Divergent clonal differentiation trajectories establish CD8+ memory T cell heterogeneity during acute viral infections in humans

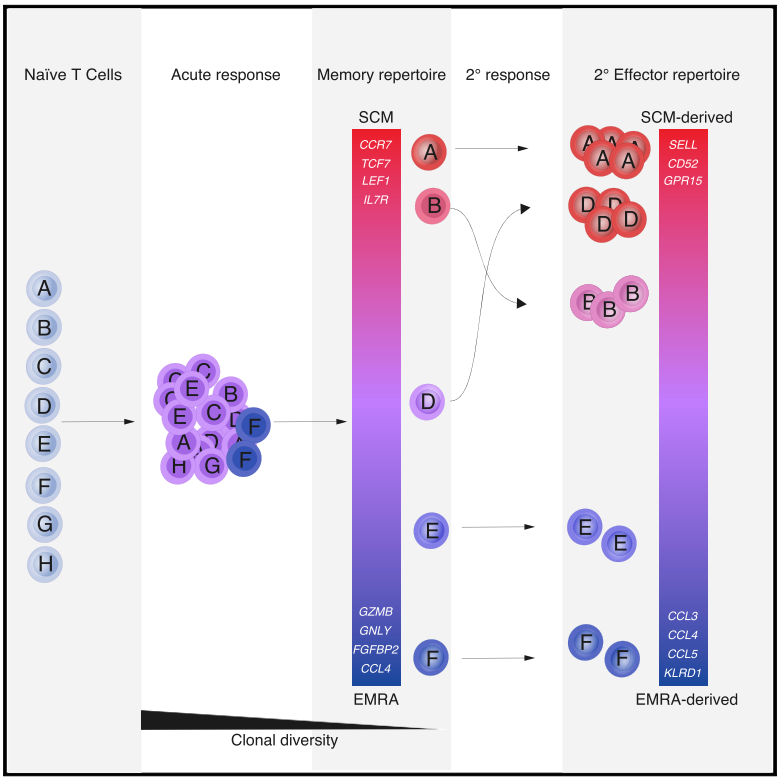
Jeff Mold, Laurent Modolo, Joanna Hård, Margherita Zamboni, Anton J.M. Larsson, Moa Stenudd, Carl-Johan Eriksson, Ghislain Durif, Patrik Ståhl, Erik Borgström, et al.

Cell Reports

Article

Divergent clonal differentiation trajectories establish CD8+ memory T cell heterogeneity during acute viral infections in humans

Graphical abstract



Highlights

- Decrease in human CD8+ T cell clonal diversity from acute to memory phase
- Clonally biased differentiation trajectories of CD8+ T cells in primary responses
- Clonally skewed CD8+ memory T cell differentiation in secondary recall responses

Authors

Jeff E. Mold, Laurent Modolo, Joanna Hård, ..., Franck Picard, Jakob Michaëlsson, Jonas Frisén

Correspondence

jakob.michaelsson@ki.se (J.M.), jonas.frisen@ki.se (J.F.)

In brief

Mold et al. investigate the clonal architecture of the human CD8+ T cell response to vaccination with live attenuated yellow fever virus, and they demonstrate that the multifaceted T cell response to an acute viral infection is created by the sum of distinct clonal phenotypes.

Mold et al., 2021, Cell Reports *35*, 109174
May 25, 2021
https://doi.org/10.1016/j.celrep.2021.109174

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Divergent clonal differentiation trajectories establish CD8$^+$ memory T cell heterogeneity during acute viral infections in humans

Jeff E. Mold,[1,6] Laurent Modolo,[2,6] Joanna Hård,[1,7] Margherita Zamboni,[1,7] Anton J.M. Larsson,[1] Moa Stenudd,[1] Carl-Johan Eriksson,[1] Ghislain Durif,[2] Patrik L. Ståhl,[1,3] Erik Borgström,[3] Simone Picelli,[1] Björn Reinius,[1,5] Rickard Sandberg,[1] Pedro Réu,[1] Carlos Talavera-Lopez,[1] Björn Andersson,[1] Kim Blom,[4] Johan K. Sandberg,[4] Franck Picard,[2] Jakob Michaëlsson,[4,8,9,*] and Jonas Frisén[1,8,*]

[1]Department of Cell and Molecular Biology, Karolinska Institutet, 171 77 Stockholm, Sweden
[2]LBBE, UMR CNRS 5558, Université Lyon 1, Villeurbanne, France LBMC UMR 5239 CNRS/ENS Lyon, Lyon, France
[3]Science for Life Laboratory, Department of Gene Technology, KTH Royal Institute of Technology, 106 91 Stockholm, Sweden
[4]Center for Infectious Medicine, Department of Medicine, Karolinska Institutet, Karolinska University Hospital Huddinge, 141 86 Stockholm, Sweden
[5]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 171 77 Stockholm, Sweden
[6]These authors contributed equally
[7]These authors contributed equally
[8]These authors contributed equally
[9]Lead contact
*Correspondence: jakob.michaelsson@ki.se (J.M.), jonas.frisen@ki.se (J.F.)
https://doi.org/10.1016/j.celrep.2021.109174

SUMMARY

The CD8$^+$ T cell response to an antigen is composed of many T cell clones with unique T cell receptors, together forming a heterogeneous repertoire of effector and memory cells. How individual T cell clones contribute to this heterogeneity throughout immune responses remains largely unknown. In this study, we longitudinally track human CD8$^+$ T cell clones expanding in response to yellow fever virus (YFV) vaccination at the single-cell level. We observed a drop in clonal diversity in blood from the acute to memory phase, suggesting that clonal selection shapes the circulating memory repertoire. Clones in the memory phase display biased differentiation trajectories along a gradient from stem cell to terminally differentiated effector memory fates. In secondary responses, YFV- and influenza-specific CD8$^+$ T cell clones are poised to recapitulate skewed differentiation trajectories. Collectively, we show that the sum of distinct clonal phenotypes results in the multifaceted human T cell response to acute viral infections.

INTRODUCTION

Primary adaptive immune responses lead to clonal expansion of rare antigen-specific naive CD8$^+$ T cells and generation of long-lived memory cells, which guard against subsequent infections (Kaech and Ahmed, 2001). During this response, CD8$^+$ T cells differentiate into a diverse array of effector and memory cells that exhibit distinct phenotypic and functional properties (Champagne et al., 2001; Gillespie et al., 2000; Jameson and Masopust, 2009, 2018; Sallusto et al., 1999; Willinger et al., 2005). Human memory T cells are classified into four major subsets primarily distinguished by expression of CCR7 and CD45RA, where CCR7$^+$CD45RA$^+$ and CCR7$^+$CD45RA$^-$ memory T cells are termed T_{SCM} and T_{CM}, respectively, and CCR7$^-$CD45RA$^+$ and CCR7$^-$CD45RA$^+$ memory T cells are termed T_{EM} and T_{EMRA}, respectively (Sallusto et al., 1999). T_{SCM} and T_{CM} are theorized to be multipotent memory cells, which re-establish the complete repertoire of CD8$^+$ effectors upon reinfection, while T_{EM} and T_{EMRA} are considered to be primed cytotoxic effectors that exist in the circulation and peripheral non-lymphoid tissues, serving as a first line of defense against subsequent infections (Fearon et al., 2001; Masopust et al., 2001; Reinhardt et al., 2001). At the molecular level, T_{SCM} and T_{CM} are similar to naive T cells, albeit with features of effectors and epigenetic signatures indicative of prior effector status (Akondy et al., 2017; Fuertes Marraco et al., 2015; Gattinoni et al., 2011). T_{EM} and T_{EMRA} have transcriptomic profiles similar to activated effector CD8$^+$ T cells, suggesting that these cells are poised for rapid cytotoxicity (Willinger et al., 2005). Transfer of each cell type in mice revealed that T_{SCM} expand to greater numbers, are better at controlling tumor growth than either T_{CM} or T_{EM}, and that individual T_{CM} can have stem cell-like capacity in serial transfer models (Gattinoni et al., 2011; Graef et al., 2014). On the other hand, T_{EM} have been shown to be more protective than T_{CM} against reinfection with vaccinia virus, despite being less protective than T_{CM} against lymphocytic choriomeningitis virus (LCMV) infection in adoptive transfer models (Bachmann et al., 2005), suggesting that the protective roles of these subsets are context-dependent.

Table 1. Summary of YFV-specific CD8$^+$ T cells analyzed *ex vivo*

Donor	HLA	Time (days)	Cells	Clones
A	A2	15	291	216
A	A2	136	485	234
A	A2	593	188	109
A	B7	15	154	67
A	B7	136	118	63
B	A2	15	86	82
B	A2	90	87	80
B	B7	10	124	56
B	B7	15	285	94
B	B7	30	213	80
B	B7	90	123	61
B	B7	148	155	37
C	A2	15	875	445
C	A2	90	303	137
D	A2	15	511	387
D	A2	90	494	260
D	A2	720	286	181

The relative proportions of the memory T cell subsets vary between different infections. Acute infections typically generate robust effector responses, which gradually give rise to more abundant T_{CM} (or T_{SCM}) after resolution of the response (Blom et al., 2013; Kaech and Ahmed, 2001; Miller et al., 2008). Recurrent or chronic infections are known to progressively generate larger T_{EM} populations, likely due to constant reactivation of existing memory populations (Appay et al., 2002; Champagne et al., 2001). Studies examining memory T cell heterogeneity following primary responses to yellow fever virus (YFV) vaccination in humans observed an emergence of T_{SCM} and T_{EMRA} that persisted for years after vaccination (Akondy et al., 2017; Fuertes Marraco et al., 2015). Similarly, both T_{SCM} and T_{EMRA} emerge after acute dengue virus infection (Chng et al., 2019). The clonal composition and functional significance of each population however remain unknown.

Many of the markers used to characterize CD8$^+$ T cells, e.g., CCR7, CD45RA, and CD27, have continuous rather than strict binary expression patterns (Appay et al., 2002; Fuertes Marraco et al., 2015; Gattinoni et al., 2011; Miller et al., 2008; Sallusto et al., 1999). Advances in high-dimensional phenotyping have also revealed a greater T cell heterogeneity, suggesting that T cell identity can be viewed as a continuum of cell states ranging from naive cells to highly differentiated effectors (Jameson and Masopust, 2018; Newell et al., 2012). This makes classification of a given single cell as a specific "cell type" challenging, as many cells have intermediate phenotypes based on conventional marker profiles. In contrast, viewing the total repertoire of CD8$^+$ T cells as continuum of cellular identities could capture both the heterogeneity of the total CD8$^+$ T cell repertoire, as well as the unique identity of a given CD8$^+$ T cell along this continuum.

In theory, the diversity of CD8$^+$ T cells in response to a given antigen can be generated from one naive CD8$^+$ T cell, or through the concerted actions of many distinct T cell clones, each of which produces a population of phenotypically similar progeny along a continuum of CD8$^+$ T cell identities. Several attempts have been made to address the clonal origins of discrete memory T cell types in mice and humans. In humans, T cell receptor (TCR) profiling of purified CD8$^+$ T_{CM} and T_{EM} populations identified biased representation of T cell clones with specificity for influenza virus (Baron et al., 2003). Analysis of human CD4$^+$ memory T cells, reactivated with fungal or bacterial antigens *in vitro*, however, indicated that most clones give rise to heterogeneous CD4$^+$ T helper cell fates with respect to cytokine production (Becattini et al., 2015). Adoptive transfer of single naive OT-1 TCR transgenic CD8$^+$ T cells in mice revealed that a single naive CD8$^+$ T cell can give rise to multiple T cell fates (Gerlach et al., 2010; Stemberger et al., 2007), while adoptive transfer of multiple barcoded naive OT-1 TCR transgenic CD8$^+$ T cells in mice revealed clearly biased clonal differentiation patterns according to cell surface phenotype (Buchholz et al., 2013; Gerlach et al., 2013). Additional mouse fate mapping studies have revealed clonal bias in CD4$^+$ T helper cell differentiation based on comparisons of multiple transgenic TCR models (Tubo et al., 2013). Taken together, although it is clear that individual naive T cells can give rise to both clonally biased progeny and heterogeneous T cell responses in experimental models, it remains largely unknown to what extent clonal bias of CD8$^+$ T cell differentiation occurs in a polyclonal response to a natural infection *in vivo*, both in mice and in humans.

In this study, we investigated the clonal contributions to memory T cell heterogeneity by longitudinally tracking individual CD8$^+$ T cell clones responding to two distinct epitopes of the YFV vaccine (YFV-17D), as well as by analyses of *in vitro* recall responses by single virus-specific CD8$^+$ T cells. We show that the heterogeneity of primary YFV-specific human memory CD8$^+$ T cells is created by the sum of distinct clonal T cell identities along a continuum from T_{SCM} to T_{EMRA}. Furthermore, we demonstrate that secondary CD8$^+$ T cell responses are clonally biased, with distinct effector progeny generated from CD8$^+$ T_{SCM} and T_{EMRA} founder cells. Our findings thus highlight an underlying clonal heterogeneity that leads to the generation of diverse memory T cell subsets in primary human immune responses and identify distinct functional attributes of memory T cell clones in recall responses.

RESULTS

Clonal diversity in circulating YFV-specific CD8$^+$ T cells drops from acute to memory phase of the response

To assess the clonal contribution to the circulating repertoire of YFV-specific CD8$^+$ T cells throughout a primary anti-viral immune response, we analyzed 4,778 single antigen-specific CD8$^+$ T cells longitudinally after primary vaccination of four healthy donors with YFV-17D (donors A–D) (Table 1). We tracked CD8$^+$ T cells specific for an immunodominant human leukocyte antigen (HLA)-A2-restricted epitope (NS4b) and a subdominant HLA-B7-restricted epitope (NS5) during acute (days 10–30), early memory (days 90–150), and late memory (day 500+) phases of the response (Figures 1A and S1A) (Blom et al., 2013). We identified productive TCRβ sequences for 3,058 cells using a nested PCR strategy, targeting the TCRβ chain locus in DNA

CellPress
OPEN ACCESS

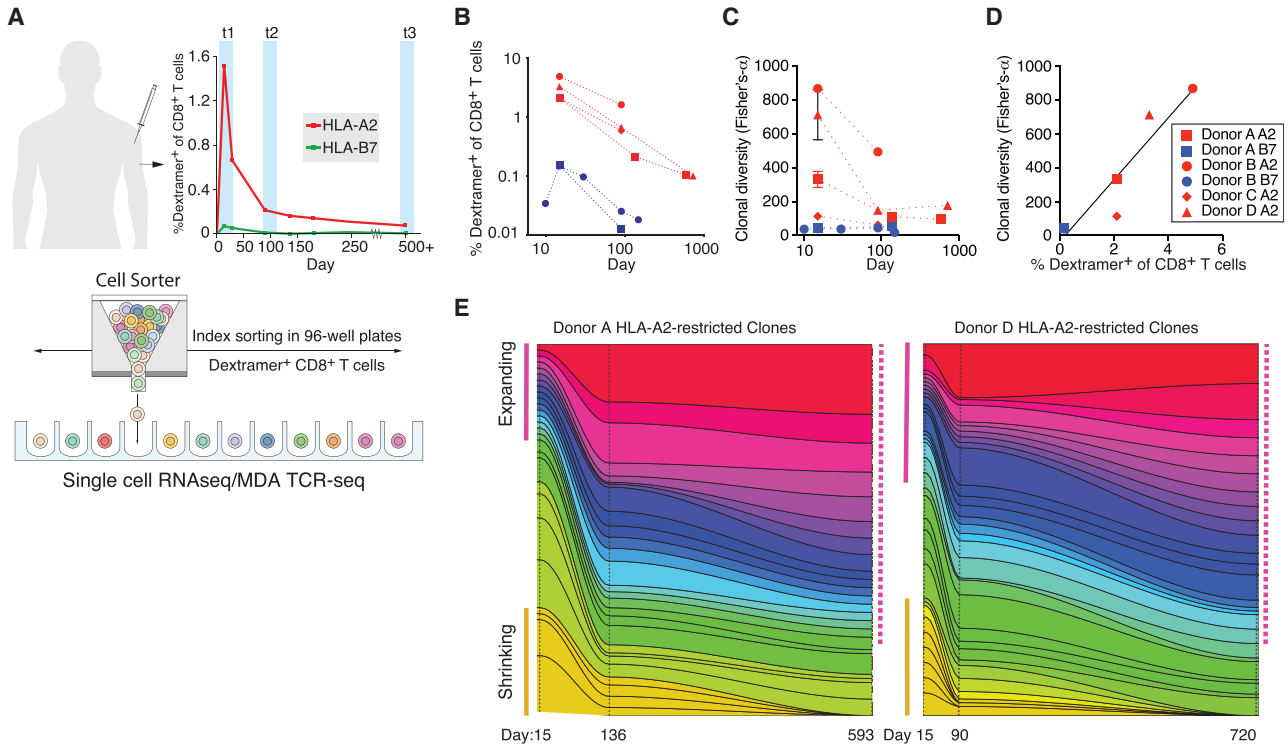


Figure 1. Clonal contributions to primary CD8$^+$ T cell responses to yellow fever vaccine

(A) Schematic diagram for experimental setup. Percent HLA-A2/NS4B- and HLA-B7/NS5-dextramer$^+$ cells of total CD8$^+$ T cells for donor A is shown as an example.

