Overline: Systems Immunology

Editor's summary

Committing to cytotoxicity

Cytotoxic CD4+ T cells (CD4-CTL) were initially identified in patients with chronic viral infections, including dengue virus (DENV) infection and these cells have been demonstrated to have a protective function in the context of DENV infection. Here Patil et al. have carried out single RNA-seq and sequenced the T cell receptors (TCR) of CD4+ T cells from human blood to identify precursors that give rise to CD4-CTL cells. They report that CD4-CTL cells undergo significant clonal expansion and that CD4-CTL precursor cells are characterized by high expression of interleukin-7 receptor. By defining the gene expression signature of CD4-CTLs and their precursors, their studies should facilitate improved vaccine design in the context of chronic viral infections.

Precursors of human CD4$^+$ cytotoxic T lymphocytes identified by single-cell transcriptome analysis

Veena S. Patil[1], Ariel Madrigal[1], Benjamin J. Schmiedel[1], James Clarke[1,2], Patrick O'Rourke[1], Aruna D. de Silva[1,3], Eva Harris[4], Bjoern Peters[1,5], Gregory Seumois[1], Daniela Weiskopf[1], Alessandro Sette[1,5] and Pandurangan Vijayanand[1,5,6]*

[1]Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, La Jolla, CA, 92037, United States.

[2]Cancer Sciences Unit, Faculty of Medicine, University of Southampton, Southampton, United Kingdom.

[3]Genetech Research Institute, Colombo, Sri Lanka, #present address: Department of Paraclinical Sciences, Faculty of Medicine, Kotelawala Defense University, Ratmalana, Sri Lanka.

[4]Division of Infectious Diseases and Vaccinology, School of Public Health, University of California, Berkeley, Berkeley, California, USA.

[5]Department of Medicine, University of California San Diego, 9500 Gilman Dr #0656, La Jolla, CA 92093, United States.

[6]Clinical and Experimental Sciences, Sir Henry Wellcome Laboratories, Faculty of Medicine University of Southampton, Southampton, United Kingdom.

*Corresponding author. Email: vijay@lji.org

ABSTRACT

$CD4^+$ cytotoxic T lymphocytes (CD4-CTLs) have been reported to play a protective role in several viral infections. However, little is known in humans about the biology of CD4-CTL generation, their functional properties and heterogeneity, especially in relation to other well-described $CD4^+$ memory T cell subsets. We performed single-cell RNA-seq in over 9000 cells to unravel CD4-CTL heterogeneity, transcriptional profile and clonality in humans. Single-cell differential gene expression analysis revealed a spectrum of known transcripts, including several linked to cytotoxic and co-stimulatory function that are expressed at higher levels in the T_{EMRA} (effector memory T cells expressing CD45RA) subset, which is highly enriched for CD4-CTLs, compared to $CD4^+$ T cells in the central and effector memory subsets (T_{CM}, T_{EM}). Simultaneous T cell antigen receptor (TCR) analysis in single cells and bulk subsets revealed that CD4-T_{EMRA} cells show marked clonal expansion compared to T_{CM} and T_{EM} cells and that the majority of CD4-T_{EMRA} were dengue virus (DENV)-specific in subjects with previous DENV infection. The profile of CD4-T_{EMRA} was highly heterogeneous across subjects, with four distinct clusters identified by the single-cell analysis. Most importantly, we identified distinct clusters of CD4-CTL effector and precursor cells in the T_{EMRA} subset; the precursor cells shared TCR clonotypes with CD4-CTL effectors and were distinguished by high expression of the interleukin-7 receptor. Our identification of a CD4-CTL precursor population may allow further investigation of how CD4-CTLs arise in humans and thus could provide insights into the mechanisms that may be utilized to generate durable and effective CD4-CTL immunity.

One sentence summary: Human $CD4^+$ cytotoxic T lymphocytes that play a vital role in anti-viral immunity are highly heterogeneous.

INTRODUCTION

Following exposure to pathogens, naive $CD4^+$ helper T lymphocytes (T_H) differentiate into memory and effector T_H cell subsets: tissue resident memory (T_{RM}) cells, which are mainly retained in the tissues, and central memory (T_{CM}) and effector memory (T_{EM}) cells, which recirculate between the blood and lymphoid organs or tissues, respectively (*1, 2*). In addition, T_H cell subsets have been classified based on their cytokine profile and functional properties into T_H1, T_H2, T_H17, T_H* ($T_H1/17$), regulatory (T_{REG}) and follicular helper (T_{FH}) T cell subsets (*3, 4*). While T lymphocytes with cytotoxic function (CTLs) are predominantly restricted to conventional MHC class I-restricted $CD8^+$ T lymphocytes, the existence of MHC class II-restricted T_H cells with cytotoxic potential (CD4-CTLs) in humans, non-human primates and mice has been reported for many decades (*5, 6*). However, compared to the other T_H subsets, the molecular and epigenetic mechanisms that drive the differentiation, maintenance and function of human CD4-CTLs are poorly understood, mainly because of the lack of precise definition of the nature of this subset in humans.

CD4-CTLs were initially reported in humans with chronic viral infections such as human cytomegalovirus (hCMV), human immunodeficiency virus (HIV), dengue virus (DENV) and hepatitis C virus (HCV) (*5, 7-15*). CD4-CTLs have also been detected in mouse lungs as early as one week after acute influenza viral infection (*16, 17*). Most importantly, the magnitude of the CD4-CTL response has been associated with better clinical outcomes in both acute and chronic viral infections, implying that CD4-CTLs are an important component of the protective immune responses to viral infections (*6*). Furthermore, expansion of CD4-CTLs has been observed in subjects carrying HLA alleles associated with protection from severe dengue disease (*7*). Thus, eliciting a strong CD4-CTL response is considered an important goal of vaccination against certain viral infections (*16, 18-20*). Notably, the highly effective yellow fever vaccine has been shown to elicit a strong CD4-CTL response, which is required for protection against fatal infection in mouse models (*18*). CD4-CTLs have also been linked to protective anti-tumor immune responses, especially in virally-induced tumors (*18*).

Given the importance of CD4-CTLs in acquired cellular immunity, we present here the single-cell transcriptomic and T cell antigen receptor (TCR) analysis of circulating human CD4-CTLs. CD4-CTLs were highly enriched in the effector memory T cells expressing CD45RA (CD4-T_{EMRA}) subset, and displayed

significant intra- and inter-subject heterogeneity. We show that the magnitude of CD4-T_{EMRA} response is linked to the degree of clonal expansion and cytotoxicity profile of CD4-CTLs. Besides a comprehensive definition of the transcriptional program of conventional CD4-CTL effectors, we identified precursor cells sharing TCR clonotypes with CD4-CTL effectors that were distinguished by higher expression of the interleukin (IL)-7 receptor (IL7R).

RESULTS

Cytotoxicity-related genes are enriched in CD4-T_{EMRA} subset

Human CD4-CTLs are enriched in CD4-T_{EMRA} subset (defined as $CD3^+CD4^+CD45RA^+CCR7^-$ cells) (**Fig. 1A**), most notably in subjects with previous DENV and CMV infection (*5, 7-10*). To capture the extent of cellular heterogeneity among human CD4-CTLs, we performed single-cell RNA-seq in over 9000 cells isolated *ex vivo* from the T_{EMRA} subset, and, as a control, in T_H subsets that contain relatively fewer CD4-CTLs, such as effector memory ($CD3^+CD4^+CD45RA^-CCR7^-$ cells, T_{EM}) and central memory ($CD3^+CD4^+CD45RA^-CCR7^+$ cells, T_{CM}) subsets (**Fig. 1A** and table S1) (*21*). Using complementary methods of single-cell differential gene expression analysis (supplementary materials and methods, (*22, 23*)), we compared the full-length transcriptome of $CD4^+$ T lymphocytes present in T_{EMRA}, T_{EM} and T_{CM} subsets from three subjects with previous DENV infection, carrying the HLA allele (DRB1*0401) previously reported to be protective against severe dengue disease (**Fig. 1B** and table S1) (*7, 24*).

We found 111 'T_{EMRA}-enriched' transcripts with significantly higher mean expression in single cells from the T_{EMRA} subset compared to the T_{EM} and T_{CM} subsets (supplementary materials and methods, **Fig. 1B** and table S2). These included several transcripts linked to the cytotoxic function of $CD8^+$ T lymphocytes and natural killer (NK) cells such as *GZMB*, *PRF1*, *GZMH*, *GNLY*, *CCL4*, *CTSW*, *FCRL6*, *SPON2*, *CX3CR1*, *S1PR5*, *NKG7*, *CD244* (*6, 25*) (**Fig. 1**, B and C); we confirmed the expression of some of these transcripts (*CX3CR1*, *GPR56*, *CD244*, *CD314*, *KLRG1*, *GZMB* and *PRF1*) at the protein level (**Fig. 1D**). Gene set enrichment analysis (GSEA) (*26*) and ingenuity pathway analysis (IPA) of 'T_{EMRA}-enriched' transcripts also revealed significant overrepresentation of cytotoxicity signature genes in T_{EMRA} subset (**Fig. 1**, E and F). Transcripts encoding transcription factors (TFs) related to CTL function such as ZNF683 (Hobit), Eomes and

T-bet, encoded by *TBX21*, were also expressed at higher levels in single cells from the T_{EMRA} subset (**Fig. 1,** B and C). ZNF683 has recently been shown to identify human CD4-CTLs (*24*), and Eomes and T-bet appear to be important in the development of CD4-CTLs (*27, 28*). These results confirm that human CD4-CTLs are highly enriched in the T_{EMRA} subset. Co-expression analysis of 'T_{EMRA}-enriched' transcripts also revealed a number of genes (*PFN1*, *PFN1P1*, *EFHD2*, *VCL*, *DIP2A*, *SYNE1*, *PLEK*) (*29-32*) whose expression was highly correlated with cytotoxicity signature genes, suggesting that products of these genes may also play important roles in the development or function of CD4-CTLs (**Fig. 1G**).

CD4-CTLs show marked clonal expansion

Given that memory CD4-CTLs are mainly generated following exposure to certain viruses such as DENV or CMV (*5, 7-10*), we expected to see a more restricted TCR repertoire, *i.e.*, greater clonal expansion in CD4-T_{EMRA} subset compared to $CD4^+$ T lymphocytes in the T_{EM} or T_{CM} subsets, which harbor a more common memory pool. We performed parallel analysis of the TCR repertoire in single cells by decoding the full-length transcriptome profiles generated by the Smart-seq2 assay (*33, 34*). Using the TraCer software (*35*), we reconstructed the TCRβ chains in 41-89% of single cells, the TCRα chain in 31-81%, and both chains in 22-70% of cells across all memory subsets (table S3). As expected, a greater clonal expansion was observed in T_{EMRA} subset compared to other subsets as shown by highly interconnected clonotype network graphs for single cells from subject #6 (**Fig. 2A**). Further, the analysis of single cells that shared TCRα or TCRβ chain clonotypes showed that over 50% of cells in the T_{EMRA} subset were clonally expanded (**Fig. 2B**, left 2 graphs). To address the rare possibility of independent cells sharing one of the TCR chains, we also analyzed single cells that shared both TCRα and TCRβ chain clonotypes and found that ~46% of cells in the T_{EMRA} subset (CD4-CTL enriched) were clonally expanded compared to only ~5% and none of the cells in T_{EM} and T_{CM} subsets, respectively (**Fig. 2B**, right graph and table S4). Together these results suggested a highly restricted TCR repertoire in T_{EMRA} subset (**Fig. 2**, A and B). The clonally expanded cells had a higher mean expression of cytotoxic signature genes ('T_{EMRA}-enriched' gene set) (**Fig. 2**, C and D), suggestive of greater effector potential (*6, 25, 36*).

To probe the pathogen specificity of clonally expanded CD4-T_{EMRA} cells (CD4-CTL enriched), we first determined the TCR clonotype of single cells in the CD4-T_{EMRA} subset that responded *ex vivo* to a pool of DENV-specific peptides (*7, 37*) from four subjects with previous DENV infection (**Fig. 2**, E and F and fig. S1, and table S4). Next, we asked what fraction of these DENV-specific TCR clonotypes were present in the general pool of cells present in the T_{EMRA} subset from the same subject. Strikingly, on an average 64% (n=4) of the clonally expanded cells in the general T_{EMRA} population carried the DENV-specific TCR clonotypes, and in all subjects one or both of the top two clonally expanded clonotypes in the T_{EMRA} population was always DENV-specific (**Fig. 2**, E and F, fig. S1, and table S4), which suggested that the majority of clonally expanded cells in the T_{EMRA} population in these individuals were specific for DENV.

CD4-CTLs (CD4-T_{EMRA} cells) are heterogeneous across subjects

We next asked if the clonality and transcriptome of CD4-CTLs differed between subjects with and without previous DENV infection, or among subjects across different geographical locations (Sri Lanka (Asia), The Americas (Nicaragua and San Diego, California)). We observed a wide range in the proportion (0-88%, median 44%) of clonally expanded cells across the 12 subjects with no major differences in their proportion when classifying subjects based on prior DENV infection status or geographical location (Sri Lanka (Asia) *versus* The Americas), Table S1). Interestingly, when we compared the percentage of clonally expanded cells between subjects with higher *versus* lower proportion of $CD4^+$ T lymphocytes in the T_{EMRA} subset (classified as T_{EMRA}^{high} or T_{EMRA}^{low} subjects), we observed greater clonal expansion in T_{EMRA}^{high} subjects (**Fig. 3A**). Consistent with this finding, there was a positive correlation between the proportion of $CD4^+$ T lymphocytes in the T_{EMRA} subset and the percentage of clonally expanded cells, suggesting that subjects with a larger T_{EMRA} pool have a greater degree of clonal expansion (**Fig. 3B**).

In most subjects, we observed sharing of a single unique TCRα and TCR β chain clonotype in a large fraction of CD4-T_{EMRA} cells, as examples in subject #4 58% (11/19), subject #1 50% (21/42), subject #2 41% (24/58), subject #5 37% (11/30), subject #8 22% (6/27), subject #3 16% (4/25) and subject #6 10% (7/67) of cells (**Fig. 3C**). Considering some of these subjects were previously infected with DENV ($DENV^+$ subjects) raised the hypothesis that selection and expansion of such TCR clonotypes may be linked to the DENV

infection. Even in subjects without previous DENV infection (DENV$^-$ subjects), we observed very high levels of clonal expansion (**Fig. 3C**), suggesting that other infections, perhaps CMV, which is common in the general population, may also contribute to the preferential expansion of some CD4-CTL clones.

Besides the heterogeneity in clonality, we also observed marked variability in the expression of 'T_{EMRA}-enriched' transcripts in CD4-T_{EMRA} cells across the study subjects (**Fig. 3**, D and E). For several cytotoxicity-related transcripts, the expression pattern was highly variable across the 12 subjects (**Fig. 3**, D and E). Notably, CD4-T_{EMRA} cells from subjects with a larger pre-existing T_{EMRA} pool and greater clonal expansion displayed more cytotoxic features (**Fig. 3D**). In other words, cytotoxicity-related transcripts were expressed in a greater fraction of single cells or at higher mean levels; other 'T_{EMRA}-enriched' transcripts such as *ZNF683*, *PRSS23*, *FCRL6*, *IFIT2* also showed a similar pattern (**Fig. 3**, D and E). We confirmed at the protein level that greater proportion of CD4-T_{EMRA} cells expressed cytotoxicity-related molecules CD244, GPR56, GZMB and PRF1 in subjects with larger pre-existing T_{EMRA} pool (**Fig. 3F**). Therefore, our combined transcriptomic, protein and TCR analysis suggest that the CD4-T_{EMRA} subset exhibit quantitative as well as qualitative differences across different subjects irrespective of their DENV infection status or geographical location, and in many instances related to the CD4-CTLs function.

