeneous and contains unbiased progenitors together with cells that appear predisposed to developing along either the P4 or P8 trajectories.

Convergence between the memory and effector like trajectories. the most partially enriched for the MP cell of effector-like paths develop from cluster P4.

Two developmental pathways diverged from cells in P9. DEG-clusters I-III which positively associate with cells undergiong proliferation were enriched along the trajectories to PAGA-clusters P4, P7 and P1 whereas, cells in the PAGA-clusters P8 and P2 inversely correlated with DEG- II and III. Together, this implies that cells in the trajectory through P4 continue proliferating, while many fewer cells in clusters P8 and P2 do so.

Il-2ra

WT and *Il2ra*-/- P14 CD8 T cells differentially contributed to specific clusters, but the overall partitioning was not distorted by co-analysis of *Il2ra*-/- P14 CD8 T cells, because nearly identical clusters were evident when only WT naïve and LCMV-activated P14 cells were analyzed separately (**Fig. S1A, dot plots and PAGA connectivies**), or when endogenous LCMV-specific CD8 T cells were analyzed (**Fig. 12B, dot plots and PAGA connectivies**), emphasizing the robustness of the approach.

##############################################################################

############################## In vivo CRF screen ##############################

##############################################################################

Prior to their first cell division, TCR-activated naïve CD8 T cells undergo global chromatin remodeling of their cis-regulatory landscape that establishes both de novo accessibility of cis-regulatory regions that are accessible in mature CD8 T cell lineages and condensation of regions normally accessible in naïve CD8 T cells.

Mammalians have an expanded repertoire of genes encoding chromatin regulatory (CR) factors compared to the orthologous genes in single-cell organisms and invertebrates, and their physiological roles during the specification of mammalian cell development in vivo are incompletely understood. We designed a pooled in vivo loss-of-function screen to comprehensively discern the requirements of virtually all genes encoding chromatin regulatory factors in CD8 T cells as they differentiate into classically-defined memory precursor effector and terminal effector CD8 T cell populations that arise at the peak of the response to acute infection with *Lymphocytic choriomeningitis virus* (LCMV). We cloned and arrayed a library of short hairpin RNAs embedded in a microRNA context (shRNAmirs) that targeted all murine genes encoding CRs with ~ 4 shRNAmirs per (1,5XX clones) in the context of an optimized, non-immunogenic, retroviral vector (**Fig. 1A** and **table S1**). Retroviral supernatants prepared from the arrayed library was used to transduce briefly activated naïve P14 TCR transgenic CD8 T cells, which express a TCR that recognizes the GP33-41 peptide of lymphocytic choriomeningitis virus (LCMV) in H2-Db. Thus, P14 cells in each well were transduced with a single shRNAmir specificity and subsequently cells from separate wells were pooled and adoptively transferred to wildtype hosts which were then infected with a low dose of LCMV Clone 13 strain (LCMVCl13)(**Fig. 1A**) [Chen et al. Immunity]. Under these conditions, systemic exposure to LCMVCl13 causes an acute infection and drives sustained accumulation and acquisition of distinct effector and memory precursor phenotypes of the transferred P14 cells (**fig. S2A**) [Chen]. Near the peak accumulation of CD8 T cells responding to various acute intracellular infections, such as LCMV, differential expression of Killer-like lectin receptor 1 (KLRG1) and interleukin-7 receptor (IL-7R) on cells discriminates phenotypic subsets that differ in their potential to form memory CTLs. KRLG1hi, CD127lo (Terminal effector, TE) cells least efficiently persist over time, whereas KRLG1lo, CD127hi (memory precursor, MP) effector cells efficiently generate central memory CTLs, which predominantly reside in lymphoid organs and exhibit self-renewing capacity. KRLG1hi, CD127hi (Double positive, DP) effector cells contain cells that are more effector-like, but many persist as memory cells. KRLG1lo, CD127lo (early effector, EE) cells are relatively uncommitted that can give rise to all four KLRG1/CD127 phenotypic subsets as the infection resolves.