(B) Percent YFV-specific CD8$^+$ T cells of total CD8$^+$ T cells (donors A–D, n = 4 for HLA-A2, n = 2 for HLA-B7).

(C) Clonal diversity measured by Fisher's alpha throughout the acute-to-memory transition in donors A–D. Fischer's alpha calculated by 20× random sampling without replacement of 75 cells from the pool of analyzed cells for each donor/antigen/time point. Error bars represent standard error of mean of repeated sampling.

(D) Correlation between percent YFV-specific (dextramer$^+$) CD8$^+$ T cells versus clonal diversity (Fisher's alpha) at day 15 ($r^2 = 0.89$, $p < 0.01$ by two-tailed Pearson test) (donors A–D, n = 4 for HLA-A2, n = 2 for HLA-B7).

(E) Evolution of clone sizes throughout the acute (day 15)-to-memory (days 90/136 and 593/720) transition. All HLA-A2-restricted clones from donors A and D with n > 3 cells at a single time point are shown for visualization purposes. Red dashed lines indicate "expanding" clones at late time points.

A summary of statistical results using all clones with $n \geq 1$ for at least at two time points is shown in Table S2.

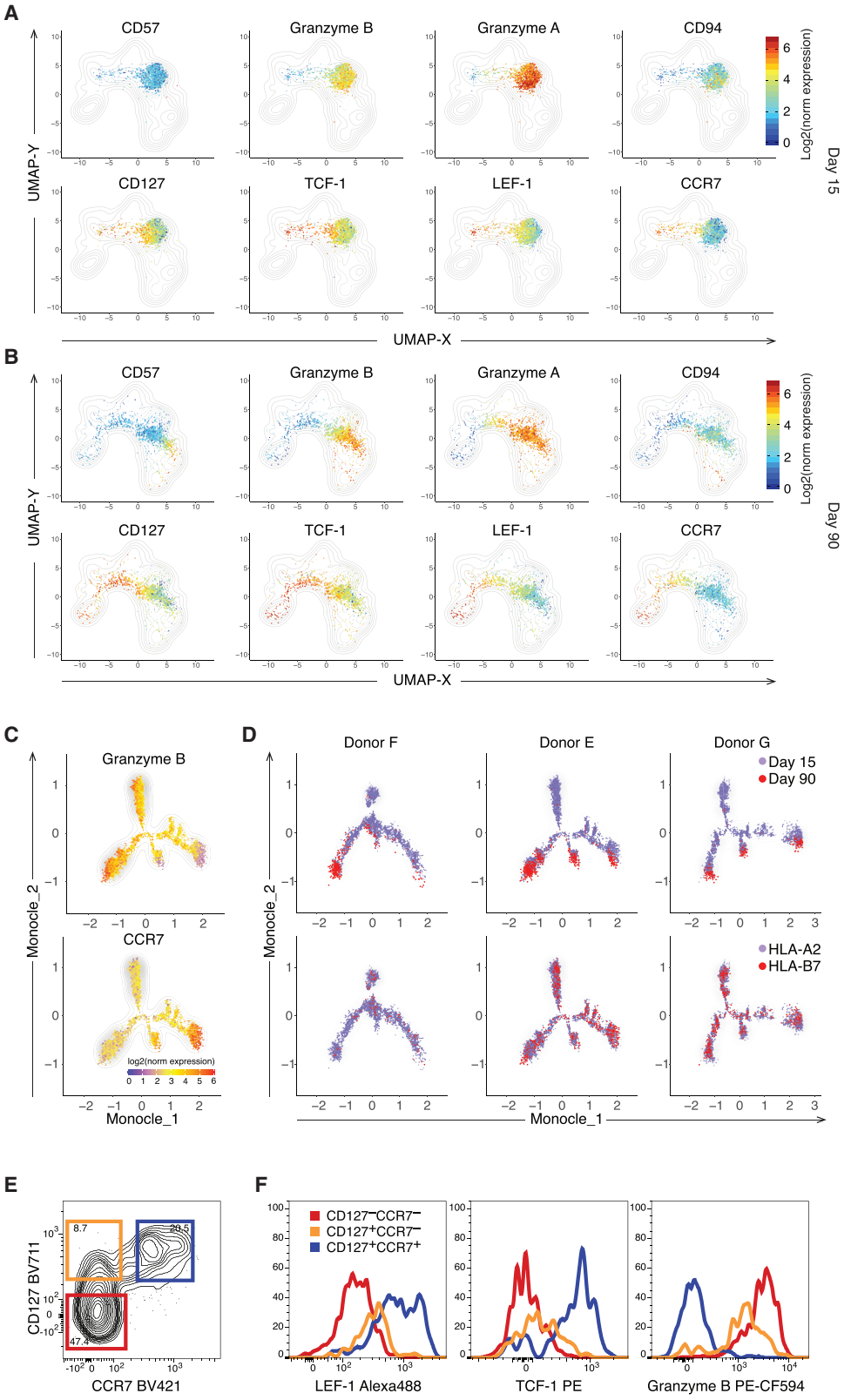
from whole genome-amplified single cells (Balazs et al., 2013)(; Hård et al., 2019), and an additional 1,720 cells using full-length single-cell RNA sequencing (scRNA-seq) and TCRα and/or β chain (Table S1) (Bolotin et al., 2015; Picelli et al., 2013). Despite an approximately 10-fold difference in response magnitude (Blom et al., 2013), HLA-A2 and HLA-B7 responses displayed similar patterns of expansion and contraction (Figures 1B, S1B, and S1C).

We next investigated whether clones exhibited uniform or biased expansion and contraction throughout the acute to memory phase of the response. We accounted for sampling variability between donors, antigens, and time points by bootstrapping analyses (Figures S1D and S1E). We measured clonal diversity in terms of the number of individual clones observed per cells sampled (Figure S1D), as well as Fisher's alpha index (Figures 1C and S1E), a commonly used measure of population diversity that is relatively insensitive to sampling depth (Fisher et al., 1943). We observed a significant decline in clonal diversity from the acute (day 15) to memory (day 90+) phase of the response for all donors (three-way ANOVA, $p < 0.001$; Fisher's alpha, acute: 270, late/memory: 195) (Figures 1C and S1E).

The clonal diversity at the acute phase of the response correlated with the total frequency of antigen-specific CD8$^+$ T cells in blood at this time (Figure 1D), indicating that the variability in overall response magnitude across donors and viral epitopes is at least partially dictated by the number of clones participating in the response. While this finding is consistent with conclusions made in some mouse models (Moon et al., 2007; Obar et al., 2008), further studies are needed to confirm whether this is generalizable to CD8$^+$ T cell responses to other viral infections and in larger cohorts.

Peripheral clonal expansions during the acute response do not predict clonal contributions to the circulating memory T cell repertoire

The longitudinal decline in diversity suggests that clonal selection plays a role in shaping the circulating memory repertoire after YF-17D vaccination. Consistent with this, contraction of large clones and expansion of smaller clones was apparent during the



(legend on next page)

acute-to-memory transition in the two donors with large datasets spanning the acute to late (day 500+) memory phases of the HLA-A2/NS4B response (Figure 1E; clones with n > 3 cells used for visualization purposes only). We assessed the relationship between clone size and persistence to the late memory phase (day 593 or 720) on clones where we observed at least one cell present in multiple different time points. Large clone sizes at day 15 were negatively associated with long-term survival (log odds ratio: −1.76 donor A, −1.83 donor D; $p < 0.05$; statistical analyses were performed on all cells), while clone sizes at day 90/136 were positively correlated with persistence to day 500+ (log odds ratio: 0.51 donor A, 0.73 donor D; $p < 0.05$) (Table S2). This suggests that clones persisting in the memory phase of the response are stably maintained in circulation, while increased clonal expansion in the acute response is inversely correlated with the likelihood of differentiation into a long-lived memory T cell clone. While this does not preclude the possibility that long-lived memory cells derive from effector precursors, as recently suggested (Akondy et al., 2017; Youngblood et al., 2017), it supports the conclusion from studies in mice that acute phase effector expansion is not necessarily a primary determinant of eventual memory fate (Grassmann et al., 2020).

Diverse memory phenotypes emerge after resolution of the acute phase of the response to YFV vaccination

To characterize the phenotypic evolution of the primary YFV-specific $CD8^+$ T cell response in peripheral blood, we examined cells from the acute and memory response for $CD8^+$ T cells specific for the HLA-A2/NS4b and HLA-B7/NS5 YFV epitopes in three donors using high-dimensional flow cytometry. We measured the expression of proteins characteristic of $T_{CM/SCM}$ (CCR7, CD127, TCF-1, LEF1, CD27) (Gattinoni et al., 2011; Kratchmarov et al., 2018; Sallusto et al., 1999; Xing et al., 2016) and $T_{EM/EMRA}$ (granzyme A, granzyme B, CD94, CD57) (Chattopadhyay et al., 2009; Willinger et al., 2005), in addition to CD45RA, KLRG1, CXCR4, and Ki67, and analyzed population-level heterogeneity using dimensionality reduction (Becht et al., 2019) (Figures 2A, S2A, and S2B). This allowed us to identify the emergence of polarized memory T cell populations throughout the acute-to-memory transition of the response similarly to recent characterizations of the T cell response to acute dengue infection (Chng et al., 2019).

Based on UMAP clustering, the vast majority of acutely responding cells grouped together, centrally between highly polarized T_{EMRA} and T_{SCM} populations, and expressed effector proteins (granzyme A and B), yet they were distinguished from memory populations on the basis of having lower expression of CD45RA and CXCR4 (Figures 2A and S2A). As expected, a large fraction of YFV-specific $CD8^+$ T cells expressed Ki67 at day 15, indicative of proliferation (Figures S2A, S2B, and S2D). In contrast, most cells at day 90 regained CD45RA expression and displayed increased phenotypic heterogeneity according to memory markers, covering the range of phenotypes between highly polarized T_{SCM} and T_{EMRA} fates (Figures 2B and S2A). Trajectory analysis of YFV-specific $CD8^+$ T cells at day 15 and 90 using Monocle (v2) revealed the emergence of cells with biased $CD45RA^+CCR7^-GZMB^+$ T_{EMRA} or $CD45RA^+CCR7^+GZMB^-$ T_{SCM} fates throughout the acute-to-memory transition (Figures 2C and 2D; bulk $CD8^+$ T cells in Figure S2C), consistent with what was seen in $CD8^+$ T cell immunity to dengue infection (Chng et al., 2019; Qiu et al., 2017). While we noted clear inter-individual differences in distribution of YFV-specific $CD8^+$ T cells along the range of memory phenotypes, no obvious differences were observed between the HLA-A2/NS4B and HLA-B7/NS5 epitopes (Figure 2D). We observed an intermediate population of cells in the trajectories similar to those recently observed in late-stage dengue-specific T cells (Figure 2D) (Chng et al., 2019). Closer analysis of individual proteins revealed the existence of intermediate phenotypes, e.g., a population of $CD127^+CCR7^-$ $CD45RA^{+/-}CD8^+$ T cells with intermediate expression of TCF-1 and LEF-1 (Figures 2E and 2F). Notably, in mouse models $CD127^+CD62L^-(CCR7^-)$ cells are generally classified as T_{EM} (Huster et al., 2004), which differs from the definition in most human studies ($CCR7^-$ $CD45RA^-$), and thus we will term these cells $CD127^+$ T_{EMRA} hereafter to avoid confusion. Our findings overall support the view that virus-specific memory $CD8^+$ T cells, when examined at a single-cell level, are likely to exhibit a wide range of variable cell states rather than discrete identities (Jameson and Masopust, 2018).

Coupling clonality to cellular identity by scRNA-seq identifies unique clonal origins of long-lived T_{EMRA} and T_{SCM}

To understand how $CD8^+$ T cell clones contribute to different phenotypes emerging during the acute-to-memory transition of the response, we tracked the phenotypic evolution of cells from individual T cell clones longitudinally. We combined cell surface protein expression analysis (Figure S3A; Table S1) with full-transcript scRNA-seq (Picelli et al., 2013). This allowed us to identify distinct clones on the basis of shared TCRα and TCRβ sequences, and to further refine the classification of

Figure 2. High-dimensional phenotypic characterization of YFV-specific $CD8^+$ T cells at acute and memory stages

(A and B) YFV-specific $CD8^+$ T cell phenotypic distribution by UMAP dimensionality reduction during acute (A, day 15) or memory (B, day 90) phases of the primary response to YFV vaccination. Concatenated data from HLA-A2/NS4B- and HLA-B7/NS5-dextramer$^+$ $CD8^+$ T cells from three donors are shown. Bulk peripheral blood $CD8^+$ T cells from the same donors are included as contour plots (gray). Expression levels of individual protein markers are displayed as heatmaps (log_2-scaled values). Individual data and UMAPs with protein expression for bulk $CD8^+$ T cells are shown in Figure S2.

(C) Monocle (v2) trajectory analysis of YFV-specific cells (concatenated data from HLA-A2 and B7-dextramer$^+$ cells from donor E) with $T_{EM/EMRA}$ (GZMB) and $T_{CM/SCM}$ (CCR7) markers shown to indicate phenotype of different branches.

(D) Monocle trajectory analysis showing distribution of $CD8^+$ T cells according to time point (top panels) and HLA-A2/B7 epitope restriction (bottom panels) for three donors (E, F, and G).

(E)Visualization of polarized $T_{CM/SCM}$ (blue square, $CD127^+CCR7^+$) and $T_{EM/EMRA}$ (red square, $CD127^-CCR7^-$) and an intermediate population of $CD8^+$ T cells (green square, $CD127^+CCR7^-$) corresponding to cells with mixed identities observed in UMAP and Monocle visualization in (A)–(D) above.

(F) Expression of LEF-1, TCF-1, and granzyme B in $T_{CM/SCM}$ (blue), $T_{EM/EMRA}$ (red), and $CD127^+CCR7^-$ (green) YFV-specific $CD8^+$ T cells.

cellular marker profiles, coupling clonality to phenotype (Table S1). In order to phenotypically classify single CD8$^+$ T cells, we used a partial least-squares (PLS) regression approach with CCR7 protein expression as a starting point to train a classifier for all sequenced YFV-specific CD8$^+$ T cells (Figure S3B; see "classifier approach" in STAR Methods) (Durif et al., 2018). The PLS-based classification strategy works as a sparse principal component analysis (PCA) based on genes identified as being differentially expressed between CCR7^{high} and CCR7$^-$ cells (Table S3). We adopted this strategy because many antigen-specific CD8$^+$ T cells are not easily classified as a discrete "cell type" on an individual basis, despite the fact that we did observe that cells gradually acquired phenotypic properties associated with the classical $T_{CM/SCM}$ and $T_{EM/EMRA}$ types over time (Figures S3B and S3C). We ranked and assigned a score to each CD8$^+$ T cell based on transcriptome differences according to the likelihood that they were strong effectors (probability of a single cell being a central memory cell [pCM] = 0) or strong $T_{CM/SCM}$ (pCM = 1.0), allowing us to also classify cells having intermediate identities (Figure S3C). We validated the pCM score on cells sorted from bulk CD8$^+$ T cell populations on the basis of classically defined phenotypes (naive, T_{SCM}, T_{CM}, T_{EM}, T_{EMRA}) (Figure S4). Notably, many of the top genes identified by our classification strategy (*TCF7*, *CCR7*, *LTB*, *SELL*, *GZMK*, *GZMB*, *GZMH*, *ZEB2*, *CCL4*, *GNLY*) (Figures 3A and S4C) play established roles in specifying $T_{CM/SCM}$ and $T_{EM/EMRA}$ identities in mice and humans (Jameson and Masopust, 2018; Kaech and Cui, 2012; Omilusik et al., 2015; Xing et al., 2016). In both donors, we found that the pCM score accurately ordered cells according to expression of these genes (Figures 3A and 3B).

For donor A, we analyzed three time points and examined how cells from individual clones varied with respect to pCM score longitudinally throughout the response (Figures 3C and 3D). In the acute phase, most clones consisted of cells with low pCM scores, indicating an effector identity. In the early (day 136) and late (day 593) memory phase, we observed a gradual increase in pCM scores over time for cells in many of the clones, consistent with the hypothesis that clones go through an effector stage prior to adopting a T_{SCM} fate (Akondy et al., 2017). Focusing on the memory phase we found that not all clones, however, transitioned toward a T_{SCM} fate, as some gave rise exclusively to CD8$^+$ memory T cells with low pCM scores (clones 127, 75, 240, 118, 230) (Figures 3C and 3D). Our findings thus reveal that distinct clonal fates, which often are skewed toward a T_{SCM} or T_{EMRA} phenotype in the memory phase of the response, contribute to the overall phenotypic heterogeneity observed in long-lived CD8$^+$ T cell memory repertoires. This finding is corroborated by a recent report, documenting skewed clonal contributions to transcriptionally distinct memory T cell subsets in a single donor surveyed 18 months after YFV vaccination (Minervina et al., 2020).