CD4-CTL effector cells revealed by single-cell analysis

We next asked if the heterogeneity observed in the single-cell transcriptomes of CD4-T_{EMRA} was due to the presence of multiple distinct subsets. Four clusters were revealed by unbiased clustering of CD4-T_{EMRA} cells from all 12 study subjects (917 cells) using Seurat software (*38*) (**Fig. 4A**); other methods of clustering also revealed a similar pattern and no major changes were introduced by technical and batch effects (fig. S2 and S3, and table S1**,** supplementary materials and methods) (*39*). The proportion of cells in each cluster varied greatly among subjects, with marked differences observed between subjects from Sri Lanka and The Americas (**Fig. 4B**, left graph), suggesting that the nature and type of infections may shape the molecular profiles of CD4-CTLs. Notably, clonal expansion was observed more frequently in cells from cluster 1 and 2 (**Fig. 4B**, right graph and table S4).

Single-cell differential gene expression analysis among the four clusters revealed striking differences in their molecular profiles (**Fig. 4C** and fig. S4, A, B, C and D and table S5). IPA and differential expression analysis of the transcripts enriched in clusters 1 and 2 compared to 3 and 4 showed significant overrepresentation of genes encoding products related to cytotoxicity, such as *GZMB*, *GZMH*, *PRF1*, *GNLY*, *NKG7*, *ZNF683* and *FGFBP2* (fig. S4, C and D, fig. S5 and table S5). Among these two clusters (1 and 2), cluster 2 shows further enrichment for the cytotoxicity-related genes (fig. S4A and fig. S5A), which suggested that cells in cluster 2 had higher cytotoxic potential. Pairwise comparison of the clusters (fig. S4D and table S5) showed differences in the enrichment of cytotoxicity-related genes amongst cluster 1 and 2; while several of them were expressed by both the clusters (*GZMB*, *GZMH*, *GNLY*, *NKG7*, *FASLG*, *CASP10*), many were expressed in a cluster specific manner (**Fig. 4**, D and E, fig. S4D and table S5). Interestingly, based on the differential expression pattern of cytotoxicity-related transcripts in cluster 1 (*FADD*, *IFNG*, *TNF*, *IFIT2*, *LMNA*, *CD69*, *FOS*, *JUN*, *DUSP1* and *DUSP2*) and cluster 2 (*FGFBP2*, *SPON2*, *CX3CR1*, *GPR56*, *PRF1*), we hypothesize that their cytotoxic function is preferentially mediated by the FAS/death receptor and perforin pathways, respectively, which requires further functional verification (**Fig. 4**, D and E, fig. S4, A and D and table S5).

Cells in clusters 1 and 2 expressed higher levels of *KLRG1 (40-42)* and *CX3CR1* (*7, 43*) transcripts, which are linked to effector status of memory cells, and lower levels of *CD27* and *CD28* transcripts, which encode co-stimulation molecules (*44, 45*) (**Fig. 4E** and **Fig. 5**, A and B). These results indicate that cells in clusters 1 and 2 have a terminal effector state and as such are CD4-CTL effectors with high cytotoxic potential. Genes linked to cell survival and CD4-CTL function such as *PRSS23*, *SPON2*, *TCF7*, *CRTAM*, *ZNF683* (encodes for Hobit) (*24, 46-51*) were also highly expressed in single cells from clusters 1 and 2 (**Fig. 4E**). The role of *ZNF683* (Hobit) (*24, 46, 49*) and other transcripts (*PRSS23*, *SPON2*, *TCF7*) (*47, 48*) in the survival and long-term persistence of these high cytotoxic terminal effector cells deserves further investigation (*45*).

The T_{EMRA} subset includes CD4-CTL precursor cells

We next examined the transcriptional profiles of cells in clusters 3 and 4 to evaluate their clonal origin and functional relationship to the other cell populations. Compared to cells in clusters 1 and 2 (referred as CD4-

CTL effectors), those in clusters 3 and 4 expressed lower levels of *KLRG1* transcripts and higher levels of transcripts encoding the IL-7 receptor (*IL7R*, *CD127*), which is known to play an important role in mediating the long-term homeostatic survival of naive as well as memory T lymphocytes (*52-55*) (**Fig. 5A**). Notably, the mean expression of *IL7R* transcripts in cluster 4 cells was even higher than that in T_{CM} cells (**Fig. 5**, A and B). The opposing expression pattern of *KLRG1* and *IL7R* transcripts in clusters 3 and 4 is similar to that observed in memory-precursor effector cells (MPECs, IL-7R^{high} and KLRG1^{-}) described in murine models of resolving acute infections (*52, 56*). Transcripts encoding CD27 and CD28 co-stimulation molecules, lymphotoxin beta (*LTB*) and JUNB transcription factor were also expressed at higher levels in clusters 3 and 4, similar to the pattern observed in T_{CM} cells (**Fig. 5**, B and C). However, cells in both cluster 3 and 4 were still markedly different from T_{CM} cells in that they expressed several transcripts that are specifically enriched in T_{EMRA} cells (*KLRG1*, *TBX21*, *S1PR5*, *FGFBP2*, *CCL4*, *PRF1*, *GZMH*, *GNLY*, *NKG7*, *ZEB2*, *GPR56*), albeit at lower levels compared to CD4-CTL effectors (cells in cluster 1 and 2) (**Fig. 5C** and **Fig. 4E**). Based on these results, we hypothesize that cells in cluster 3 and 4 likely represent memory precursor cells for the cells present in cluster 1 or 2 that have a terminal CD4-CTL effector phenotype.

Given that surface expression of IL-7 receptor (CD127) can be readily determined by flow cytometry, we looked for the presence of potential CD4-CTL precursors within the CD4-T_{EMRA} subset from an independent cohort of healthy subjects. For these studies, we capitalized on the La Jolla Institute cohort of 89 healthy subjects of which 15 subjects had longitudinal samples (2 time points) (104 samples in total). As expected there was a large variation (0.12 to 15.2%) in the proportion of CD4-T_{EMRA} subset across the study subjects (**Fig. 5D**). Based on the surface expression pattern of IL-7 receptor, cells in the CD4-T_{EMRA} subset were classified into IL7R^{high} and IL7R^{-} (**Fig. 5E**). The expression level of IL7 receptor on IL7R^{high} T_{EMRA} cells was similar to that observed in T_{CM} and T_N CD4^{+} T lymphocytes (**Fig. 5E** and fig. S5B). These cells (IL7R^{high} T_{EMRA} cells) may represent the CD4-CTL precursors defined by our single-cell transcriptome analysis. In longitudinal samples obtained at 3-6 month intervals, the proportion of cells in the CD4-T_{EMRA} subset was remarkably stable (**Fig. 5F**), however, in one subject, we noted a marked (>8-fold) expansion of the CD4-T_{EMRA} subset; interestingly, this expansion was mainly confined to the IL7R^{-} T_{EMRA} cells (CD4-CTL effectors),

and not the IL7R^{high} T_{EMRA} (CD4-CTL precursors), perhaps reflecting an expansion of effectors in response to an infection that occurred during the interval period (**Fig. 5**, F and G).

To further confirm that IL7R^{high} T_{EMRA} subset (CD4-CTL precursors) shared the molecular profile of both T_{CM} (long term memory cells) and cytotoxic cells, we isolated IL7R^{high} T_{EMRA} and IL7R^{-} T_{EMRA} (CD4-CTL effectors) along with T_{CM} and T_{EM} subsets and performed RNA-seq in two longitudinal samples from five subjects (**Fig. 5H**). Consistent with our single-cell cluster analysis (**Fig. 5C**), IL7R^{high} T_{EMRA} subset shared molecular features of both T_{CM} (memory precursor cells) and cytotoxic cells (IL7R^{-} T_{EMRA}, CD4-CTL effectors) that were stable over two longitudinal visits (**Fig. 5H** and table S6). The IL7R^{high} T_{EMRA} subset (CD4-CTL precursors) expressed at the protein level both cytotoxicity related molecules (GPR56 and CD244) and a co-stimulatory molecule (CD28, T_{CM} enriched) (**Fig. 5I**). Together these results show that cells in IL7R^{high} T_{EMRA} subset have properties of memory precursor cells and cytotoxic cells.

CD4-CTL precursors share TCRs with CD4-CTL effectors

In order to verify our hypothesis, we analyzed the overlap in the TCR clonotypes of CD4-CTL effectors (clusters 1 and 2) with those from CD4-CTL precursors (clusters 3 and 4). In a total of 5/12 subjects, TCR clonotypes were shared between CD4-CTL precursors and effectors (**Fig. 2A** and **Fig. 6A** and fig. S6A and table S4). As examples, in subject #1 where 23/80 T_{EMRA} cells had the same TCRα chain clonotype (fig. S6A) as a single cell from precursor cluster (cluster 3); in subject #12, four TCRβ chain clonotypes were shared between CD4-CTL precursors and effectors (**Fig. 6A,** clones highlighted in dotted lines with the precursor cells indicated with red arrow). A similar pattern was observed in subjects #3, #5 and #6 (**Fig. 6A** and **Fig. 2A** and table S4). To further support our inference, we constructed cell state hierarchy maps using SINCELL software (*39*) for cells in the T_{CM}, T_{EM} and T_{EMRA} subsets, and observed that CD4-CTL effectors clustered closest to the CD4-CTL precursors (**Fig. 6B** and fig. S6B). Together, these data lend support to the hypothesis that CD4-CTL effectors were mainly generated *in vivo* from distinct CD4-CTL precursor cells, although the possibility of arising from memory T_{EM} or T_{CM} cells cannot be excluded.

To support our hypothesis that CD4-CTL effectors are mainly derived from the precursor population, we performed a unique molecular identifier (UMI)-based TCR-seq assays (*57*) in T_{CM}, T_{EM}, IL7R^{high} T_{EMRA}

(CD4-CTL precursors) and IL7R^{-} T_{EMRA} (CD4-CTL effectors) subsets from 14 subjects (table S7). Consistent with single-cell TCR analysis (**Fig. 2,** A and B), the CD4-CTL precursor and effector cells in T_{EMRA} subset had highly restricted TCR repertoire, compared to cells in the T_{CM} and T_{EM} subset as shown by lower Shannon-Wiener diversity index, (**Fig. 6C**) (*58*), and contained higher percentage of expanded clonotypes (**Fig. 6**, D, E and F left panels and fig.S7 left panels). Most importantly, a larger fraction of the clonotypes in the CD4-CTL precursor cells relative to T_{EM} and T_{CM} was comprised of the expanded clonotypes (frequency ≥3) present in the CD4-CTL effector cells (**Fig. 6E** middle panel and fig. S7 middle panel). Further, the most expanded clonotype in the CD4-CTL effector cells was observed more often and at higher frequency in CD4-CTL precursor cells relative to T_{EM} and T_{CM} (**Fig. 6E** graphs right panel and fig. S7 graphs right panel).

We also performed TCR-seq in CD4 subsets from 5 subjects who provided longitudinal samples obtained at 3-6 month intervals to assess the kinetics of precursor-effector relation. To better assess the relationship between effectors and putative precursors for each donor, we determined the proportion of expanded CD4-CTL effectors clonotypes from 2nd visit (V2) that were present in the different CD4 subsets at the 1st visit (V1) (table S8). In 4 of the 5 subjects analyzed, a greater fraction of the TCR clonotypes detected in CD4-CTL effectors at V2 was shared with cells in the precursor population (IL7R^{high} T_{EMRA} subset) relative to the T_{EM} or T_{CM} subset at V1 (**Fig. 6F** middle panel, fig. S7 and table S8). Most importantly, a total of 12 of the CD4-CTL effector clonotypes from V2 were only found in CD4-CTL precursors at V1 but not in the T_{EM}, T_{CM} or CD4-CTL effectors (table S8). Together, the kinetic data from the longitudinal samples provided stronger evidence to support the precursor-effector relationship and suggested that IL7R^{high} T_{EMRA} subset is the predominant precursor/progenitor for the CD4-CTL effector cells.

Variable number of CD4-CTL precursors across subjects

The proportion of CD4-CTL precursors varied greatly among subjects, ranging from 3% to 92% of cells in the T_{EMRA} subset (fig. S8A). To determine if additional CD4-CTL subsets exist and to definitively assess the presence of CD4-CTL precursors, we analyzed single-cell transcriptomes of ten times as many cells (>6000 cells) from two subjects using the high-throughput 10X genomics platform. The sensitivity of the 10X platform, which uses beads to capture transcripts and also sequences only their 3' ends (*59*), is two-fold lower than the

Smart-seq2 protocol employed to study full-length transcriptomes of the same subjects (fig. S8, B and C). In subject #2, where 24/87 cells (28%) (fig. S8A) were previously assessed as CD4-CTL precursors, unbiased clustering of over 3000 single-cell transcriptomes revealed two dominant clusters of ~2000 and ~1000 cells each, implying that sequencing more cells does not necessarily reveal additional clusters (**Fig. 7A** left panel). These two clusters differentially expressed transcripts characteristic of CD4-CTL effectors (*GZMB*, *NKG7*) and precursors (*IL7R*, *JUNB*, *LTB*), respectively (**Fig. 7B** left panel and table S9). Even in subject #1, where only 3/87 cells (3.4%) (fig. S8A) were previously found to be CD4-CTL precursors by full-length transcriptome analysis, we observed a similar proportion of CD4-CTL precursors (116/3664 cells, 3.2%) when analyzing more cells by the 10X platform (**Fig. 7**, A and B right panels and table S9). Notably, at the single-cell level, CD4-CTL precursors in the CD4-T_{EMRA} subset were distinguished by the high expression levels of *IL7R* transcripts. The expression pattern of molecules in CD4-CTL precursor and effector cells was further confirmed by single-cell RNA-seq analysis of purified population of $IL7R^{high}$ and $IL7R^{-}$ T_{EMRA} cells (**Fig. 7C** and table S1).

DISCUSSION

CD4-CTLs have long been considered to be terminal effector cells derived from effector memory cells following persistent or repeated (long-term) antigen stimulation in the context of certain viral infections, particularly CMV and DENV (*7, 8, 14*). Consistent with the notion, our single-cell transcriptome studies of the T_{EMRA} subset did identify cells that have features of terminal CD4-CTL effector cells ($KLRG1^{high}$ and $IL7R^{low}$, cluster 1 and 2). However, we observed another population of cells in the T_{EMRA} subset that displayed a molecular program ($KLRG1^{low}$ and $IL7R^{high}$, cluster 3 and 4) indicative of memory precursor cells, intermixed with several features of CD4-CTLs (*KLRG1*, *TBX21*, *S1PR5*, *FGFBP2*, *CCL4*, *PRF1*, *GZMH*, *GNLY*, *NKG7*, *ZEB2*, *GPR56*), albeit less prominently expressed than in the CD4-CTL effectors. Bulk and single-cell transcriptome analysis of the purified $IL7R^{high}$ and $IL7R^{-}$ T_{EMRA} subset further confirmed our observation from single-cell transcriptome analysis of total T_{EMRA} subset. In several subjects, we found cells in the CD4-CTL effector subset that shared their TCR clonotypes with cells in the precursor subset. TCR-seq analysis in longitudinal samples further confirmed this finding and provided more evidence for the relation between CD4-

CTL precursors and effectors. Though the kinetic studies support the development of effectors from precursor cells, the nature of our studies in human subjects limits the definitive assessment of directionality, hence we cannot completely exclude the possibility that the progenitor cell generated immediately after infection is the cytotoxic effector cell that eventually reverts. By defining the single-cell transcriptional program of CD4-CTL precursor cells, we have identified a number of new molecules potentially important for their differentiation and function, and represent attractive targets for further validation studies.

We showed that surface expression of the IL7R (CD127) in T_{EMRA} cells defines a subset of cells with expression of IL7R as high as that observed in the T_{CM} subset. Given the established role of the IL7R signaling pathway in homeostatic T cell proliferation and survival, such $IL7R^{high}$ T_{EMRA} cells are likely to represent the precursors of CD4-T_{EMRA}. The isolation of a CD4-CTL precursor subset based on surface expression of IL7R would enable detailed epigenetic studies to define the nature and extent of CD4-CTL reprogramming in such precursor cells. These future studies are likely to provide insights into the molecular mechanisms that govern the early development, differentiation and function of CD4-CTLs in humans as they transit from naive to memory precursor and effector cells.

