BRIEF DEFINITIVE REPORT

Single-cell lineage mapping of a diverse virus-specific naive CD4 T cell repertoire

Achia Khatun[1,2], Moujtaba Y. Kasmani[1,2], Ryan Zander[2], David M. Schauder[1,2], Jeremy P. Snook[6], Jian Shen[1,2], Xiaopeng Wu[2], Robert Burns[2], Yi-Guang Chen[1,4,5], Chien-Wei Lin[3], Matthew A. Williams[6], and Weiguo Cui[1,2]

Tracking how individual naive T cells from a natural TCR repertoire clonally expand, differentiate, and make lineage choices in response to an infection has not previously been possible. Here, using single-cell sequencing technology to identify clones by their unique TCR sequences, we were able to trace the clonal expansion, differentiation trajectory, and lineage commitment of individual virus-specific CD4 T cells during an acute lymphocytic choriomeningitis virus (LCMV) infection. Notably, we found previously unappreciated clonal diversity and cellular heterogeneity among virus-specific helper T cells. Interestingly, although most naive CD4 T cells gave rise to multiple lineages at the clonal level, ~28% of naive cells exhibited a preferred lineage choice toward either Th1 or T_{FH} cells. Mechanistically, we found that TCR structure, in particular the CDR3 motif of the TCR α chain, skewed lineage decisions toward the T_{FH} cell fate.

Introduction

In response to an acute infection, pathogen-specific CD4 T cells differentiate into type 1 helper T cell (Th1) and T follicular helper cell (T_{FH}) subsets to orchestrate cellular and humoral defense, respectively (Marshall et al., 2011; Pepper and Jenkins, 2011; Pepper et al., 2011). A long-lasting debate (Cho et al., 2017; Lönnberg et al., 2017; Tubo et al., 2013) has been centered on whether one pathogen-specific naive T cell can give rise to both Th1 and T_{FH} cells or develop into only one lineage. Moreover, whether T cell receptor (TCR) structure, which defines a T cell clone, can itself influence this differentiation process remains unclear. Using T cell adoptive transfer, previous studies have elegantly demonstrated that one naive T cell can indeed differentiate into both Th1 and T_{FH} cells following clonal expansion (Cho et al., 2017; Tubo et al., 2013). However, due to the technical limitations, only a small number of naive CD4 T cell clones were examined in these studies, making it difficult to exclude the possibility that rare T cell clones may have a preferred lineage choice to make. Furthermore, some of these studies used TCR transgenic T cells (Cho et al., 2017; Lönnberg et al., 2017; Tubo et al., 2013) which inevitably negates the contribution of the TCR itself to T cell lineage commitment. In addition, when considering the differentiation trajectory of a single T cell clone over time, it remains controversial whether memory precursor T cells arise first and then further develop into terminal Th1 and

T_{FH} effector cells, or whether memory precursor and effector T cells bifurcate at an early stage of activation and go down distinct developmental paths thereafter (Badovinac et al., 2005; Buchholz et al., 2016; Corbin and Harty, 2004; Gett et al., 2003; Sallusto et al., 2004).

Results and discussion

In an attempt to answer these fundamental questions, we used single-cell TCR sequencing (scTCR-seq) to identify T cell clones at the peak of expansion during acute lymphocytic choriomeningitis virus (LCMV) infection by their unique paired TCR α and β chain sequences (Fig. 1 A). T cells with identical TCR α and β chains are almost certainly clones that derive from a single naive T cell ancestor, as the chances of two unrelated T cells sharing the same TCR are ~1 in 10^{15} to 1 in 10^{20} (Davis and Bjorkman, 1988; Lieber, 1991). Simultaneously, we performed single-cell RNA sequencing (scRNA-seq) to determine the phenotype of every T cell, which was uniquely linked to each cell's TCR clonotype by a barcoded cell identifier (Fig. 1 A). This technology enabled us to study the lineage choice and differentiation trajectory of thousands of cells at the clonal level from a naturally arising CD4 T cell repertoire.

To increase the specificity of our analysis, we restricted our sequencing to GP_{66-77}-specific CD4 T cells at day 10 post-infection

Figure 1. **Broad clonal diversity of naive antigen-specific CD4 T cells revealed by scTCR-seq during acute LCMV infection. (A)** General outline for this study to trace CD4 T cell clonal fate at the single-cell level during acute LCMV infection. GP_{66-77}-specific CD4 T cells were sorted from five C57BL/6 mice on day

10 following acute LCMV infection to perform the study. All the results in this study are based on the observations made in these five infected mice from three independent experiments. **(B)** Bar graphs showing frequency of the top 50 clones (two or more cells/clone) in each mouse. **(C)** Chord diagram showing TCR α and β V-J gene usage in five mice. Colors for V and J genes are consistent among the diagrams. **(D)** Upset plot showing clonal overlap based on TCR α and β chain among five mice. Clones are defined by paired TCR α and β chain CDR3 nucleotide sequences. Each color indicates one individual mouse; dots below each bar denotes which mouse or mice the clones were found in. Gray bars with more than one dot represent clones found in more than one mouse.

(p.i.) with LCMV Armstrong, which induces an acute infection (Fig. S1 A). A total of 18,000 CD4 T cells from five mice with productive TCR α and β chains were found. However, to limit any possible FACS sorting errors, only clones with at least two cells with identical paired TCR α and β chains were included in this study. By this criterion, we recovered 10,492 cells giving rise to 686 clones in five mice, with a range of 100–160 clones per mouse. Taken together, these results suggest that our single-cell sequencing approach likely recovered the majority of the GP_{66-77}-specific CD4 T cell clonal repertoire, as these values are in a similar range as previously published work (Jenkins and Moon, 2012; Kotturi et al., 2008). In addition, we found that as few as 7–10 dominant clones accounted for ∼40–50% of expanded GP_{66-77}-specific T cells per mouse (Fig. 1 B), suggesting that these could be the naive T cells recruited early during the initial priming phase (Vezys et al., 2006).

In agreement with traditional bulk TCR sequencing studies (Izraelson et al., 2018; Qu et al., 2016), we found that the five mice shared very similar V gene usage with TRAV14D-3DV8*02 or TRAV14-3*01 and TRBV31*01 or TRBV13-3*01 genes predominantly used for TCR α and β chains, respectively (Fig. S1, B and C). Likewise, we also identified similar V-J gene combinations for both chains among the mice, although a more diverse pattern was seen in TCR α chains compared with TCR β chains (Fig. 1 C; Turner et al., 2006). When each chain was separately used to define a clone, we identified 38 TCR α and 11 TCR β clones shared among at least two mice (Fig. S1, D and E; Bousso et al., 1998; Madi et al., 2014), with multiple overlapping events for both individual chains among the same group of mice. To our surprise, when we instead defined clones by paired TCR α and β CDR3 nucleotide sequences, we detected only four instances of clonal overlap among five mice, which had a total of 686 clones (0.6% overlap rate; Fig. 1 D). This demonstrates an immense TCR diversity even in genetically identical hosts, which has not been previously appreciated (Izraelson et al., 2018; Madi et al., 2017; Shifrut et al., 2013; Tong et al., 2016).

Dimensionality reduction analysis performed on single-cell gene expression data using the top 20 principal components revealed seven distinct clusters by t-distributed stochastic neighbor embedding (t-SNE) visualization (Figs. 2 A and S2 A) after regressing out cell cycle gene expression (Fig. S2 B). Only cells belonging to clones with two or more cells, as mentioned above, were included in this analysis. Each mouse showed a similar distribution of cells among the seven clusters (Fig. 2 B). Based on the expression of known lineage specific markers (Fig. 2, C and D; and Fig. S2 C) and module scores for published memory, Th1, and germinal center (GC) T_{FH}/T_{FH} CD4 T cell gene sets (Fig. 2 E; Ciucci et al., 2019), cells in these clusters were largely divided into Th1, T_{FH}, T central memory precursor (Tcmp), and regulatory T (T reg) cells, showing consistency with