Individual CCR7$^+$ T_{SCM}, CD127$^+$ T_{EMRA}, and CD127$^-$ T_{EMRA} memory T cells give rise to distinct effector progeny following secondary reactivation

Currently, T_{CM}/T_{SCM} are positioned at the apex of a differentiation cascade, acting as multipotent, self-renewing cells capable of generating all effector cell classes (Gattinoni et al., 2011; Graef et al., 2014; Lugli et al., 2013), whereas T_{EM}/T_{EMRA} only generate short bursts of highly cytotoxic effector cells (Lalvani et al., 1997; Sallusto et al., 1999; Schenkel et al., 2013). This model has been established and tested primarily on the basis of adoptive transfer studies in mice and *in vitro* re-stimulation assays of bulk populations of cells sorted according to conventional memory and effector phenotypes in humans (Gattinoni et al., 2011).

In this study, we analyzed recall potential by single memory CD8$^+$ T cells in an *in vitro* model of antigen-specific T cell reactivation. We performed antigen-specific reactivation of index-sorted, single YFV-specific memory CD8$^+$ T cells and monitored expansion potential by measuring viable HLA-A2/NS4b-specific CD8$^+$ T cell numbers after 2–3 weeks of culture (Figure 4A). We characterized the effector progeny of nine individual memory T cell founders identified as either T_{SCM} (clones B4, E4, H2), CD127$^+$ T_{EMRA} (clones A7, A8, G6), or CD127$^-$ T_{EMRA} (clones F3, G8, H9), representing the general spectrum of memory identities observed in our study (Figures 4B and 4C). After an 18-day expansion period, we detected thousands of progeny from each founder clone, all of which were CCR7$^-$CD127$^-$ and which expressed variable levels of PD-1 and CD57, indicating their status as highly activated effector cells (Figure 4D). TCR sequencing confirmed the clonal origins of each population of expanded cells (Table S1). In general, we observed higher PD-1 expression in progeny derived from T_{EMRA} compared to T_{SCM}, indicating that T_{EMRA}-derived effectors are predisposed to stronger regulation after initial rounds of expansion (Figure 4D).

We performed high-coverage scRNA-seq on 36–45 clonally expanded cells from each clone to determine the diversity of phenotypes arising from individual memory CD8$^+$ T cells. In contrast to the CD127$^-$ T_{EMRA}-derived effectors, some T_{SCM}-derived and CD127$^+$ T_{EMRA}-derived effector progeny displayed signs of ongoing proliferation even 18 days after activation (Figure 4E), consistent with previous findings that T_{SCM} undergo greater expansion relative to $T_{EM/EMRA}$ (Gattinoni et al., 2011). Because clonally expanded progeny from all memory T cells exhibited a stereotypic effector profile (CCR7$^-$CD45RA^{low}), we could not use conventional protein marker expression as a guide for interpreting cell type heterogeneity. Therefore, after accounting for cell cycle differences, we examined intra- and inter-clonal variability of effector progeny on the basis of gene expression differences by diffusion map dimensionality reduction (Figures 4F and S5A; Table S4) (DM1 genes and Gene Ontology analysis for each cluster) (Haghverdi et al., 2015). We observed that clonal effector progeny generally separated by founder cell phenotype, with some cells exhibiting more intermediate phenotypic identities based on diffusion map scores (Figure 4F). There were, however, exceptions; clone E4, which had a clear CD45RA$^+$CCR7$^+$ T_{SCM} founder phenotype (Figure 4B), produced progeny more similar to progeny from CD127$^+$ T_{EMRA} founders, while clone A7, a clone derived from a CD45RA$^+$CCR7$^-$CD127$^+$ T_{EMRA}, produced progeny similar to those from T_{SCM} founders. The observation that different memory T cells could yield distinct effector progeny, despite exhibiting stereotypical T_{SCM} or T_{EMRA} phenotypes, provides a possible explanation for how heterogeneous recall responses can be generated after reactivation of bulk T_{SCM} populations (Gattinoni et al., 2011; Kratchmarov et al., 2018).

CellPress
OPEN ACCESS

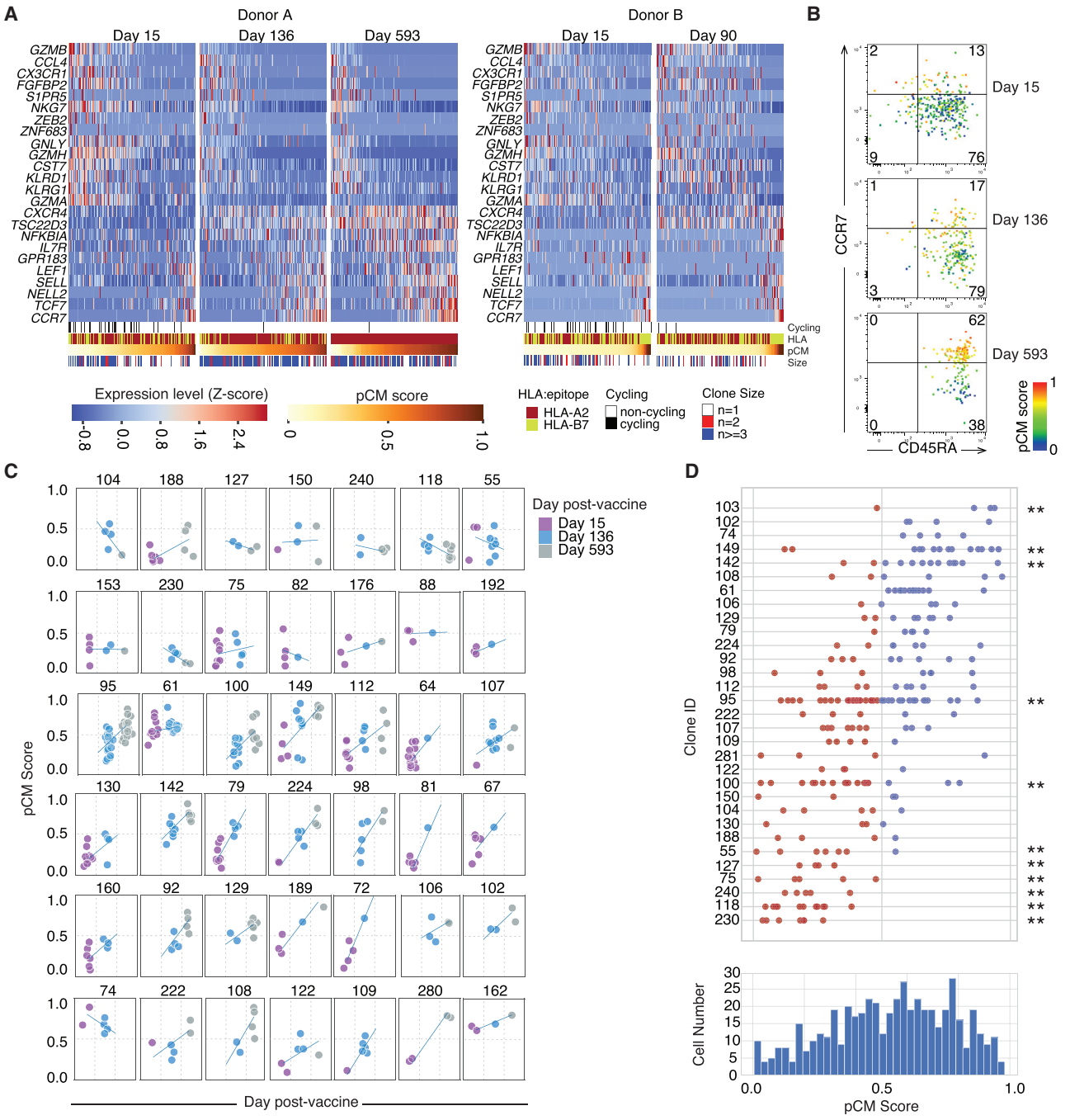


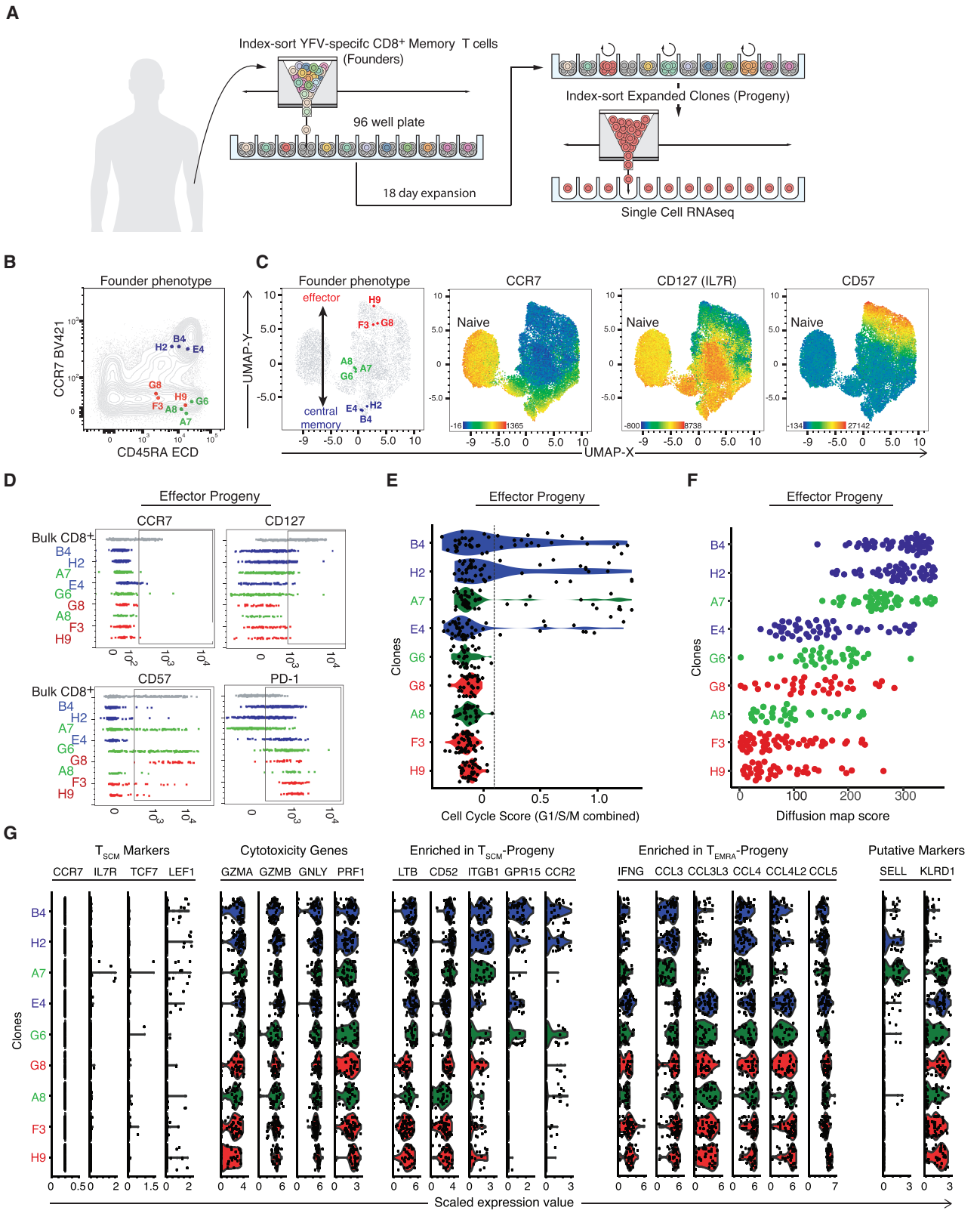
Figure 3. High-resolution phenotypic characterization of individual CD8⁺ T cell clones throughout primary YFV vaccination

(A) Heatmaps (*Z* scored) depicting selected genes that are differentially expressed between CD8⁺ T cell subsets in high-coverage scRNA-seq analysis of single CD8⁺ T cells from donors A and B taken at acute (day 15) and late (day 90, 136, or 593) time points post-vaccination. Cells are ordered according to score (pCM score) (Figure S3), with pCM = 0 being effectors and pCM = 1 indicating polarized $T_{CM/SCM}$. HLA restriction and cell cycle status are indicated in bars below individual heatmaps. The number of clonally related cells identified for each cell is indicated by size.

(B) Flow cytometry plots for donor A depicting individual CD8⁺ T cells (HLA-A2 and B7 combined) colored according to pCM score.

(C) Individual CD8⁺ T cells from clones (n > 3 cells at two or more time points) from donor A plotted according to pCM score versus time.

(D) Distribution of single-cell phenotypes for clones during the memory (day 136/593) phase of the response according to pCM score (donor A). Clones with n > 3 cells found at either time point for both HLA-A2- and B7-restricted responses are shown. Histogram represents the distribution of all cells in clones with n > 3 cells along the gradient of pCM scores. **Clones that significantly diverge from expected distributions based on the total population ($p < 0.05$, Wilcoxon rank-sum test).



(legend on next page)

The progeny from all CD8$^+$ memory T cell founders lacked *CCR7* expression, and only rarely expressed *TCF7* or *IL7R* mRNA (Figure 4G). In contrast, we observed *LEF1* expression in multiple cells across both T_{SCM}- and T_{EMRA}-derived progeny (Figure 4G), suggesting that this Wnt family transcription factor might be able to sustain memory potential. We noted that effector progeny generated from all memory T cells expressed high levels of *GZMA*, *GZMB*, *GNLY*, and *PRF1* transcripts, supporting their classification as putative cytotoxic effector cells (Figure 4G). We observed an enrichment of genes involved in lymphocyte homing (*CXCR3*, *GPR15*, *ITGB1*, *CCR2*, *SELL*) in T_{SCM}-derived progeny, while T_{EMRA}-derived effectors showed significantly higher expression of inflammatory chemokines (e.g., *CCL3*, *CCL4*, *CCL5*) and *IFNG* expression (Figures 4G and S5B; Table S3). We additionally noted that T_{SCM}- and T_{EMRA}-derived progeny could partially be distinguished based on the expression of *SELL* and *KLRD1*. CD62L (the protein encoded by *SELL*) is a widely used marker of T_{CM} in mice (Buchholz et al., 2013; Gerlach et al., 2013), and it is enriched in $T_{CM/SCM}$ in our *in vivo* data, whereas *KLRD1* is highly enriched in $T_{EM/EMRA}$ (Figures 3A and S4C). We therefore reasoned that the cell surface proteins encoded by these genes (CD62L and CD94) might be useful for identifying effector T cell subpopulations with distinct memory T cell founders in humans (Figure 4G).

CD94 and CD62L delineate heterogeneous effector T cell subsets generated in recall responses *in vitro*

We next validated whether KLRD1 (CD94) and SELL (CD62L) were potential markers to classify distinct subtypes of effector T cells generated in memory T cell recall assays. We replicated our *in vitro* re-stimulation experiments using single memory T cells isolated from two donors (donors A and B) at different times throughout the memory phase of the response (days 90, 136, 593, and 1,401) (Figure 5A). Founder cell phenotypes were once again defined on the basis of index sorting of HLA-A2/NS4b-specific CD8$^+$ memory T cells with CD45RA, CCR7, CD127, and CD57. Similar frequencies of each memory T cell type exhibited evidence of some recall potential (Figure S5D), with single cells generating as much as 100,000 effector progeny in 2–3 weeks (Figure 5B). We also noted that single T_{SCM} typically gave rise to more progeny in these assays than did CD127$^-$ and CD127$^+$ T_{EMRA} memory T cells (Figure 5B), consistent with our observation of increased cell cycling in this subset (Figure 4E) and with reports of greater proliferative potential by T_{SCM} in mice and humans (Gattinoni et al., 2011; Gattinoni et al., 2009).

We analyzed the effector progeny of nine additional single memory T cells (donor B, HLA-A2/NS4b-specific, day 1,401) with respect to expression of CD62L and CD94. Similar to our results based on RNA expression (Figure 4G), the progeny had distinct patterns of CD62L or CD94 expression at the protein level (Figure 5C), indicating that expression of these proteins is indeed clonally biased. Summing up the phenotypic diversity of the progeny of the individual clones largely recreated the pattern of phenotypic heterogeneity observed in bulk re-stimulation (Figure 5D). In this experiment five of nine memory T founder cells were phenotypically identified as T_{SCM} (CD45RA$^+$ CCR7$^+$CD127$^+$CD57$^-$) and gave rise to CD94$^-$CD62L$^{+/-}$ (clones D09, H04, H05, H06) and CD94$^-$CD62L$^-$ (clone F02) progeny. Two founders exhibited a clear T_{EMRA} phenotype (CD45RA$^+$ CCR7$^-$CD57$^+$) (clones F04 and D12) (Figure 5E) and gave rise to CD62L$^-$CD94$^+$ progeny. One CCR7$^+$CD45RA$^+$ founder cell (clone A08) produced CD62L$^-$CD94$^+$ progeny similar to the T_{EMRA} founders, but it was distinguished from T_{SCM} founder cells based on lack of CD127 expression (Figure 5E). These results again indicate that although founder phenotype generally correlated well with the phenotype of the progeny, there were also exceptions. More importantly, these findings suggest that heterogeneity within the effector populations generated in recall responses may in part derive from the collective behavior of different T cell clones, as opposed to any individual memory T cell possessing pluripotent recall capacity. However, future studies are needed to examine whether clonally biased recall responses can be observed *in vivo* in humans.