The CD4-CTL effector cells displayed a number of known cytotoxicity molecules and new players (*PFN1*, *PFN1P1*, *EFHD2*, *VCL*, *DIP2A*, *PLEK*, *SYNE1*) whose single-cell co-expression pattern suggests an important biological role for these molecules in CD4-CTL function. Further, their transcriptional program suggests that CD4-CTL effectors are terminally differentiated and likely short-lived, as they express low levels of the co-stimulatory molecules CD27 and CD28 and high levels of *KLRG1* transcripts (*41, 52, 60-62*). However, we observed marked clonal expansion of CD4-CTL effectors in several subjects, suggesting that compared to other conventional T_H effector cells, different molecular mechanisms may operate in CD4-CTL effectors to promote their long-term survival. $IL7R^{-}$ T_{EMRA} CD4-CTL effectors highly expressed several molecules that are linked to cell survival such as *CRTAM*, *ZNF683* (Hobit), *PRSS23*, *SPON2* and *TCF1* (*24, 46-51*). Therefore, we hypothesize that these candidate molecules are likely to confer long-term survival properties to CD4-CTL effectors, which warrants further functional investigation in model organisms.

Overall, our single-cell transcriptomic studies in the human CD4-T_{EMRA} population have uncovered an unprecedented level of heterogeneity, presumably created by the diverse nature of infections, and the timing

of exposures coupled with genetic diversity among our study subjects. Importantly, we identified a T_{EMRA} subset of precursor CD4-CTLs, whose isolation and further characterization may open avenues for investigating the mechanisms that govern the generation of CD4-CTLs in humans. The stem-like virus specific long lived human memory T cell subset (T_{SCM}) originating from naive T cells has been described for CD8 compartment where these cells share the molecular profiles of T_N and T_{CM} subsets (*63*). Considering the expansion of T_{EMRA} in response to viral infections such as DENV and CMV, it is possible that IL7R^{high} T_{EMRA} subset (CD4-CTL precursors) may develop from such compartment. Understanding the origins and biology of potentially long-lived CD4-CTL precursors may pave the way for developing strategies to boost durable CD4-CTL immune responses following vaccination against viral infections and cancer. A comprehensive assessment of heterogeneity in pathogen- or vaccine-epitope-specific CD4-CTLs by single-cell approaches is likely to yield insights into the nature of protective CD4-CTL response generated against specific pathogens or vaccines.

MATERIALS AND METHODS

Study design

The goal of this study was to utilize single-cell RNA-seq assay to capture the transcriptome of individual cells in CD4$^+$ memory subsets in human peripheral blood mononuclear cells (PBMC). Details on the sample collection and processing are described in supplementary materials and methods.

Flow cytometry

Human PBMCs were isolated and stained as described in supplementary materials and methods.

Single-cell RNA-seq

Single-cell RNA-seq was performed as described previously (*33, 34*) with some modifications that are described in supplementary materials and methods.

Single-cell RNA-seq and statistical analysis

Data processing and analysis was performed using R, Qlucore omics and GraphPad Prism as described in supplementary materials and methods.

TCR-seq and analysis

TCR-seq was performed and analyzed as described previously (*57, 58*). Details are described in supplementary materials and methods.

SUPPLEMENTARY MATERIALS

Materials and Methods

Fig. S1. Majority of the expanded TCR clonotypes are DENV-specific.

Fig. S2. T_{EMRA} cells cluster into four distinct clusters.

Fig. S3. Batch analysis of T_{EMRA} cells.

Fig. S4. T_{EMRA} clusters express distinct set of transcripts.

Fig. S5. Clusters 1 and 2 are enriched for transcripts involved in cytotoxicity-related pathways.

Fig. S6. Precursor cells in T_{EMRA} subset.

Fig. S7. CD4-CTL precursors share clonotypes with CD4-CTL effectors.

Fig. S8. Smart-seq2 is more sensitive compared to droplet-based approach (10X genomics).

Table S1. Summary of study subjects.

Table S2. List of differentially expressed genes between memory subsets (for data in **Fig. 1**, B and C).

Table S3. Summary of TCRα and TCRβ chains recovered from full-length single-cell transcriptomes of all single cells in the 15 subjects.

Table S4. TCRα and TCRβ chain reconstruction from full-length single-cell transcriptome of every single cell sequenced from the 15 subjects.

Table S5. List of differentially expressed genes between clusters (for data in **Fig. 4C** and fig. S4).

Table S6. List of differentially expressed genes between T_{CM} and CD4-CTL effectors (for data in **Fig. 5H**).

Table S7: TCRα and TCRβ chain sequences derived from TCR-seq analysis of 14 subjects (for data in **Fig. 6**, E and F and fig. S7).

Table S8. TCRα and TCRβ chain clonotype sharing derived from TCR-seq analysis of longitudinally collected samples from 5 subjects (for data in **Fig. 6F** and fig. S7).

Table S9. List of differentially expressed genes between the two clusters in subject #1 and 2 (for data in **Fig. 7,** A and B).

Table S10. List of antibodies used in the study.

Table S11. Primers used for TCR-seq.

Table S12. List of gene sets used for gene set enrichment analysis (GSEA) (for data in **Fig. 1E**).

Excel sheet with source data for figures shown.

REFERENCES AND NOTES

1. M. D. Rosenblum, S. S. Way, A. K. Abbas, Regulatory T cell memory. *Nature reviews. Immunology* **16**, 90-101 (2016).
2. S. L. Swain, K. K. McKinstry, T. M. Strutt, Expanding roles for CD4(+) T cells in immunity to viruses. *Nature reviews. Immunology* **12**, 136-148 (2012).
3. J. J. O'Shea, W. E. Paul, Mechanisms underlying lineage commitment and plasticity of helper CD4+ T cells. *Science* **327**, 1098-1102 (2010).
4. C. L. Arlehamn, G. Seumois, A. Gerasimova, C. Huang, Z. Fu, X. Yue, A. Sette, P. Vijayanand, B. Peters, Transcriptional profile of tuberculosis antigen-specific T cells reveals novel multifunctional features. *Journal of immunology* **193**, 2931-2940 (2014).
5. H. Cheroutre, M. M. Husain, CD4 CTL: living up to the challenge. *Semin Immunol* **25**, 273-281 (2013).
6. J. A. Juno, D. van Bockel, S. J. Kent, A. D. Kelleher, J. J. Zaunders, C. M. Munier, Cytotoxic CD4 T Cells-Friend or Foe during Viral Infection? *Frontiers in immunology* **8**, 19 (2017).
7. D. Weiskopf, D. J. Bangs, J. Sidney, R. V. Kolla, A. D. De Silva, A. M. de Silva, S. Crotty, B. Peters, A. Sette, Dengue virus infection elicits highly polarized CX3CR1+ cytotoxic CD4+ T cells associated

with protective immunity. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E4256-4263 (2015).

8. E. Derhovanessian, A. B. Maier, K. Hahnel, R. Beck, A. J. de Craen, E. P. Slagboom, R. G. Westendorp, G. Pawelec, Infection with cytomegalovirus but not herpes simplex virus induces the accumulation of late-differentiated CD4+ and CD8+ T-cells in humans. *J Gen Virol* **92**, 2746-2756 (2011).
9. A. M. Intlekofer, N. Takemoto, E. J. Wherry, S. A. Longworth, J. T. Northrup, V. R. Palanivel, A. C. Mullen, C. R. Gasink, S. M. Kaech, J. D. Miller, L. Gapin, K. Ryan, A. P. Russ, T. Lindsten, J. S. Orange, A. W. Goldrath, R. Ahmed, S. L. Reiner, Effector and memory CD8+ T cell fate coupled by T-bet and eomesodermin. *Nat Immunol* **6**, 1236-1244 (2005).
10. N. Takemoto, A. M. Intlekofer, J. T. Northrup, E. J. Wherry, S. L. Reiner, Cutting Edge: IL-12 inversely regulates T-bet and eomesodermin expression during pathogen-induced CD8+ T cell differentiation. *Journal of immunology* **177**, 7515-7519 (2006).
11. V. Appay, J. J. Zaunders, L. Papagno, J. Sutton, A. Jaramillo, A. Waters, P. Easterbrook, P. Grey, D. Smith, A. J. McMichael, D. A. Cooper, S. L. Rowland-Jones, A. D. Kelleher, Characterization of CD4(+) CTLs ex vivo. *Journal of immunology* **168**, 5954-5958 (2002).
12. J. J. Zaunders, W. B. Dyer, B. Wang, M. L. Munier, M. Miranda-Saksena, R. Newton, J. Moore, C. R. Mackay, D. A. Cooper, N. K. Saksena, A. D. Kelleher, Identification of circulating antigen-specific CD4+ T lymphocytes with a CCR5+, cytotoxic phenotype in an HIV-1 long-term nonprogressor and in CMV infection. *Blood* **103**, 2238-2247 (2004).
13. P. J. Norris, H. F. Moffett, O. O. Yang, D. E. Kaufmann, M. J. Clark, M. M. Addo, E. S. Rosenberg, Beyond help: direct effector functions of human immunodeficiency virus type 1-specific CD4(+) T cells. *Journal of virology* **78**, 8844-8851 (2004).
14. E. M. van Leeuwen, E. B. Remmerswaal, M. T. Vossen, A. T. Rowshani, P. M. Wertheim-van Dillen, R. A. van Lier, I. J. ten Berge, Emergence of a CD4+CD28- granzyme B+, cytomegalovirus-specific T cell subset after recovery of primary cytomegalovirus infection. *Journal of immunology* **173**, 1834-1841 (2004).

15. N. Aslan, C. Yurdaydin, J. Wiegand, T. Greten, A. Ciner, M. F. Meyer, H. Heiken, B. Kuhlmann, T. Kaiser, H. Bozkaya, H. L. Tillmann, A. M. Bozdayi, M. P. Manns, H. Wedemeyer, Cytotoxic CD4 T cells in viral hepatitis. *Journal of viral hepatitis* **13**, 505-514 (2006).

16. D. M. Brown, S. Lee, L. Garcia-Hernandez Mde, S. L. Swain, Multifunctional CD4 cells expressing gamma interferon and perforin mediate protection against lethal influenza virus infection. *Journal of virology* **86**, 6792-6803 (2012).

17. K. K. McKinstry, T. M. Strutt, Y. Kuang, D. M. Brown, S. Sell, R. W. Dutton, S. L. Swain, Memory CD4+ T cells protect against influenza through multiple synergizing mechanisms. *The Journal of clinical investigation* **122**, 2847-2856 (2012).

18. A. M. Watson, L. K. Lam, W. B. Klimstra, K. D. Ryman, The 17D-204 Vaccine Strain-Induced Protection against Virulent Yellow Fever Virus Is Mediated by Humoral Immunity and CD4+ but not CD8+ T Cells. *PLoS pathogens* **12**, e1005786 (2016).

19. A. J. Vogel, D. M. Brown, Single-Dose CpG Immunization Protects Against a Heterosubtypic Challenge and Generates Antigen-Specific Memory T Cells. *Frontiers in immunology* **6**, 327 (2015).

20. K. Terahara, H. Ishii, T. Nomura, N. Takahashi, A. Takeda, T. Shiino, Y. Tsunetsugu-Yokota, T. Matano, Vaccine-induced CD107a+ CD4+ T cells are resistant to depletion following AIDS virus infection. *Journal of virology* **88**, 14232-14240 (2014).

21. Y. D. Mahnke, T. M. Brodie, F. Sallusto, M. Roederer, E. Lugli, The who's who of T-cell differentiation: human memory T-cell subsets. *European journal of immunology* **43**, 2797-2809 (2013).

22. G. Finak, A. McDavid, M. Yajima, J. Deng, V. Gersuk, A. K. Shalek, C. K. Slichter, H. W. Miller, M. J. McElrath, M. Prlic, P. S. Linsley, R. Gottardo, MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* **16**, 278 (2015).

23. P. V. Kharchenko, L. Silberstein, D. T. Scadden, Bayesian approach to single-cell differential expression analysis. *Nat Methods* **11**, 740-742 (2014).

24. A. E. Oja, F. A. Vieira Braga, E. B. Remmerswaal, N. A. Kragten, K. M. Hertoghs, J. Zuo, P. A. Moss, R. A. van Lier, K. P. van Gisbergen, P. Hombrink, The Transcription Factor Hobit Identifies Human Cytotoxic CD4+ T Cells. *Frontiers in immunology* **8**, 325 (2017).

25. L. G. Hidalgo, G. Einecke, K. Allanach, P. F. Halloran, The transcriptome of human cytotoxic T cells: similarities and disparities among allostimulated CD4(+) CTL, CD8(+) CTL and NK cells. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* **8**, 627-636 (2008).

26. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550 (2005).

27. E. L. Pearce, A. C. Mullen, G. A. Martins, C. M. Krawczyk, A. S. Hutchins, V. P. Zediak, M. Banica, C. B. DiCioccio, D. A. Gross, C. A. Mao, H. Shen, N. Cereb, S. Y. Yang, T. Lindsten, J. Rossant, C. A. Hunter, S. L. Reiner, Control of effector CD8+ T cell function by the transcription factor Eomesodermin. *Science* **302**, 1041-1043 (2003).

28. K. Eshima, S. Chiba, H. Suzuki, K. Kokubo, H. Kobayashi, M. Iizuka, K. Iwabuchi, N. Shinohara, Ectopic expression of a T-box transcription factor, eomesodermin, renders CD4(+) Th cells cytotoxic by activating both perforin- and FasL-pathways. *Immunol Lett* **144**, 7-15 (2012).

29. W. Witke, The role of profilin complexes in cell motility and other cellular processes. *Trends in cell biology* **14**, 461-469 (2004).

30. H. Chen, D. M. Choudhury, S. W. Craig, Coincidence of actin filaments and talin is required to activate vinculin. *The Journal of biological chemistry* **281**, 40389-40398 (2006).

31. D. M. Cohen, B. Kutscher, H. Chen, D. B. Murphy, S. W. Craig, A conformational switch in vinculin drives formation and dynamics of a talin-vinculin complex at focal adhesions. *The Journal of biological chemistry* **281**, 16006-16015 (2006).

32. N. Hoppmann, C. Graetz, M. Paterka, L. Poisa-Beiro, C. Larochelle, M. Hasan, C. M. Lill, F. Zipp, V. Siffrin, New candidates for CD4 T cell pathogenicity in experimental neuroinflammation and multiple sclerosis. *Brain : a journal of neurology* **138**, 902-917 (2015).

33. S. Picelli, O. R. Faridani, A. K. Bjorklund, G. Winberg, S. Sagasser, R. Sandberg, Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181 (2014).

34. I. Engel, G. Seumois, L. Chavez, D. Samaniego-Castruita, B. White, A. Chawla, D. Mock, P. Vijayanand, M. Kronenberg, Innate-like functions of natural killer T cell subsets result from highly divergent gene programs. *Nat Immunol* **17**, 728-739 (2016).

35. M. J. Stubbington, T. Lonnberg, V. Proserpio, S. Clare, A. O. Speak, G. Dougan, S. A. Teichmann, T cell fate and clonality inference from single-cell transcriptomes. *Nat Methods* **13**, 329-332 (2016).

36. A. Takeuchi, S. Badr Mel, K. Miyauchi, C. Ishihara, R. Onishi, Z. Guo, Y. Sasaki, H. Ike, A. Takumi, N. M. Tsuji, Y. Murakami, T. Katakai, M. Kubo, T. Saito, CRTAM determines the CD4+ cytotoxic T lymphocyte lineage. *The Journal of experimental medicine* **213**, 123-138 (2016).

37. D. Weiskopf, M. A. Angelo, A. Grifoni, P. H. O'Rourke, J. Sidney, S. Paul, A. D. De Silva, E. Phillips, S. Mallal, S. Premawansa, G. Premawansa, A. Wijewickrama, B. Peters, A. Sette, HLA-DRB1 Alleles Are Associated With Different Magnitudes of Dengue Virus-Specific CD4+ T-Cell Responses. *The Journal of infectious diseases* **214**, 1117-1124 (2016).

38. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502 (2015).

39. M. Julia, A. Telenti, A. Rausell, Sincell: an R/Bioconductor package for statistical assessment of cell-state hierarchies from single-cell RNA-seq. *Bioinformatics* **31**, 3380-3382 (2015).

40. T. Hanke, L. Corral, R. E. Vance, D. H. Raulet, 2F1 antigen, the mouse homolog of the rat "mast cell function-associated antigen", is a lectin-like type II transmembrane receptor expressed by natural killer cells. *European journal of immunology* **28**, 4409-4417 (1998).