In the scereening approach, the integrated retroviral shRNAmir-encoding provirus served as a barcode to quantify the representation of cells carrying individual shRNAmirs among the initial pool of donor P14 cells before transfer (input), all donor CD8 T cells in the spleen at the peak response to the infection, and in FACS-isolated CTL subsets based on staining of KLRG1 and CD127. We expected that reduced representation of shRNAmirs in one phenotypic CTL subset relative to another resulted from suppression of their cognate gene target, and reflected a requirement in that gene for development of cells manifesting the phenotypic subset wherein the shRNAmirs were underrepresented. To identify genes with *bona fide* RNAi-mediated affects, two biological replicates of the entire screen were performed, and changes in the relative representation of each shRNAmir in pairwise comparisons between distinct CTL populations were calculated and then summarized to quantify an average apparent effect of all shRNAmirs specific to each cognate gene for each replicate (**fig. S2A**). The values were converted to Z-scores and were adjusted by their associated p-values, to de-prioritize genes with RNAi-effects that were driven by distinct cognate shRNAmirs producing variable effects (**Fig. 1C** heat map of hits, and consistency scores). Notably, virtually all shRNAmirs introduced in the input were recovered in the output, demonstrating that nearly all RNAi-effects were non-lethal, and thus that most phenotypes would result from biologically meaningful reductions in gene expression, which we anticipated as an advantage over generating null alleles. Consistent with this notion, the approach identified many genes in a manner consistent with their known functions from previous studies in CD8 T cells, and in other biological settings. For example, shRNAmirs specific for factors that control cell proliferation (*Kdm2b, Rpa3* and *Runx3*) were depleted {He, 2008 #142;McJunkin, 2011 #143;Wang, 2018 #19}, or enriched (*Ing2, Ing3, Ing4, Ing5* and *Bop1*)from total output cells; those specific for factors that drive TE CTL differentiation (*Tbx21*, *Prdm1, Id2, Ccnt1*, and *Cdk9*) were depleted from that subset {Chen, 2014 #70;Joshi, 2007 #33;Kallies, 2009 #129;Rutishauser, 2009 #81;Yang, 2011 #116}, and conversely, those that promote MP CTL differentiation (*Runx3* and *Ncor1*) were depleted from MP CTL {Milner, 2018 #39;Wang, 2018 #19;Yu, 2017 #17}; and finally, those that promote overall maturation of EE cells into LE CTL subsets of TE, MP, and/or DP CTL (*Tbx21, Runx3, Cdk9,* and *Ccnt1*) were depleted from those subsets {Chen, 2014 #70;Joshi, 2007 #33;Olson, 2013 #132;Wang, 2018 #19}.

Based on this framework, we classified all CRs according to their loss-of-function phenotype during infection (**table S2**). Chromatin factors identified within each classification were enriched for experimentally verified and predicted physical or genetic interactions (**fig. S2A** to **S2D**), which suggests they likely work concertedly to promote distinct aspects of CTL differentiation, as might be expected. This included differential and potentially antagonistic functions for multiple members of well-recognized chromatin remodeling complexes including BAF, NuRD/HDAC, HBO1 (HAT complexes), PRP1 and 2, MLL/Set, as well as, transcriptional regulatory complexes such as P-TEFb and RNA Pol II associated factors, and less well-defined factors that promote covalently modify arginine residues. Collectively, these results indicated that chromatin factors known to promote nucleosome disassembly, histone hyperacetylation, histone H3K27 tri-methylation and transcriptional elongation were necessary for EE cells to differentiate into one more LE CTL subset, and were preferentially necessary for TE CTL differentiation. Conversely, factors known to promote histone deacetylation, histone H3K4 mono and tri-methylation, Ncor??, and histone readers that recognize H3K36me3 (a mark of transcriptional elongation) appeared to be necessary to restrain activated CD8+ T cells from dif Remarkably, XX genes encoding subunits of BAF complexes were defined as factors required for TE CTL differentiation, and inspection of all shRNAmirs specific to XX BAF-complex members indicated they were preferentially depleted from TE CTL.

We confirmed the phenotypes of multiple factors by analyzing their phenotypes using individual analyses (Carm1, Prmt5, Chd4, Dot1l, Chd7, Smarca4, SmarcP2, Arid1a, Mll2 (have Adam or Ugoma do this), what else). These results demonstrate that multiple chromatin remodeling machines, chromatin modifying ezymes and chromatin readers orchestrate the differential patterning of activated CD8 T cell subsets that posses distinct propensities to develop into memory CTLs.

##############################################################################

############################## Tragetory analysis ##############################

##############################################################################

**Description of results from trajectory analysis:**

We anticipated that utilization of distinct CR factors and complexes to regulate specific transcriptional outputs that alter the differentiation of activated CD8 T cells would be governed in part by their mRNA expression

##############################################################################

########################### Correlate with In vitro screen ##########################

##############################################################################

To identify CR responsible for establishing early CTL progenitor phenotype (such as P8 and P9 from Fig2), an in vitro RNAi screen was conducted to uncover the effect of CRs on some import surface markers such as Il2ra (**Fig. 3A**).