Diverse effector CD8$^+$ T cell progeny generated in recall responses to YFV and influenza antigens are composed of distinct T cell clones

Recall potential of different human memory T cell subpopulations is commonly studied by *in vitro* reactivation of sorted bulk T cell populations. We wanted to address whether effector potential for a memory T cell could be encoded at the level of individual clones, rather than within populations of phenotypically similar cell types. Because we observed that the clonal effector progeny of single T_{SCM} and T_{EMRA} tended to exhibit distinct patterns of CD62L and CD94 expression *in vitro*, we reasoned that we may be able to isolate unique T cell clones from bulk

Figure 4. Secondary reactivation of single CD8$^+$ T cell clones yields phenotypically distinct effector progeny

(A) Experimental setup showing sorting of memory CD8$^+$ T cells, expansion for 18 days, and sorting single cells from each expanded clone for downstream scRNA-seq.

(B and C) Phenotype of index-sorted CD8$^+$ T cell memory clones from donor A at day 136 overlaid on bulk CD8$^+$ T cells from matched blood, using (B) classical $T_{EM/EMRA}$ and $T_{CM/SCM}$ classification (CD45RA versus CCR7, contour plot) or (C) UMAP plots depicting bulk CD8$^+$ T cells (gray) and individual sorted T_{SCM} (CCR7$^+$CD127$^+$; blue), CD127$^+$ T_{EMRA} (CCR7$^-$CD127$^+$CD57$^-$; green), and CD127$^-$ T_{EMRA} (CCR7$^-$CD127$^-$CD57$^+$; red) YFV-specific founder CD8$^+$ T cells from donor A (day 136) (far left). Color gradients in UMAPs indicate protein expression levels for CCR7, CD127, and CD57 on total CD8$^+$ T cells.

(D) Cell surface protein expression of CCR7, CD127 ($T_{SCM/CM}$ markers), CD57 and PD-1 (effector/activation markers) on expanded T cell clones (B4-H9) from each sorted single founder T cell shown in (B). Clones are colored according to founder phenotype (B and C).

(E) Cell cycle score based on scRNA-seq analysis combining score from G_1/S/M phases for single cells from each clone (Butler et al., 2018).

(F) Diffusion map score (using DM1) for expanded single CD8$^+$ T cells from each clone based on variable genes identified by scRNA-seq from each clone (n = 32–45 cells/clone).

(G) Violin plots showing scaled expression levels for individual genes associated with (1) T_{SCM}, (2) cytotoxic effectors, and enriched in progeny of (3) T_{SCM} or (4) CD127$^-$ T_{EMRA}. (5) Putative markers *SELL* and *KLRD1* were identified as surface proteins that differed between T_{SCM} and T_{EMRA} progeny.

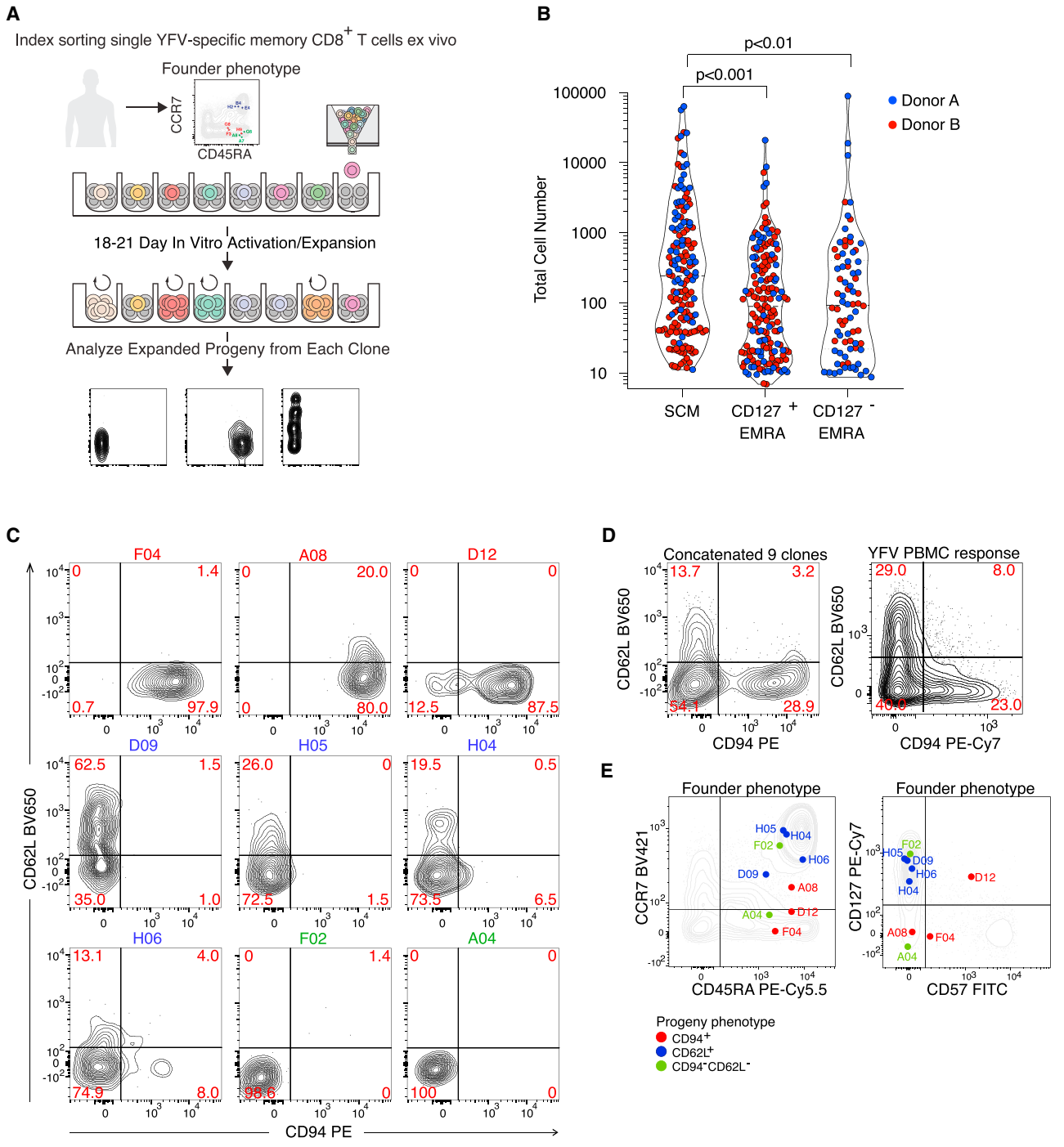


Figure 5. Individual CD8⁺ memory T cells give rise to progeny characterized by CD62L or CD94 expression

(A) Schematic illustrating the process for monitoring expansion potential and effector phenotypes for single HLA-A2/NS4B-specific CD8⁺ memory T cells.

(B) Number of progeny generated from single-cell expansions according to "founder" memory T cell phenotype: T_{SCM} (n = 164, CD45RA⁺CCR7⁺CD127⁺), CD127⁺ T_{EMRA} (n = 170, CD45RA⁺CCR7⁻CD127⁻), and CD127⁻ T_{EMRA} (n = 79, CD45RA⁺CCR7⁻CD127⁻). Statistical significance was assessed by a Kruskal-Wallis test and Dunn's multiple comparison test.

(C) Surface marker expression profiles of progeny from individual CD8⁺ YFV-specific memory T cells isolated at day 1,401 post-vaccination and restimulated as shown in (A). Different expanded clones are labeled above plots and colored according to phenotype: CD94⁺ progeny (red: A08, F04, D12), CD62L⁺/⁻CD94⁻ progeny (blue: D09, H05, H04, H06), and CD62L⁻CD94⁻ progeny (green: F02, A04).

(legend continued on next page)

reactivation assays on the basis of these markers after effector expansion.

We analyzed bulk CD8$^+$ T cell responses against YFV, influenza, and the chronic viral pathogen cytomegalovirus (CMV). After 2 weeks of reactivation, we observed expanded, antigen-specific CD8$^+$ T cells for each viral antigen (Figure S6A). We phenotyped the expanded T cell populations for each donor/virus combination using CD62L and CD94 expression and isolated triplicates of 50 sorted CellTrace Violet (CTV)$^-$dextramer$^+$ CD8$^+$ T cells from CD62L$^+$CD94$^-$ and CD62L$^-$CD94$^+$ subsets for RNA-seq analysis (Figures 6A and S6B). Comparing gene expression in CD62L$^+$CD94$^-$ versus CD62L$^-$CD94$^+$ CD8$^+$ T cells specific for YFV and influenza revealed a clear segregation of the two populations across donors and anti-viral responses (Figure 6B; Table S3). Gene expression differences were similar, although not identical, to our clonal scRNA-seq experiments, with CD62L$^+$CD94$^-$ cells expressing elevated levels of genes related to homing (*GPR15*, *S1PR1*, *KLF2*) and CD62L$^-$CD94$^+$ cells expressing elevated levels of effector-associated genes (*CCL3*, *CCL4*, *CCL5*, *TNF, NKG7*) (Figure 6B). In contrast to the YFV and influenza-specific CD8$^+$ T cells, we detected substantially fewer differences between CD62L$^+$CD94$^-$ and CD62L$^-$CD94$^+$ CMV-specific CD8$^+$ T cells, possibly reflecting a larger overlap in effector identities in the setting of chronic viral infections (Figure S6C).

To address whether CD62L$^+$CD94$^-$ and CD62L$^-$CD94$^+$ cells among the expanded virus-specific CD8$^+$ T cells had distinct clonal compositions, we analyzed the distribution of TCRβ sequences between samples (Figure 6C; Table S5). YFV-specific responses were characterized by a higher number of unique clones relative to influenza and CMV specific responses in all donors (Figure 6C). The CD62L$^+$CD94$^-$ and CD62L$^-$CD94$^+$ populations, in both YFV- and influenza-specific effector cells, largely contained unique and non-overlapping clones (Figure 6D). This supported our hypothesis that distinct populations of T cell clones in the memory pool give rise to phenotypically distinct populations of effector progeny during reactivation. In contrast to YFV and influenza, CMV responses were characterized by substantial clonal overlap between the two populations (Figure 6D), suggesting that chronic antigen exposure may lead to subclonal diversification and thus increased clonal heterogeneity, or possibly promote the differentiation of polyfunctional clones.

DISCUSSION

In this study, we provide evidence that distinct T cell clones participate to different extents in the effector and memory phase of the immune response, at least in the circulating T cell repertoire. Combining clonal tracing and transcriptomic analysis we additionally found that individual clones exhibited skewed differentiation trajectories throughout the response, giving rise to long-lived memory T cell subsets with different clonal compositions. Finally, we provide evidence that distinct memory T cell types can give rise to phenotypically distinct effector progeny, suggesting that the diversification of clones into distinct memory T cell states likely impacts the clonal contributions to heterogeneous recall responses.

Our findings corroborate and build on multiple lines of evidence from fate mapping studies performed in transgenic mice (Buchholz et al., 2013; Gerlach et al., 2013; Plumlee et al., 2013), where mouse CD8$^+$ T cells derived from adoptively transferred single naive OT-1 transgenic CD8$^+$ T cells displayed biased differentiation patterns with respect to expression of CD27, CD62L, and KLRG1 after infection *in vivo*. We extend these findings to a primary immune response to a live viral vaccine in humans, which activates a polyclonal CD8$^+$ T cell population, and demonstrate that the biased phenotype of individual clones extends beyond the expression of a limited set of cell surface markers. By defining clonal T cells at a single-cell level, according to transcriptome profiles rather than selected markers, we could further demonstrate that clones exhibit variable identities within the continuous distribution of cell states ranging from T_{SCM} to T_{EMRA} phenotypes (Jameson and Masopust, 2018).

Our findings that clonal diversity decreases from the acute to memory phase of the response and that clonal size at the acute phase is negatively associated with persistence as long-lived memory clones are consistent with results from clonal fate mapping in mouse models, where the size of clones observed in the primary response was not correlated with persistence as long-lived memory cells (Buchholz et al., 2013; Gerlach et al., 2013; Grassmann et al., 2020). Focusing of the TCR repertoire has also been observed in acute to chronic phases of CMV responses (Day et al., 2007). A recent study tracking changes in clonal diversity during primary or secondary responses to YFV vaccination in two donors found that in short-term (18 months) recall immunity there was a general preservation of overall clonal richness, but a marked variability with respect to clonal expansions in each response (Minervina et al., 2020). In a second donor, re-vaccinated 30 years after the primary vaccine, a strikingly lower number of unique clones was observed during recall responses. Therefore, loss of acutely responding clones and variable participation of acutely responding clones to recall responses have both been observed. Whether our observations concerning changes in clonal composition reflect loss of certain clones, unevenness in the dynamics of clonal expansion and contraction during the acute and memory phase of the response, or clonal differences in migration to tissues throughout the response, however, remains to be investigated. Given the likely importance of tissue-resident memory T cells in combatting secondary infections, this will be important to address in settings where tissue biopsies can be monitored in parallel.

Studies tracking clonal CD8$^+$ T cell differentiation after adoptive transfer have shown that clones adopting a T_{CM} fate divide

(D) Expression of CD62L and CD94 on HLA-A2/NS4B-specific CD8$^+$ T cells from nine concatenated clones (left) and after restimulation of 10^6 total peripheral blood mononuclear cells (PBMCs) (right). Bulk PBMC stimulation was done on donor B for 12 days, and responding cells were identified based on CTV dilution and HLA-A2 dextramer staining.

(E) Expression levels of CCR7, CD45RA, CD127, and CD57 on founder CD8$^+$ T cells depicted in (C) show that most T_{SCM} give rise to CD94$^-$CD62L$^+$ progeny, while T_{EMRA} give rise to CD94$^+$ progeny.

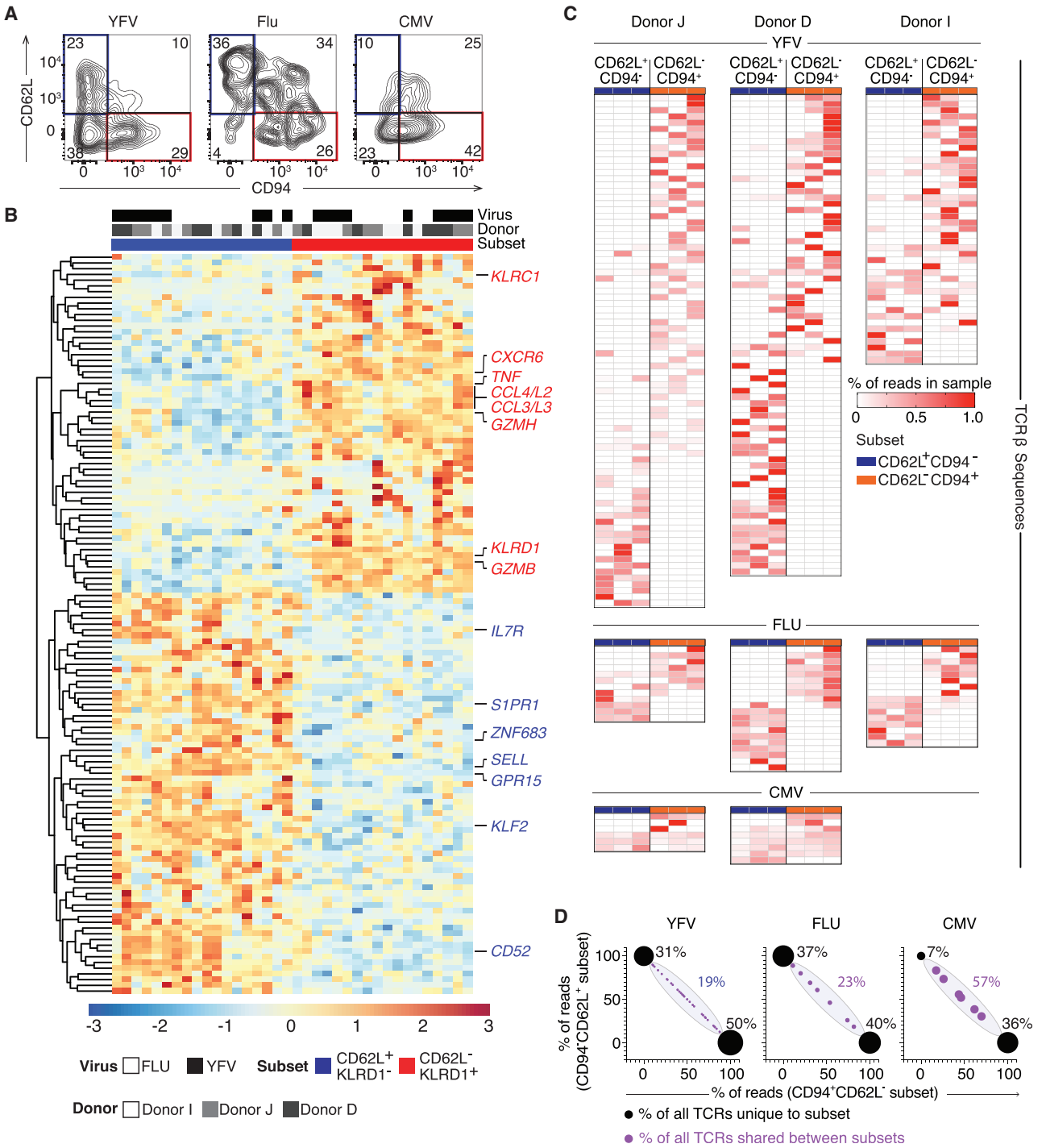


Figure 6. Clonally distinct effector populations emerge upon secondary activation of bulk YFV- and influenza-specific memory CD8⁺ T cells

(A) Expression of CD62L and CD94 on expanded virus-specific (HLA-A2/virus-dextramer⁺CTV⁻) CD8⁺ T cells 2 weeks after stimulation with YFV, influenza, or CMV from a single donor (donor I; donors J and D are shown in Figure S6B).