41. D. Voehringer, M. Koschella, H. Pircher, Lack of proliferative capacity of human effector and memory T cells expressing killer cell lectinlike receptor G1 (KLRG1). *Blood* **100**, 3698-3702 (2002).

42. C. Blaser, M. Kaufmann, H. Pircher, Virus-activated CD8 T cells and lymphokine-activated NK cells express the mast cell function-associated antigen, an inhibitory C-type lectin. *Journal of immunology* **161**, 6451-6454 (1998).

43. J. P. Bottcher, M. Beyer, F. Meissner, Z. Abdullah, J. Sander, B. Hochst, S. Eickhoff, J. C. Rieckmann, C. Russo, T. Bauer, T. Flecken, D. Giesen, D. Engel, S. Jung, D. H. Busch, U. Protzer, R. Thimme, M. Mann, C. Kurts, J. L. Schultze, W. Kastenmuller, P. A. Knolle, Functional classification of memory CD8(+) T cells by CX3CR1 expression. *Nature communications* **6**, 8306 (2015).

44. V. Appay, P. R. Dunbar, M. Callan, P. Klenerman, G. M. Gillespie, L. Papagno, G. S. Ogg, A. King, F. Lechner, C. A. Spina, S. Little, D. V. Havlir, D. D. Richman, N. Gruener, G. Pape, A. Waters, P. Easterbrook, M. Salio, V. Cerundolo, A. J. McMichael, S. L. Rowland-Jones, Memory CD8+ T cells vary in differentiation phenotype in different persistent virus infections. *Nat Med* **8**, 379-385 (2002).

45. V. Appay, S. L. Rowland-Jones, Lessons from the study of T-cell differentiation in persistent human virus infection. *Semin Immunol* **16**, 205-212 (2004).

46. J. Braun, M. Frentsch, A. Thiel, Hobit and human effector T-cell differentiation: The beginning of a long journey. *European journal of immunology* **45**, 2762-2765 (2015).

47. H. S. Chan, S. J. Chang, T. Y. Wang, H. J. Ko, Y. C. Lin, K. T. Lin, K. M. Chang, Y. J. Chuang, Serine protease PRSS23 is upregulated by estrogen receptor alpha and associated with proliferation of breast cancer cells. *PLoS One* **7**, e30397 (2012).

48. F. Schmid, Q. Wang, M. R. Huska, M. A. Andrade-Navarro, M. Lemm, I. Fichtner, M. Dahlmann, D. Kobelt, W. Walther, J. Smith, P. M. Schlag, U. Stein, SPON2, a newly identified target gene of MACC1, drives colorectal cancer metastasis in mice and is prognostic for colorectal cancer patient survival. *Oncogene* **35**, 5942-5952 (2016).

49. F. A. Vieira Braga, K. M. Hertoghs, N. A. Kragten, G. M. Doody, N. A. Barnes, E. B. Remmerswaal, C. C. Hsiao, P. D. Moerland, D. Wouters, I. A. Derks, A. van Stijn, M. Demkes, J. Hamann, E. Eldering, M. A. Nolte, R. M. Tooze, I. J. ten Berge, K. P. van Gisbergen, R. A. van Lier, Blimp-1 homolog Hobit identifies effector-type lymphocytes in humans. *European journal of immunology* **45**, 2945-2958 (2015).

50. G. Jeannet, C. Boudousquie, N. Gardiol, J. Kang, J. Huelsken, W. Held, Essential role of the Wnt pathway effector Tcf-1 for the establishment of functional CD8 T cell memory. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 9777-9782 (2010).

51. X. Zhou, S. Yu, D. M. Zhao, J. T. Harty, V. P. Badovinac, H. H. Xue, Differentiation and persistence of memory CD8(+) T cells depend on T cell factor 1. *Immunity* **33**, 229-240 (2010).

52. S. M. Kaech, J. T. Tan, E. J. Wherry, B. T. Konieczny, C. D. Surh, R. Ahmed, Selective expression of the interleukin 7 receptor identifies effector CD8 T cells that give rise to long-lived memory cells. *Nat Immunol* **4**, 1191-1198 (2003).

53. C. C. Ibegbu, Y. X. Xu, W. Harris, D. Maggio, J. D. Miller, A. P. Kourtis, Expression of killer cell lectin-like receptor G1 on antigen-specific human CD8+ T lymphocytes during active, latent, and resolved infection and its relation with CD57. *Journal of immunology* **174**, 6088-6094 (2005).

54. B. Bengsch, H. C. Spangenberg, N. Kersting, C. Neumann-Haefelin, E. Panther, F. von Weizsacker, H. E. Blum, H. Pircher, R. Thimme, Analysis of CD127 and KLRG1 expression on hepatitis C virus-specific CD8+ T cells reveals the existence of different memory T-cell subsets in the peripheral blood and liver. *Journal of virology* **81**, 945-953 (2007).

55. S. H. Robbins, S. C. Terrizzi, B. C. Sydora, T. Mikayama, L. Brossay, Differential regulation of killer cell lectin-like receptor G1 expression on T cells. *Journal of immunology* **170**, 5876-5885 (2003).

56. S. M. Kaech, W. Cui, Transcriptional control of effector and memory CD8+ T cell differentiation. *Nature reviews. Immunology* **12**, 749-761 (2012).

57. M. Shugay, O. V. Britanova, E. M. Merzlyak, M. A. Turchaninova, I. Z. Mamedov, T. R. Tuganbaev, D. A. Bolotin, D. B. Staroverov, E. V. Putintseva, K. Plevova, C. Linnemann, D. Shagin, S. Pospisilova, S. Lukyanov, T. N. Schumacher, D. M. Chudakov, Towards error-free profiling of immune repertoires. *Nature methods* **11**, 653-655 (2014).

58. M. Shugay, D. V. Bagaev, M. A. Turchaninova, D. A. Bolotin, O. V. Britanova, E. V. Putintseva, M. V. Pogorelyy, V. I. Nazarov, I. V. Zvyagin, V. I. Kirgizova, K. I. Kirgizov, E. V. Skorobogatova, D. M. Chudakov, VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS computational biology* **11**, e1004503 (2015).

59. G. X. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, J. H. Bielas, Massively parallel digital transcriptional profiling of single cells. *Nature communications* **8**, 14049 (2017).

60. T. Boettler, E. Panther, B. Bengsch, N. Nazarova, H. C. Spangenberg, H. E. Blum, R. Thimme, Expression of the interleukin-7 receptor alpha chain (CD127) on virus-specific CD8+ T cells identifies functionally and phenotypically defined memory T cells during acute resolving hepatitis B virus infection. *Journal of virology* **80**, 3532-3540 (2006).

61. M. J. Fuller, D. A. Hildeman, S. Sabbaj, D. E. Gaddis, A. E. Tebo, L. Shang, P. A. Goepfert, A. J. Zajac, Cutting edge: emergence of CD127high functionally competent memory T cells is compromised by high viral loads and inadequate T cell help. *Journal of immunology* **174**, 5926-5930 (2005).

62. K. S. Lang, M. Recher, A. A. Navarini, N. L. Harris, M. Lohning, T. Junt, H. C. Probst, H. Hengartner, R. M. Zinkernagel, Inverse correlation between IL-7 receptor expression and CD8 T cell exhaustion during persistent antigen stimulation. *European journal of immunology* **35**, 738-745 (2005).

63. L. Gattinoni, E. Lugli, Y. Ji, Z. Pos, C. M. Paulos, M. F. Quigley, J. R. Almeida, E. Gostick, Z. Yu, C. Carpenito, E. Wang, D. C. Douek, D. A. Price, C. H. June, F. M. Marincola, M. Roederer, N. P. Restifo, A human memory T cell subset with stem cell-like properties. *Nature medicine* **17**, 1290-1297 (2011).

64. D. Weiskopf, M. A. Angelo, E. L. de Azeredo, J. Sidney, J. A. Greenbaum, A. N. Fernando, A. Broadwater, R. V. Kolla, A. D. De Silva, A. M. de Silva, K. A. Mattia, B. J. Doranz, H. M. Grey, S. Shresta, B. Peters, A. Sette, Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. *Proceedings of the National Academy of Sciences of the United States of America* **110**, E2046-2053 (2013).

65. N. Kanakaratne, W. M. Wahala, W. B. Messer, H. A. Tissera, A. Shahani, N. Abeysinghe, A. M. de-Silva, M. Gunasekera, Severe dengue epidemics in Sri Lanka, 2003-2006. *Emerg Infect Dis* **15**, 192-199 (2009).

66. R. J. Fernandez, S. Vazquez, Serological diagnosis of dengue by an ELISA inhibition method (EIM). *Memorias do Instituto Oswaldo Cruz* **85**, 347-351 (1990).

67. A. A. Kraus, W. Messer, L. B. Haymore, A. M. de Silva, Comparison of plaque- and flow cytometry-based methods for measuring dengue virus neutralization. *J Clin Microbiol* **45**, 3777-3780 (2007).

68. J. A. Swanstrom, J. A. Plante, K. S. Plante, E. F. Young, E. McGowan, E. N. Gallichotte, D. G. Widman, M. T. Heise, A. M. de Silva, R. S. Baric, Dengue Virus Envelope Dimer Epitope Monoclonal Antibodies Isolated from Dengue Patients Are Protective against Zika Virus. *mBio* **7**, (2016).

69. J. T. Gaublomme, N. Yosef, Y. Lee, R. S. Gertner, L. V. Yang, C. Wu, P. P. Pandolfi, T. Mak, R. Satija, A. K. Shalek, V. K. Kuchroo, H. Park, A. Regev, Single-Cell Genomics Unveils Critical Regulators of Th17 Cell Pathogenicity. *Cell* **163**, 1400-1412 (2015).

70. C. Zheng, L. Zheng, J. K. Yoo, H. Guo, Y. Zhang, X. Guo, B. Kang, R. Hu, J. Y. Huang, Q. Zhang, Z. Liu, M. Dong, X. Hu, W. Ouyang, J. Peng, Z. Zhang, Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. *Cell* **169**, 1342-1356 e1316 (2017).

71. G. Seumois, J. Zapardiel-Gonzalo, B. White, D. Singh, V. Schulten, M. Dillon, D. Hinz, D. H. Broide, A. Sette, B. Peters, P. Vijayanand, Transcriptional Profiling of Th2 Cells Identifies Pathogenic Features Associated with Asthma. *Journal of immunology* **197**, 655-664 (2016).

72. A. P. Ganesan, J. Clarke, O. Wood, E. M. Garrido-Martin, S. J. Chee, T. Mellows, D. Samaniego-Castruita, D. Singh, G. Seumois, A. Alzetani, E. Woo, P. S. Friedmann, E. V. King, G. J. Thomas, T. Sanchez-Elsner, P. Vijayanand, C. H. Ottensmeier, Tissue-resident memory features are linked to the magnitude of cytotoxic T cell responses in human lung cancer. *Nature immunology* **18**, 940-950 (2017).

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

Acknowledgments: We thank members of the Vijayanand's laboratory for critical reading of the manuscript and discussions, and the Human Immune Profiling Consortium team (HIPC) at the La Jolla Institute for valuable discussions. We thank Sandy Rosales, Shu Liang and Divya Singh from Vijayanand's group for help

with sequencing runs. We thank Cristhiam Cerpas and Angel Balmaseda for the preparation of PBMCs in Nicaragua. We thank Dr. Dmitriy M. Chudakov for sharing the detailed protocol for TCR-seq. **Funding**: This work was supported by NIH grants U19AI118626, U19AI118610 (P.V., A.S.), R01HL114093, R24AI108564, William K. Bowes Jr Foundation (P.V.), and NIH contracts no. HHSN272200900042C, HHSN27220140045C (A.S.). **Author contributions**: V.S.P., G.S., A.S., B.P., and P.V. conceived of the work; V.S.P. and P.V. designed the study and wrote the paper; V.S.P. designed and performed experiments and analyzed the data under the supervision of P.V.; A.M. assisted V.S.P. in the RNA-seq analysis and performed all R based programming; J.C. assisted V.S.P. in single-cell RNA-seq experiments; P.O'R. assisted V.S.P. and D.W. in DENV peptide stimulation experiments; B.J.S. performed IL7R FACS experiments shown in Fig. 5D-G; E.H. and A.D.dS. collected blood samples in Nicaragua and Sri Lanka and determined dengue virus sero-status respectively. **Competing interests**: The authors declare that they have no competing interests. **Data and materials availability**: The sequence data sets reported in this paper have been deposited in the Gene Expression Omnibus with accession no. GSE106544.

FIGURE LEGENDS

Fig. 1. Cytotoxicity-related transcripts are enriched in CD4-T_{EMRA} subset. (**A**) Schematic representation of the $CD4^+$ T cell subsets and their surface markers used for the study. T_N, naive T cell; T_{CM}, central memory T cell; T_{EM}, effector memory T cell; T_{EMRA}, effector memory T cell expressing CD45RA. (**B**) Single-cell RNA-seq analysis showing row-wise *z*-scores of normalized TPM (transcripts per million) counts of cells in each subset (indicated at the top) for each differentially expressed transcript (rows) obtained by pairwise comparison of T_{EMRA} *versus* T_{EM}, T_{EMRA} *versus* T_{CM} and T_{EM} *versus* T_{CM} (SCDE and MAST analysis, Benjamini-Hochberg adjusted $P < 0.05$ and ≥2-fold change). (**C**) Violin plots show the single-cell expression pattern of the indicated 'T_{EMRA}-enriched' transcripts (cytotoxicity-related and transcription factors) in the indicated subsets. The shapes represent the distribution of cells based on their $\log_2$(TPM+1) expression values (y-axis). The color scale represents the mean expression. (**D**) The contour plots show the surface expression of CD45RA, CX3CR1, GPR56, CD244, CD314 and KLRG1 and intracellular expression of GZMB and PRF1 in

singlet gated $CD3^+CD4^+CCR7^-$ cells. The numbers denote percentage of cells in each quadrant. Bar graphs beneath show the average percentages. Error bars are mean ± SEM from 9 (GZMB), 10 (CX3CR1) and 11 (other indicated proteins) subjects. ** $P < 0.001$, *** $P < 0.0005$, **** $P < 0.0001$ from Student's paired two tailed t-test. (**E**) GSEA (gene set enrichment analysis) enrichment plots for the indicated gene sets in the transcriptome of T_{EMRA} *versus* T_{EM} and T_{CM} (supplementary materials and methods). The top portion of the plot shows the running enrichment score (RES) for the gene set as the analysis walks down the ranked list of genes and reflects the degree to which the gene set is overrepresented at the top or bottom of the ranked list of genes. The middle portion of the plot shows where the members of the gene set (indicated as blue lines) appear in the ranked list of genes. The bottom portion of the plot shows the value of the ranking metric. *NES*, normalized enrichment score. (**F**) Ingenuity pathway analysis (IPA) of canonical pathways enriched in T_{EMRA} subset, *P* values calculated by Fisher Exact Test (supplementary materials and methods). (**G**) Spearman correlation plot showing co-expression of the 111 'T_{EMRA}-enriched' transcripts. Dotted black box shows a cluster of transcripts showing high correlation; the list of these transcripts is shown in the text box on the right, stars highlight the genes not previously reported to have cytotoxic function.