previous research (Choi et al., 2011; Ciucci et al., 2019; Lee et al., 2011; Lönnberg et al., 2017; Marshall et al., 2011; Pepper et al., 2011; Tubo et al., 2013). Among these clusters, clusters 1 and 3, which comprised 35% of the GP_{66-77}-specific CD4 T cell pool, showed higher expression of Th1-specific markers *Cxcr6*, *Tbx21* (encodes T-box expressed on T cells), *Ifng*, *Gzmb*, and *Id2* compared with other clusters (Fig. 2, C and D; and Fig. S2 C). In addition, clusters 2, 4, and 5 showed higher expression of T_{FH}-specific genes *Cxcr5*, *Bcl6*, *Il21*, *Il4*, *Icos*, and *Pdcd1* (encodes PD-1; Fig. 2, C and D; and Fig. S2 C) and together comprised ∼42% of cells. Cluster 0, which accounted for ∼23% of cells, had higher expression of Tcmp-associated genes *Ccr7*, *Il7r*, *Lef1*, *Id3*, and *Slamf6* (Fig. 2, C and D; and Fig. S2 C; Blander et al., 2003; Ciucci et al., 2019; Colpitts et al., 2009; Marshall et al., 2011; Pepper and Jenkins, 2011; Pepper et al., 2011; Utzschneider et al., 2016). Lastly, cluster 6, which constituted only 0.4% of the cells, showed higher expression of T reg–specific genes, namely *Foxp3*, *Il2ra*, *Gzmb*, *Il10*, *Cd81*, and *Cd74* (Fig. 2, C and D; and Fig. S2 C) as reported in previous studies (Marshall et al., 2011; Moon et al., 2011). Intriguingly, additional heterogeneity was revealed within both the Th1 and T_{FH} compartments (Fig. 2, C–E), as has been shown in a recent study as well (Ciucci et al., 2019). Both clusters 1 and 3 showed significantly higher Th1 gene set module scores compared with other clusters (Fig. 2 E, middle). Th1-like cluster 1, here onward designated as the $Ly6c^{Hi}$-Th1 cell subset, showed higher expression of *Selplg* (encodes PSGL1) and *Ly6c2* and also displayed high expression of chemokine receptors *Cx3cr1*, *S1pr1*, and *S1pr5*, which are associated with terminal differentiation (Fig. S2 D; Garris et al., 2014; Marshall et al., 2011). Cluster 3 exhibited lower expression of *Ly6c2* and higher expression of CTL markers *Lag3*, *Pdcd1*, *Havcr2* (encodes Tim3), and *Slamf1* (Fig. S2 D), suggesting its possible similarity with a previously described CD4 CTL subset (Donnarumma et al., 2016; Jellison et al., 2005). We thus designated this cluster as the $Lag3^{Hi}$-Th1 cell subset. On the other hand, clusters 2, 4, and 5 showed significantly higher GC T_{FH} gene set module scores compared with other clusters (Fig. 2 E, right). Within the three T_{FH}-like clusters, cluster 2 showed higher expression of *Selplg* (encodes PSGL1) and other T_{FH}-specific genes (Fig. 2, C and D; and Fig. S2, C and E) but comparatively lower expression than the other two T_{FH} subsets (clusters 4 and 5). This was further validated by a GC T_{FH} gene set module score (Fig. 2 E, right), which was significantly higher in cluster 2 compared with non-T_{FH} clusters (0, 1, 3, and 6) but lower compared with the other two T_{FH} subsets (clusters 4 and 5). Thus, this cluster was designated as the precursor T_{FH} subset (Pre-T_{FH}), which may potentially down-regulate PSGL1 expression to initiate T_{FH} differentiation while being in a transitioning state (Crotty, 2014). Despite having some transcriptomic similarities with Tcmp cells, the Pre-T_{FH} cluster differs from Tcmp cells in that it has higher expression of *Pdcd1* and lower
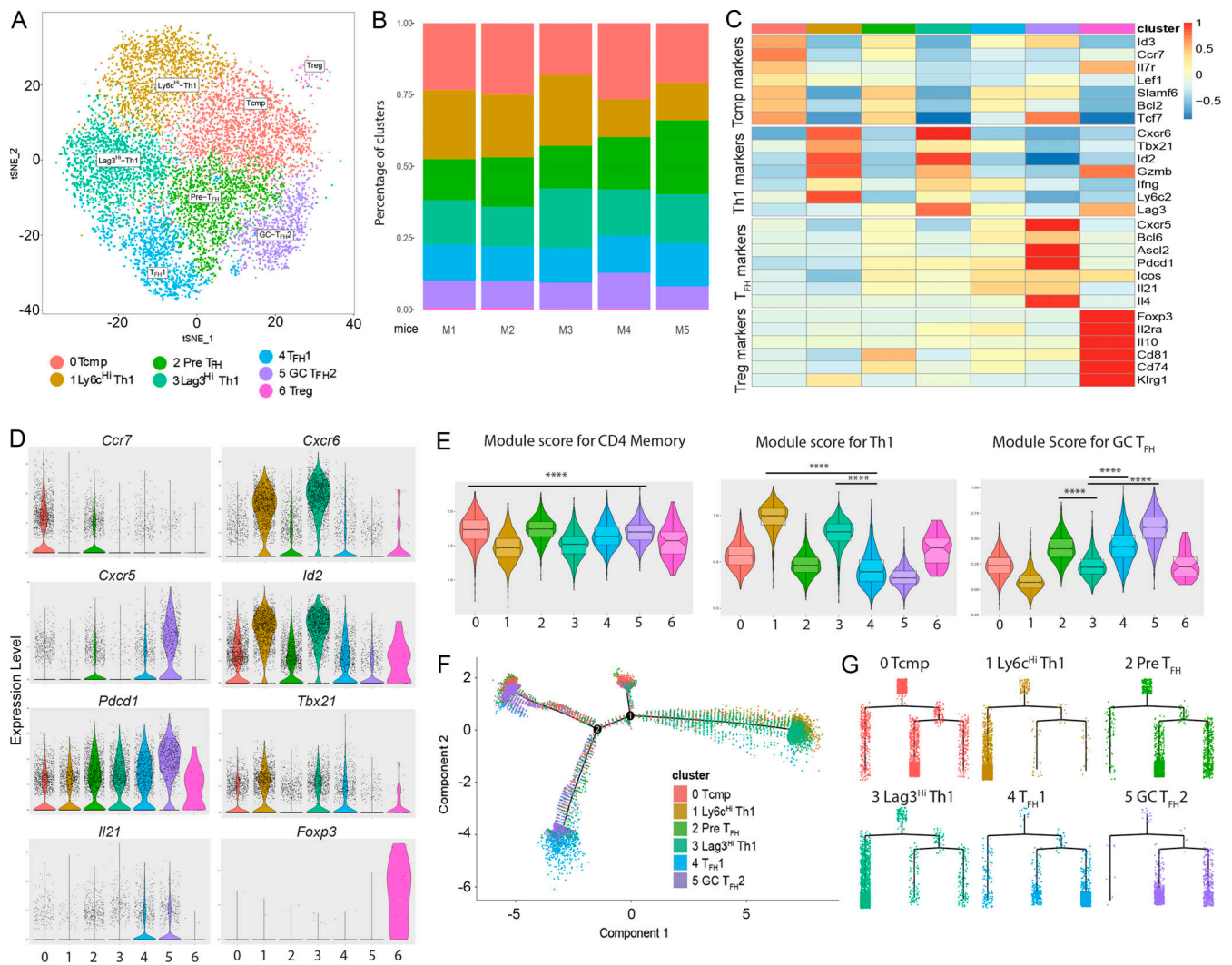
Figure 2. **Cellular heterogeneity and lineage relationship among naive polyclonal CD4 T cells responding to acute LCMV infection. (A)** t-SNE plot showing clusters identified by variable gene expression from polyclonal LCMV-specific CD4 T cells in five mice at day 10 p.i. from three independent experiments. Each dot is an individual CD4 T cell. t-SNE plot was calculated using top 20 principal components. **(B)** Bar plot showing percentage of cells in each cluster across five mice. **(C)** Heatmap showing Z scores for the average expression of CD4 T cell subset–specific genes in each cluster. **(D)** Violin plots showing expression of cell type–specific genes *Ccr7, Cxcr6, Cxcr5, Id2, Pdcd1, Tbx21, Il21,* and *Foxp3* across all clusters. **(E)** Violin plots showing module scores for memory, Th1, and GC T_{FH} CD4 T cell gene sets across all clusters in five mice; P value was calculated using the Wilcoxon test with Holm-Sidak correction. **(F)** Single-cell trajectory plot made by Monocle 2 showing lineage relationships among major CD4 T cell subsets during acute LCMV infection (cluster 6 T reg cells excluded). **(G)** Tree trajectory plot made by Monocle 2 showing transcriptional progression of different CD4 T cells cellular subsets at day 10 p.i.

expression of *Il7r* (Fig. 2 C). Along with typical T_{FH} markers (Fig. 2, C and D), cluster 4 showed higher expression of genes like *Ran* and *Cdk4*, which are associated with cell cycle progression and metabolic fitness (Fig. S2, E–G). Certain transcription factors like *Batf* and *Bhlhe40* were highly expressed in this cluster compared with the other two T_{FH} clusters (Fig. S2 E). More interestingly, cluster 4 also expressed higher levels of Th1 genes such as *Tbx21* and *Ifng* (Fig. S2 E), suggesting its possible association in providing help to IgG2a- and IgG2c-producing B cells during a type I immune response (Crotty, 2014; Iwata et al., 2017; Weinstein et al., 2018; Zhang et al., 2019). We thus designated this subset as $T_{FH}1$. Lastly, we found that cluster 5, designated here as GC $T_{FH}2$, displayed higher expression of canonical GC T_{FH} cell markers *Cxcr5, Bcl6, Il21, Il4, Icos, Ascl2,*

Tox2, Sh2d1a, and *Pdcd1* (Fig. 2, C and D; and Fig. S2, C and E; Cannons et al., 2010; Xu et al., 2019). This suggests cluster 5 may specialize in providing critical help signals to promote the differentiation of IgG1-producing B cells (Ozaki et al., 2002; Reinhardt et al., 2009).

To better understand the differentiation state of, and lineage relationships among, these CD4 T cell clusters (with the exception of T reg cells), we employed the R package Monocle 2, which uses unsupervised machine learning algorithms to predict differentiation trajectories (Qiu et al., 2017). Notably, Monocle 2 predicted a differentiation trajectory with two major branch points (Fig. 2 F). This included an overlapping differentiation state observed for clusters 1 and 3 (Ly6c^Hi-Th1 and Lag3^Hi-Th1 subsets, respectively) emerging from one branch point, but two

distinct fates for clusters 4 and 5 ($T_{FH}1$ and GC $T_{FH}2$ subsets, respectively) emerging from another point. The cells from cluster 0 (Tcmp) converged on both of these branches, while cells from cluster 2 (Pre-T_{FH}) converged more toward T_{FH} branches, respectively. This finding may suggest that Pre-T_{FH} cells hold the potential to further differentiate into either one of these two T_{FH} subsets as previously reported (Crotty, 2014). Notably, cells from cluster 0 (Tcmp) were enriched in Monocle state 2 compared with any other cluster (Table S1 and Fig. S2 H). Based on this fact and the potential of Tcmp cells to give rise to both Th1 and T_{FH} subsets, as reported before (Pepper et al., 2011), Monocle state 2 was set as the root for pseudo-temporal ordering of differentiation progression (Fig. S2 I). We then performed Monocle tree trajectory analysis. This tree trajectory analysis found cells from clusters 1 (Ly6c^{Hi}-Th1), 3 (Lag3^{Hi}-Th1), 4 ($T_{FH}1$), and 5 (GC $T_{FH}2$) to be highly concentrated at the tips of the branches with distinct lineage commitment (∼90%, ∼80%, ∼75%, and ∼65%, respectively), indicating a terminally differentiated state with reduced memory potential (Fig. 2, F and G). Conversely, cells from cluster 0 (Tcmp) displayed multiple lineage choices (either Th1 or T_{FH} subsets) with some cells concentrated at the root of the tree trajectory, suggesting enhanced memory potential (Fig. 2, F and G), as previously shown (Pepper and Jenkins, 2011; Pepper et al., 2011). Cluster 2 (Pre-T_{FH}) cells instead had a preferred lineage bias toward T_{FH} subsets. Interestingly, tree trajectory predicted early lineage specification within the Tcmp and Pre-T_{FH} clusters, although they appeared more homogenous in the prior t-SNE analysis (Fig. 2, A and G). Taken together, single-cell transcriptomics revealed a higher degree of transcriptional heterogeneity than previously appreciated and supported a bifurcated differentiation trajectory from Tcmp to either Th1 or T_{FH} effector cells at the population level (Lönnberg et al., 2017; Marshall et al., 2011; Pepper and Jenkins, 2011; Pepper et al., 2011; Sallusto et al., 2004; Tubo et al., 2013).

Merging single-cell clonal and transcriptomic datasets together empowered us to track cellular lineage commitment and differentiation trajectory of 673 individual clones from the GP$_{66-77}$-specific CD4 T cell repertoire (Table S2). We excluded the T reg cell cluster from this analysis, as only 14 clones had membership in both the T reg cell cluster and another cluster (one or two T reg cells per clone) and there were only 13 clones found exclusively in the T reg cell cluster (Table S2). By examining some of the dominant clones (≥10 cells/clone; Table S2), we first identified two obvious patterns: one naive cell giving rise to multiple lineages (Fig. 3 A) and one naive cell preferentially giving rise to one lineage, either Th1 or T_{FH} (Fig. 3 B). Interestingly, a number of clones, ranging between 5 and 30, showed cellular enrichment in terminally differentiated cell subsets, in particular the Ly6c2^{Hi}-Th1 and Lag3^{Hi}-Th1 subsets, suggesting a possible clonal burst associated with lineage specification (Table S3). Intriguingly, although 27 clones exclusively contained Tcmp cells (Table S3), Monocle analysis revealed their differentiation progression toward a distinct lineage choice. For this reason and to understand the dynamic lineage relationship among CD4 T cell clones, we decided to use differentiation states predicted by Monocle (Fig. S2 H) to more precisely identify the lineage choices (Th1, T_{FH}, or Tcmp) of the 673 clones (Table S2).