Since the intermediate cluster P9 showed reduction of Il7r expression and elevated expression of Il2ra, we focused on CRs which can be responsible for the regulation of those two surface markers. We found that 2 CRs, Smarca4 and HdaP2, might be essential for both downregulation of Il7r and upregulation of Il2ra (when deficient, Il2ra expression reduced and Il7r expression increased, both p-val <= 0.05).

CR that control expression the genes that are divergence of the cells

CD25KO nacent RNA-seq (differential genes in Cluster9, Blimp1 & Bcl6)

GSEA analysis of Cluster9 cells in Cluster 4 or cluster 8 signature.

References and Notes:

**Acknowledgments:**

**Fig. 1**. **Developmental paths comprising early memory CD8 T cell ontogeny during viral infection.**

(**A**) Experimental scheme for scRNA-seq samples.

(**B**) PAGA-initialized clusters identified using 900 randomly selected cells from each sample (A) are projected in 2-dimensions (UMAP). The number of cells contributing to each cluster in each condition are quantified.

(**C**)

by analyzing enrichement of defined sets of kinetically grouped dynamically expressed genes, and “signature” gene sets derived from mature TE, MP, and TMEM subset

(**A**) Scheme for arrayed retroviral transduction and in vivo screening. Arrayed DNA libraries encoding retroviral genomes carrying shRNAmirs were packaged into retroviral particles and used to transduce arrays of naïve P14 CD8+ T cells (Thy1.1+) after brief activation. Four hours post-transduction the cells were pooled and aliquots of 500,000 cells representing ~ 500 unique shRNAmirs were transferred i.v. to groups of wildtype naïve congenic host mice (Thy1.2+), that were subsequently infected with LCMV-cl13. On day 8 post infection, spleens from infected mice were removed, donor P14 cells were enriched by positive selection, transduced (Amt+) CTL-subsets were isolated based on KLRG-1 and CD127 staining by FACS, genomic DNA was extracted, next generation sequencing libraries were prepared by PCR of proviral sequences and the numbers of shRNAmirs in the different sorted subsets were quantified. The entire screen was performed independently twice. (**B**)

**Fig. 2. Separate stem like memory and effector cell pathways diverge from common effector memory precursors that require IL-2Ra and both give rise to memory CD8 T cells.**

**Fig. 3. Non-uniform single cell RNA velocities direct differentiation in unpredicatable trajectories within the developmental paths.**

**Fig. 4. Identification of chromatin remodeling factors that drive differentiation of effector and memory precursor subsets during LCMV infection.**

**Fig. 5. The BAF-complex cooperates with IL-2Ra signals to establish nascent CTL precursors**

**Fig. 6. Distinct BAF-complex subunits and regulation of transcriptional pausing governs divergence of stem like memory and effector cell pathways.**

**(A** to **H**)

**Fig. 3.** **Brg1-driven BAF-complexes are essential for initiating CTL differentiation**

**Fig. 4.Chd7 is essential for antiviral and anti-tumor immunity**

**Fig. 5 Chd7 integrate IL-2R and inflammatory signals to drive TE CTL differentiation**

**Fig. 6 Chd7 controls chromatin accessibility during initialization of memory CTL differentiation**

Supplementary Materials:

Materials and Methods

Figures S1-S19

Tables S1-S#

External Databases S1-S#

References (*##-##*)

Supplementary Materials:

This section includes the actual text of the Supplementary Materials, which can include any or all of the preceding items, and figure captions and tables that can easily be incorporated into one supplementary material file. Please edit the list above as appropriate and include it at the end of your main paper. If there are additional files that cannot be easily accommodates (e.g., movies or large tables), please include captions here.

**Figure S1.**

**Figure S2.**

**Figure S3.**

**Figure S4.**

**Figure S5.**

**Figure S6.**

**Figure S7.**

**Figure S8.**

**Figure S9.**

**Materials and Methods: Animals and Viruses**

P57BL/6 (B6) mice were purchased from the Jackson Laboratory company. P14 Thy1.1, dLCK-Cre and Ert2-Cre mice were bred in-house. All mice were maintained in specific-pathogen free facilities and used according to protocols approved by the Institutional Animal Care and Use Committee of TSRI-FL. LCMV-Armstrong and LCMV-Clone13 were produced from Vero-E6 and Vero cell lines respectively. LM-GP33 Listeria monocytogenes was grown as described.