Fig. 2. CD4-T_{EMRA} cells show marked clonal expansion. (**A**) Clonotype network graphs of cells from the T_{EMRA}, T_{EM}, T_{CM} subset, and DENV-specific T_{EMRA} and DENV-specific T_{EM} cells from subject #6. Each circle represents a single cell; the reconstructed TCRα and TCRβ sequences for each cell are depicted as red and blue colored bars, respectively, inside each circle. Dark and light colored bars represent productive, and nonproductive TCRs, respectively. Red connecting lines indicate shared TCRα sequences, blue lines indicate shared TCRβ sequences. Red arrow indicates a precursor cell. Purple dotted line indicates a precursor cell and an effector cell sharing the same unique TCRα and TCRβ chain clonotype. (**B**) Bar graph shows the percentage of cells with TCRα (left), TCRβ (middle) and TCRαβ (right) chain clonotype frequency ≥2 (clonally expanded cells) in T_{EMRA}, T_{EM} and T_{CM} subsets. Error bars are mean ± SEM from 3 subjects with same HLA allele. * $P < 0.05$ from Student's paired two tailed t-test. (**C**) Single-cell RNA-seq analysis shows the row-wise z-score of normalized TPM counts for the indicated T_{EMRA}-enriched transcripts (SCDE and MAST analysis comparing T_{EMRA} cells with clonotype frequency ≥2 or =1, Benjamini-Hochberg adjusted $P < 0.05$ and ≥2-fold

change) in clonally expanded (unique TCRα and TCRβ chain clonotype frequency ≥2; magenta) and not expanded cells (unique TCRα and TCRβ chain clonotype frequency =1; green). The cells are arranged based the clonotype frequency and the numbers on the top in black boxes represent clonotype frequency. (**D**) Violin plots show the single-cell expression pattern of the indicated cytotoxicity-related transcripts in clonally expanded *versus* not expanded T_{EMRA} cells. The shapes represent the distribution of cells based on their log_2(TPM+1) expression values (y-axis). The color scale represents the mean expression. (**E**) Bar graph shows the number of T_{EMRA} cells with a unique TCRα and TCRβ chain clonotype for subject #6. Stars in red color indicate the clonotypes shared by DENV-specific T_{EMRA}. Pie chart shows the percentage of expanded T_{EMRA} cells that share the unique TCRα and TCTβ chain clonotypes of DENV-specific T_{EMRA} cells (pink). (**F**) Bar graph shows the percentage of the expanded T_{EMRA} cells that share TCRαβ clonotypes of DENV-specific T_{EMRA} cells. Each dot represents a subject. Error bars are mean ± SEM from 4 subjects.

Fig. 3. CD4-T_{EMRA} cells are heterogeneous across subjects. (**A**) Bar graph shows the percentage of clonally expanded T_{EMRA} cells (cells having a unique TCRα and TCRβ chain clonotype with frequency ≥2) across 12 subjects (left), and between subjects classified based on their status of previous DENV infection ($DENV^+$ and $DENV^-$), geographical location (Sri Lanka, Nicaragua and San Diego (The Americas)) and CD4-T_{EMRA} proportion (T_{EMRA}^{high}; subjects #1-6, T_{EMRA}^{low}; subjects #7-12). Each dot represents data from a single subject. Error bars are mean ± SEM. ****P* = 0.001 and ns, not significant from Student's unpaired two tailed t-test. (**B**) Correlation between the proportions of CD4-T_{EMRA} subset in $CD4^+$ T cells and clonally expanded T_{EMRA} cells (cells having a unique TCRα and TCRβ chain clonotype with frequency ≥2). *r* = Spearman correlation coefficient and *P* by Student's two tailed paired t-test. (**C**) Pie-charts show the percentage of T_{EMRA} cells with the first (clone 1), second (clone 2) most frequent unique TCRα and TCTβ chain clonotype, and other clonally expanded cells (all others) and rest of the TCRα and TCRβ chain clonotypes present in only one cell (unique clones). (**D**) Single-cell RNA-seq analysis of T_{EMRA} (blue), T_{EM} (yellow) and T_{CM} (brown) cells shows the row-wise normalized mean TPM counts of 111 'T_{EMRA}-enriched' transcripts (from Fig. 1, B and C) (rows) for each subject (column). Panel above heat map identifies the cell type and subject # (1 to 12, ordered based on the frequency of CD4-T_{EMRA} subset). (**E**) Violin plots show the single-cell expression pattern of the

indicated transcripts in T_{EMRA} subset across 12 subjects. The shapes represent the distribution of cells based on their log_2(TPM+1) expression values (y-axis). The color scale represents the mean expression. (**F**) Matched data from the same subjects showing the proportion of T_{EMRA} cells in $CD4^+$ T cell subset (blue) and the proportion of $CD244^+$ (red), $GPR56^+$ (green), $GZMB^+$ (pink) and $PRF1^+$ (dark blue) T_{EMRA} cells from 9 to 11 subjects. Each dot represents a subject. ** $P < 0.001$, *** $P < 0.0005$, **** $P < 0.0001$ from Student's paired two tailed t-test.

Fig. 4. Four distinct subsets of CD4-CTLs revealed by single-cell RNA-seq analysis. (**A**) tSNE 2D plot of single-cell RNA-seq data (n=917 cells, based on 1000 most variable transcripts) shows 4 distinct clusters of cells in the T_{EMRA} subset. (**B**) Bar graph shows the distribution of cells across 4 clusters in each indicated subject (subject numbers match those in Table 1 and table S1) classified based on geographical location (The Americas and Sri Lanka, left), and the percentage of clonally expanded cells (cells having a unique TCRα and TCRβ chain clonotype with frequency ≥2) in each cluster (right). (**C**) Single-cell RNA-seq analysis shows row-wise *z*-scores of mean normalized TPM counts of cells in each cluster for each differentially expressed transcripts (rows) obtained by pairwise comparison of each cluster *versus* the rest of the clusters (Seurat analysis, Benjamini-Hochberg adjusted $P < 0.05$) for the four clusters. (**D**) tSNE 2D plots of single-cell RNA-seq data show the expression of the indicated differentially expressed transcripts (SCDE and MAST analysis, Benjamini-Hochberg adjusted $P < 0.05$ and ≥2-fold change). *B2M* is shown as a control house-keeping gene. (**E**) Violin plots show the single-cell expression pattern of the indicated differentially expressed transcripts across four T_{EMRA} clusters along with T_{CM}. The shapes represent the distribution of cells based on their log_2(TPM+1) expression values (y-axis). The color scale represents the mean expression.

Fig. 5. CD4-CTL precursor cells in T_{EMRA} subset. (**A**) tSNE 2D plots of single-cell RNA-seq data show the expression of the indicated differentially expressed transcripts between the four clusters (SCDE and MAST analysis, adjusted $P < 0.05$ and ≥2-fold change). (**B**) Violin plots show the single-cell expression pattern of the indicated differentially expressed transcripts across four T_{EMRA} clusters along with T_{CM}. Color scale shows the log_2(TPM+1) expression values (A) and mean expression values (B). The shapes represent the

distribution of cells based on their log_2(TPM+1) expression values (y-axis) (B). (**C**) Venn diagram comparing the transcripts enriched in precursor clusters (clusters 3 and 4) *versus* effector clusters (clusters 1 and 2) and transcripts enriched in T_{CM} *versus* T_{EMRA} (left); transcripts enriched in precursor clusters (clusters 3 and 4) *versus* T_{CM} and transcripts enriched in T_{EMRA} *versus* T_{CM} (right) (gene list obtained from SCDE and MAST analysis, Benjamini-Hochberg adjusted $P < 0.05$ and ≥2-fold change). Examples of overlapping genes are shown to the right. (**D**) The percentage of CD4-T_{EMRA} subset in the $CD4^+$ T cell population across 104 samples from 89 individual subjects. (**E**) Contour plots show the surface expression of CD45RA and CCR7 (left) or CD127 (IL7R) (right) in live and singlet-gated $CD3^+CD4^+$ T cells obtained from PBMCs (peripheral blood mononuclear cells) (left). Contour plots show co-expression of CD45RA and CD127 in T_{EMRA} and T_{CM} subsets (right). The gating strategy shows $CD127^-$ and $CD127^{high}$ proportion. (**F**) The proportion of total CD4-T_{EMRA} (left), $CD127^-$ CD4-T_{EMRA} (middle), $CD127^{high}$ CD4-T_{EMRA} cells (right) in $CD4^+$ T cells from 15 subjects with two longitudinal samples (visits 1 and 2). The subject #16, marked in red color, shows marked increase in the proportion of CD4-T_{EMRA} and $CD127^-$ CD4-T_{EMRA} (effectors) cells from visit 1 to 2. (**G**) Contour plots show the surface expression of CD45RA and CCR7 (top) or CD127 (bottom) in live and singlet-gated $CD3^+CD4^+$ cells obtained from PBMCs of subject #16 at two time points (168 days apart) (visit 1, left and visit 2, right). Bottom panels show the expression of CD127 in CD4-T_{EMRA} subset for the same subject. ns, not significant by Student's two tailed paired t-test. Number inside the contour plots show the proportion of the indicated cell type. (**H**) Bulk RNA-seq analysis of $IL7R^{high}$ T_{EMRA} subset (CD4-CTL precursors), $IL7R^-$ T_{EMRA} (CD4-CTL effectors), and T_{CM} cells shows the row-wise normalized TPM counts of top 200 (100 upregulated and down regulated, based on fold change) differentially expressed transcripts obtained by pairwise comparison of $IL7R^-$ T_{EMRA} (CD4-CTL effectors) *versus* T_{CM} from the 5 subjects (DEseq2 analysis, Benjamini-Hochberg adjusted $P < 0.05$ and ≥2-fold change). Panel above heat map identifies the cell type, subject, and visit. (**I**) Contour plots show co-expression of CD244 or GPR56 with CD28 in T_{CM} (red), $IL7R^{high}$ T_{EMRA} (CD4-CTL precursors) (magenta) and $IL7R^-$ T_{EMRA} (CD4-CTL effectors) (blue) in singlet gated $CD3^+CD4^+$ T lymphocytes.

Fig. 6. CD4-CTL effectors share TCR clonotypes with CD4-CTL precursors. (**A**) Clonotype network graphs of single cells from T_{EMRA} subsets from subject #12 (top) and subject #3 (bottom). Each circle

represents a single cell; the reconstructed TCRα and TCRβ sequences for each cell are depicted as red and blue colored bars, respectively, inside each circle. Dark and light colored bars represent productive, and nonproductive TCRs, respectively. Red connecting lines indicate shared TCRα sequences, blue lines indicate shared TCRβ sequences. Red arrow indicates a precursor cell. Purple dotted line indicates a cluster of cells with intermixed precursor and effector cells sharing the same clonotype. (**B**) Cell state hierarchy map constructed for precursor cluster-T_{EMRA} (light green), effector cluster-T_{EMRA} (orange), T_{CM} (brown) and T_{EM} (yellow) cells from subject #12. The network shows connection (black line) between cells (circle). (**C** and **D**) TCR-seq assays in T_{CM}, T_{EM}, IL7R^{high} T_{EMRA} (CD4-CTL precursors) and IL7R^{-} T_{EMRA} (CD4-CTL effectors) subset from 14 subjects; bar graphs show the Shannon-Wiener diversity index obtained using V(D)J tools (C) and percentage of expanded TCRβ clonotypes (clonotype frequency ≥3) (D); error bars are mean ± SEM from 13-14 subjects. * $P < 0.05$, ** $P < 0.005$, *** $P < 0.001$, ns, not significant from Student's paired two tailed t-test. (**E**) Pie charts (left panel) show the distribution of TCRβ clonotypes based on clonal frequency, the most clonally expanded (top expanded clone, blue) and the next most clonally expanded (second most expanded clone, orange), rest of the expanded clonotypes (≥3) (all other expanded clones, yellow) and not expanded (frequency ≤2, grey). The pie charts (middle panel) show the distribution of shared (overlapping) expanded effector TCRβ clonotypes (frequency ≥3) within the other indicated subsets. The graphs (right) show the percentage overlap of the most expanded TCRβ clonotype from effector subset with the other indicated subsets. (**F**) Pie charts (left panel) show the distribution of TCRβ clonotypes based on clonal frequency, the most clonally expanded (top expanded clone, blue) and the next most clonally expanded (second most expanded clone, orange), rest of the expanded clonotypes (≥3) (all other expanded clones, yellow) and not expanded (frequency ≤2, grey) in visit 2 samples (V2) from the longitudinal samples. The pie charts (middle panel) show the distribution of shared (overlapping) expanded V2 effector clonotypes (frequency ≥3) within other subsets from V1 (visit 1 samples). The graphs (right) show the percentage overlap of the most expanded TCRβ clonotype from effector subset at V2 with other subsets at V1.

Fig. 7. Variable number of CD4-CTL precursors across subjects. (**A**) tSNE 2D plot of single-cell RNA-seq data obtained using Seurat software package for subject #2 (left) and subject #1 (right) from the 3'

transcriptome of single cells (RNA-seq done using the 10x genomics platform) for the most variable transcripts (1799, subject #2 and 1569, subject #1). (**B**) tSNE 2D plots (left) and violin plots (right) of single-cell RNA-seq data show the expression (log_2(UPM+1) expression, UMI (unique molecular identifier) per million) of the indicated differentially expressed transcripts (SCDE and MAST analysis, Benjamini-Hochberg adjusted $P < 0.05$ and ≥2-fold change) for the indicated subjects. Precursor cluster is marked with a black dotted line. The shapes in violin plots represent the distribution of cells based on their log_2(UPM+1) expression values (y-axis). The color scale represents the expression of log_2(UPM+1) (tSNE 2D plots) and mean expression (violin plots). (**C**) Violin plots show the single-cell expression pattern of the indicated differentially expressed transcripts between IL7R^{high} T_{EMRA} (CD4-CTL precursors) and IL7R^{-} T_{EMRA} (CD4-CTL effectors) subsets. The shapes represent the distribution of cells based on their log_2(TPM+1) expression values (y-axis). Color scale shows the mean expression values.