This demonstrated differentiation progression of clones toward the Th1 and T_{FH} lineages. We then used a linage plasticity index (LPI) that quantitively scored the diversified lineage choice for each lineage by evaluating all clones with constituent cells in that particular lineage (Zhang et al., 2018; see Materials and methods). In this analysis, the Tcmp lineage showed significantly higher plasticity than both the Th1 and T_{FH} lineages (Fig. 3 C). This suggests that cells in the Tcmp cluster possess a high propensity to develop into other cellular lineages (Pepper and Jenkins, 2011; Pepper et al., 2011), whereas terminally differentiated Th1 and T_{FH} cells gradually lose their lineage flexibility and memory-forming potential, further supporting the findings shown in Monocle tree trajectory analysis (Fig. 2 G). To validate this at the clonal level, we checked the cellular distribution for each of the 673 clones (two or more cells/clone) in our data among the three different lineages, as determined by Monocle states (Table S2; see Materials and methods for details). Consistent with our LPI score analysis, a large proportion (∼44%) of CD4 T cell clones across the five mice had cells shared among the Tcmp and Th1 and/or T_{FH} lineages, based on lineage classification of clones by Monocle states (Fig. 3 D, Fig. S2 H, and Table S2). Surprisingly, 56% of the clones had cells that were shared between the Th1 and T_{FH} lineages or were exclusive to one lineage (Th1 or T_{FH}), and thus did not include Tcmp cells at all. This may suggest the ability of a large number of CD4 T cell clones to bypass the Tcmp stage of differentiation at the clonal level following priming during acute LCMV infection (Fig. 3 D). To rule out the possibility that small clones may preferentially be restricted to a single cell fate and larger clones may preferentially have multiple fates, cellular fates across all clones with at least four cells per clone (total of 352 clones) were analyzed among the five mice (Table S4). A single-fate clone was defined more rigidly as having at least 65% of its cells belonging to a specific cellular fate (either the Th1, T_{FH}, or Tcmp Monocle state), with all clones not meeting this criterion designated multifate clones. A relatively even distribution of single versus multiple cellular fates was observed across all clones, regardless of their size (Fig. 3 E). A similar pattern was seen even when the 65% threshold classifying clones as single fate or multifate was changed to 55% or 75% (Fig. S2, J and K). This further suggested that some naive cells made early cell fate decisions following their clonal proliferation and bypassed the Tcmp stage to become terminally differentiated effector cells, as suggested before (Fig. 3 D). To validate the 65% threshold used to classify 352 clones (four or more cells each) as single fate or multifate, Th1 and T_{FH} gene set module scores were calculated for each clone (Table S4). Using logistic regression, a significant correlation was observed between high Th1 or T_{FH} clonal module score and Th1 (Fig. 3 F, left) or T_{FH} (Fig. 3 F, right) clonal fate, respectively. Because the sets of Th1 and T_{FH} module scores have different means and ranges, direct comparisons are not valid. Instead, we calculated the percentile rank of each cell compared with all other cells for Th1 and T_{FH} module scores separately (Table S4). Clonal percentile module scores were calculated by averaging those of all cells within that clone. Categorization of each clone's fate bias using percentile module scores had a high degree of agreement with our previous assignment based on the 65% threshold.
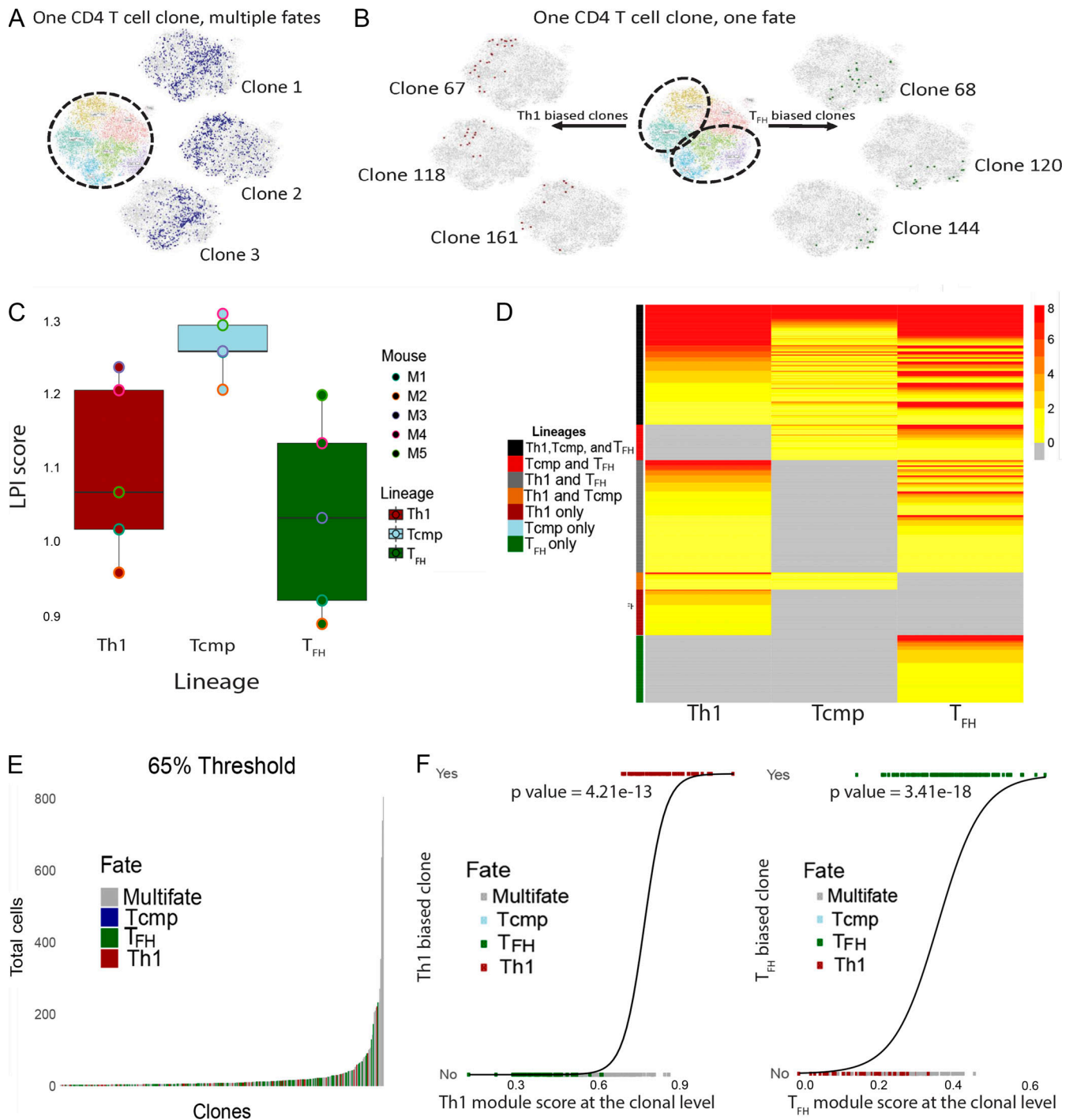
Figure 3. **Both single and multiple cellular fates are apparent at the single CD4 T cell clonal level. (A)** t-SNE plot showing T cells with multiple cellular fates. A few representative clones out of total 673 clones (two or more cells/clone) in five mice from three independent experiments are shown. Each dot represents one T cell. **(B)** Similar to A but showing CD4 T cell clones biased to a single cell fate, either Th1 (left) or T_{FH} (right). **(C)** Box plot showing LPI for 673 clones (two or more cells/clone, excluding T reg cells) from five mice across three different cellular lineages (Th1, Tcmp, and T_{FH}) defined by Monocle state IDs. **(D)** Heatmap showing distribution and frequency of 673 clones (two or more cells/clone, excluding T reg cells) among indicated CD4 T cell lineages (Th1, Tcmp, and T_{FH}) defined by Monocle state IDs in five mice. **(E)** Bar graph showing distribution of single cell fate (Th1, T_{FH}, or Tcmp biased) and multifate clones by clone size (four or more cells/clone; n = 352 clones) in five mice. A 65% threshold was used to define single cell fate clones (see Materials and methods). Colors denote clonal fate bias. **(F)** Logistic regression analysis comparing Th1 (left) or T_{FH} (right) single cell fate–biased clones versus subset–specific (Th1, left; T_{FH}, right) module scores calculated at the clonal level. P values were calculated using Wald test.

The unexpected finding that some CD4 T cell clones preferentially developed into one particular lineage (either Th1 or T_{FH}) led us to consider the possible contribution of TCR structure to cell fate decisions, as the TCR is the major contact point for recognizing peptide MHC complexes (Garcia and Adams, 2005; Rudolph and Wilson, 2002; Turner et al., 2006). Although CDR3 length has been shown to play an important role in some cell fate decisions (Lin and Welsh, 1998; Miconnet et al., 2011; Sourdive et al., 1998), we did not find any significant association between lineage commitment and CDR3 length of either the TCR α or β chain (data not shown). Next, we examined possible contributions of V-J gene usage to cell fate commitment. No distinguishable features were identified in the TCR β chain, with TRBV13-3 and TRBV13-2 being used by both Th1 and T_{FH} cells and pairing very similarly with a set of J genes (TRBJ1-2 to TRBJ1-5, TRBJ2-7, and TRBJ2-5; Fig. 4 A, left and right, respectively), as previously identified by bulk TCR sequencing in LCMV-specific CD4 T cells (Kim et al., 2013). Conversely, three major TCR α V-J gene combinations (TRAV14D-3-DV8–TRAJ21, TRAV14N-3–TRAJ15*01, and TRAV3-1*01–TRAJ9*02) were preferentially used by Th1 cells (Fig. 4 A, left). In a similar way, five TCR α V-J gene combinations (TRAV14-3–TRAJ32, TRAV14D-3-DV8–TRAJ37, TRAV14D-3-DV8–TRAJ26, TRAV14-3*01–TRAJ26*01, and TRAV14-3*01–TRAJ12*01) were preferentially used by T_{FH} cells (Fig. 4 A, right), suggesting a possible role of the TCR α chain structure in lineage commitment.

Given this apparent bias in TCR α chain gene usage, we focused our analysis on CDR3 amino acid motifs of the TCR α chain in lineage-specific clones to investigate how this could influence cellular fate within one clone. All clones with four or more cells from three randomly chosen mice (total of 239 clones) were used as a training set, and all clones with four or more cells from the other two mice (total of 122 clones) were used as a test set. Using sensitivity analysis tests across all clones, different thresholds (55%, 65%, and 75%) were used to define single-fate clones as before (either Th1 or T_{FH}; see Materials and methods for details). Based on this test, the 55% and 65% thresholds were found to provide similar results when defining lineage-specific (Th1 or T_{FH}) motifs. Analysis was continued by selecting 65% as the threshold to define single-fate clones (either Th1 or T_{FH}) and was validated by Th1 and GC T_{FH} gene set module scores at the clonal level, as mentioned before (Fig. 3 F). Next, from all Th1 and T_{FH} subset-specific clones (44 and 76 clones, respectively) in training mice, 11 Th1-specific and 16 T_{FH}-specific motifs were identified via position weight matrices (PWMs; Fig. S3 A; Bailey et al., 2009). 2 out of 11 motifs (Motifs 6 and 7) showed significant Th1 bias in training mice (Fig. 4 B; Fig. S3, B and C; and Table S5). Similarly, 5 out of 16 T_{FH}-specific motifs (Motifs 1, 2, 5, 10, and 16) showed significant T_{FH} bias in training mice (Fig. 4 B; Fig. S3, D–H; and Table S5). We then tried to use these validated TCR α CDR3 motifs to predict cell fate bias by scanning across clones in test mice (total of 122 clones with four or more cells/clone). The threshold score was set at the cellular level for each motif in training mice and kept consistent in test mice during the analysis. We found that neither of the Th1-specific motifs (Motifs 6 and 7) could predict Th1 clonal bias in test mice (Fig. 4 C and Table S5). While looking at T_{FH}-specific motifs, Motif 2 was not found to be significant, Motif 10 showed significant preference toward Th1 rather than T_{FH} bias in test mice, and analysis for Motif 16 in test mice was out of scope as the threshold score could not be reached (Fig. S3, F–H; and Table S5). However, Motifs 1 and 5 were able to predict a significant T_{FH} bias in test mice, with a total of 10 clones and 11 clones derived from training and test mice, respectively (Fig. 4 D and Table S5). Together, these results suggest that some TCR structures, an inherent property of naive T cells, may skew CD4 T cell fate decisions toward the T_{FH} lineage.