(B) Heatmap depicting differentially expressed genes identified by RNA-seq (adjusted p [p_{adj}] < 0.05, $\log_2$ fold change > 1) between CD62L⁺CD94⁻ and CD62L⁻CD94⁺ expanded YFV- and influenza-specific CD8⁺ T cells from three donors (three replicates with 50 cells each per donor/virus combination). Scale represents *Z* score, and virus and donor are depicted in top bars. Euclidean clustering segregates effector T cells on the basis of phenotype, but not on the basis of donor origin or viral specificity. Heatmap for CMV-specific CD8⁺ T cells is shown in Figure S6C.

(legend continued on next page)

slower (Kretschmer et al., 2020), which may explain the appearance of larger T_{SCM}-biased clones in the memory phase of the YFV response that were not detected in the acute phase of the response. Given the phenotypic and functional similarities of T_{SCM} to naive T cells, it is possible that a subset of T_{SCM} may represent clones that fail to fully undergo differentiation in acute responses. Inhibition of the mitogen-activated protein kinase pathway, a key signaling pathway for T cell activation, was recently shown to favor T_{SCM} differentiation, suggesting that limited acute activation may lead to clonal skewing toward a T_{SCM} fate (Verma et al., 2021). Conversely, mouse studies have shown that repeated vaccinations lead to substantial memory T cell expansions consisting primarily of T_{EM}, rather than T_{CM} (Masopust et al., 2006; Vezys et al., 2009). A more detailed analysis of the efficacy of polarized $T_{CM/SCM}$ versus $T_{EM/EMRA}$ both on a per cell and population level may help to further elucidate the efficacy of each for long-term memory immunity and the potential for additional heterogeneity of T_{SCM} cells with distinct activation histories.

We observed similar clonal differentiation trajectories for both HLA-A2- and HLA-B7-restricted YFV-specific $CD8^+$ T cells, despite an approximately 10-fold difference in the number of clones participating in the response, arguing that the frequency of naive precursors does not strongly influence the ability of individual clones to adopt a specific memory fate (Marzo et al., 2005; Obar et al., 2008). Additionally, we were unable to identify clear differences in gene expression in $CD8^+$ T cells at day 15 between clones that subsequently developed into either memory cells with a clear T_{SCM} or T_{EMRA} phenotype, and as such were not able to predict the memory fate of acutely responding $CD8^+$ T cells. The observation that many clones had a clear effector phenotype early in the response, and subsequently transitioned into a T_{SCM} phenotype in the memory phase of the response (Figure 3C), is consistent with the model that human $CD8^+$ T_{SCM} cells undergo linear differentiation from effector cells (Ahmed et al., 2009; Akondy et al., 2017; Bannard et al., 2009), but it does not rule out asymmetric division early in the response as an explanation for memory formation (Chang et al., 2007; Lin et al., 2016). Recent evidence in mice suggests that some long-lived T_{CM} clones undergo limited effector differentiation during the primary response (Grassmann et al., 2020). We observed some clones where all cells exhibited high pCM scores at both acute and memory stages of the response. Thus, our results are consistent with the existence of multiple modes of generating long-lived memory T cell subsets.

We focused much of our attention on T cells present in the memory phase of the response, where we observed less clonal diversity, more stability of clones between time points, and more phenotypic heterogeneity according to protein and RNA expression patterns. In our *in vitro* recall assays we observed that individual T_{SCM} may not act as multipotent stem cell-like cells with the capacity to generate a diverse repertoire of effector and memory cells upon secondary activation (Fearon et al., 2001; Gattinoni et al., 2011; Kratchmarov et al., 2018). We observed that individual T_{SCM} and T_{EMRA} can each produce thousands of activated progeny upon antigen-specific re-stimulation, all of which exhibit a highly differentiated effector phenotype ($CD45RA^-CCR7^-CD127^-GZMB^+$) (Figures 4 and 5), but they still generated effector cells with distinct gene expression profiles. Notably, they exhibited limited intra-clonal variability, suggesting that each cell type gave rise to distinct effector types that were relatively similar within each clone. This is reminiscent of findings in hematopoietic stem and progenitor assays, where phenotypically similar cells transferred in bulk generate heterogeneity, while individual cells show more restricted potential to give rise to mature hematopoietic lineages (Naik et al., 2013; Weinreb et al., 2020). This suggests that the appearance of plasticity in bulk adoptive transfer studies may in fact reflect the cumulative behaviors of many restricted individual cells, consistent with observations of distinct patterns of activation after single-cell adoptive transfer of T_{CM} in mice (Graef et al., 2014). This observation is not inconsistent with the notion that an individual T cell clone, defined by all progeny of a single naive T cell, could give rise to a full range of progeny on reactivation, but argues that high degrees of plasticity are not likely to be typically encoded in a single memory T cell.

In our experiments T_{SCM} generally gave rise to effector progeny with higher expression of genes regulating tissue homing, but lower levels of inflammatory chemokines. In contrast, T_{EMRA}-derived effectors expressed higher levels of inflammatory chemokines, e.g., *CCL3* and *CCL4*. A greater understanding of the roles of T_{EMRA} *in vivo* will be necessary to define what their specific roles are in mediating secondary immunity but, given that they appear poised to express inflammatory chemokines, it is tempting to speculate that they could serve as "alarmins" in peripheral tissues, which potentially recruit progeny of T_{SCM} that are thought to be generated in secondary lymphoid tissues. If this were the case, the fact that T_{SCM} and T_{EMRA} contain clonally distinct populations of T cells may then have implications for how T cells bearing unique TCRs would have to signal from different tissues and cooperate to maximize secondary immunity. Such a two-step mechanism could also serve to prevent potentially harmful autoreactivity caused by a single cross-reactive clone.

In our clonal restimulation experiments we did not observe evidence of T_{CM}- or T_{SCM}-like cells emerging within the 20-day expansion period after *in vitro* activation of T_{SCM} clones. This stands in contrast to bulk *in vitro* restimulation assays performed on these cells where non-specific TCR crosslinking has been used to activate sorted populations of T_{SCM} or $T_{EM/EMRA}$ that were monitored for proliferative responses (Gattinoni et al., 2011). We also observed minimal gene expression of the

(C) Heatmaps depicting distribution of unique TCRβ chains in $CD62L^+CD94^-$ and $CD62L^-CD94^+$ populations for each donor/virus combination identified in bulk RNA-seq replicates. Each row represents a unique TCRβ chain. Triplicates with 50 cells for each donor/virus/subpopulation were used to estimate clonal abundance in each group.

(D) Distribution of TCRβ chains between $CD62L^+CD94^-$ and $CD62L^-CD94^+$ subsets in YFV, influenza, and CMV-responding $CD8^+$ T cells. y and x axes represent percent of reads found in $CD62L^+CD94^-$ and $CD62L^-CD94^+$ subsets, respectively. Size of circles and percentages represent percent of all clones found in either the $CD62L^+CD94^-$ or the $CD62L^-CD94^+$ subsets (black circles) or that are shared between subsets (blue circles).

$T_{CM/SCM}$-associated transcription factor *TCF7*, which was previously shown to be retained in a fraction of dividing cells after bulk activation of naive and T_{CM} cells *in vitro* as evidence of asymmetric differentiation driving phenotypic diversity in memory populations (Kratchmarov et al., 2018). It is possible that our activating conditions preclude the ability of such cells to persist or expand; however, the use of antigen-specific stimulation rather than non-specific TCR crosslinking should in theory represent a more physiological model system to study this response. An alternative explanation is that effectors generated after T_{SCM} reactivation retain epigenetic signatures that allow them to return to a T_{SCM} state once rested for extended periods after the resolution of the response (Abdelsamed et al., 2017; Youngblood et al., 2017). Interestingly, *LEF1* gene expression was observed in a fraction of responding cells from each clone (Figure 4G). Given that *LEF1* and *TCF7* are typically co-expressed in long-lived memory cells, this could be early evidence of a re-acquisition of a quiescent memory state by at least a fraction of effector cells in our assays (Zhou and Xue, 2012). The fact that some memory T cells appeared to generate cells that continue to express CD62L after secondary restimulation is reminiscent of findings in mouse models demonstrating that T_{CM} reacquire CD62L expression after secondary and tertiary infections, possibly reflecting an early re-adoption of a $T_{CM/SCM}$ phenotype (Gerlach et al., 2016).

The observation that some T_{SCM} behaved more similarly to T_{EMRA} in our reactivation assays, despite clearly expressing CCR7 protein, while some T_{EMRA} behaved as T_{SCM}-like cells, highlights that CCR7 expression alone does not always predict the quality of secondary recall responses at the clonal level. This is supported by a study in mice that observed heterogeneity in recall responses after transfer of purified T_{CM} or T_{EM}, despite a clear bias for T_{CM} to have greater expansion potential and persistence overall (Graef et al., 2014). In general, the presence of cells with distinct fates within these phenotypically defined populations would support the observation that adoptive transfer of sort-purified memory subsets yields varying degrees plasticity within responding T cells. Focusing future efforts on tracking the fate and function of T cells at the clonal level should help to better define the mechanisms determining the generation of memory T cell diversity, and perhaps more importantly how distinct clonal identities impact the quality of secondary immune responses, which may guide the development of new vaccines and adoptive transfer strategies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Flow cytometry
  - Index sorting and single-cell RNA-Seq
  - Single T cell multiple displacement amplification (MDA)
  - Amplification of TCR β chains from scMDA libraries
  - *In vitro* restimulation assays for single-cell responses
  - Bulk virus stimulation assays
  - Bulk RNA-seq
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Sequencing TCR β chains from scMDA libraries
  - Single-cell RNA-seq mapping
  - Single-cell RNA-seq normalization and batch correction
  - Classifier approach for assigning probability of a single cell being a "central memory" cell (pCM) based on protein and transcriptome analysis
  - Analysis of single-cell data based on cell type
  - *In vitro* clonal single-cell RNA-Seq analysis
  - Bulk RNA-Seq analysis
  - Ecological analysis of longitudinal clonal information in Figure 1

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.109174.

ACKNOWLEDGMENTS

We thank the Sequencing Core Facility at SciLife Labs for help with sequencing and data management, and M. Toro and S. Giatrellis (Department of Cell and Molecular Biology) and the flow cytometry core facility (Department of Medicine Huddinge) at Karolinska Institutet for assistance with FACS. This study was supported by grants from the Swedish Research Council, the Swedish Cancer Society, the Karolinska Institute, the Strategic Research Programme in Stem Cells and Regenerative Medicine at Karolinska Institutet (StratRegen), the Swedish Society for Strategic Research, and Knut och Alice Wallenbergs Stiftelse. J.E.M. was supported by a Human Frontiers Science Program Long-Term Fellowship (LT-000231/2011-L).

AUTHOR CONTRIBUTIONS

J.E.M., J.M., and J.F. designed the study and wrote the paper. J.E.M. and J.M. performed all experiments. J.E.M., J.M., L.M., J.H. M.Z., A.J.M.L., M.S., C.-J.E., P.L.S., E.B., C.T.-L., B.A., P.R., and F.P. were involved in data analysis and provided support for data analysis. J.E.M., K.B., and J.K.S. collected samples and data relevant to YFV-17D longitudinal study. B.R., S.P., and R.S. assisted with single-cell RNA-seq studies.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 28, 2020
Revised: February 15, 2021
Accepted: May 4, 2021
Published: May 25, 2021

REFERENCES

Abdelsamed, H.A., Moustaki, A., Fan, Y., Dogra, P., Ghoneim, H.E., Zebley, C.C., Triplett, B.M., Sekaly, R.P., and Youngblood, B. (2017). Human memory CD8 T cell effector potential is epigenetically preserved during in vivo homeostasis. J. Exp. Med. *214*, 1593–1606.

Ahmed, R., Bevan, M.J., Reiner, S.L., and Fearon, D.T. (2009). The precursors of memory: models and controversies. Nat. Rev. Immunol. *9*, 662–668.

Akondy, R.S., Fitch, M., Edupuganti, S., Yang, S., Kissick, H.T., Li, K.W., Youngblood, B.A., Abdelsamed, H.A., McGuire, D.J., Cohen, K.W., et al. (2017). Origin and differentiation of human memory CD8 T cells after vaccination. Nature *552*, 362–367.

Angerer, P., Haghverdi, L., Büttner, M., Theis, F.J., Marr, C., and Buettner, F. (2016). *destiny*: Diffusion maps for large-scale single-cell data in R. Bioinformatics *32*, 1241–1243.

Appay, V., Dunbar, P.R., Callan, M., Klenerman, P., Gillespie, G.M., Papagno, L., Ogg, G.S., King, A., Lechner, F., Spina, C.A., et al. (2002). Memory $CD8^+$ T cells vary in differentiation phenotype in different persistent virus infections. Nat. Med. *8*, 379–385.

Bacher, R., Chu, L.F., Leng, N., Gasch, A.P., Thomson, J.A., Stewart, R.M., Newton, M., and Kendziorski, C. (2017). SCnorm: Robust normalization of single-cell RNA-seq data. Nat. Methods *14*, 584–586.

Bachmann, M.F., Wolint, P., Schwarz, K., Jäger, P., and Oxenius, A. (2005). Functional properties and lineage relationship of $CD8^+$ T cell subsets identified by expression of IL-7 receptor α and CD62L. J. Immunol. *175*, 4686–4696.

Bannard, O., Kraman, M., and Fearon, D.T. (2009). Secondary replicative function of $CD8^+$ T cells that had developed an effector phenotype. Science *323*, 505–509.

Baron, V., Bouneaud, C., Cumano, A., Lim, A., Arstila, T.P., Kourilsky, P., Ferradini, L., and Pannetier, C. (2003). The repertoires of circulating human $CD8^+$ central and effector memory T cell subsets are largely distinct. Immunity *18*, 193–204.

Becattini, S., Latorre, D., Mele, F., Foglierini, M., De Gregorio, C., Cassotta, A., Fernandez, B., Kelderman, S., Schumacher, T.N., Corti, D., et al. (2015). T cell immunity. Functional heterogeneity of human memory $CD4^+$ T cell clones primed by pathogens or vaccines. Science *347*, 400–406.

Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. *37*, 38–44.

Balazs, A.B., Tsai, J.M., and Baltimore, D. (2013). Isolation of unknown rearranged T-cell receptors from single cells. US patent US8497071 B2, filed June 28, 2010, and granted July 30, 2013.

Benjamin, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. A Stat. Soc. *57*, 289–300.

Blom, K., Braun, M., Ivarsson, M.A., Gonzalez, V.D., Falconer, K., Moll, M., Ljunggren, H.G., Michaëlsson, J., and Sandberg, J.K. (2013). Temporal dynamics of the primary human T cell response to yellow fever virus 17D as it matures from an effector- to a memory-type response. J. Immunol. *190*, 2150–2158.

Bolotin, D.A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I.Z., Putintseva, E.V., and Chudakov, D.M. (2015). MiXCR: software for comprehensive adaptive immunity profiling. Nat. Methods *12*, 380–381.

Buchholz, V.R., Flossdorf, M., Hensel, I., Kretschmer, L., Weissbrich, B., Gräf, P., Verschoor, A., Schiemann, M., Höfer, T., and Busch, D.H. (2013). Disparate individual fates compose robust $CD8^+$ T cell immunity. Science *340*, 630–635.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Champagne, P., Ogg, G.S., King, A.S., Knabenhans, C., Ellefsen, K., Nobile, M., Appay, V., Rizzardi, G.P., Fleury, S., Lipp, M., et al. (2001). Skewed maturation of memory HIV-specific CD8 T lymphocytes. Nature *410*, 106–111.

Chang, J.T., Palanivel, V.R., Kinjyo, I., Schambach, F., Intlekofer, A.M., Banerjee, A., Longworth, S.A., Vinup, K.E., Mrass, P., Oliaro, J., et al. (2007). Asymmetric T lymphocyte division in the initiation of adaptive immune responses. Science *315*, 1687–1691.

Chattopadhyay, P.K., Betts, M.R., Price, D.A., Gostick, E., Horton, H., Roederer, M., and De Rosa, S.C. (2009). The cytolytic enzymes granyzme A, granzyme B, and perforin: Expression patterns, cell distribution, and their relationship to cell maturity and bright CD57 expression. J. Leukoc. Biol. *85*, 88–97.