Fig. 1

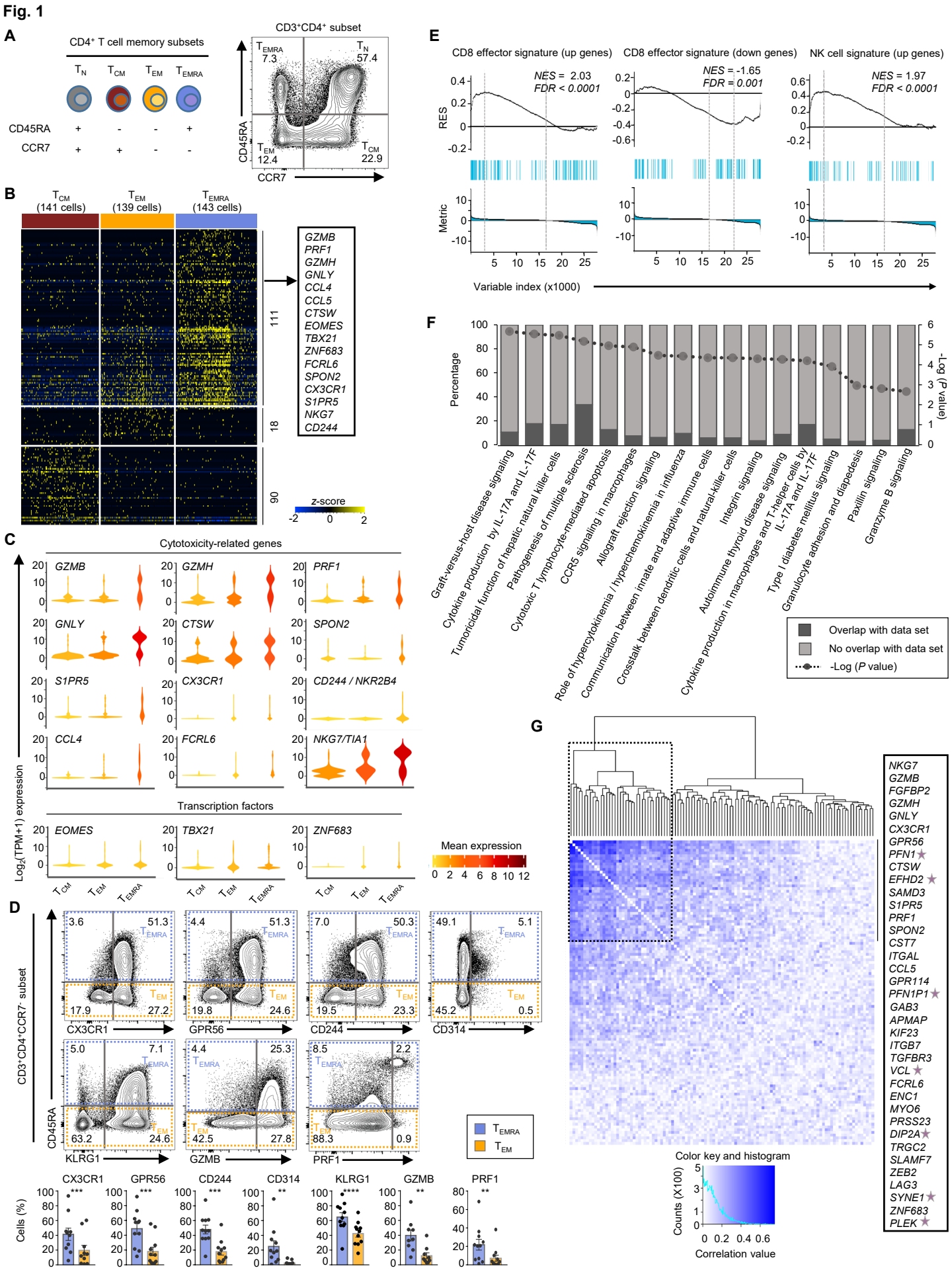

A
CD4+ T cell memory subsets
TN TCM TEM TEMRA
CD45RA + - - +
CCR7 + + - -
CD3+CD4+ subset
TEMRA 7.3
TN 57.4
TEM 12.4
TCM 22.9
CD45RA
CCR7
B
TCM (141 cells)
TEM (139 cells)
TEMRA (143 cells)
111
18
90
z-score
-2 0 2
GZMB
PRF1
GZMH
GNLY
CCL4
CCL5
CTSW
EOMES
TBX21
ZNF683
FCRL6
SPON2
CX3CR1
S1PR5
NKG7
CD244
C
Cytotoxicity-related genes
GZMB
GZMH
PRF1
GNLY
CTSW
SPON2
S1PR5
CX3CR1
CD244 / NKR2B4
CCL4
FCRL6
NKG7/TIA1
Transcription factors
EOMES
TBX21
ZNF683
Log2(TPM+1) expression
TCM TEM TEMRA
Mean expression
0 2 4 6 8 10 12
D
CD3+CD4+CCR7- subset
CD45RA
CX3CR1
GPR56
CD244
CD314
KLRG1
GZMB
PRF1
TEMRA
TEM
Cells (%)
E
CD8 effector signature (up genes)
CD8 effector signature (down genes)
NK cell signature (up genes)
NES = 2.03
FDR < 0.0001
NES = -1.65
FDR = 0.001
NES = 1.97
FDR < 0.0001
RES
Metric
Variable index (x1000)
F
Percentage
-Log (P value)
Graft-versus-host disease signaling
Cytokine production by IL-17A and IL-17F
Tumoricidal function of hepatic natural killer cells
Pathogenesis of multiple sclerosis
Cytotoxic T lymphocyte-mediated apoptosis
CCR5 signaling in macrophages
Allograft rejection signaling
Role of hypercytokinemia / hyperchemokinemia in influenza
Communication between innate and adaptive immune cells
Crosstalk between dendritic cells and natural-killer cells
Integrin signaling
Autoimmune thyroid disease signaling
Cytokine production in macrophages and T-helper cells by IL-17A and IL-17F
Type I diabetes mellitus signaling
Granulocyte adhesion and diapedesis
Paxillin signaling
Granzyme B signaling
Overlap with data set
No overlap with data set
-Log (P value)
G
NKG7
GZMB
FGFBP2
GZMH
GNLY
CX3CR1
GPR56
PFN1
CTSW
EFHD2
SAMD3
S1PR5
PRF1
SPON2
CST7
ITGAL
CCL5
GPR114
PFN1P1
GAB3
APMAP
KIF23
ITGB7
TGFBR3
VCL
FCRL6
ENC1
MYO6
PRSS23
DIP2A
TRGC2
SLAMF7
ZEB2
LAG3
SYNE1
ZNF683
PLEK
Color key and histogram
Counts (X100)
Correlation value

Fig. 2

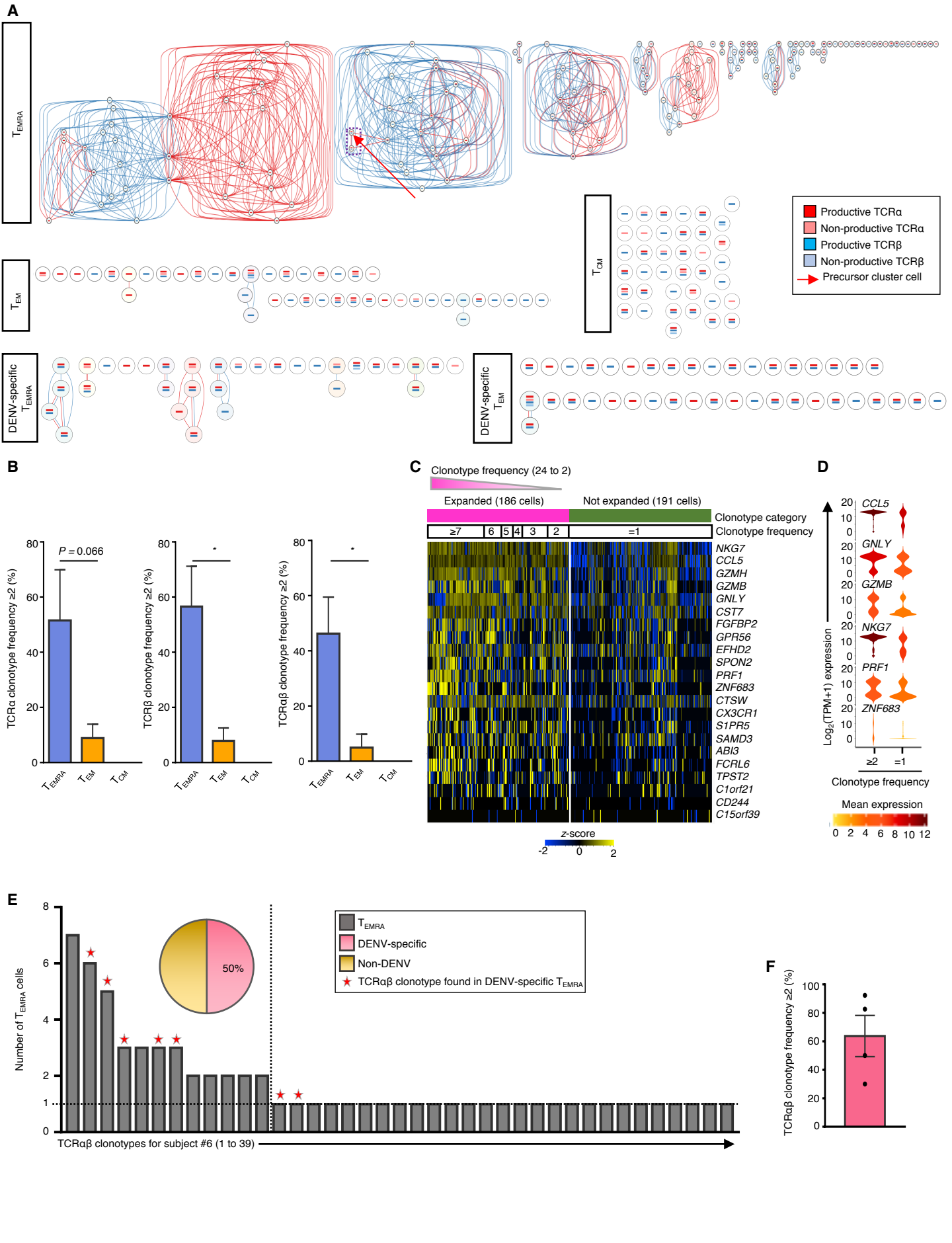

A
T_EMRA
T_EM
T_CM
DENV-specific T_EMRA
DENV-specific T_EM
Productive TCRα
Non-productive TCRα
Productive TCRβ
Non-productive TCRβ
Precursor cluster cell
B
TCRα clonotype frequency ≥2 (%)
TCRβ clonotype frequency ≥2 (%)
TCRαβ clonotype frequency ≥2 (%)
P = 0.066
*
*
T_EMRA
T_EM
T_CM
C
Clonotype frequency (24 to 2)
Expanded (186 cells)
Not expanded (191 cells)
Clonotype category
Clonotype frequency
≥7
6
5
4
3
2
=1
NKG7
CCL5
GZMH
GZMB
GNLY
CST7
FGFBP2
GPR56
EFHD2
SPON2
PRF1
ZNF683
CTSW
CX3CR1
S1PR5
SAMD3
ABI3
FCRL6
TPST2
C1orf21
CD244
C15orf39
z-score
-2
0
2
D
CCL5
GNLY
GZMB
NKG7
PRF1
ZNF683
Log2(TPM+1) expression
≥2
=1
Clonotype frequency
Mean expression
0 2 4 6 8 10 12
E
Number of T_EMRA cells
50%
T_EMRA
DENV-specific
Non-DENV
TCRαβ clonotype found in DENV-specific T_EMRA
TCRαβ clonotypes for subject #6 (1 to 39)
F
TCRαβ clonotype frequency ≥2 (%)

Fig. 3

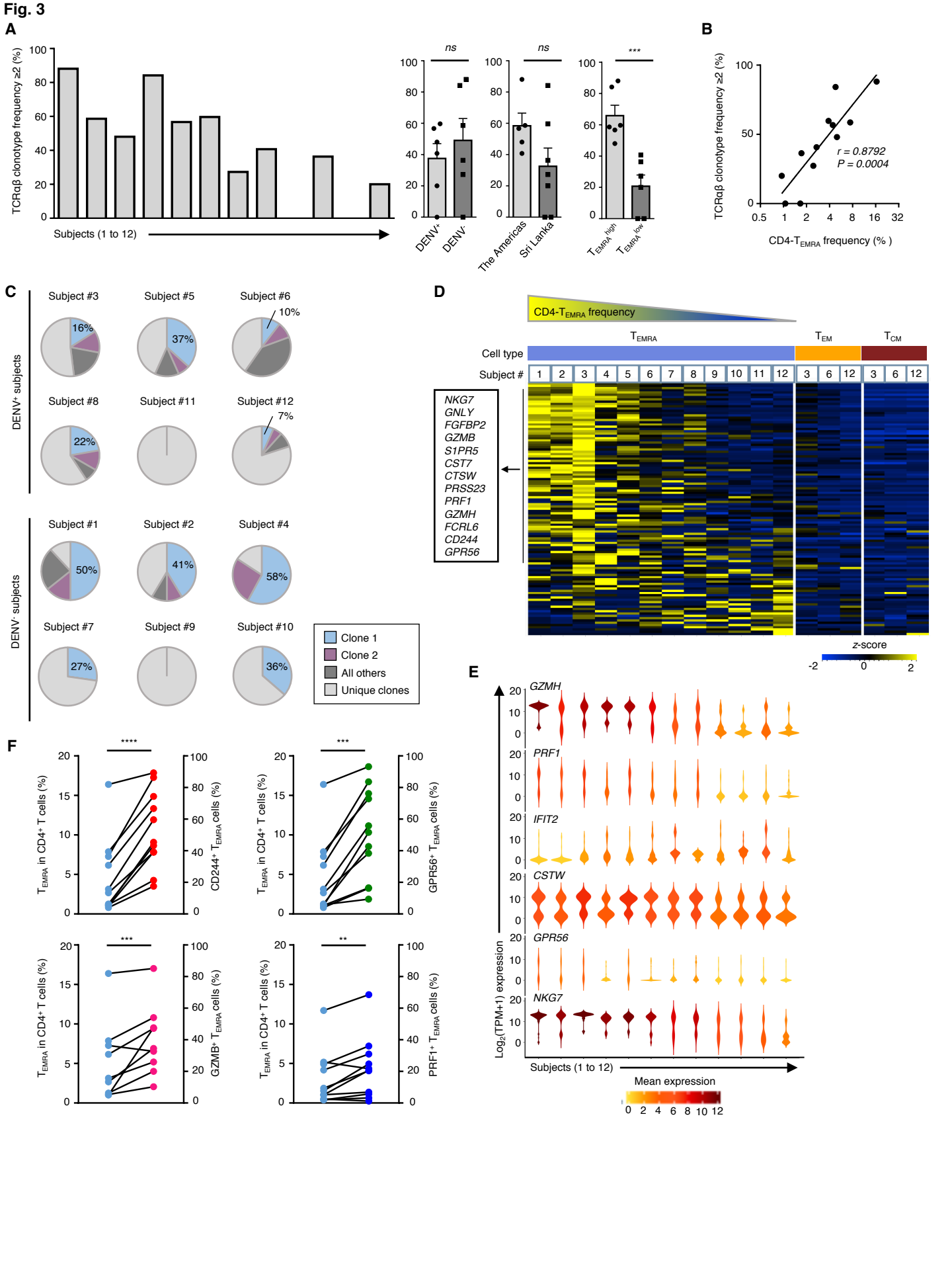

A
TCRαβ clonotype frequency ≥2 (%)
Subjects (1 to 12)
ns
ns
***
DENV+
DENV-
The Americas
Sri Lanka
TEMRA high
TEMRA low
B
TCRαβ clonotype frequency ≥2 (%)
r = 0.8792
P = 0.0004
CD4-TEMRA frequency (%)
C
DENV+ subjects
DENV- subjects
Subject #3
Subject #5
Subject #6
Subject #8
Subject #11
Subject #12
Subject #1
Subject #2
Subject #4
Subject #7
Subject #9
Subject #10
16%
37%
10%
22%
7%
50%
41%
58%
27%
36%
Clone 1
Clone 2
All others
Unique clones
D
CD4-TEMRA frequency
Cell type
TEMRA
TEM
TCM
Subject #
NKG7
GNLY
FGFBP2
GZMB
S1PR5
CST7
CTSW
PRSS23
PRF1
GZMH
FCRL6
CD244
GPR56
z-score
E
GZMH
PRF1
IFIT2
CSTW
GPR56
NKG7
Log2(TPM+1) expression
Subjects (1 to 12)
Mean expression
F
****
***
***
**
TEMRA in CD4+ T cells (%)
CD244+ TEMRA cells (%)
GPR56+ TEMRA cells (%)
GZMB+ TEMRA cells (%)
PRF1+ TEMRA cells (%)