Using recently developed scRNA- and scTCR-seq technologies and bioinformatic analyses, we uncovered unprecedented clonal diversity and cellular heterogeneity of nontransgenic LCMV GP66 tetramer-specific naive CD4 T cells in an acute viral infection model. Although we could not entirely rule out the possibility of contamination with antigen nonspecific or virtual memory CD4 T cells (Marusina et al., 2017) during FACS sorting, we minimized contamination by restricting our analyses to CD4 T cell clones with at least two cells. Based on our analysis, although most CD4 T cell clones can differentiate into multiple lineages as previously described (Lönnberg et al., 2017; Tubo et al., 2013), we found that some clones may adopt a preferred lineage choice. More intriguingly, we found that TCR structure may skew this decision toward the T_{FH} lineage. Theoretically, a few lineage-specific motifs identified in this study could exhibit altered affinity and/or dwell time for their cognate GP_{66-77}-I-A^b peptide MHC complex, which may in turn affect TCR signaling and cell fate decisions (Tubo et al., 2013). Detailed delineation of these possibilities will be needed in future studies using TCR retrogenic mice (Holst et al., 2006; Kong et al., 2019). However, TCR structure is only one of many factors impacting CD4 T cell fate determination, which is also heavily influenced by well-established external factors during infection (Choi et al., 2011; Pepper et al., 2011). In addition, our study further reveals that many T cell clones do not necessarily follow a bifurcated differentiation trajectory from memory precursor cells to effector cells (either Th1 or T_{FH}), but rather bypass memory cell fate at an early stage and progress to terminal differentiation. Collectively, our new findings from this work could provide insight and frameworks for vaccine designs that can generate more tailored cellular or humoral immunity against infection and cancers.
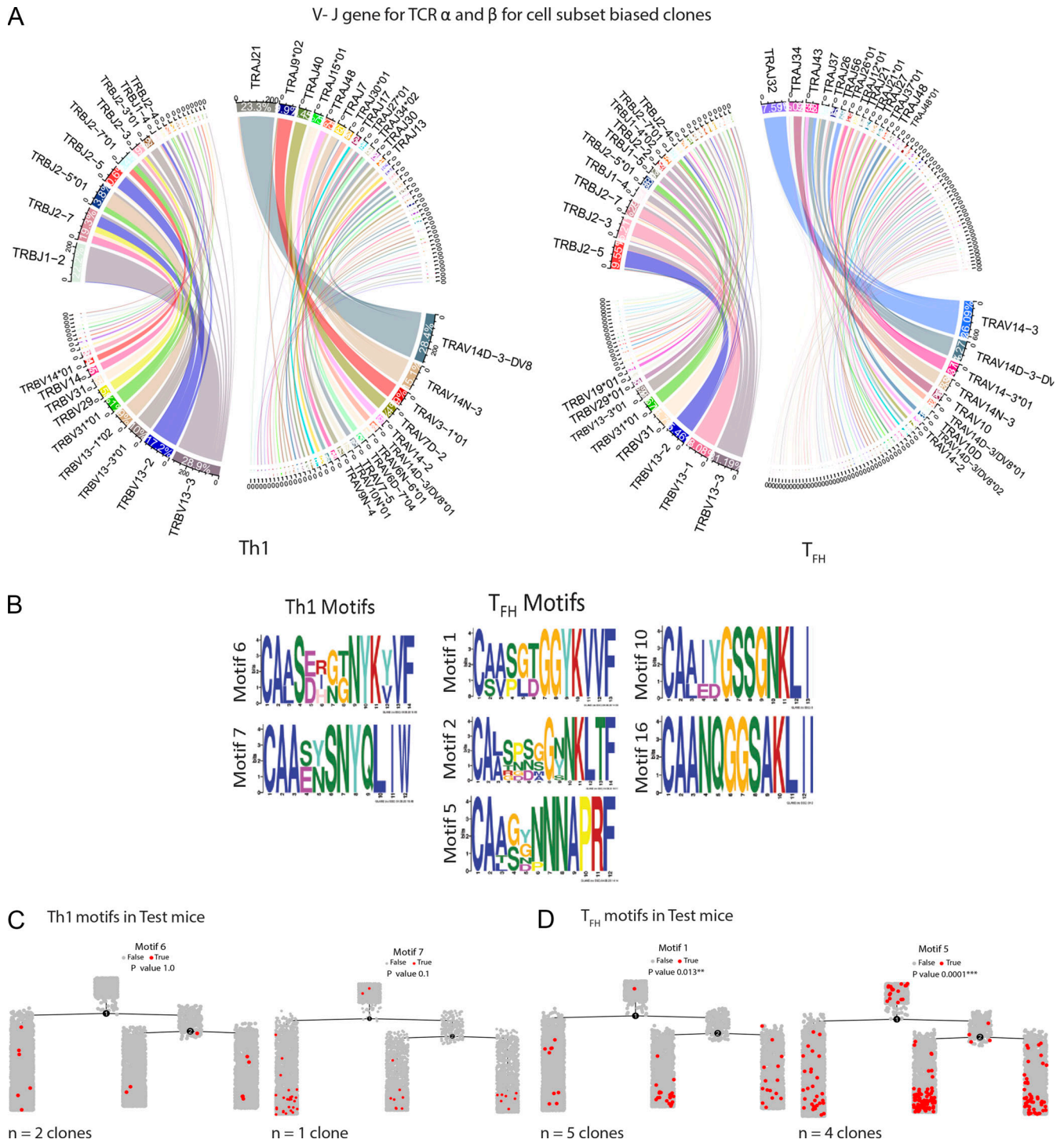
Materials and methods

Mice and LCMV Armstrong infection
Five C57BL/6 mice, 6–8 wk old, were obtained through the National Cancer Institute grantees program (Frederick, MD). Mice were bred and maintained in a closed breeding facility, and mouse handling conformed to the requirements of the Institutional Animal Care and Use Committee guidelines of the Medical College of Wisconsin. The mice were infected with 2×10^5 PFUs of LCMV Armstrong strain by intraperitoneal injection to establish acute infection in three independent experiments.

GP66 tetramer staining
CD4 T cells from the infected mice was isolated using the EasySep mouse CD4 T cell isolation kit (STEMCELL; Cat#19852).

Khatun et al.
Clonal fate mapping of CD4 T cells

Journal of Experimental Medicine 7 of 13
https://doi.org/10.1084/jem.20200650

Figure 4. **Role of TCR α CDR3 motifs in predicting CD4 T cell fates following acute LCMV infection. (A)** Chord diagram showing TCR α and β chain V-J gene combinations in Th1 (clusters 1 and 3; left) and T_{FH} (clusters 2, 4, and 5; right) subset–biased clones defined by Monocle state IDs and 65% cellular threshold in five mice from three independent experiments. **(B)** Dominant and significant TCR α CDR3 motifs in Th1- (left) and T_{FH}-biased clones (right; four or more cells/clone), each represented as a PWM and defined by using a 65% threshold to define single cell fate–biased clones in three training mice. **(C)** Monocle tree trajectory plots showing cellular fates of polyclonal CD4 T cell clones (four or more cells/clone) in two test mice, bearing Th1 subset–specific TCR α CDR3 motifs, Motif 6 (left), and Motif 7 (right) in acute LCMV infection at day 10 p.i. **(D)** Similar to C but showing T_{FH} subset–specific TCR α CDR3 motifs, Motif 1 (left) and Motif 5 (right) in two test mice. Each red dot indicates a cell and *n* indicates the number of clones found positive for that respective motif. For C and D, P values indicate the significance of each motif to define clonal bias toward the Th1 or T_{FH} fate. P values were calculated using Fisher's exact test.

Khatun et al.
Clonal fate mapping of CD4 T cells

Journal of Experimental Medicine   8 of 13
https://doi.org/10.1084/jem.20200650

Table 1. **Primers used for PCR-based cDNA amplification with TCR-specific primers**

Primer name	Sequence (5′–3′)
Common forward primer	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC
Reverse primers	
TCR α outer primer	TGAAGATATCTTGGCAGGTGA
TCR α inner primer with R2 adduct	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTCCTGAGACCGAGGATCT
TCR β outer primer	GCGGAAGTGGTTTCGAGGAT
TCR β inner primer with R2 adduct	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGACCTTGGGTGGAGTCACA
Common reverse primer with 10× SCS sample ID	CAAGCAGAAGACGGCATACGAGATNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

After that, enriched CD4 T cells were stained using either LCMV-specific GP$_{66-77}$ PE tetramer or human CLIP$_{103-117}$ PE tetramer (negative control; stock conc 1.3 mg/ml; from National Institutes of Health) along with CD4 APC (Biolegend; Cat#100411) and CD44 Pacific Blue (Biolegend; Cat#103019) in FACS buffer. The staining was performed in the dark for 1 h at room temperature, followed by three washes with FACS buffer.

scTCR-seq
scTCR-seq was adapted from the 10× Genomics Single Cell V(D)J Reagent kit, which was available for humans and has been optimized for mice in our system. As per the protocol, an aliquot of the amplified cDNA (used for scRNA-seq) was used for TCR α and β gene enrichment separately by doing three rounds of semi-nested PCR amplification.

A list of primers used for PCR-based cDNA amplification with TCR-specific primers is given in Table 1.

In the first round of PCR, an aliquot of amplified barcoded cDNA (used for scRNA-seq as well) was used for further TCR α and β gene enrichment using a common forward primer and gene-specific reverse primers. In the next two rounds of enrichment, similar to nested PCR, purified PCR products from a previous round of PCR were used as a template, using primers in a similar way as round 1 PCR. The final purified samples for both TCR α and β chains were quantified using the Kapa Library Quantification Kit. The final pooled libraries from all mice were sequenced using the MiSeq Reagent Kit v3 (600-cycle; MS-102-3003; Illumina) with the following cycle distributions: 300 cycles for read 1, 300 cycles for read 2, and 8 cycles for the i7 index read. The raw sequencing data were downloaded using the Python Run Downloader (Illumina) and then demultiplexed and converted to files containing information about V(D)J clonotypes, consensus sequences, and contigs using the Cellranger count and Cellranger VDJ functions (version 2.1). In total, 19,224 cells with V(D)J gene information for either the TCR α chain, β chain, or both chains were recovered from five mice. Further analysis was performed in R (version 3.6). For some cells, multiple α and β chain sequences were recovered from the 10× Cell Ranger VDJ pipeline. For these cells, only those chains that were annotated as full-length and functional were retained for analysis. In cases where a cell contained multiple full-length functional sequences for a given chain, the sequence supported by the most unique molecular identifiers (UMIs) was retained for analysis. To avoid any contamination from sorting, only clones having at least two cells from a single mouse were considered for further analysis.

scRNA-seq
LCMV-specific (GP$_{66-77}$ tetramer$^+$) activated CD4 T cells were harvested from the spleens of five LCMV Armstrong–infected mice on day 10 after infection and were FACS sorted using a BD-FACS-Melody Sorter (BD Biosciences). Sorted cells were loaded onto the 10× Chromium Controller with a target cell number of 5,000–10,000 per mouse. scRNA-seq libraries were prepared using the Chromium Single Cell 5′ v2 and v3 Reagent Kit (10× Genomics) according to the manufacturer's protocol. Five libraries were then quantified using the Kapa Library Quantification Kit and then were loaded onto an Illumina NextSeq 500 sequencer with the NextSeq 500/550 High Output Kit v2 (150 cycles; FC-404-2002; Illumina) with the following conditions: 26 cycles for read 1, 98 cycles for read 2, and 8 cycles for the i7 index read. Raw sequencing data were downloaded using the Python Run Downloader (Illumina) and then demultiplexed and converted to gene-barcode matrices using the Cellranger (version 2.1) mkfastq and count functions, respectively (10× Genomics). A total of ~18,000 cells were recovered from five mice. The downstream analysis was performed in R (version 3.6.0) using the package Seurat (version 2.3.3; Satija et al., 2015) and Monocle 2 (Qiu et al., 2017). The number of genes detected per cell, number of UMIs, and percentage of mitochondrial genes were plotted and outliers were removed (number of genes >3,000, number of UMIs over ~20,000, or percentage of mitochondrial genes >5%) to filter out doublets and dead cells. Cell cycle score was calculated for all cells and regressed out.