Chng, M.H.Y., Lim, M.Q., Rouers, A., Becht, E., Lee, B., MacAry, P.A., Lye, D.C., Leo, Y.S., Chen, J., Fink, K., et al. (2019). Large-scale HLA tetramer tracking of T Cells during dengue infection reveals broad acute activation and differentiation into two memory cell fates. Immunity *51*, 1119–1135.e5.

Day, E.K., Carmichael, A.J., ten Berge, I.J., Waller, E.C., Sissons, J.G., and Wills, M.R. (2007). Rapid CD8+ T cell repertoire focusing and selection of high-affinity clones into memory following primary infection with a persistent human virus: human cytomegalovirus. J. Immunol. *179*, 3203–3213.

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. J. Veg. Sci. *14*, 927–930.

Durif, G., Modolo, L., Michaelsson, J., Mold, J.E., Lambert-Lacroix, S., and Picard, F. (2018). High dimensional classification with combined adaptive sparse PLS and logistic regression. Bioinformatics *34*, 485–493.

Evrony, G.D., Cai, X., Lee, E., Hills, L.B., Elhosary, P.C., Lehmann, H.S., Parker, J.J., Atabay, K.D., Gilmore, E.C., Poduri, A., et al. (2012). Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. Cell *151*, 483–496.

Fearon, D.T., Manders, P., and Wagner, S.D. (2001). Arrested differentiation, the self-renewing memory lymphocyte, and vaccination. Science *293*, 248–250.

Fisher, R.A., Corbet, A.S., and Williams, C.B. (1943). The relation between the number of species and the number of individuals in a random sample. J. Anim. Ecol. *12*, 42–58.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielson, A., and Sibert, J. (2012). AD model builder: using automatic differentiation for statistical inference of highly paramterized complex nonlinear models. Optimization Meth. Software *27*, 233–249.

Fuertes Marraco, S.A., Soneson, C., Cagnon, L., Gannon, P.O., Allard, M., Abed Maillard, S., Montandon, N., Rufer, N., Waldvogel, S., Delorenzi, M., and Speiser, D.E. (2015). Long-lasting stem cell-like memory $CD8^+$ T cells with a naïve-like profile upon yellow fever vaccination. Sci. Transl. Med. *7*, 282ra48.

Gattinoni, L., Zhong, X.S., Palmer, D.C., Ji, Y., Hinrichs, C.S., Yu, Z., Wrzesinski, C., Boni, A., Cassard, L., Garvin, L.M., et al. (2009). Wnt signaling arrests effector T cell differentiation and generates $CD8^+$ memory stem cells. Nat. Med. *15*, 808–813.

Gattinoni, L., Lugli, E., Ji, Y., Pos, Z., Paulos, C.M., Quigley, M.F., Almeida, J.R., Gostick, E., Yu, Z., Carpenito, C., et al. (2011). A human memory T cell subset with stem cell-like properties. Nat. Med. *17*, 1290–1297.

Gerlach, C., van Heijst, J.W., Swart, E., Sie, D., Armstrong, N., Kerkhoven, R.M., Zehn, D., Bevan, M.J., Schepers, K., and Schumacher, T.N. (2010). One naive T cell, multiple fates in $CD8^+$ T cell differentiation. J. Exp. Med. *207*, 1235–1246.

Gerlach, C., Rohr, J.C., Perié, L., van Rooij, N., van Heijst, J.W., Velds, A., Urbanus, J., Naik, S.H., Jacobs, H., Beltman, J.B., et al. (2013). Heterogeneous differentiation patterns of individual $CD8^+$ T cells. Science *340*, 635–639.

Gerlach, C., Moseman, E.A., Loughhead, S.M., Alvarez, D., Zwijnenburg, A.J., Waanders, L., Garg, R., de la Torre, J.C., and von Andrian, U.H. (2016). The chemokine receptor CX3CR1 defines three antigen-experienced CD8 T cell subsets with distinct roles in immune surveillance and homeostasis. Immunity *45*, 1270–1284.

Gillespie, G.M., Wills, M.R., Appay, V., O'Callaghan, C., Murphy, M., Smith, N., Sissons, P., Rowland-Jones, S., Bell, J.I., and Moss, P.A. (2000). Functional heterogeneity and high frequencies of cytomegalovirus-specific $CD8^+$ T lymphocytes in healthy seropositive donors. J. Virol. *74*, 8140–8150.

Graef, P., Buchholz, V.R., Stemberger, C., Flossdorf, M., Henkel, L., Schiemann, M., Drexler, I., Höfer, T., Riddell, S.R., and Busch, D.H. (2014). Serial transfer of single-cell-derived immunocompetence reveals stemness of $CD8^+$ central memory T cells. Immunity *41*, 116–126.

Grassmann, S., Mihatsch, L., Mir, J., Kazeroonian, A., Rahimi, R., Flommersfeld, S., Schober, K., Hensel, I., Leube, J., Pachmayr, L.O., et al. (2020). Early

emergence of T central memory precursors programs clonal dominance during chronic viral infection. Nat. Immunol. *21*, 1563–1573.

Haghverdi, L., Buettner, F., and Theis, F.J. (2015). Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics *31*, 2989–2998.

Hård, J., Al Hakim, E., Kindblom, M., Björklund, A.K., Sennblad, B., Demirci, I., Paterlini, M., Reu, P., Borgström, E., Ståhl, P.L., et al. (2019). Conbase: A software for unsupervised discovery of clonal somatic mutations in single cells through read phasing. Genome Biol. *20*, 68.

Huster, K.M., Busch, V., Schiemann, M., Linkemann, K., Kerksiek, K.M., Wagner, H., and Busch, D.H. (2004). Selective expression of IL-7 receptor on memory T cells identifies early CD40L-dependent generation of distinct $CD8^+$ memory T cell subsets. Proc. Natl. Acad. Sci. USA *101*, 5610–5615.

Jameson, S.C., and Masopust, D. (2009). Diversity in T cell memory: An embarrassment of riches. Immunity *31*, 859–871.

Jameson, S.C., and Masopust, D. (2018). Understanding subset diversity in T cell memory. Immunity *48*, 214–226.

Kaech, S.M., and Ahmed, R. (2001). Memory $CD8^+$ T cell differentiation: initial antigen encounter triggers a developmental program in naïve cells. Nat. Immunol. *2*, 415–422.

Kaech, S.M., and Cui, W. (2012). Transcriptional control of effector and memory $CD8^+$ T cell differentiation. Nat. Rev. Immunol. *12*, 749–761.

Kratchmarov, R., Magun, A.M., and Reiner, S.L. (2018). TCF1 expression marks self-renewing human $CD8^+$ T cells. Blood Adv. *2*, 1685–1690.

Kretschmer, L., Flossdorf, M., Mir, J., Cho, Y.L., Plambeck, M., Treise, I., Toska, A., Heinzel, S., Schiemann, M., Busch, D.H., and Buchholz, V.R. (2020). Differential expansion of T central memory precursor and effector subsets is regulated by division speed. Nat. Commun. *11*, 113.

Lalvani, A., Brookes, R., Hambleton, S., Britton, W.J., Hill, A.V., and McMichael, A.J. (1997). Rapid effector function in $CD8^+$ memory T cells. J. Exp. Med. *186*, 859–865.

Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics *28*, 882–883.

Lefranc, M.P., Giudicelli, V., Duroux, P., Jabado-Michaloud, J., Folch, G., Aouinti, S., Carillon, E., Duvergey, H., Houles, A., Paysan-Lafosse, T., et al. (2015). IMGT®, the international ImMunoGeneTics information system® 25 years on. Nucleic Acids Res. *43*, D413–D422.

Li, H., van der Leun, A.M., Yofe, I., Lubling, Y., Gelbard-Solodkin, D., van Akkooi, A.C.J., van den Braber, M., Rozeman, E.A., Haanen, J., Blank, C.U., et al. (2019). Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated Compartment within Human Melanoma. Cell *176*, 775–789.

Lin, W.W., Nish, S.A., Yen, B., Chen, Y.H., Adams, W.C., Kratchmarov, R., Rothman, N.J., Bhandoola, A., Xue, H.H., and Reiner, S.L. (2016). $CD8^+$ T lymphocyte self-renewal during effector cell determination. Cell Rep. *17*, 1773–1782.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lugli, E., Dominguez, M.H., Gattinoni, L., Chattopadhyay, P.K., Bolton, D.L., Song, K., Klatt, N.R., Brenchley, J.M., Vaccari, M., Gostick, E., et al. (2013). Superior T memory stem cell persistence supports long-lived T cell memory. J. Clin. Invest. *123*, 594–599.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. *17*, 10–12.

Marzo, A.L., Klonowski, K.D., Le Bon, A., Borrow, P., Tough, D.F., and Lefrançois, L. (2005). Initial T cell frequency dictates memory $CD8^+$ T cell lineage commitment. Nat. Immunol. *6*, 793–799.

Masopust, D., Vezys, V., Marzo, A.L., and Lefrançois, L. (2001). Preferential localization of effector memory cells in nonlymphoid tissue. Science *291*, 2413–2417.

Masopust, D., Ha, S.J., Vezys, V., and Ahmed, R. (2006). Stimulation history dictates memory CD8 T cell phenotype: implications for prime-boost vaccination. J. Immunol. *177*, 831–839.

Miller, J.D., van der Most, R.G., Akondy, R.S., Glidewell, J.T., Albott, S., Masopust, D., Murali-Krishna, K., Mahar, P.L., Edupuganti, S., Lalor, S., et al. (2008). Human effector and memory $CD8^+$ T cell responses to smallpox and yellow fever vaccines. Immunity *28*, 710–722.

Minervina, A.A., Pogorelyy, M.V., Komech, E.A., Karnaukhov, V.K., Bacher, P., Rosati, E., Franke, A., Chudakov, D.M., Mamedov, I.Z., Lebedev, Y.B., et al. (2020). Primary and secondary anti-viral response captured by the dynamics and phenotype of individual T cell clones. eLife *9*, e53704.

Modolo, L., and Lerat, E. (2015). UrQt: An efficient software for the unsupervised quality trimming of NGS data. BMC Bioinformatics *16*, 137.

Moon, J.J., Chu, H.H., Pepper, M., McSorley, S.J., Jameson, S.C., Kedl, R.M., and Jenkins, M.K. (2007). Naive $CD4^+$ T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. Immunity *27*, 203–213.

Naik, S.H., Perié, L., Swart, E., Gerlach, C., van Rooij, N., de Boer, R.J., and Schumacher, T.N. (2013). Diverse and heritable lineage imprinting of early haematopoietic progenitors. Nature *496*, 229–232.

Newell, E.W., Sigal, N., Bendall, S.C., Nolan, G.P., and Davis, M.M. (2012). Cytometry by time-of-flight shows combinatorial cytokine expression and virus-specific cell niches within a continuum of $CD8^+$ T cell phenotypes. Immunity *36*, 142–152.

Obar, J.J., Khanna, K.M., and Lefrançois, L. (2008). Endogenous naive $CD8^+$ T cell precursor frequency regulates primary and memory responses to infection. Immunity *28*, 859–869.

Omilusik, K.D., Best, J.A., Yu, B., Goossens, S., Weidemann, A., Nguyen, J.V., Seuntjens, E., Stryjewska, A., Zweier, C., Roychoudhuri, R., et al. (2015). Transcriptional repressor ZEB2 promotes terminal differentiation of $CD8^+$ effector and memory T cell populations during infection. J. Exp. Med. *212*, 2027–2039.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.

Picelli, S., Björklund, A.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

Picelli, S., Björklund, A.K., Reinius, B., Sagasser, S., Winberg, G., and Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. *24*, 2033–2040.

Plumlee, C.R., Sheridan, B.S., Cicek, B.B., and Lefrançois, L. (2013). Environmental cues dictate the fate of individual CD8+ T cells responding to infection. Immunity *39*, 347–356.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*, 979–982.

Reinhardt, R.L., Khoruts, A., Merica, R., Zell, T., and Jenkins, M.K. (2001). Visualizing the generation of memory CD4 T cells in the whole body. Nature *410*, 101–105.

Reinius, B., Mold, J.E., Ramsköld, D., Deng, Q., Johnsson, P., Michaëlsson, J., Frisén, J., and Sandberg, R. (2016). Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. Nat. Genet. *48*, 1430–1435.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Sallusto, F., Lenig, D., Förster, R., Lipp, M., and Lanzavecchia, A. (1999). Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. Nature *401*, 708–712.

Schenkel, J.M., Fraser, K.A., Vezys, V., and Masopust, D. (2013). Sensing and alarm function of resident memory $CD8^+$ T cells. Nat. Immunol. *14*, 509–513.

Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N., et al. (2014). Single-cell

RNA-seq reveals dynamic paracrine control of cellular variation. Nature *510*, 363–369.

Skaug, H.J., Fournier, D.A., Nielson, A., Magnusson, A., and Bolker, B.M. (2016). glmmADMB: generalized linear mixed models using AD model builder. R Package version 0.8.3.3.

Stemberger, C., Huster, K.M., Koffler, M., Anderl, F., Schiemann, M., Wagner, H., and Busch, D.H. (2007). A single naive $CD8^+$ T cell precursor can develop into diverse effector and memory subsets. Immunity *27*, 985–997.

Tubo, N.J., Pagán, A.J., Taylor, J.J., Nelson, R.W., Linehan, J.L., Ertelt, J.M., Huseby, E.S., Way, S.S., and Jenkins, M.K. (2013). Single naive $CD4^+$ T cells from a diverse repertoire produce different effector cell types during infection. Cell *153*, 785–796.

van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell *174*, 716–729.e27.

Verma, V., Jafarzadeh, N., Boi, S., Kundu, S., Jiang, Z., Fan, Y., Lopez, J., Nandre, R., Zeng, P., Alolaqi, F., et al. (2021). MEK inhibition reprograms CD8+ T lymphocytes into memory stem cells with potent antitumor effects. Nat. Immunol. *22*, 53–66.

Vezys, V., Yates, A., Casey, K.A., Lanier, G., Ahmed, R., Antia, R., and Masopust, D. (2009). Memory CD8 T-cell compartment grows in size with immunological experience. Nature *457*, 196–199.

Vuong, Q.A. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica *57*, 26.

Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F.D., and Klein, A.M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. Science *367*, eaaw3381.

Willinger, T., Freeman, T., Hasegawa, H., McMichael, A.J., and Callan, M.F. (2005). Molecular signatures distinguish human central memory from effector memory CD8 T cell subsets. J. Immunol. *175*, 5895–5903.

Xing, S., Li, F., Zeng, Z., Zhao, Y., Yu, S., Shan, Q., Li, Y., Phillips, F.C., Maina, P.K., Qi, H.H., et al. (2016). Tcf1 and Lef1 transcription factors establish CD8(+) T cell identity through intrinsic HDAC activity. Nat. Immunol. *17*, 695–703.

Youngblood, B., Hale, J.S., Kissick, H.T., Ahn, E., Xu, X., Wieland, A., Araki, K., West, E.E., Ghoneim, H.E., Fan, Y., et al. (2017). Effector CD8 T cells dedifferentiate into long-lived memory cells. Nature *552*, 404–409.