Fig. 4

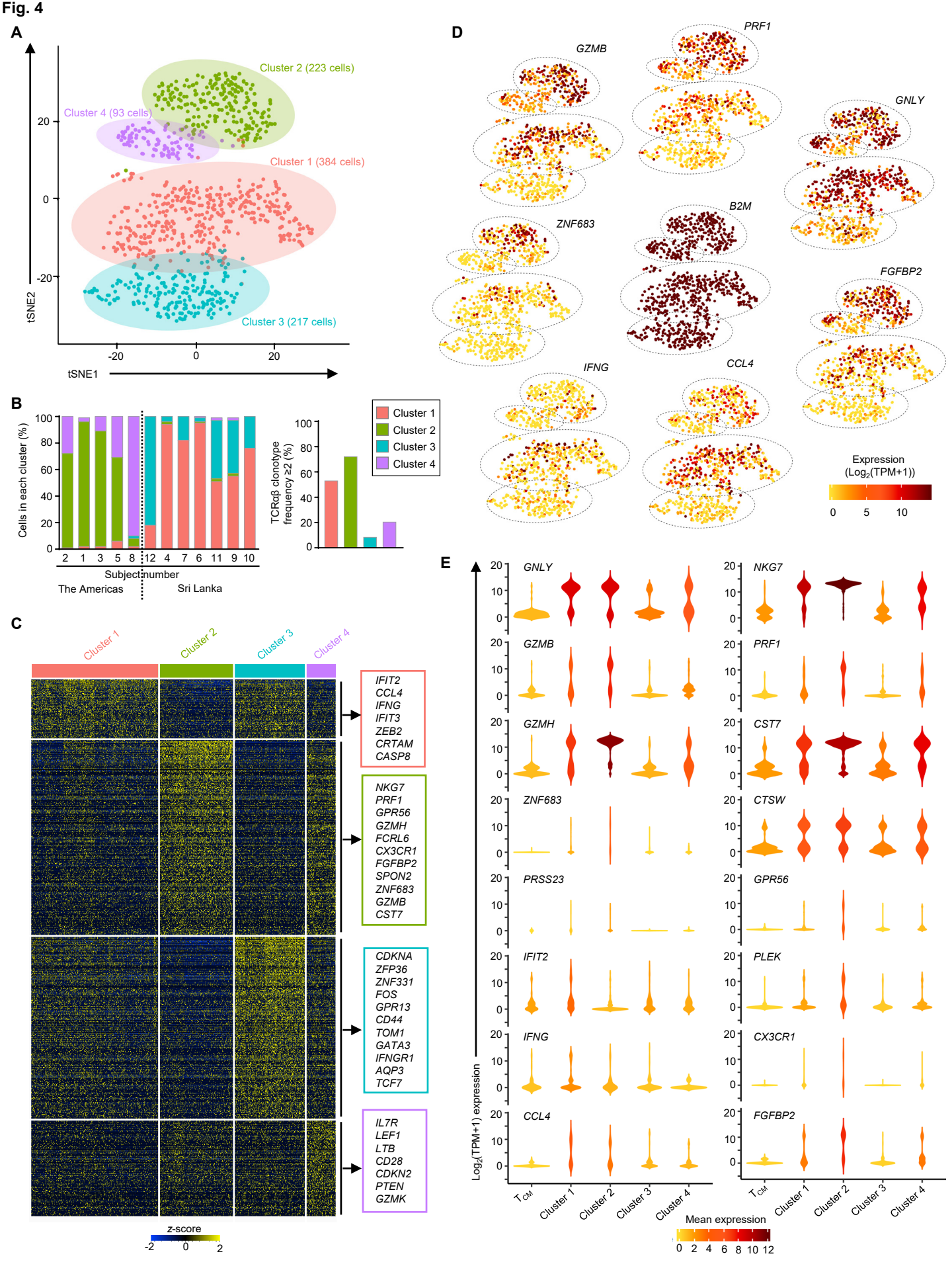

A
Cluster 2 (223 cells)
Cluster 4 (93 cells)
Cluster 1 (384 cells)
Cluster 3 (217 cells)
tSNE2
tSNE1
B
Cells in each cluster (%)
Subject number
The Americas
Sri Lanka
TCRαβ clonotype frequency ≥2 (%)
Cluster 1
Cluster 2
Cluster 3
Cluster 4
C
Cluster 1
Cluster 2
Cluster 3
Cluster 4
IFIT2 CCL4 IFNG IFIT3 ZEB2 CRTAM CASP8
NKG7 PRF1 GPR56 GZMH FCRL6 CX3CR1 FGFBP2 SPON2 ZNF683 GZMB CST7
CDKNA ZFP36 ZNF331 FOS GPR13 CD44 TOM1 GATA3 IFNGR1 AQP3 TCF7
IL7R LEF1 LTB CD28 CDKN2 PTEN GZMK
z-score
D
GZMB
PRF1
GNLY
ZNF683
B2M
FGFBP2
IFNG
CCL4
Expression (Log2(TPM+1))
E
GNLY
NKG7
GZMB
PRF1
GZMH
CST7
ZNF683
CTSW
PRSS23
GPR56
IFIT2
PLEK
IFNG
CX3CR1
CCL4
FGFBP2
Log2(TPM+1) expression
TCM
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Mean expression

Fig. 5

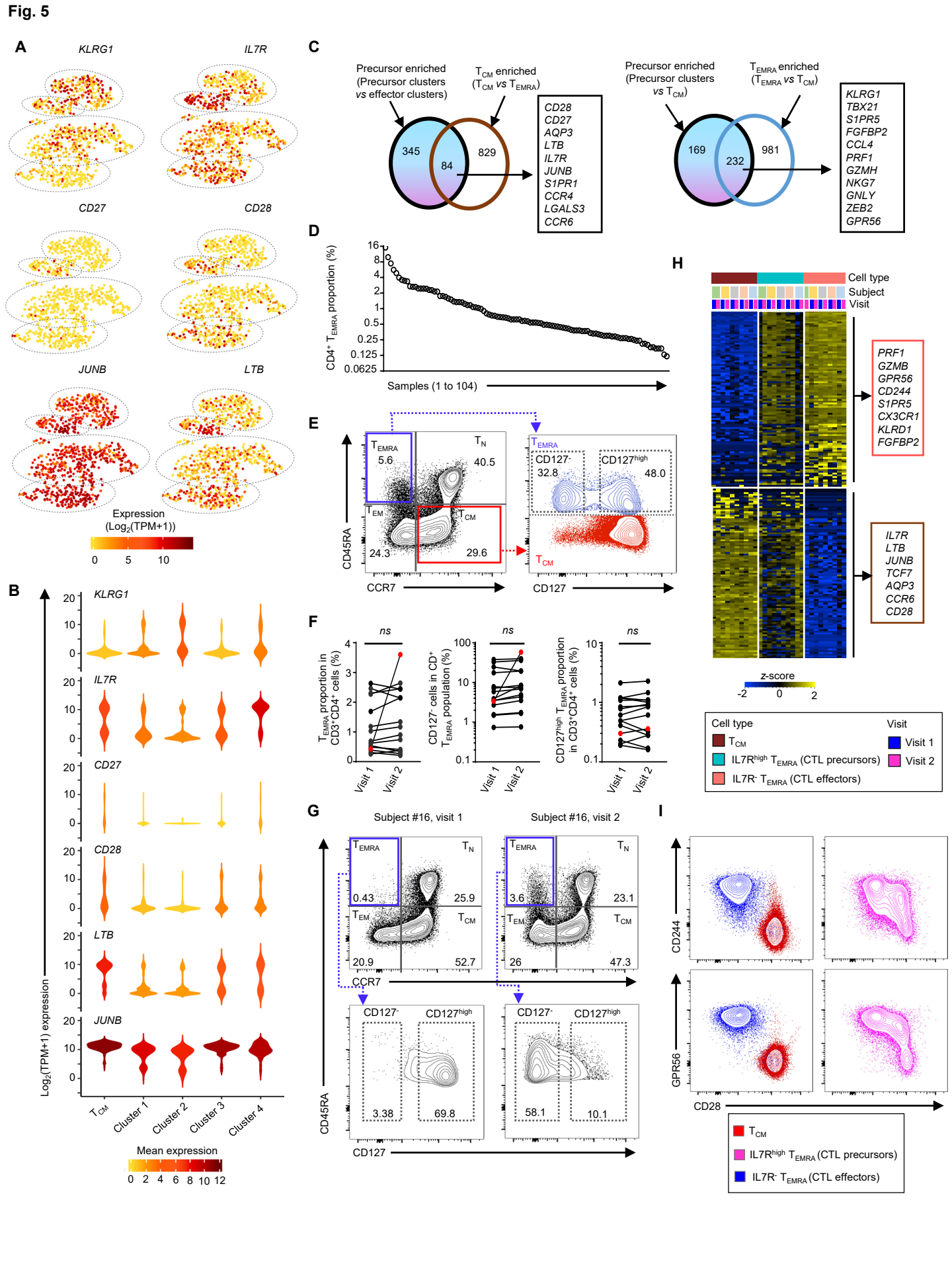

A
KLRG1
IL7R
CD27
CD28
JUNB
LTB
Expression (Log2(TPM+1))
0 5 10
B
KLRG1
IL7R
CD27
CD28
LTB
JUNB
Log2(TPM+1) expression
TCM
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Mean expression
0 2 4 6 8 10 12
C
Precursor enriched (Precursor clusters vs effector clusters)
TCM enriched (TCM vs TEMRA)
345
84
829
CD28
CD27
AQP3
LTB
IL7R
JUNB
S1PR1
CCR4
LGALS3
CCR6
Precursor enriched (Precursor clusters vs TCM)
TEMRA enriched (TEMRA vs TCM)
169
232
981
KLRG1
TBX21
S1PR5
FGFBP2
CCL4
PRF1
GZMH
NKG7
GNLY
ZEB2
GPR56
D
CD4+ TEMRA proportion (%)
16
8
4
2
1
0.5
0.25
0.125
0.0625
Samples (1 to 104)
E
TEMRA
5.6
TN
40.5
TEM
24.3
TCM
29.6
CD45RA
CCR7
TEMRA
CD127-
32.8
CD127high
48.0
TCM
CD127
F
ns
TEMRA proportion in CD3+CD4+ cells (%)
CD127- cells in CD+ TEMRA population (%)
CD127high TEMRA proportion in CD3+CD4+ cells (%)
Visit 1
Visit 2
G
Subject #16, visit 1
Subject #16, visit 2
TEMRA
TN
0.43
25.9
TEM
TCM
20.9
52.7
3.6
23.1
26
47.3
CCR7
CD127-
CD127high
3.38
69.8
58.1
10.1
CD45RA
CD127
H
Cell type
Subject
Visit
PRF1
GZMB
GPR56
CD244
S1PR5
CX3CR1
KLRD1
FGFBP2
IL7R
LTB
JUNB
TCF7
AQP3
CCR6
CD28
z-score
-2 0 2
Cell type
TCM
IL7Rhigh TEMRA (CTL precursors)
IL7R- TEMRA (CTL effectors)
Visit
Visit 1
Visit 2
I
CD244
GPR56
CD28
TCM
IL7Rhigh TEMRA (CTL precursors)
IL7R- TEMRA (CTL effectors)

Fig. 6

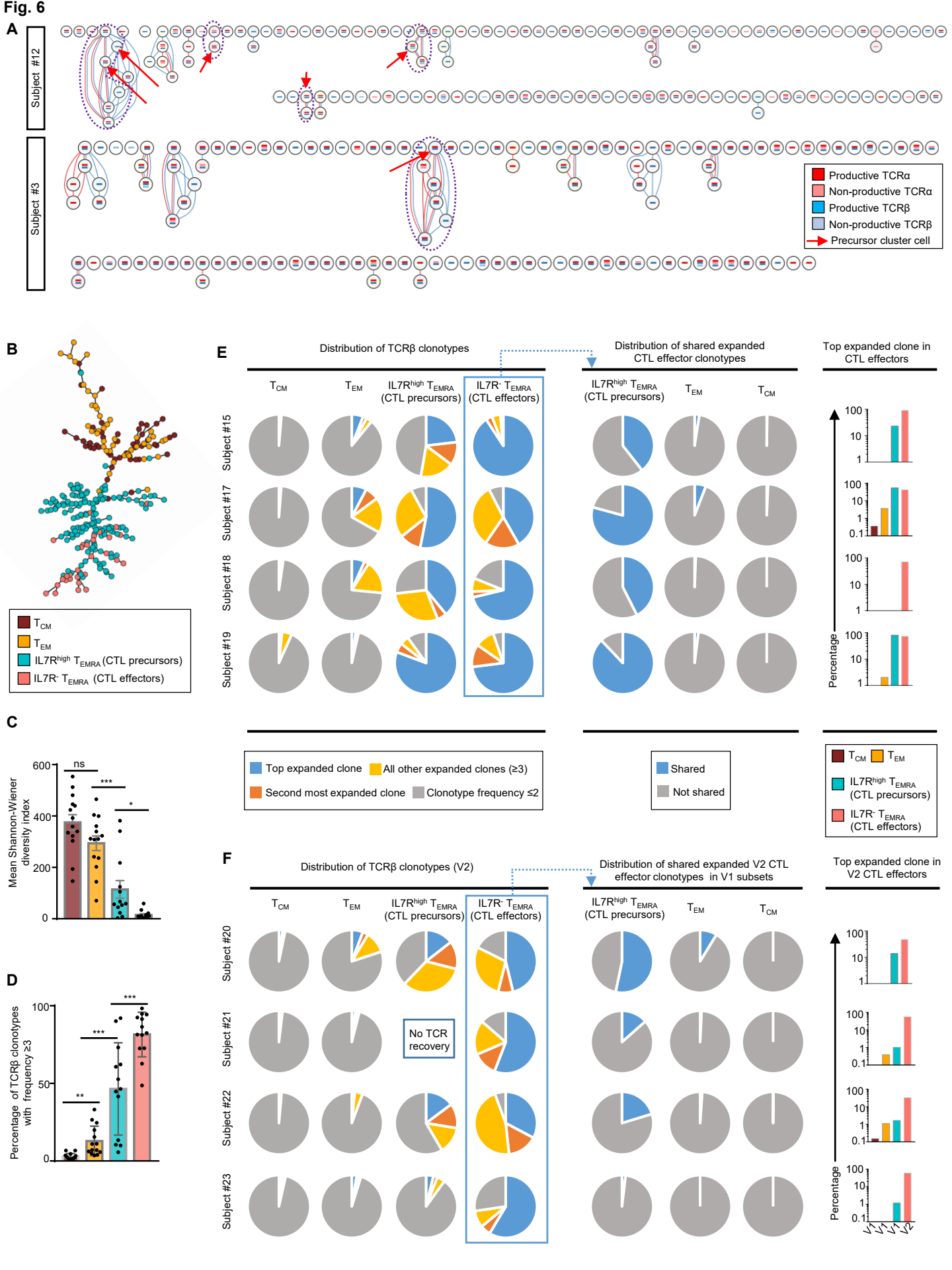
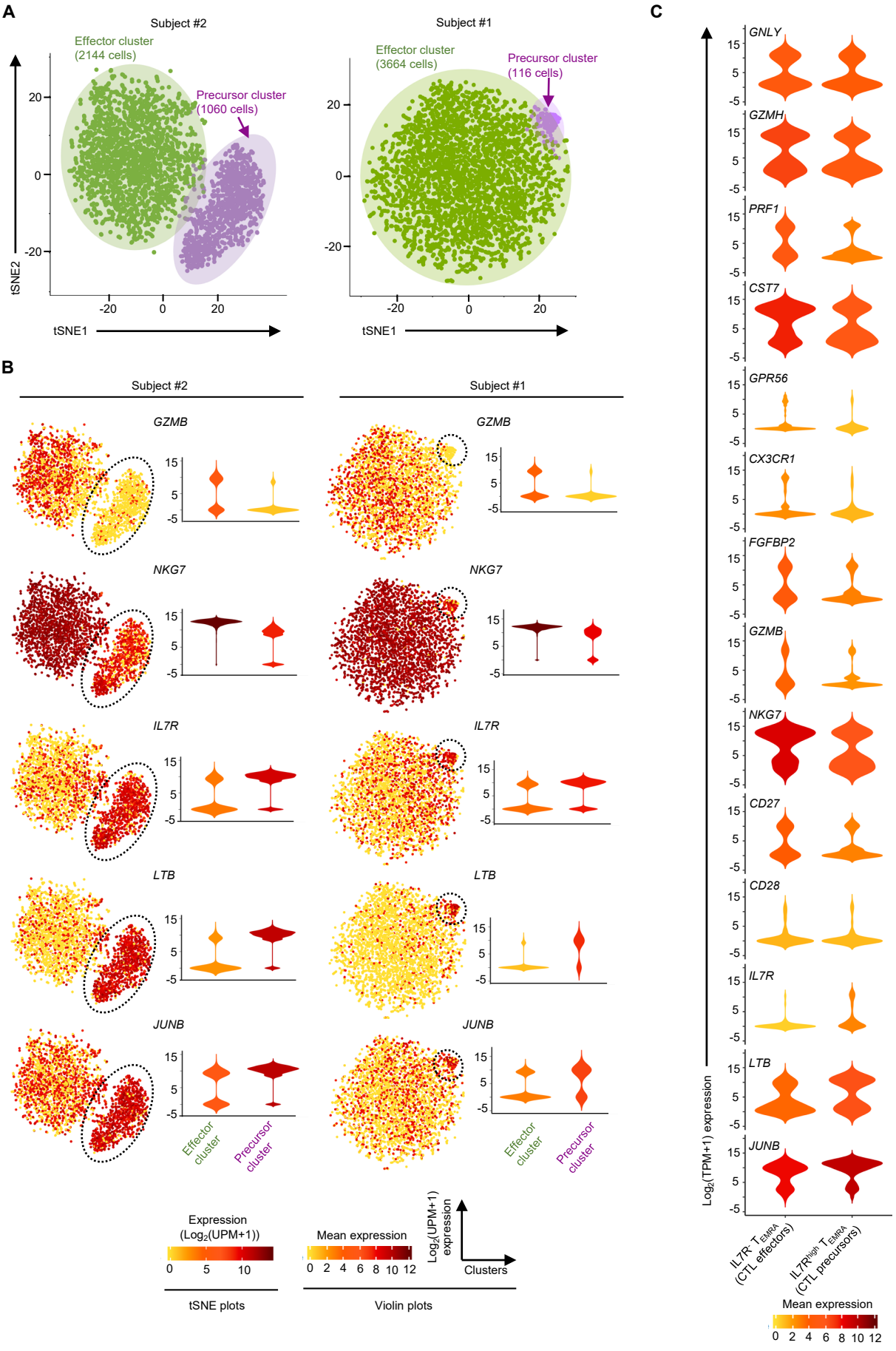