Combined analysis of scRNA-seq and scTCR-seq data
Using R (version 3.6), both scRNA-seq data (Seurat expression data) and scTCR-seq (TCR sequence) data were combined based on shared 10× cell barcodes. Only those cells with productive TCR sequences (both α and β chain) were retained for analysis. Further analysis of cellular heterogeneity at the single-cell level and clonal level were performed with cells belonging to clones with at least two cells. Canonical correlation analysis of Seurat was performed and cells were then clustered based on highly variable gene expression, using shared nearest neighbor clustering with an input of the top 20 principal components and visualized using t-SNE. All cells and genes remaining in the final Seurat analysis, except cells from the T reg cluster, were passed

into the Monocle 2 analysis using log-normalized UMI counts as gene expression values. The Monocle 2 trajectory plot was calculated based on the top 150 statistically significant, differentially expressed genes within each cluster with an average log fold change (logFC) of at least 0.25. The effects of the sample identity and the number of genes expressed per cell were also regressed from the data before calculating the trajectory. The cells were ordered in pseudotime by assigning the root state to the Monocle state 2, which had the highest proportion of Tcmp cells (Seurat cluster 0) due to its higher memory potential (Pepper and Jenkins, 2011; Pepper et al., 2011).

Gene set enrichment analysis (GSEA)

A preranked analysis module in GSEA (Subramanian et al., 2005) was used and the gene sets were found in MSigDB (Liberzon et al., 2011). For single cluster enrichment analysis, differentially expressed genes were identified first (logFC >0.1) and the average logFC expression of each gene was calculated. Then, this average gene expression data for each cluster was used as an input for GSEA analysis. After this, a gene set variation analysis score (Hänzelmann et al., 2013) was calculated using log normalized expression data for each cluster and chosen gene sets, as identified by GSEA analysis. Seurat clusters were scored by GSEA gene sets using the Addmodulescore function. Gene sets for Th1, GC T_{FH}, and memory CD4 T cells were obtained from published data (Ciucci et al., 2019).

LPI score analysis

To understand cellular plasticity in a clone-specific manner, LPI was used. Clones with cells evenly distributed among different cellular lineages will have a higher plasticity potential compared with clones having cells with restricted cellular lineages or having an uneven distribution of cells among different lineages. To perform this analysis, cellular lineages were identified using Monocle states (Th1, Tcmp, or T_{FH}). This was done because some cells belonging to the Tcmp cluster on the t-SNE plot have already made their cellular fate decision in terms of their transcriptional progression, which allows them to be segregated into their final lineages by Monocle analysis. Following this, LPI scores were calculated at the lineage level (V) to define the plasticity of each cellular lineage. The input of LPI for each lineage was the cellular lineage distribution of all clones with constituent cells in that lineage:

$$V_{lin} = \sum_{x=1}^{X} p_{Lin}^{x} Z_x.$$

The lineage-level LPI score, V, for a particular lineage, Lin, with X total clones was calculated by multiplying the LPI Z score of each clone x by the ratio of the number of cells with clonotype x in lineage Lin to the total number of cells in lineage Lin (p_{Lin}^{x}; Zhang et al., 2018). LPI score is fundamentally defined by Shannon entropy, so higher values indicate higher plasticity to differentiate into multiple cellular lineages at the lineage level.

Cellular distribution at the clonal level for three lineages

The cellular distribution of all clones (two or more cells/clone; $n = 673$ clones; Table S2) among the three different lineages defined by Monocle states (Th1, Tcmp, and T_{FH}; described above)

was shown using the pheatmap R package (Fig. 3 D). This allowed for visualization of cellular distributions of each clone among the lineages, thus demonstrating the differentiation preference of each clone (i.e., single fate or multifate).

Single versus multiple cellular fate determination at the clonal level

To show the clone size distribution (Fig. 3 E and Fig. S2, J and K) in relationship to their defined single fate (Th1/T_{FH}/Tcmp) versus multiple cellular fates, a total of 352 clones with four or more cells (Table S4) from five mice were used. The data frame used for this analysis is a subset of the data frame used for Fig. 3, A–D, by filtering clones with four or more cells. After that, single-fated clones (Th1, T_{FH}, or Tcmp) were defined based on sensitivity analysis using different thresholds: ≥55%, 65%, or 75% of constituent cells being present in a single cellular fate. Multifated clones were defined as those which did not meet the criterion to be considered a single-fate clone. This was validated by performing a logistic regression analysis in R comparing clone fate versus subset-specific module scores (Th1 or T_{FH}) at the clonal level, including all 352 clones from all five mice (Fig. 3 F and Table S4). The range of distribution of Th1 and T_{FH} gene set module scores are different. This makes direct comparison of these scores for the same clone infeasible. To enable this comparison, we instead calculated the percentile ranks of Th1 and T_{FH} gene set module scores for each cell. Each of these scores were then calculated at the clonal level, averaging the scores of all cells belonging to that particular clone (Table S4).

TCR CDR3 amino acid (AA) motif analysis

For the motif analysis, only TCR α chain CDR3 motifs from defined single-fate (either Th1 or T_{FH}) clones (four or more cells/clone) as mentioned above were considered. For this analysis, three out of five mice (M2, M4, and M5) with a total of 230 clones and 6,405 cells were used as a training set to define Th1 or T_{FH} subset–specific motifs. Only the motifs that were found to be significant in training mice were used to predict the biased cellular fates of clones in test mice. The remaining two mice (M1 and M3) were used as the test mice (122 clones with 3,295 cells). Mice were randomly assigned to be training or test mice.

First, a FASTA file containing barcodes for all cells at the clonal level and each clone's TCR α chain CDR3 AA sequence was created, separately, for both training and test mice, using the "Biostrings" R package. Only functional TCR sequences were included for this analysis. Next, the glam2 function of Multiple EM for Motif Elicitation was used to find dominant CDR3 AA motifs as position weight matrices (PWMs) out of all motifs (Bailey et al., 2009). Only Th1 and T_{FH} cell subset–specific clones (four or more cells/clone) were used. These were defined as single-fate clones using different thresholds (55%, 65%, and 75%), as mentioned above. The analysis was performed individually with each of these cellular thresholds, and top-scored motifs for each of the cell subsets were further used to make a PWM using the same function. Following this, the glam2scan function was used to scan all these PWMs individually, across all cellular TCR sequences (including all clones in training mouse with 230 clones having four or more cells) using the FASTA file

for TCR α chain, mentioned before. In longer TCR chain sequences with multiple motif matches, the best score was used for downstream analysis. These scores were then added to all clonal cells (clones with four or more cells) in a Monocle 2 object, and then a best match threshold score was determined by examining a score distribution and setting thresholding under the upper mode for each motif. Finally, the cellular fates of clones in the Monocle 2 object having higher threshold scores for that motif were plotted using the complex tree trajectory function in Monocle 2. This analysis also provided a table detailing the number of cells from a clone in each Monocle state, with a true or false positive match for that motif, which was used for statistical analysis (Fisher's exact test) to understand the ability of the motif to predict cellular fate (either Th1 or T_{FH}) of a clone. A Th1 or T_{FH} subset–specific motif was considered to be significant if that particular motif was found in at least two out of the three training mice. Only significant motifs were used to scan across all clones in testing mice (total of 122 clones with four or more cells) in the same way as mentioned before. The threshold score for each motif was kept at the same level as in training mice. Single-fate clones (either Th1 or T_{FH} biased) were further validated using the percentile rank of their Th1 and T_{FH} gene set module scores at the clonal level. Percentile ranks were used instead of raw module scores because each module score has different ranges, thus making direct comparisons between raw module scores invalid.

The 55% and 65% thresholds for defining single-fate clones produced similar significant motifs in the training mice. However, when the threshold was raised to 75%, the analysis was limited by the small number of single-fate clones in training mice as shown in Fig. S2 K. Due to this limitation, the motif analysis was performed using the 65% threshold to define single-fate clones followed by validation using subset-specific gene set module scores. A total of 44 Th1-fated and 76 T_{FH}-fated clones in training mice were included in this study. All clones found to be significant for either Th1- or T_{FH}-specific motifs in training or test mice were validated by the percentile rank of their Th1 and T_{FH} gene set module scores at the clone level (Table S4) compared with all other clones (Table S5).

Data availability
The scRNA-seq and scTCR-seq data have been deposited in the GEO database (accession no. GSE158896). All other R code and analyses are available from the corresponding author upon request.

Statistical analysis
P values for violin plots showing gene set–specific module scores across clusters were calculated by the Wilcoxon test with Holm-Sidak correction. For logistic regression analysis, P values were calculated using the Wald test. P values for TCR CDR3 AA motif analysis were calculated using Fisher's exact test.

Online supplemental material
Fig. S1 shows FACS sorting strategy, variable gene usage, and clonal overlap based on single TCR chains for five mice. Fig. S2 demonstrates intercluster heterogeneity based on differential gene expression among CD4 T cell clusters, along with the size distribution of cell fate–biased clones. Fig. S3 shows Th1 and T_{FH}

subset–specific TCR α CDR3 motif analysis in training and test mice. Table S1 shows cellular distributions across different clusters based on their Monocle states. Table S2 lists cellular distributions and percentages of all clones across different clusters. Table S3 shows cellular distributions and percentages of cluster-specific clones. Table S4 shows cellular distributions and percentages of clones with least four cells across five mice along with their respective subset-specific module scores and percentile module scores. Table S5 lists cellular distributions based on Monocle states for Th1 and T_{FH} motif–specific clones in training and test mice.