Zhou, X., and Xue, H.H. (2012). Cutting edge: generation of memory precursors and functional memory CD8+ T cells depends on T cell factor-1 and lymphoid enhancer-binding factor-1. J. Immunol. *189*, 2722–2726.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti- Human CD3e A700	BD Biosciences	Clone UCHT1, RRID:AB_396952
Anti- Human CD3e BUV805	BD Biosciences	Clone SK7, RRID:AB_2870181
Anti-Human CD3e PE-Cy5	BD Biosciences	Clone UCHT1, RRID:AB_395741
Anti- Human CD4 PE-Cy5	BD Biosciences	Clone SK3, RRID:AB_2833103
Anti- Human CD4 BUV615	BD Biosciences	Clone SK3, RRID:AB_2870258
Anti- Human CD8a APC-Cy7	BD Biosciences	Clone SK1, RRID:AB_396892
Anti- Human CD8a BV570	BioLegend	Clone RPA-T8, RRID:AB_2563213
Anti- Human CD14 V500	BD Biosciences	Clone MφP9, RRID:AB_2737727
Anti- Human CD16 BUV496	BD Biosciences	Clone 3G2, RRID:AB_2870224
Anti- Human CD19 V500	BD Biosciences	Clone HIB19, RRID:AB_10562391
Anti- Human CD27 BV786	BD Biosciences	Clone L128, RRID:AB_2744353
Anti- Human CD27 BV750	BioLegend	Clone O323, RRID:AB_2810436
Anti- Human CD45RA PE-CF594	BD Biosciences	Clone HI100, RRID:AB_11154413
Anti- Human CD45RA BUV563	BD Biosciences	Clone HI100, RRID:AB_2870211
Anti- Human CD49f BV650	BD Biosciences	Clone GoH3, RRID:AB_2744415
Anti- Human CD57 FITC	BD Biosciences	Clone NK1, RRID:AB_395986
Anti- Human CD57 BV605	BioLegend	Clone QA17A04, RRID:AB_2728426
Anti- Human CD62L (L-Selectin) BV650	BioLegend	Clone DREG-56, RRID:AB_2563821
Anti- Human CD94 (KLRD1) PE	BioLegend	Clone DX22, RRID:AB_314536
Anti- Human CD94 (KLRD1) PE-Cy7	BioLegend	Clone DX22, RRID:AB_2632753
Anti- Human CD127 PE-Cy7	BeckmanCoulter	Clone R34.34, RRID:AB_2833031
Anti- Human CD127 BV711	BioLegend	Clone A019D5, RRID:AB_2562908
Anti- Human CD184 (CXCR4) PE-Cy5	BioLegend	Clone 12G5, RRID:AB_314614
Anti- Human CD197 (CCR7) BV421	BD Biosciences	Clone 150503, RRID:AB_2728119
Anti- Human CD279 (PD-1) BV711	BD Biosciences	Clone EH12.1, RRID:AB_2738543
Anti- Human Ki67 BV786	BD Biosciences	Clone B56, RRID:AB_2732007
Anti- Human Granzyme A A700	BioLegend	Clone CB9, RRID:AB_961343
Anti- Human Granzyme B PE-CF594	BD Biosciences	Clone GB11, RRID:AB_2737618
Anti- Human TCF-1 PE	BioLegend	Clone 7F11A10, RRID:AB_2728492
Anti- Human LEF1 A488	Cell Signaling Technology	Clone C12A5, RRID:AB_10949502
Chemicals, peptides, and recombinant proteins		
HLA-A*0201-NLVPMVATV-FITC Dextramer (CMV)	Immudex	WB2132
HLA-A*0201-GILGFVFTL-APC Dextramer (FLU)	Immudex	WB2161
HLA-A*0201-LLWNGPMAV-APC Dextramer (YFV, HLA-A2/NS4B)	Immudex	WB3584
HLA-B*0702/RPIDDRFGL-APC Dextramer (YFV, HLA-B7/NS5)	Immudex	custom order
Recombinant Tn5 Transposase	Produced in house by Sandberg Lab	N/A
Peptide: LLWNGPMAV (HLA-A2/NS4B)	JPT Peptide Technologies	SP-MHCI-0063
Peptide: GILGFVFVTL (HLA-A2/Flu M1)	JPT Peptide Technologies	SP-MHCI-0004
Peptide: NLVPMVATV (HLA-A2/CMV pp65)	JPT Peptide Technologies	SP-MHCI-0005
CellTrace Violet (CTV) cell proliferation kit	Invitrogen	C34557

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant human IL-2	Peprtotech	200-02
Lymphoprep	Stem Cell Technologies	07851
Recombinant RNase Inhibitor	Takara	2313B
Triton X-100	Sigma-Aldrich	T9284
dNTP mix (10mM)	Thermo-Fisher	R0192
Agencourt Ampure XP Beads	BeckmanCoulter	A63881
EDTA (0.5M)	Thermo-Fisher	AM9261
Sodium Dodecyl sulfate 20%	Thermo-Fisher	AM9820
TE Buffer (pH 8.0) RNase Free	Thermo-Fisher	AM9849
Recombinant Phi29 DNA Polymerase	Lucigen	No Longer Produced
Phi29 DNA Polymerase	Thermo-Fisher	EP0094
Deposited data		
TCR Sequencing FASTQ Files	N/A	N/A
Single Cell RNaseq Datasets	N/A	N/A
Bulk RNaseq Datasets	N/A	N/A
Critical commercial assays		
CD8 T cell Negative Selection Kit (Human)	Miltenyi Biotec	130-096-495
KAPA Hotstart Ready Mix	Roche	KK2602
KAPA HiFi Polymerase	Roche	KK2102
Oligonucleotides		
Template Switching Oligonucleotide (TSO) 5′ AAGCAGTGGTATCAACGCAGA GTACATrGrG+G 3′	Exiqon	Picelli et al., 2013
Nextera XT 24-index kit, 96 samples	Illumina	FC-131-1001
ISPCR Primer 5′AAGCAGTGGTATCAACGCAGAGT 3′	IDT	Picelli et al., 2013
Oligo dT30VN 5′ AAGCAGTGGTATCAACGCAGA GTACT$_{30}$VN 3′	IDT	Picelli et al., 2013
TCR Sequencing Primers	IDT	Hård et al., 2019
Random Hexamer (3′ phosphorothioated) 5′ NpNpNp^{S}Np^{S}N 3′	IDT	N/A
Software and algorithms		
Salmon Light Weight Mapping Algorithm	Patro et al., 2017	https://combine-lab.github.io/salmon/
CutAdapt (v1.14)	Martin, 2011	https://cutadapt.readthedocs.io/en/stable/guide.html
UrQt	Modolo and Lerat, 2015	https://github.com/l-modolo/UrQt
FASTX-Tookit		http://hannonlab.cshl.edu/fastx_toolkit/
ImMunoGeneTics information system (IMGT)	Lefranc et al., 2015	http://www.imgt.org
Limma	Ritchie et al., 2015	https://www.bioconductor.org/packages/devel/bioc/vignettes/limma/inst/doc/usersguide.pdf
Sparse Partial Least-squares Analysis	Durif et al., 2018	https://gdurif.perso.math.cnrs.fr/tags/sparse-pls/
R studio (v1.20)	R Core Team	https://www.R-project.org/
Seurat R package (v2.3.4)	Butler et al., 2018	https://satijalab.org/seurat/
Vegan R package	Dixon, 2003	https://www.rdocumentation.org/packages/vegan/versions/2.4-2

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DeSeq2	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
MiXCR	Bolotin et al., 2015	https://mixcr.readthedocs.io/en/master/
Link to Relevant Custom Code	N/A	https://gitbio.ens-lyon.fr/lmodolo/yellow_fever

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and can be fulfilled by the Lead Contact, Jakob Michaëlsson (jakob.michaelsson@ki.se).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Custom R scripts relating to PLS and Diversity analysis are deposited at: https://gitbio.ens-lyon.fr/lmodolo/yellow_fever.

The sequencing data reported in this paper is deposited on a secure Swedish server (https://doi.org/10.17044/scilifelab.14376104). Data access requests may be submitted to the Science for Life Laboratory Data Centre through the DOI link.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human volunteers were identified from an ongoing study examining the longitudinal immune response to yellow fever vaccine YFV-17D (approved by the Regional Ethical Review Board in Stockholm, Sweden: 2008/1881-31/4, 2013/216-32, and 2014/1890-32). Informed consent was obtained from all subjects prior to study start. Two subjects (Donor 'A' (male, 32 years), Donor 'B' (female, 33 years)) were selected for scRNA-seq analysis on the basis of being positive for both HLA-A2 and HLA-B7 and having T cell responses against two yellow fever virus epitopes (HLA-A2/LLWNGPMAV and HLA-B7/RPIDDRFGL) presented on the respective HLA types. Peripheral blood was collected at days 15, 136, and 593 after vaccination for Donor 'A' and days 10, 15, 30, 90, 148 and 1401 for Donor B. For experiments using DNA amplification and TCR sequencing from single cell libraries Donor 'B' was used as well as additional donors (Donor 'C' (female, 28 years) and 'D' (male, 25 years)) with time points at days 15 and 90 for Donor 'C' and days 15, 90 and 720 for Donor 'D'. Phenotyping based on flow cytometry was performed on an additional set of donors (Donor 'E' (male, 38 years), Donor 'F' (male, 36 years), and Donor 'G' (female, 25 years)) at days 15 and 90. Viral restimulation assays were performed using donors Donor 'D', Donor 'I' (female, 23 years), and Donor 'J' (male, 44 years) and bulk RNA-seq was used to classify cell types. In all cases, peripheral blood mononuclear cells (PBMCs) were isolated by density centrifugation (Lymphoprep, StemCell Technologies, Inc., 07801) and stored in liquid nitrogen in 90% FCS and 10% DMSO until sorting. The number of cells used for TCR sequencing *ex vivo* are summarized in Table 1.

METHOD DETAILS

Flow cytometry

For all analysis of *ex vivo* $CD8^+$ T cells, cryopreserved PBMC were used for flow cytometry analysis. PBMC were thawed at 37°C and washed in FACS buffer (PBS with 2% fetal bovine serum and 2mM EDTA). Prior to staining and sorting $CD8^+$ T cells were first enriched by negative isolation of $CD8^+$ T cells by magnetic microbeads (Miltenyi). For index sorting of YFV-specific $CD8^+$ T cells, isolated $CD8^+$ T cells were first stained with HLA-A2/YFV NS4B LLW- or HLA-B7/YF NS5 RPI -dextramer APC (Immudex) for 15 min on ice and subsequently stained with CCR7 BV421 (clone 150503), CD14 V500 (clone MφP9), CD19 V500 (clone HIB19), CD27 BV786 (clone L128), CD3 Alexa700 (clone UCHT1), CD4 PE-Cy5 (clone SK3), CD45RA PE-CF594 (clone HI100), CD49f BV650 (clone GoH3), CD57 FITC (clone NK1), CD8a APC-Cy7 (clone SK1), CD95 PE (clone DX2), PD1 BV711 (clone EH12.1) (all from BD Biosciences), CD127 PE-Cy7 (clone R34.34, BeckmanCoulter), and Live/Dead fixable Aqua (ThermoFisher). After staining, the cells were washed twice and used for index-sorting as detailed below. For sorting of *in vitro* expanded single $CD8^+$ T cell clones, cells were stained with the same antibody panel as above.

For the extended phenotyping of YFV-specific $CD8^+$ T cells (shown in Figures 2A, 2B, and S2) thawed PBMC from donors E, F, and G were stained with HLA-A2/YFV- or HLA-B7/YFV-dextramer APC for 15 min on ice, and subsequently cell surface stained with the following antibodies: CD16 BUV495 (clone 3G2), CD45RA BUV563 (clone HI100), (all from BD Biosciences), CD127 BV711 (clone A019D5), CD27 BV750 (clone O323), CD57 BV605 (clone QA17A04), CD8 BV570 (clone RPA-T8), CD94 PE-Cy7 (clone DX22),

CXCR4 PE-Cy5 (clone 12G5), KLRG1 APC-Fire750 (cloneSA231A2) (all from BioLegend), and Live/Dead fixable Aqua (ThermoFisher). Subsequently, the cells were washed twice in FACS wash, fixed and permeabilized using BD transcription factor kit (BD Bioscience) and stained with CD3 BUV805 (clone SK7), CD4 BUV615 (clone SK3), Granzyme B PE-CF594 (clone GB11), Ki67 BV786 (clone B56) (all from BD Biosciences), Granzyme A Alexa700 (clone CB9), TCF-1 PE (clone 7F11A10) (both from BioLegend), and LEF-1 Alexa488 (clone C12A5, Cell Signaling Technology). After washing twice in permeabilization buffer (BD Biosciences), the cells were resuspended in FACS buffer and analyzed on a FACSymphony (BD Biosciences).

For sorting of CMV-, Flu-, and YFV-specific $CD8^+$ T cells (Figure 5), stimulated PBMC were stained with HLA-A2/CMV-dextramer FITC, HLA-A2/Influenza-dextramer APC, or HLA-A2/YFV-dextramer APC for 15min on ice, prior to staining with CD14 V500 (clone MφP9), CD19 V500 (clone HIB19), CD27 BV786 (clone L128), CD3e PE-Cy5 (clone UCHT1), CD8a APC-Cy7 (clone SK1), PD1 BV711 (clone EH12.1) (all from BD Biosciences), CD57 BV605 (clone QA17A04), CD62L BV650 (clone DREG-56), CD94 PE (clone DX22) (all from BioLegend), CD45RA PE-Cy5.5 (MEM-56, ThermoFisher), and Live/Dead fixable Aqua (ThermoFisher).

Flow cytometry data were analyzed using FlowJo (v10, FlowJo) using regular bi-variate gating. UMAP-analysis was performed in FlowJo v. 10.5.3. In Figures 2A, 2B, and S2B, 8000 bulk $CD8^+$ T cells and 2400 YFV-specific $CD8^+$ T cells collected at day 15 and 90 after vaccination were used for the UMAP analysis based on expression of CCR7, CD16, CD27, CD45RA, CD57, CD94, CD127, CXCR4, granzyme A, granzyme B, KLRG1, LEF-1, and TCF-1. In Figure 4C, UMAP analysis was based on expression of CCR7, CD27, CD45RA, CD57, CD95, CD127, and PD-1.

Index sorting and single-cell RNA-Seq

Cryopreserved PBMC isolated from donors at different time points post-vaccination were stained as described above (Flow cytometry). For sorting, single live $CD14^-CD19^-CD3^+CD8^+$Dextramer/YFV^+ cells were index-sorted in single-cell mode on a BD FACS Aria III (BD Biosciences). All other parameters (CD57, CD95, CD45RA, CD127, CD27, CCR7, CD49f) were also recorded for each sorted cell. Cells were sorted into 96-well V-bottom plates (Thermo) containing lysis buffer (0.1%Triton X-100, 2.5mM dNTP, 2.5uM Oligo-dT, 0.1U RNase inhibitor (Takara)) and immediately stored on dry ice or transferred to a −80°C freezer for long-term storage. Single cell RNA-seq was performed using the Smart-Seq2 protocol as described previously (Picelli et al., 2013). 24 cycles of amplification were used to account for the low RNA content of single lymphocytes. Final libraries containing 96-individually indexed cells were generated based on a custom version of the Nextera XT protocol using an in house Tn5 transposase (Picelli et al., 2014) and were sequenced to high depth (Table S1) on an Illumina HiSeq2000 on high output mode (Illumina).

Single T cell multiple displacement amplification (MDA)

Cryopreserved PBMC isolated from donors at different time points post-vaccination were thawed at 37°C and washed in FACS buffer. The cell pellet was re-suspended in FACS buffer containing antibodies for CD3E, CD8A, and HLA-A2/YFV- or HLA-B7/YF-dextramer APC. After a 30-minute incubation cells were washed and sorted by FACS on an Influx using index sort mode (BD Biosciences). Single YFV-specific $CD8^+$ T cells were sorted into 96-well v-bottom plates (as for scRNA-seq) containing a lysis buffer (200mM KOH, 4mM EDTA, 40mM DTT) and immediately transferred to ice for 10-minutes after which a neutralization buffer (400mM HCL, 600mM Tris-HCL pH 7.5) was added followed by an additional 10-minute incubation on ice as previously described (Evrony et al., 2012). MDA was carried out by adding a prepared cocktail containing Phi29 polymerase [Final concentration: 2U Phi29 (epicenter), 50 μM random hexamer primers (IDT), 2mM dNTP, 1x Phi29 reaction buffer (epicenter)] and incubating plates at 30°C for 6-10 hours followed by a 65°C incubation for 3 minutes to terminate the reaction. Random samples were screened to determine successful reactions using a Qubit to detect dsDNA concentration (BR Kit, ThermoFisher). Plates were subsequently stored at −20°C for downstream screening. In some cases plates were stored at −80°C after neutralization until reactions could be performed.

Amplification of TCR β chains from scMDA libraries

Approximately 100ng of amplified DNA was used to amplify TCRβ-chains from single T cell MDA libraries as previously described (Benjamin Balazs, 2013; Hård et al., 2019). Briefly, a two-step (nested) PCR was used to generate amplicons of approximately 300-400bp in length in which the 3′ end represented the junctional region (J segments) of the TCR adjacent to the complementary determining region (CDR). Primers used for the second PCR contained handles for downstream indexing steps (Hård et al., 2019). A third PCR was then used to index individual samples corresponding to wells (1-96) of the 96-well plate to which the single cells were originally sorted. After indexing samples were pooled (each plate into a single sample) and a second indexing step was performed during the addition of Illumina sequencing handles (TruSeq protocol). Primers used for the initial amplification steps were adapted from a published source (Benjamin Balazs, 2013). A summary of detected TCRs can be found in Table S1.

In vitro restimulation assays for single-cell responses

Cryopreserved peripheral blood samples from donor Donor A were washed and total $CD8^+$ T cells were processed as described for single cell RNA-seq and in a previous publication (Reinius et al., 2016). Index sorting was also performed based on the same antibody panel and protocol described for single cell RNA-seq. Single live $CD8^+CD3^+$HLA-A2/YFV-$dextramer^+$ cells were sorted directly into 96 well U-bottom plates containing 2 μg/ml peptide (LLWNGPMAV), 20U/ml IL-2, and 50.000 irradiated (40Gy) CD3-depleted autologous PBMCs in T cell media (RPMI1640 with 10% heat inactivated human AB sera, 1mM sodium pyruvate, 10mM HEPES, 50 μM 2-mercaptoethanol, 1mM L-glutamine, 100U/ml penicillin and 50 μg/ml streptomycin) and were cultured for 18 days. Every 4–5 days

half of the media was replaced with fresh T cell media containing 50U/ml IL-2 and 2 μg/ml peptide, and the wells were visually inspected for proliferation. After 18 days, nine expanded clones were selected for single-cell sorting for RNA-seq based on the presence of sufficient cell numbers (> 100) and no evidence of contaminating cells (all live cells were $CD8^+$ and bound the HLA-A2/NS4B dextramer). Single-cell RNA-seq libraries were generated as described above for the *in vivo* studies on YFV-specific $CD8^+$ T cells. A list of TCRs detected, as well as cell surface protein index data can be found in Table S1 (Index scRNaseq Data tab).