**Runx3 knockout model and infection**

Mice bearing alleles in which *LoxP* sites flanked exon 4 of *Runx3*, which contains most of its DNA-binding domain [REF], were crossed to mice carrying a YFP reporter in the *Rosa26* locus in which a *LoxP*-flanked transcriptional terminator is upstream of EYFP and either dLck-*Cre* or ERT2-*Cre* transgenes, or to P14 Thy1.1+ mice. Thus, cells that experienced Cre-recombinase activity were enumerated based on YFP expression. In dLck-Cre mice, Cre expression is induced after positive selection [REF], whereas in CreERT2 mice [REF], Cre expression is constitutively localized in the cytoplasm and translocates to the nucleus upon provision of estrogen-receptor analogs.

For Runx3 constitutive knockout model, Runx3+/+, Runx3+/fl and Runx3fl/fl mice carrying both dLCK-Cre and Rosa26-YFP were set up. These mice were infected with 2×105 PFU of LCMV-Arm. For Runx3 inducible knockout model, Runx3+/+, Runx3+/fl and Runx3fl/fl mice carrying both Ert2-Cre and Rosa26-YFP were set up. These mice were first treated with 1.5mg of Tamoxifen per mouse for three successive days. After another 4 days, 2×105 PFU of LCMV-Arm was used to challenge these mice. :

**CD8 T cell isolation, activation and transduction**

Naïve CD8 T cells were isolated using negative selection methods. After activation with 1ng/ml of anti-CD3 and ant-iCD28 antibodies for 16 hours, cells were transduced with MSCV-based retrovirus carrying GFP or Ametrine reporter for 4 hours. Then, 5×105 of cells were transferred to naïve P57/B6 mice. Thus, purified naïve P14 Runx3+/+, Runx3+/fl and Runx3fl/fl CD8 T cells were stimulated via their TCRs, transduced with cDNAs encoding either *Runx3* or *Cre*-recombinase 16 hours later, and then analyzed after 6 days in culture with IL-2 [REF].

**Adoptive transfer and infection**

For Naïve CD8 T cells adoptive transfer, 1×105 of Naïve CD8 T cells isolated from P14+ Thy1.1+ mice were adoptively transferred into naïve P57/B6 mice. 2×105 PFU of LCMV-Arm was used to infect the mice one hour post cell transfer. For activated CD8 T cells adoptive transfer, 5×105 of activated CD8 T cells isolated from P14+ Thy1.1+ mice were transferred into naïve P57/B6 mice. 1.5×105 PFU of LCMV-Clone 13 was used to infect the mice one hour post cell transfer.

**Flow cytometry**

Splenocytes or blood cells were treated with RBC lysis buffer and then subject to surface staining. All of the antibodies used in this study were from Biolegend or eBioscience company. LCMV GP33, GP276 and NP396 tetramers obtained from NIH were used to label LCMV antigen reactive CD8 T cells before routine surface staining.

**Plaque assay**

Serum and splenocytes from infected mice were harvested and subjected to dounce disruption for homogenization. The extracts were serially diluted and used to infect Vero cells. Plaques were counted 4 days after infection

**LM-GP33 protection assay**

CD8 cells were isolated from Runx3+/+ or Runx3fl/fl mice carrying Rosa26-YFP. The cells were activated for two days before transduction with Cre-GFP virus to induce Runx3 wild type or knockout model. After culture with 10U/ml of IL2 for another 4 days, 25×104 of YFP+ cells were transferred to naïve P57/B6 mice. One day later, 5×104 of CFU of LM-GP33 was used to challenge the mice. Colony formation assay of spleens were performed on day 4 post-infection.

**Effector CD8 T cells sorting, qPCR and immunoblot**

Splenocytes prepared from infected mice were used to isolate transferred P14 cells by streptavidin-conjugated magnetic beads. After staining with CD127-PE and KLRG-1-APC, different effector subsets were sorted. One aliquot of the sorted cells was subjective to RNA isolation, cDNA synthesis and qPCR using Tagman probes (from Life Technology). The other aliquot was used for immunoblot analysis.

**RNA-Seq**

**Any Additional Author notes:** For example, author contributions or a list of group authors.

Extra text:

A linear differentiation model suggests naïve cells pass through an effector cell stage, and some of these cells ultimately become TMEM cells, whereas another linear model indicates naïve cells initially develop into TCM precursors, some of which progressively develop into effector memory (TEM) and ultimately TE cells. A bifurcating fate model postulates that separate effector or memory fates are established early in the response as a result of assymetric cell division.