References
Badovinac, V.P., K.A. Messingham, A. Jabbari, J.S. Haring, and J.T. Harty. 2005. Accelerated CD8+ T-cell memory and prime-boost response after dendritic-cell vaccination. *Nat. Med.* 11:748–756. https://doi.org/10.1038/nm1257

Bailey, T.L., M. Boden, F.A. Buske, M. Frith, C.E. Grant, L. Clementi, J. Ren, W.W. Li, and W.S. Noble. 2009. MEME SUITE: tools for motif discovery

and searching. *Nucleic Acids Res.* 37(Web Server):W202–W208. https://doi.org/10.1093/nar/gkp335

Blander, J.M., D.B. Sant'Angelo, D. Metz, S.W. Kim, R.A. Flavell, K. Bottomly, and C.A. Janeway Jr. 2003. A pool of central memory-like CD4 T cells contains effector memory precursors. *J. Immunol.* 170:2940–2948. https://doi.org/10.4049/jimmunol.170.6.2940

Bousso, P., A. Casrouge, J.D. Altman, M. Haury, J. Kanellopoulos, J.-P. Abastado, and P. Kourilsky. 1998. Individual variations in the murine T cell response to a specific peptide reflect variability in naive repertoires. *Immunity.* 9:169–178. https://doi.org/10.1016/S1074-7613(00)80599-3

Buchholz, V.R., T.N. Schumacher, and D.H. Busch. 2016. T Cell Fate at the Single-Cell Level. *Annu. Rev. Immunol.* 34:65–92. https://doi.org/10.1146/annurev-immunol-032414-112014

Cannons, J.L., H. Qi, K.T. Lu, M. Dutta, J. Gomez-Rodriguez, J. Cheng, E.K. Wakeland, R.N. Germain, and P.L.J.I. Schwartzberg. 2010. Optimal germinal center responses require a multistage T cell:B cell adhesion process involving integrins, SLAM-associated protein, and CD84. *Immunity.* 32:253–265. https://doi.org/10.1016/j.immuni.2010.01.010

Cho, Y.L., M. Flossdorf, L. Kretschmer, T. Höfer, D.H. Busch, and V.R. Buchholz. 2017. TCR Signal Quality Modulates Fate Decisions of Single CD4+ T Cells in a Probabilistic Manner. *Cell Rep.* 20:806–818. https://doi.org/10.1016/j.celrep.2017.07.005

Choi, Y.S., R. Kageyama, D. Eto, T.C. Escobar, R.J. Johnston, L. Monticelli, C. Lao, and S. Crotty. 2011. ICOS receptor instructs T follicular helper cell versus effector cell differentiation via induction of the transcriptional repressor Bcl6. *Immunity.* 34:932–946. https://doi.org/10.1016/j.immuni.2011.03.023

Ciucci, T., M.S. Vacchio, Y. Gao, F. Tomassoni Ardori, J. Candia, M. Mehta, Y. Zhao, B. Tran, M. Pepper, L. Tessarollo, et al. 2019. The Emergence and Functional Fitness of Memory CD4+ T Cells Require the Transcription Factor Thpok. *Immunity.* 50:91–105.e4. https://doi.org/10.1016/j.immuni.2018.12.019

Colpitts, S.L., N.M. Dalton, and P. Scott. 2009. IL-7 receptor expression provides the potential for long-term survival of both CD62Lhigh central memory T cells and Th1 effector cells during Leishmania major infection. *J. Immunol.* 182:5702–5711. https://doi.org/10.4049/jimmunol.0803450

Corbin, G.A., and J.T. Harty. 2004. Duration of infection and antigen display have minimal influence on the kinetics of the CD4+ T cell response to Listeria monocytogenes infection. *J. Immunol.* 173:5679–5687. https://doi.org/10.4049/jimmunol.173.9.5679

Crotty, S. 2014. T follicular helper cell differentiation, function, and roles in disease. *Immunity.* 41:529–542. https://doi.org/10.1016/j.immuni.2014.10.004

Davis, M.M., and P.J. Bjorkman. 1988. T-cell antigen receptor genes and T-cell recognition. *Nature.* 334:395–402. https://doi.org/10.1038/334395a0

Donnarumma, T., G.R. Young, J. Merkenschlager, U. Eksmond, N. Bongard, S.L. Nutt, C. Boyer, U. Dittmer, V.T. Le-Trilling, M. Trilling, et al. 2016. Opposing Development of Cytotoxic and Follicular Helper CD4 T Cells Controlled by the TCF-1-Bcl6 Nexus. *Cell Rep.* 17:1571–1583. https://doi.org/10.1016/j.celrep.2016.10.013

Garcia, K.C., and E.J. Adams. 2005. How the T cell receptor sees antigen--a structural view. *Cell.* 122:333–336. https://doi.org/10.1016/j.cell.2005.07.015

Garris, C.S., V.A. Blaho, T. Hla, and M.H. Han. 2014. Sphingosine-1-phosphate receptor 1 signalling in T cells: trafficking and beyond. *Immunology.* 142:347–353. https://doi.org/10.1111/imm.12272

Gett, A.V., F. Sallusto, A. Lanzavecchia, and J. Geginat. 2003. T cell fitness determined by signal strength. *Nat. Immunol.* 4:355–360. https://doi.org/10.1038/ni908

Hänzelmann, S., R. Castelo, and J. Guinney. 2013. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 14. https://doi.org/10.1186/1471-2105-14-7

Holst, J., A.L. Szymczak-Workman, K.M. Vignali, A.R. Burton, C.J. Workman, and D.A. Vignali. 2006. Generation of T-cell receptor retrogenic mice. *Nat. Protoc.* 1:406–417. https://doi.org/10.1038/nprot.2006.61

Iwata, S., Y. Mikami, H.W. Sun, S.R. Brooks, D. Jankovic, K. Hirahara, A. Onodera, H.Y. Shih, T. Kawabe, K. Jiang, et al. 2017. The Transcription Factor T-bet Limits Amplification of Type I IFN Transcriptome and Circuitry in T Helper 1 Cells. *Immunity.* 46:983–991.e4. https://doi.org/10.1016/j.immuni.2017.05.005

Izraelson, M., T.O. Nakonechnaya, B. Moltedo, E.S. Egorov, S.A. Kasatskaya, E.V. Putintseva, I.Z. Mamedov, D.B. Staroverov, I.I. Shemiakina, M.Y. Zakharova, et al. 2018. Comparative analysis of murine T-cell receptor repertoires. *Immunology.* 153:133–144. https://doi.org/10.1111/imm.12857

Jellison, E.R., S.K. Kim, and R.M. Welsh. 2005. Cutting edge: MHC class II-restricted killing in vivo during viral infection. *J. Immunol.* 174:614–618. https://doi.org/10.4049/jimmunol.174.2.614

Jenkins, M.K., and J.J. Moon. 2012. The role of naive T cell precursor frequency and recruitment in dictating immune response magnitude. *J. Immunol.* 188:4135–4140. https://doi.org/10.4049/jimmunol.1102661

Kim, C., T. Wilson, K.F. Fischer, and M.A. Williams. 2013. Sustained interactions between T cell receptors and antigens promote the differentiation of CD4+ memory T cells. *Immunity.* 39:508–520. https://doi.org/10.1016/j.immuni.2013.08.033

Kong, Y., Y. Jing, and M. Bettini. 2019. Generation of T Cell Receptor Retrogenic Mice. *Curr. Protoc. Immunol.* 125:e76. https://doi.org/10.1002/cpim.76

Kotturi, M.F., I. Scott, T. Wolfe, B. Peters, J. Sidney, H. Cheroutre, M.G. von Herrath, M.J. Buchmeier, H. Grey, and A. Sette. 2008. Naive precursor frequencies and MHC binding rather than the degree of epitope diversity shape CD8+ T cell immunodominance. *J. Immunol.* 181:2124–2133. https://doi.org/10.4049/jimmunol.181.3.2124

Lee, S.K., R.J. Rigby, D. Zotos, L.M. Tsai, S. Kawamoto, J.L. Marshall, R.R. Ramiscal, T.D. Chan, D. Gatto, R. Brink, et al. 2011. B cell priming for extrafollicular antibody responses requires Bcl-6 expression by T cells. *J. Exp. Med.* 208:1377–1388. https://doi.org/10.1084/jem.20102065

Liberzon, A., A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, and J.P. Mesirov. 2011. Molecular signatures database (MSigDB) 3.0. *Bioinformatics.* 27:1739–1740. https://doi.org/10.1093/bioinformatics/btr260

Lieber, M.R. 1991. Site-specific recombination in the immune system. *FASEB J.* 5:2934–2944. https://doi.org/10.1096/fasebj.5.14.1752360

Lin, M.Y., and R.M. Welsh. 1998. Stability and diversity of T cell receptor repertoire usage during lymphocytic choriomeningitis virus infection of mice. *J. Exp. Med.* 188:1993–2005. https://doi.org/10.1084/jem.188.11.1993

Lönnberg, T., V. Svensson, K.R. James, D. Fernandez-Ruiz, I. Sebina, R. Montandon, M.S. Soon, L.G. Fogg, A.S. Nair, U. Liligeto, et al. 2017. Single-cell RNA-seq and computational analysis using temporal mixture modelling resolves Th1/T$_{fh}$ fate bifurcation in malaria. *Sci. Immunol.* 2:eaal2192. https://doi.org/10.1126/sciimmunol.aal2192

Madi, A., E. Shifrut, S. Reich-Zeliger, H. Gal, K. Best, W. Ndifon, B. Chain, I.R. Cohen, and N. Friedman. 2014. T-cell receptor repertoires share a restricted set of public and abundant CDR3 sequences that are associated with self-related immunity. *Genome Res.* 24:1603–1612. https://doi.org/10.1101/gr.170753.113

Madi, A., A. Poran, E. Shifrut, S. Reich-Zeliger, E. Greenstein, I. Zaretsky, T. Arnon, F.V. Laethem, A. Singer, J. Lu, et al. 2017. T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. *eLife.* 6:e22057. https://doi.org/10.7554/eLife.22057

Marshall, H.D., A. Chandele, Y.W. Jung, H. Meng, A.C. Poholek, I.A. Parish, R. Rutishauser, W. Cui, S.H. Kleinstein, J. Craft, and S.M. Kaech. 2011. Differential expression of Ly6C and T-bet distinguish effector and memory Th1 CD4(+) cell properties during viral infection. *Immunity.* 35:633–646. https://doi.org/10.1016/j.immuni.2011.08.016

Marusina, A.I., Y. Ono, A.A. Merleev, M. Shimoda, H. Ogawa, E.A. Wang, K. Kondo, L. Olney, G. Luxardi, Y. Miyamura, et al. 2017. CD4+ virtual memory: Antigen-inexperienced T cells reside in the naïve, regulatory, and memory T cell compartments at similar frequencies, implications for autoimmunity. *J. Autoimmun.* 77:76–88. https://doi.org/10.1016/j.jaut.2016.11.001

Miconnet, I., A. Marrau, A. Farina, P. Taffé, S. Vigano, A. Harari, and G. Pantaleo. 2011. Large TCR diversity of virus-specific CD8 T cells provides the mechanistic basis for massive TCR renewal after antigen exposure. *J. Immunol.* 186:7039–7049. https://doi.org/10.4049/jimmunol.1003309

Moon, J.J., P. Dash, T.H. Oguin III, J.L. McClaren, H.H. Chu, P.G. Thomas, and M.K. Jenkins. 2011. Quantitative impact of thymic selection on Foxp3+ and Foxp3- subsets of self-peptide/MHC class II-specific CD4+ T cells. *Proc. Natl. Acad. Sci. USA.* 108:14602–14607. https://doi.org/10.1073/pnas.1109806108

Ozaki, K., R. Spolski, C.G. Feng, C.F. Qi, J. Cheng, A. Sher, H.C. Morse III, C. Liu, P.L. Schwartzberg, and W.J. Leonard. 2002. A critical role for IL-21 in regulating immunoglobulin production. *Science.* 298:1630–1634. https://doi.org/10.1126/science.1077002