Bulk virus stimulation assays

Cryopreserved PBMC from 3 individual donors (Donors D, I, J) previously vaccinated against yellow fever virus, all of which also had detectable HLA-A2-restricted responses to Influenza virus, and two of which had detectable responses to CMV, were thawed and resuspended in PBS. The cells were subsequently labeled with 2 μM CellTrace Violet (CTV) (Invitrogen) to allow tracking cell division. After labeling the cells were washed and re-suspended in T cell media. 10^6 cells/well were cultured in T cell media supplemented with 50U/ml recombinant human IL-2 and 250ng/ml peptide (LLWNGPMAV (YFV HLA-A2), GILGFVFVTL (Flu HLA-A2), or NLVPMVATV (CMV HLA-A2)) for 7 days, after which the medium was replaced with fresh T cell media supplemented with 50U IL-2/ml and the cells were cultured for an additional 7 days. After stimulation, the cells were harvested and stained as described above (Flow cytometry). Subsequently the cells were washed twice in FACS buffer and sorted on a BD ARIA III (BD Bioscience). For sorting, live, single, $CD3^+CD8^+Dextramer^+$ cells were sorted into $CD62L^+CD94^-$ and $CD62L^-CD94^+$ subsets. For each donor, virus and population 50 cells were sorted into 96-well plates for RNaseq analysis as described above.

Bulk RNA-seq

Bulk RNA-seq was performed using the Smart-Seq2 protocol modified to accommodate larger amounts of RNA. Modifications included increasing the amount of strand-switch primer used for cDNA library generation (30 μM) and decreasing the cycles for amplification of cDNA (18 cycles). 50 cells were sorted into 96-well plates and the smart-seq2 protocol was carried out with the aforementioned modifications.

QUANTIFICATION AND STATISTICAL ANALYSIS

Sequencing TCR β chains from scMDA libraries

Pooled libraries were sequenced on a MiSeq (Illumina). After demultiplexing of Illumina sample indexes, the reverse read (R2, 150bases) Fastq file was converted to Fasta format. Identical sequences were clustered using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) FASTA Collapser. Then sequences were sorted by our 96-well indexes using the FASTX barcode splitter, and the first 44 bases were finally trimmed off using the FASTA trimmer to facilitate downstream sequence analysis. Accurate assignment of individual TCRβ-chains to single cells was accomplished by first identifying the top sequences with counts above background levels (determined by examining count frequencies in 0 cell wells or wells in which no PCR product was observed). These sequences were then submitted to the International ImMunoGeneTics (IMGT) TCR sequence identifying platform (Lefranc et al., 2015) and productive or consistently observed unproductive TCR beta chains were assigned to individual cells. For Donor B one set of samples (P11259) was sequenced using 2x300bp sequencing and these samples were analyzed directly by MIXCR and top hits were screened and selected using a similar threshold (Bolotin et al., 2015). Unproductive sequences that consistently occurred were screened by performing BLAST searches and excluded if they matched known genomic regions outside of the TCR locus.

Single-cell RNA-seq mapping

Following sequencing and demultiplexing, read pairs were trimmed from Illumina adapters using Cutadapt (version 1.14) (Martin, 2011), and UrQt was used to trim all bases with a phred quality score below 20 (Modolo and Lerat, 2015). Read pairs were subsequently aligned to the protein coding sequences of the human transcriptome (gencode.v26.pc_transcripts.fa) using Salmon (version 0.8.2) (Patro et al., 2017), and gene annotation using gencode.v26.annotation.gtf. MIXCR was used to determine TCRα- and β-chains.

Single-cell RNA-seq normalization and batch correction

For Donor A we filtered poor quality libraries using an approach similar to that described in Shalek et al. (2014). Briefly we applied the SVM-bagging algorithm to all cells comparing blank libraries (libraries generated from 0 cell wells) to all single cell libraries using the R package e1071 (https://cran.r-project.org/web/packages/e1071/index.html). Under the hypothesis that each random sub-sample should contain few cells looking like blanks, we classify each time the full dataset from the fitted SVM and record the results. We can then compute the probability of a given single cell library to look like 0 cell libraries from the number of times that it has been classified as such. Finally, we ran a last SVM classification trained on the 25% of cells with the highest and 25% of the cells with the lowest probability of looking like a blank to label the cells. With less cells for the female donor *in vivo* data, and for the *in vitro* data, we used the SVM model fitted on the male donor to classify these cells.

All cells determined to be of poor quality by this procedure were removed from downstream analysis. Normalization across wells was performed using the SCnorm procedure (Bacher et al., 2017). We processed each time-point independently to avoid removing any day specific effect. Counts obtained after normalization were used for downstream analysis.

For visualization we used Combat to remove batch effect (Leek et al., 2012).

Classifier approach for assigning probability of a single cell being a "central memory" cell (pCM) based on protein and transcriptome analysis

Human T cells appear to undergo high rates of transcriptional bursting as previously judged by the fraction (approximately 80%) of genes which are expressed in a mono-allelic pattern, leading to a substantial fraction of zeros in single cell RNaseq datasets complicating downstream analysis (Reinius et al., 2016). This limits the ability to perform clustering and classification of single cells using conventional approaches. In recent publications this has been circumvented by the creation of 'meta-cells' that compile average gene expression for cells that cluster in close proximity according to standard methods (Li et al., 2019; van Dijk et al., 2018).

To overcome this barrier and avoid averaging unrelated clones, we adopted a semi-supervised approach to rank cells based on similarity. To classify the cells, we started with a manual classification of donor Donor A *in vivo* cells. This classification was constructed manually based on high or low CCR7 protein expression and identified 101 $CCR7^+$ memory cells and 101 $CCR7^-$ effector cells. Our goal was to extend this typology to all the *in vivo* cells and not just extreme CCR7 phenotypes. For this we trained a logistic PLS model with a sparsity constraint on the manually annotated cells. For computational reasons we could not train our model on the full scRNA-seq dataset. To circumvent this problem, we build a first model on a reduced set of markers, then we used this model to detect differentially expressed genes between the two predicted groups by this first model and trained our final model on the differentially expressed genes (see Figure S3B for example of manual and PLS-based selection of $CCR7^+$ and $CCR7^-$ cells in each time point). We trained our first model using the following factors:

- *GNLY*, *GZMH*, *CCL4*, *KLRD1*, *GZMB* and *ZEB2* as typical effector/effector memory T cell genes.
- *LTB*, *TCF7*, *CCR7*, *GZMK* and *SELL* as typical central memory T cell genes.
- The cell surface proteins CCR7 and CD127 as central memory T cell markers.

This procedure is described in detail in Durif et al. (2018). The sparsity constraint selected the cell surface proteins CCR7 and CD127, and the gene *CCL4* as sufficient markers for this classification.

We used this initial PLS model on the complete Donor A day 15, 136 and 593 datasets to produce a first classification of cells into memory or effector type. This classification was built on the full Donor A dataset, crossing different batches and time-points, to try to extract a memory signature independent of these two factors.

We then performed a differential expression analysis between the two predicted cell types while accounting for batch and day effects (Table S3). The following procedure describes the way we conducted differential expression analysis throughout the study: We place ourselves in the framework of the generalized linear model with over-dispersed count distributions. We tested each gene for zero-inflation with the Vuong test (Vuong, 1989) to compare a zero-inflated Negative Binomial model to a Negative Binomial model of the count distribution. In case of zero-inflation the gene expression profile between cell types was modeled with a zero-inflated Negative Binomial distribution or a Negative Binomial distribution using the R package glmmTBM (https://github.com/glmmTMB/glmmTMB). Then, we used a likelihood ratio test with the R package lmtest (https://CRAN.R-project.org/package=lmtest) between the model accounting for cell type, batch and day effect and the model only accounting for batch and day effect, for each gene. To account for the large number of batches, we used the generalized linear model framework with mixed effect to model the batch effect as a mixed effect (Fournier et al., 2012; Skaug et al., 2016).

We then used the 322 differentially expressed genes ($FDR \leq 0.05$) (Benjamin and Hochberg, 1995), in addition to the 11 genes used before identify a memory signature only relying on scRNA-seq data (without using CCR7 protein expression) (Table S3). To achieve this goal, we fit a second logistic PLS model with a sparsity constraint on these factors and the manually annotated cells. The sparsity constraint selected the genes ABI3, CD8A, COTL1, FOXP1, HNRNPA1, LTB, NUCB2, GNLY, GZMH, CCL4, KLRD1, GZMB, ZEB2, TCF7, CCR7, GZMK and SELL as sufficient for this classification. We used this model on donor A *in vivo* dataset to compute our final classification. In addition of using the binary memory or effector classification, we also used the probability of being in the memory group (pCM-score) throughout the paper, which is a byproduct of modeling the two groups with a logistic distribution. This memory probability also contains information when studying cells of intermediary expression type, which allowed us to rank all cells.

For quality filtering we used the second logistic PLS model fitted on donor A *in vivo* data to classify donor B *in vivo* data.

The final list of genes differentially expressed between cell-types was computed using the same likelihood ratio tests as before. For each time-point, we compared a model accounting for the memory probability as a continuous factor and the batch effect as a mixed effect only impacting batch differences.

Analysis of single-cell data based on cell type

For 'training plate' data on sorted $CD8^+$ T cells based on cell type in Figure S6 the following analysis was done. We selected 2580 genes based on differential expression analysis performed at each time-point on Donor A YFV-specific cells according to scRNaseq based on effector and central memory ($CCR7^+$) differences (FDR $\leq$ 0.05). The genes were further filtered to keep only genes expressed in more than 20% of the cells in the training dataset (429 genes). We then performed a PCA on the scaled and anscomb transformed reads counts for the single-cell training dataset. We transformed the bulk training dataset in the same fashion and projected them on the single-cell PCA.

For the heatmap in Figure S6 we scored cells from the training plate using the second logistic PLS model from the donor A YFV-specific $CD8^+$ T cells.

In vitro clonal single-cell RNA-Seq analysis

Single cell RNaseq data from Figures 4 and S5 were analyzed using Seurat R package (v2.3.4) (Butler et al., 2018). First, we removed two cells based on the low number of unique genes recorded (i.e., less than 3 standard deviations from the mean). Counts for the remaining 359 cells were normalized for sequencing depth, log-transformed, and scaled to a constant factor of 10,000. Next, using a mean variability plot, we identified the top 6,456 highly variable genes (HVGs), which were used for downstream principal component analysis (PCA). We further ran Seurat's CellCycleScoring function to produce a cell cycle phase score based on normalized expression of canonical markers of S and G2M phases, and combined these to define an average cell cycle score.

We then ran differential expression analysis (DEA) to compare gene expression profiles of cells originating from $CCR7^+IL7R^+$ SCM founder cells (clones H2, E4, and B4) or $CD57^+$ EMRA founder cells (clones G8, F3, and H9). For the purpose, we used Wilcoxon's rank sum test followed by Bonferroni's correction for multiple testing and identified 154 differentially expressed genes (DEGs, adjusted p value < 0.05).

After regressing out the effect of cell cycle, we performed a second PCA, and identified relevant principal components (PC) to use for subsequent non-linear dimensionality reduction. We ran diffusion map using destiny R package (v2.12.0) (Angerer et al., 2016), considering previously identified DEGs (n = 154) and top 14 PCs. The first dimension on our diffusion map (DM1) effectively ordered cells along a SCM to EMRA continuum and was therefore used to assign a score to each cell in the dataset.

Gene Ontology Biological Process was used as an ontology source to generate lists of enriched terms based on genes contributing significantly to DM1.

Bulk RNA-Seq analysis

DeSeq2 was used to analyze bulk RNA-seq data in R studio version 1.20 (Love et al., 2014). Briefly, raw count values were used as input into deSeq2, and variance stabilizing transformation was used to transform data. Data were donor normalized using Limma (Ritchie et al., 2015). A cut-off of > 50 total counts among all samples was used to filter out low expressed genes. Genes with an adjusted p value < 0.01 and an absolute log2-fold change greater than 1 were considered as differentially expressed between paired samples of $CD62L^+CD94^-$ and $CD62L^-CD94^+$ cells. Heatmaps of gene expression were generated using Pheatmap in R.

TCR$\alpha-$ and β-chains were identified in bulk RNA-seq data using the MIXCR software package (Bolotin et al., 2015). To exclude TCR sequences which represent mapping artifacts, caused by inappropriate classification of alternative transcripts in Bulk RNA-seq datasets as TCRs, we first removed any TCR sequence that was found in wells stimulated by different viral epitopes, as it would be unlikely that a cross reactive TCR would exist for two unrelated viral peptides. We used TCRβ-chains to define individual clones although our final data includes information on both chains of the receptor. An unfiltered list of all TCR sequences identified by MIXCR is included (Table S5). Relative amounts of TCR$\beta-$chain sequences obtained for each replicate sample were determined by taking the sum of all counts for a given TCR within a single donor/virus combination and then calculating the frequency of counts observed in each sample relative to the total counts. In order to visualize differences between $CD94^+CD62L^-$ and $CD94^-CD62L^+$ subsets, samples were sorted based on abundance of a given TCR observed within each group relative to the other group.

Ecological analysis of longitudinal clonal information in Figure 1

We studied the evolution of the clonal diversity through time for each donor and antigen. We assigned the cells to clones based on shared TCRβ chain sequences (or both TCRα and β for RNaseq). Single T cells expressing a TCRβ-chain not detected in any other cells were assigned as an individual clone. We then computed the size of each clone at each time point.

Estimating the clonal diversity is similar to estimating species diversity in ecology. Thus, considering different clones as different species, we computed the Fisher alpha coefficient of each sample. We made the computation in R, with the **fisher.alpha** function for the R package **vegan**, for each time-point (Dixon, 2003; Fisher et al., 1943). The Fisher alpha coefficient can be seen as a measure of biodiversity allowing us to compare the clonal diversity between time-point (31).

To evaluate the effect of the sample size on the number of clones detected we performed a bootstrap subsampling experiment. For this analysis we identifed each sample as a specific donor/antigen/time point combination. We then drew cells with replacement from each sample data. We drew 1000 of these samples for each number of cells between 20 and up to the highest number of cells across samples for an antigen and donor. We then computed the number of clones detected in each bootstrap draw. The loess regression of the bootstrapped number of detected clone for each sampling day is represented in each figure as a smooth black line and the colored ribbon represents the 90% confidence interval.

The Fisher's alpha index is a tool to measure the diversity within a population. It's a parametric diversity index which assumes that the distribution of clone abundance follows a log distribution. We identified the Fisher's alpha indexes as a scale independent indicator of diversity as the loess regression are straight line. Therefore we chose the Fisher's alpha index to test if we can detect a significant decrease in clonal diversity between early and late (after day 15) time point. For this we pooled the cell information across sampling day within each donor and antigen. We used these two groups of cells for each donor and antigen to draw 10000 bootstrap sample, of size the lowest number of cells between the early and late group. We then computed the Fisher's alpha index for each of

the draw. We computed a 3-way ANOVA on the Fisher's alpha indexes, to test for time differences while accounting for donor and antigen differences.

Clone survival analysis. For this analysis, we studied the effect of clone size on clonal survival. Here survival refers specifically to presence/persistence within the circulation as we cannot account for cells that have either not entered the circulation or have exited the circulation permanently and seeded peripheral tissues. With 3 to 4 time points per donor and antigen (A2 or B7), we assigned the cells to clone according to shared TCRβ-chain sequences (Donor D) or both TCRα- and $\beta-$chain sequences (Donor A). We then computed the size of each clone at each time point, and again assigned a size of one, for clones not sampled at earlier time-points but found at a later time-points.

To test if highly expanding clones in the circulation at early time points (day 15; ''acute response'') had a lower probability of survival in the circulation than less expanding clones, we studied the effect of the size of the clones at each time points (excluding the last one) on the presence of the clones at the last time point sampled. For Donor D (HLA-A2 response) (Donor D_HLA-A2) we studied the effect of the clone size on days 15 and 90 and their interaction on the clone survival to the last time-point (day 720). For, Donor A (HLA-A2 response) (Donor A_HLA-A2) we studied the effect of the clone size on days 15 and 136, and their interaction on the clone survival to the last time-point (day 593). We used a logistic regression using R (32) to model these effects for the two donors.

We can see that an increase in unit of size on day 15 has a deleterious effect on the survival probability, with a log odds ratio of -1.83 (SE 0.14), and -1.76 (SE 0.17) for Donor D_HLA-A2 and Donor A_HLA-A2 respectively (p value < 0.05) (Table S2). Thus, the bigger the clone is on day 15, the lower its probability of surviving to the last time-point. We can also see that an increase in unit of size at a later time-point (day 90 or day 136) lead to an increased survival probability, with a log odds ratio of 0.73 (SE 0.18) and 0.49 (SE 0.15) for Donor D_HLA-A2 (day 90) and Donor A_HLA-A2 (day 136) respectively (p value < 0.05). This means that clones that grow later have a higher probability of surviving to the last time-point. Finally, an increase unit in size at both time-points, while increasing the probability of survival at the last time-point (p value < 0.5), has a lower positive effect on the survival probability than the size at the late time point alone. With a log odds ratio of 0.41 (SE 0.11) versus 0.73 (SE 0.18) for Donor D_HLA-A2 and 0.25 (SE 0.15) versus 0.51 (SE 0.14) for Donor A_HLA-A2.