Pepper, M., and M.K. Jenkins. 2011. Origins of CD4(+) effector and central memory T cells. *Nat. Immunol.* 12:467–471. https://doi.org/10.1038/ni.2038

Pepper, M., A.J. Pagán, B.Z. Igyártó, J.J. Taylor, and M.K. Jenkins. 2011. Opposing signals from the Bcl6 transcription factor and the interleukin-2 receptor generate T helper 1 central and effector memory cells. *Immunity.* 35:583–595. https://doi.org/10.1016/j.immuni.2011.09.009

Qiu, X., Q. Mao, Y. Tang, L. Wang, R. Chawla, H.A. Pliner, and C. Trapnell. 2017. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods.* 14:979–982. https://doi.org/10.1038/nmeth.4402

Qu, Y., Y. Huang, D. Liu, Y. Huang, Z. Zhang, Z. Mi, X. An, Y. Tong, and J. Lu. 2016. High-Throughput Analysis of the T Cell Receptor Beta Chain Repertoire in PBMCs from Chronic Hepatitis B Patients with HBeAg Seroconversion. *Can. J. Infect. Dis. Med. Microbiol.* 2016. https://doi.org/10.1155/2016/8594107

Reinhardt, R.L., H.E. Liang, and R.M. Locksley. 2009. Cytokine-secreting follicular T cells shape the antibody repertoire. *Nat. Immunol.* 10: 385–393. https://doi.org/10.1038/ni.1715

Rudolph, M.G., and I.A. Wilson. 2002. The specificity of TCR/pMHC interaction. *Curr. Opin. Immunol.* 14:52–65. https://doi.org/10.1016/S0952-7915(01)00298-9

Sallusto, F., J. Geginat, and A. Lanzavecchia. 2004. Central memory and effector memory T cell subsets: function, generation, and maintenance. *Annu. Rev. Immunol.* 22:745–763. https://doi.org/10.1146/annurev.immunol.22.012703.104702

Satija, R., J.A. Farrell, D. Gennert, A.F. Schier, and A. Regev. 2015. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33: 495–502. https://doi.org/10.1038/nbt.3192

Shifrut, E., K. Baruch, H. Gal, W. Ndifon, A. Deczkowska, M. Schwartz, and N. Friedman. 2013. CD4(+) T Cell-Receptor Repertoire Diversity is Compromised in the Spleen but Not in the Bone Marrow of Aged Mice Due to Private and Sporadic Clonal Expansions. *Front. Immunol.* 4:379. https://doi.org/10.3389/fimmu.2013.00379

Sourdive, D.J., K. Murali-Krishna, J.D. Altman, A.J. Zajac, J.K. Whitmire, C. Pannetier, P. Kourilsky, B. Evavold, A. Sette, and R. Ahmed. 1998. Conserved T cell receptor repertoire in primary and memory CD8 T cell responses to an acute viral infection. *J. Exp. Med.* 188:71–82. https://doi.org/10.1084/jem.188.1.71

Subramanian, A., P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, and J.P. Mesirov. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc.*

Natl. Acad. Sci. USA. 102:15545–15550. https://doi.org/10.1073/pnas.0506580102

Tong, Y., Z. Li, H. Zhang, L. Xia, M. Zhang, Y. Xu, Z. Wang, M.W. Deem, X. Sun, and J. He. 2016. T Cell Repertoire Diversity Is Decreased in Type 1 Diabetes Patients. *Genomics Proteomics Bioinformatics.* 14:338–348. https://doi.org/10.1016/j.gpb.2016.10.003

Tubo, N.J., A.J. Pagán, J.J. Taylor, R.W. Nelson, J.L. Linehan, J.M. Ertelt, E.S. Huseby, S.S. Way, and M.K. Jenkins. 2013. Single naive CD4+ T cells from a diverse repertoire produce different effector cell types during infection. *Cell.* 153:785–796. https://doi.org/10.1016/j.cell.2013.04.007

Turner, S.J., P.C. Doherty, J. McCluskey, and J. Rossjohn. 2006. Structural determinants of T-cell receptor bias in immunity. *Nat. Rev. Immunol.* 6: 883–894. https://doi.org/10.1038/nri1977

Utzschneider, D.T., M. Charmoy, V. Chennupati, L. Pousse, D.P. Ferreira, S. Calderon-Copete, M. Danilo, F. Alfei, M. Hofmann, D. Wieland, et al. 2016. T Cell Factor 1-Expressing Memory-like CD8(+) T Cells Sustain the Immune Response to Chronic Viral Infections. *Immunity.* 45: 415–427. https://doi.org/10.1016/j.immuni.2016.07.021

Vezys, V., D. Masopust, C.C. Kemball, D.L. Barber, L.A. O'Mara, C.P. Larsen, T.C. Pearson, R. Ahmed, and A.E. Lukacher. 2006. Continuous recruitment of naive T cells contributes to heterogeneity of antiviral CD8 T cells during persistent infection. *J. Exp. Med.* 203:2263–2269. https://doi.org/10.1084/jem.20060995

Weinstein, J.S., B.J. Laidlaw, Y. Lu, J.K. Wang, V.P. Schulz, N. Li, E.I. Herman, S.M. Kaech, P.G. Gallagher, and J. Craft. 2018. Correction: STAT4 and T-bet control follicular helper T cell development in viral infections. *J. Exp. Med.* 215:999. https://doi.org/10.1084/jem.2017045702062018c

Xu, W., X. Zhao, X. Wang, H. Feng, M. Gou, W. Jin, X. Wang, X. Liu, and C. Dong. 2019. The Transcription Factor Tox2 Drives T Follicular Helper Cell Development via Regulating Chromatin Accessibility. *Immunity.* 51: 826–839.e5. https://doi.org/10.1016/j.immuni.2019.10.006

Zhang, L., X. Yu, L. Zheng, Y. Zhang, Y. Li, Q. Fang, R. Gao, B. Kang, Q. Zhang, J.Y. Huang, et al. 2018. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature.* 564:268–272. https://doi.org/10.1038/s41586-018-0694-x

Zhang, R., C.F. Qi, Y. Hu, Y. Shan, Y.P. Hsieh, F. Xu, G. Lu, J. Dai, M. Gupta, M. Cui, et al. 2019. T follicular helper cells restricted by IRF8 contribute to T cell-mediated inflammation. *J. Autoimmun.* 96:113–122. https://doi.org/10.1016/j.jaut.2018.09.001
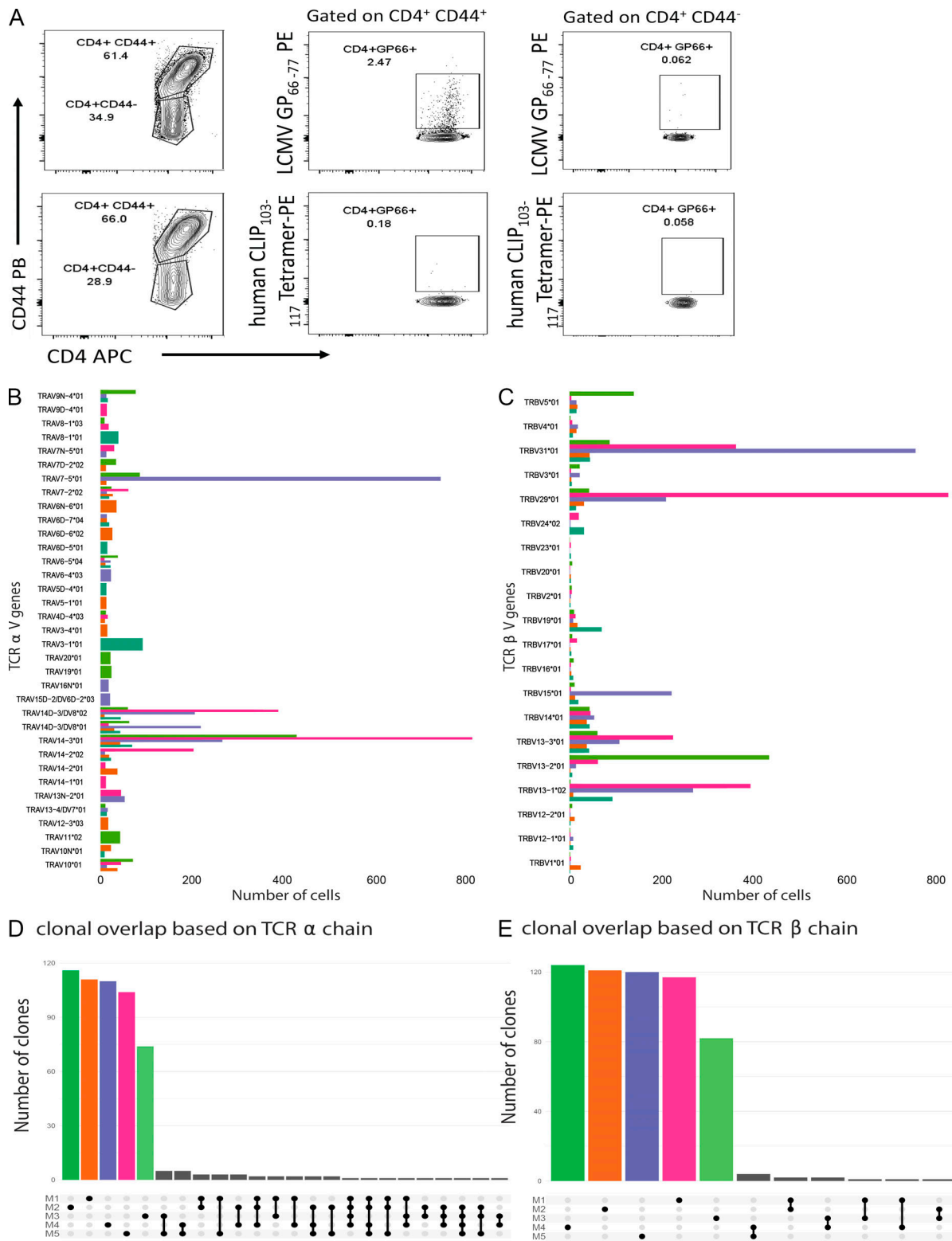
Supplemental material

Figure S1. **Sorting strategy, variable gene usage, and clonal overlap based on single TCR chains among five mice.** Related to Fig. 1. **(A)** Upper panel: flow cytometry data showing percentage of GP66 tetramer⁺ activated CD4 T cells (CD4⁺ CD44⁺ GP66⁺) and GP66 tetramer⁺ naive CD4 T cells (CD4⁺ CD44⁻ GP66⁺) at day 10 after acute LCMV infection. Only GP$_{66-77}$ tetramer⁺ activated CD4 T cells were sorted from five mice in three independent experiments to perform this study. Lower panel: similar to upper panel but showing percentage of activated CD4 T cells (CD4⁺ CD44⁺ GP66⁺) and naive CD4 T cells (CD4⁺ CD44⁻ GP66⁺) using human CLIP$_{103-117}$ tetramer as a negative control. **(B and C)** Bar plots showing TCR α and β chain variable gene usage among CD4 T cells at day 10 after acute LCMV infection. **(D and E)** Upset plot showing clonal overlap based on unpaired TCR α and β chains. Each color indicates one individual mouse. Dots below each bar denote which mouse or mice the clones were found in. Gray bars with more than one dot represent clones found in more than one mouse. All the experiments were done three times using five mice in total.
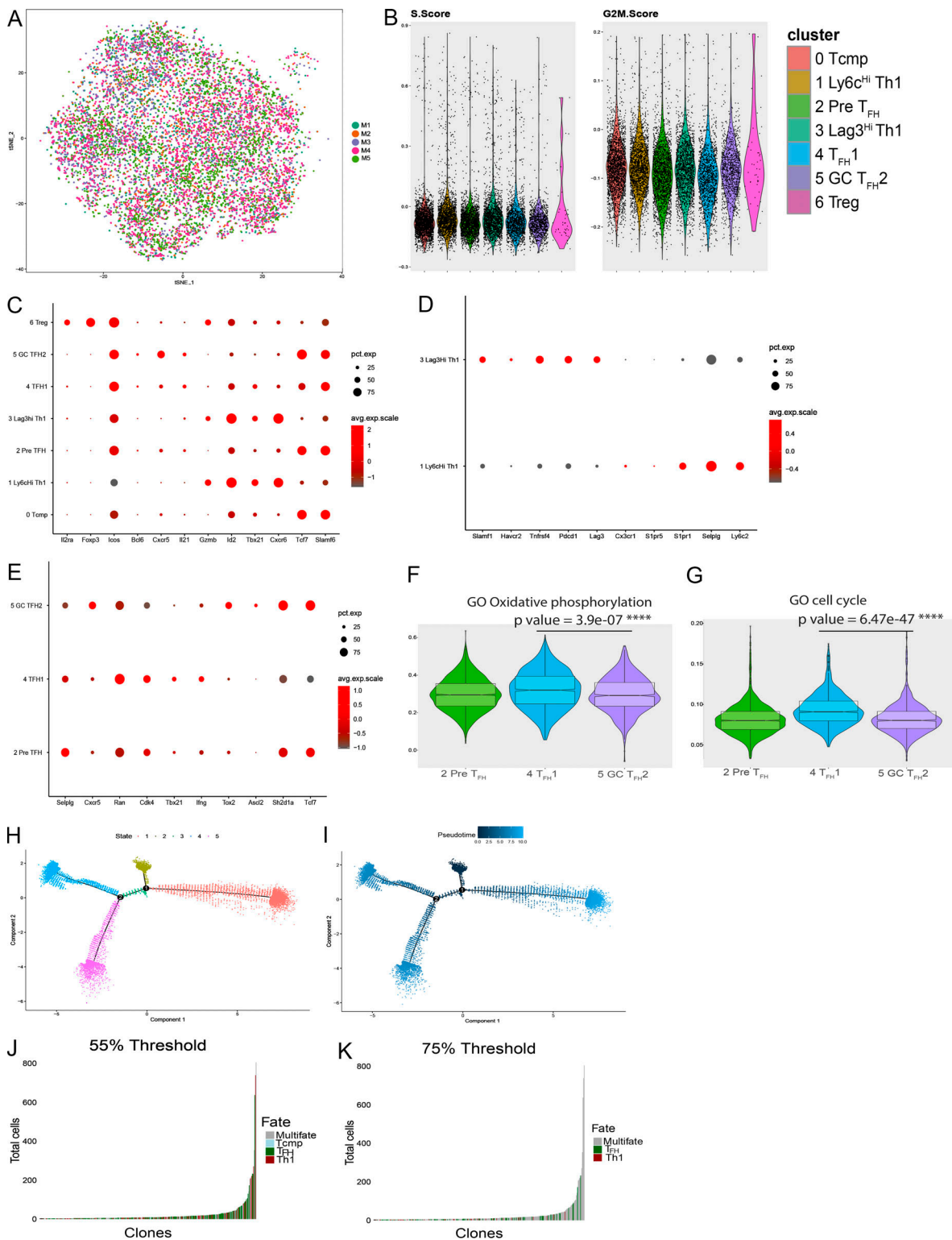
Figure S2. **Intercluster heterogeneity and clone size distribution among different CD4 T cell subsets.** Related to Figs. 2 and 3. **(A)** t-SNE plot as shown in Fig. 2 but colored by sample origin of each cell. **(B)** Violin plots showing cell cycle gene module scores for each cluster. **(C)** Dot plot showing differential expression of cell subset–specific (Tcmp, right; Th1 and GC T_{FH}, middle; and T reg, left) genes across all clusters in five mice. **(D and E)** Similar to C but showing top differentially expressed genes among Th1 (D) and T_{FH} (E) clusters. **(F and G)** Violin plots showing module scores for oxidative phosphorylation (F) and cell cycle (G) gene sets among three T_{FH} subsets; P values calculated by Wilcoxon test with Holm-Sidak correction. GO, gene ontology. **(H)** Differentiation states for Seurat clusters 0–5 predicted by Monocle single-cell trajectory plot. **(I)** Similar to H but colored by pseudotime after setting state 2 as the root state, based on comparatively higher number of Tcmp cells at this state. **(J and K)** Distribution of single cell fate (Th1, T_{FH}, or Tcmp biased) and multifate clones by clone size (four or more cells/clone; n = 352 clones) in five mice using 55% (J) and 75% (K) single cell fate thresholds (see Materials and methods). The data represent experiments done in five mice three independent times.
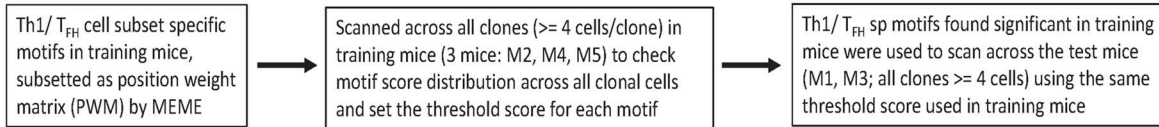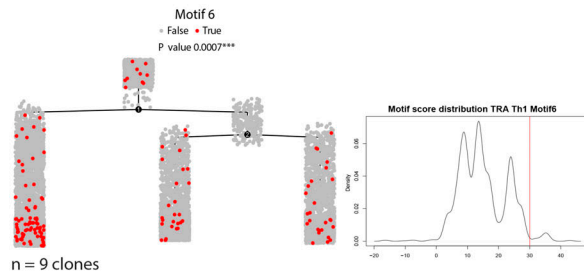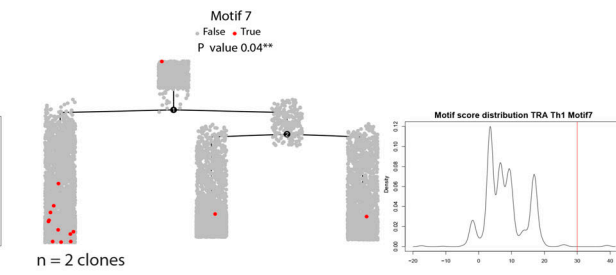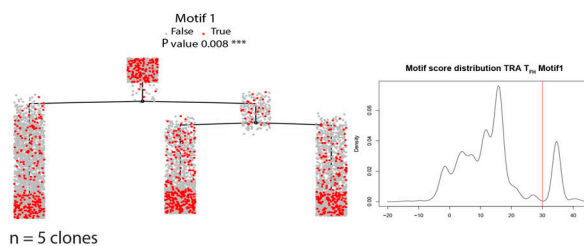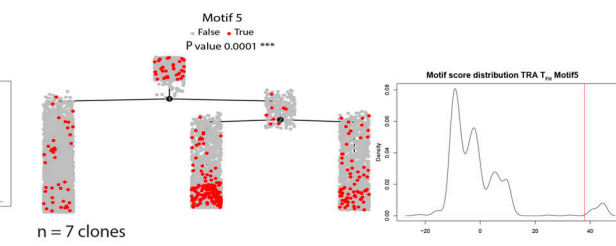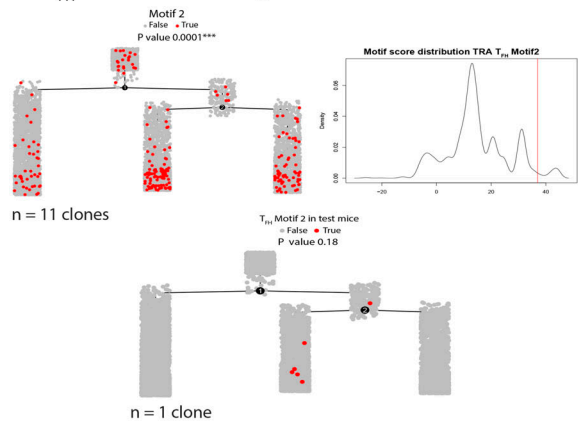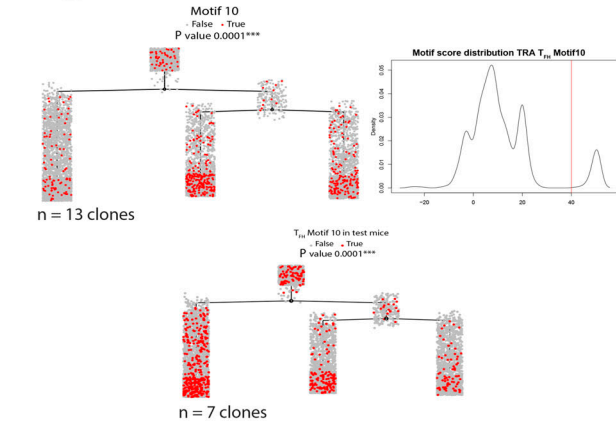
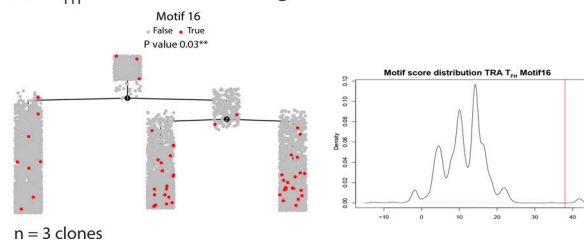Figure S3. **Th1- or T_FH-specific TCR α chain motif analysis in training and test mice.** Related to Fig. 3. **(A)** Diagram explaining motif analysis pipeline (see Materials and methods for details). sp, specific. **(B and C)** Monocle tree trajectory plots showing cellular fates of polyclonal CD4 T cell clones (four or more cells/clone) in three training mice, bearing Th1 subset–specific TCR α CDR3 motifs, Motif 6, and Motif 7, having scores greater than the threshold score (right) in acute LCMV infection at day 10 p.i. Each red dot indicates a cell and n indicates the number of clones belonging to at least two out of three training mice. **(D and E)** As in B and C but using T_FH-biased motifs, Motif 1 (D) and Motif 5 (E). **(F–H)** Monocle tree trajectory plot showing validation and testing of T_FH-biased motifs, Motif 2, Motif 10, and Motif 16 across polyclonal CD4 T cell clones (four or more cells/clone) in three training mice and two test mice, respectively, as explained in B–E. Motif 16 was not tested in test mice as the threshold score for this motif could not be reached in test mice. P values indicate the significance of each motif to define clonal bias toward Th1 or T_FH fate. P values were calculated using Fisher's exact test. TRA, TCR α chain.

Five tables are provided online as Excel files. Table S1 shows cellular distributions across different clusters based on their Monocle states. Table S2 lists cellular distributions and percentages of all clones across different clusters. Table S3 shows cellular distributions and percentages of cluster-specific clones. Table S4 shows cellular distributions and percentages of clones with least four cells across five mice along with their respective subset-specific module scores and percentile module scores. Table S5 lists cellular distributions based on Monocle states for Th1 and T_{FH} motif–specific clones in training and test mice.