A
Subject #12
Subject #3
Productive TCRα
Non-productive TCRα
Productive TCRβ
Non-productive TCRβ
Precursor cluster cell
B
TCM
TEM
IL7Rhigh TEMRA (CTL precursors)
IL7R- TEMRA (CTL effectors)
C
Mean Shannon-Wiener diversity index
ns
***
*
D
Percentage of TCRβ clonotypes with frequency ≥3
**
***
***
E
Distribution of TCRβ clonotypes
Distribution of shared expanded CTL effector clonotypes
Top expanded clone in CTL effectors
Subject #15
Subject #17
Subject #18
Subject #19
Percentage
Top expanded clone
All other expanded clones (≥3)
Second most expanded clone
Clonotype frequency ≤2
Shared
Not shared
F
Distribution of TCRβ clonotypes (V2)
Distribution of shared expanded V2 CTL effector clonotypes in V1 subsets
Top expanded clone in V2 CTL effectors
Subject #20
Subject #21
Subject #22
Subject #23
No TCR recovery
V1
V1
V1
V2

Fig. 7


A
Subject #2
Effector cluster
(2144 cells)
Precursor cluster
(1060 cells)
tSNE2
tSNE1
Subject #1
Effector cluster
(3664 cells)
Precursor cluster
(116 cells)
tSNE1
B
Subject #2
Subject #1
GZMB
NKG7
IL7R
LTB
JUNB
Effector cluster
Precursor cluster
Expression
(Log2(UPM+1))
tSNE plots
Mean expression
Log2(UPM+1) expression
Clusters
Violin plots
C
GNLY
GZMH
PRF1
CST7
GPR56
CX3CR1
FGFBP2
GZMB
NKG7
CD27
CD28
IL7R
LTB
JUNB
Log2(TPM+1) expression
IL7R- TEMRA
(CTL effectors)
IL7Rhigh TEMRA
(CTL precursors)
Mean expression

SUPPLEMENTARY MATERIALS AND METHODS

Human blood samples

Blood samples were obtained from healthy adult blood donors from the National Blood Centers of the Nicaraguan Red Cross, and Ministry of Health, Colombo, Sri Lanka in an anonymous fashion (de-identified samples only). As required by the ethics review committee (ERC) (University of Colombo ERC serving as Genetech ERC) in Sri Lanka, an MTA (material transfer agreement) has been signed between Genetech Research Institute and the La Jolla Institute of Allergy & Immunology (LJI) to restrict the sample use to LJI and their researchers. Samples obtained were discarded buffy coats from routine blood donations and thus exempt from human subject review. Donors, seronegative for DENV have been recruited in San Diego and informed consent has been given according to the guidelines of the institutional review board (IRB) of the LJI (IRB# VD-085-0616).

For the LJI cohort, blood samples were obtained by leukapheresis from healthy subjects enrolled in a separate LJI study to define the transcriptome and epigenome of human immune cells after written informed consent. Ethical approval for the use of this material was obtained from the IRB of the LJI.

PBMC isolation

Peripheral blood mononuclear cells (PBMC) were isolated by density-gradient sedimentation using Ficoll-Paque (Lymphoprep, Nycomed Pharma, Oslo, Norway) as previously described (*64*). Isolated PBMC were cryopreserved in cell recovery media containing 10% DMSO (Gibco), supplemented with 10-90% heat inactivated fetal bovine serum, depending on the processing laboratory, (FBS; Hyclone Laboratories, Logan UT) and stored in liquid nitrogen until used in the assays. PBMC collected in Sri Lanka were stored in Synth-a-Freeze Cryopreservation medium (Cat A1254201 Thermo Fisher Scientific, USA).

Serology

In general, DENV seropositivity was determined by DENV IgG or an Inhibition ELISA, as previously described (*65, 66*). Flow cytometry-based or Vero cell-based focus reduction neutralization assays were performed for further characterization of seropositive donors, as previously described (*67, 68*).

Flow cytometry

Cryopreserved PBMCs isolated from peripheral blood samples were thawed and stained with cocktails of fluorescently conjugated antibodies (table S10), sorted on a FACS Aria-III (Becton Dickinson) or protein expression was analyzed on LSR II and FACS data was analyzed using FlowJo (version 10) software. For single-cell RNA-seq experiments, live and singlet gated cells were further gated to sort T_{CM}; $CD3^+CD4^+CD8\alpha/CD14/CD19^-CD45RA^-CCR7^+$, T_{EM}; $CD3^+CD4^+CD8\alpha/CD14/CD19^-CD45RA^-CCR7^-$, T_{EMRA}; $CD3^+CD4^+CD8\alpha/CD14/CD19^-CD45RA^+CCR7^-$, $IL7R^{high}$ T_{EMRA} (CD4-CTL precursors); $CD3^+CD4^+CD8\alpha/CD14/CD19^-CD45RA^+CCR7^-CD127^{high}$, $IL7R^-$ T_{EMRA} (CD4-CTL effectors); $CD3^+CD4^+CD8\alpha/CD14/CD19^-CD45RA^+CCR7^-CD127^-$. Single cells were directly sorted into 96 well plate with lysis buffer for downstream processing using smart-seq2 method or into eppendorf tubes with 50% FBS for 10x sequencing experiments or with Trizol for RNA isolation. For analysis of IL7R surface expression, live and singlet gated cells that were $CD3^+CD4^+CD56/CD14^-CD45RA^+CCR7^-$ (T_{EMRA}), were further gated into cells that were $IL7R^-$ and $IL7R^{high}$ (**Fig. 5**, D and G). PBMCs were stimulated with DENV peptide pool following the protocol described previously (*37*) and IFNγ producing single cells were sorted for single-cell RNA-seq analysis.

Single-cell RNA-seq assay

For full-length transcriptome analysis using Smart-seq2, we followed the protocol described previously (*33, 34*) with the following modifications. The pre-amplification PCR cycle was increased to 23 to get enough amount of cDNA for downstream analysis. To eliminate any traces of primer-dimers, the PCR pre-amplified product was purified twice, using 0.8x Ampure-XP beads. 0.3 to 0.5 ng of pre-amplified cDNA was used to generate barcoded Illumina sequencing libraries (Nextera XT library preparation kit, Illumina) in 8µL reaction volume. Libraries were pooled and sequenced on the HiSeq2500 Illumina platform to obtain 50-bp single end reads. Single-cell RNA-seq using 10X genomics platform was performed using Chromium™ Single Cell 3' v2 Reagent Kits following the manufacturer's protocol (*59*). Libraries were sequenced on HiSeq2500 platform to obtain 100 and 32-bp paired end reads using the following read length; read 1, 26 cycles, read 2, 98 cycles and i7 index, 8 cycles.

Single-cell RNA-seq mapping

Single-cell RNA-seq data were mapped against the human hg19 reference genome using TopHat (v1.4.1., --library-type fr-secondstrand -C) and Gencode version 19 (GRCh37.p13) as gene model reference for alignment. Sequencing read coverage per gene was counted using HTSeq-count (-m union -s yes -t exon -i gene_id, http://www huber.embl.de/users/anders/HTSeq/). Counts per gene are obtained by counting all the transcripts mapping to a gene and is together referred to as transcript.

Post mapping quality control pipeline for single-cell RNA-seq

Quality control parameters were adopted as described previously (*69*) with the following modifications. For each library we obtained the following scores from picard tools: 1) total raw reads, 2) reads passing filters, 3) percentage of reads passing filters, 4) percentage of total mapping to spike in control, 5) percentage of total reads mapping to tRNA or rRNA, 6) total reads mapping to reference, 7) percentage of total reads mapping to reference, 8) total uniquely mapping reads, 9) percentage of total uniquely mapping reads, 10) percentage of coding bases, 11) percentage of UTR bases, 12) percentage of intronic bases, 13) percentage of intergenic bases, 14) percentage of mRNA bases, 15) median coefficient of variation of coverage, 16) median 5' bias, 17) median 3' bias, 18) median 5' to 3' bias. Following four parameters from the mapping statistics were used to classify cells as low or high quality and excluded the low quality cells from further downstream analysis (table S1); total raw counts <300,000 reads, total uniquely mapping reads <100,000, percentage of mRNA bases <40% and the threshold for the percentage of total uniquely mapping reads was calculated using formula mean(X)/2. A cell was classified as high quality if it passed the criterions for all the above four parameters. Further, we examined the high quality cells for any median 5' and 3' bias and also determined the total number of genes expressed per cell. All the high quality cells showed no median 5' and 3' bias and expressed ≥1000 genes per cell. The data was log transformed and displayed as normalized TPM (transcripts per million) counts and a value of 1 was added prior to log transformation to account for low or zero values.

Single cell differential gene expression analysis

Differentially expressed (DE) genes were identified by performing pairwise comparison using two complimentary methods SCDE (version 1.99.1) and MAST (version 0.99) as described previously (*22, 23*). To account for different number of cells the SCDE parameter min.nonfailed was set to 20% of the total cell number and other parameters were used at default settings. Two-sided *P*-values were calculated from the Benjamini-Hochberg multiple testing corrected Z-score (cZ) using the normal distribution as null hypothesis. Results are adjusted for multiple comparisons using Benjamini-Hochberg procedure. While MAST mainly identified highly expressed DE genes, SCDE also identified DE genes that are expressed at lower level but in higher proportion of cells. Hence, we obtained the DE gene list by combining the gene list from both the methods using adjusted $P < 0.05$ and $\log_2$ fold change ≥1 from each method. The cell subset enriched transcript list (**Fig. 1**, B and C) was obtained by the algorithm described previously (*34*) (table S2). Seurat software (**Fig. 4C** and table S5) as well as MAST and SCDE (fig. S4 and table S5) were used to identify cluster specific differentially expressed gene sets.

Clustering and cell state hierarchy methods

Unbiased clustering of single cells was performed using Seurat (version 1.4) (**Fig. 4A** and **Fig. 7A**) and SINCELL (version 1.6) packages (fig. S2) (*38, 39*). Principal component analysis (PCA) was performed using a set of 1000 top variable genes and then dimensionality reduction was performed using t-SNE algorithm with top 10 PCAs. We also used SINCELL package (version 1.6) to perform clustering and cell state hierarchies. For clustering, we used the following dimensionality reduction algorithms; PCA, ICA (independent component analysis), classical MDS (multidimensional scaling) and non-metric MDS. Cell-state hierarchy maps were generated using the MST (minimum spanning tree) algorithm.

Batch effect analysis

We analyzed the effect of single-cell library preparation batches on the clustering of T_{EMRA} cells (**Fig. 4A** and fig. S2 and table S1). Since batch variability is a concern in single-cell RNA-seq assays, single cells from donors in different geographical location (Nicaragua and Sri Lanka) were processed in the same 96-cell batch (48 cells from each subject). We observed that donors from different geographical location clustered

separately even when processed in the same batch (cluster 2 and 4 for Nicaragua, 1 and 3 for Sri Lanka, fig. S3A and table S1). We next examined cells from the same subjects (#6 and #12) processed in three different batches and found that in both the subjects the cells clustered together and mapped to cluster 1 and 3 (effector and precursor respectively) with approximately the same proportion (fig. S3B). Also, we found that across all batches, the clustering was more specific to subjects than to batches (**Fig. 4**, A and B and fig. S3C). Together our analysis showed no major detectable variation due to batch (fig. S3, A, B and C and table S1), and biological variability dominated the single-cell transcriptional signatures.

TCR reconstruction from single-cell RNA-seq

Recombined TCR sequences were reconstructed from full length transcriptome of single-cells using the TraCeR algorithm (version 0.1) with default settings (*35*). In brief, reads originating from TCR mRNA were extracted from the sequence data for each cell and assembled into contigs representing the full-length TCR sequences expressed in each cell. Clonally related and expanded cells were identified by finding instances where more than one cell had a unique TCRα and TCRβ chain clonotype sequence (table S4). Clonotype identity was called based on the complete CDR3 and VDJ gene sequence. Cells with >1 productive TCRα (~8%, 78 of 985 cells) and TCRβ (~8%, 97 of 1299 cells) chains were dealt based on the method described previously where the authors suggest utilizing only the productive clonotype with higher read counts (*70*). To be more stringent, we only included single-cells where the clonotype read count was also ≥2 fold higher compared to other productive clonotype (this criterion excluded ~3% of the cells for TCRβ chains and 4% of the cells for TCRα chains).

TCR-seq and data analysis

TCR-seq was performed as described previously (*57*). In brief, 1ng of total RNA was reverse transcribed with TCRα and TCRβ gene specific primers followed by PCR amplification (table S11). The reaction volume was then split into 2 parts and TCRα and TCRβ were amplified separately in a 2nd PCR amplification step. After the amplification, libraries were prepared using True-seq nano DNA library kit (Illumina). Libraries were pooled and sequenced on the MiSeq Illumina platform to obtain 200-bp paired end reads. The sequencing data was

mapped and analyzed using MIGEC software with default settings (*57*). The downstream analysis was performed using V(D)J tools with default settings (*58*). The mapping QC matrices are included in supplementary table S7.

Bulk RNA-seq, mapping and analysis

RNA-seq was performed as described previously using smart-seq2 method (*71, 72*) with 0.5ng of total RNA. The mapping was performed as described for single-cell RNA-seq data. The differentially expressed genes from bulk RNA-seq were identified by pairwise comparisons using DESeq2 (*73*) (**Fig. 5H** and table S6).

Analysis of 3' transcriptome of single-cells from 10X genomics platform

Reads from single-cell RNA-seq were aligned and collapsed into unique molecular identifier (UMI) counts using 10x genomics's cell-ranger software (version 1.2) (*59*). Counts were normalized to get UMIs per million reads (UPM) (**Fig. 7**, A and B).

Co-expression matrix

A matrix showing co-expression of transcripts was generated by calculating gene-gene correlation matrix *p* over all T_{EMRA} cells using Spearman Correlation and then converted to absolute values via the transformation M= abs(p). (**Fig. 1F**). The matrix M was clustered using Ward's method.

Statistical analysis and Data display

Comparison between two groups was assessed with two-tailed unpaired or paired Student's *t*-test using GraphPad Prism 6. Spearman correlation coefficient (*r* value) was calculated to assess significance of correlations between the levels of any two components of interest. GraphPad Prism 6 software was used for generating graphs and performing statistical significance tests and R package with custom scripts for generating violin plots. Heat maps were generated using Qlucore omics software.

Knowledge-based network generation and pathway analysis

The biological relevance of differentially expressed genes identified by SCDE and MAST analysis was further investigated using the ingenuity pathways analysis (IPA) platform. The enrichment of canonical pathways (pre-defined, well-described metabolic and signaling pathways curated from literature reviews) amongst differentially expressed genes was assessed, with significance determined by right-tailed Fisher's exact test, FDR values set to <0.05 (cut off set to –log (B-H p-value greater than 1.3).

Gene Set Enrichment Analysis (GSEA)

The Qlucore Omics Explorer 3.2 software package was used for GSEA analysis (*26*). GSEA was used to further assess whether specific signatures were significantly enriched between two groups. GSEA determines whether a prior defined 'set' of genes (such as a signature) show statistically significant cumulative changes in gene expression between phenotypic subgroups (*26*). In brief, all genes are ranked based on their differential expression between two groups. Next, a running enrichment score (RES) is calculated for a given gene set based on how often its members appear at the top or bottom of the ranked differential list. 1000 random permutations of the phenotypic subgroups are used to establish a null distribution of RES against which a normalized running enrichment score (NES) and FDR-corrected q values are calculated using Kolmogorov-Smirnov statistic. GSEA was run with a focused group of gene signatures, namely CD8 effector (up), CD8 effector (down) and natural killer (NK) cell signature gene sets obtained from Broad Institute. These gene signatures (**Fig. 1E** and table S12) were selected to test the null hypothesis.

SUPPLEMENTARY FIGURES AND TABLES.

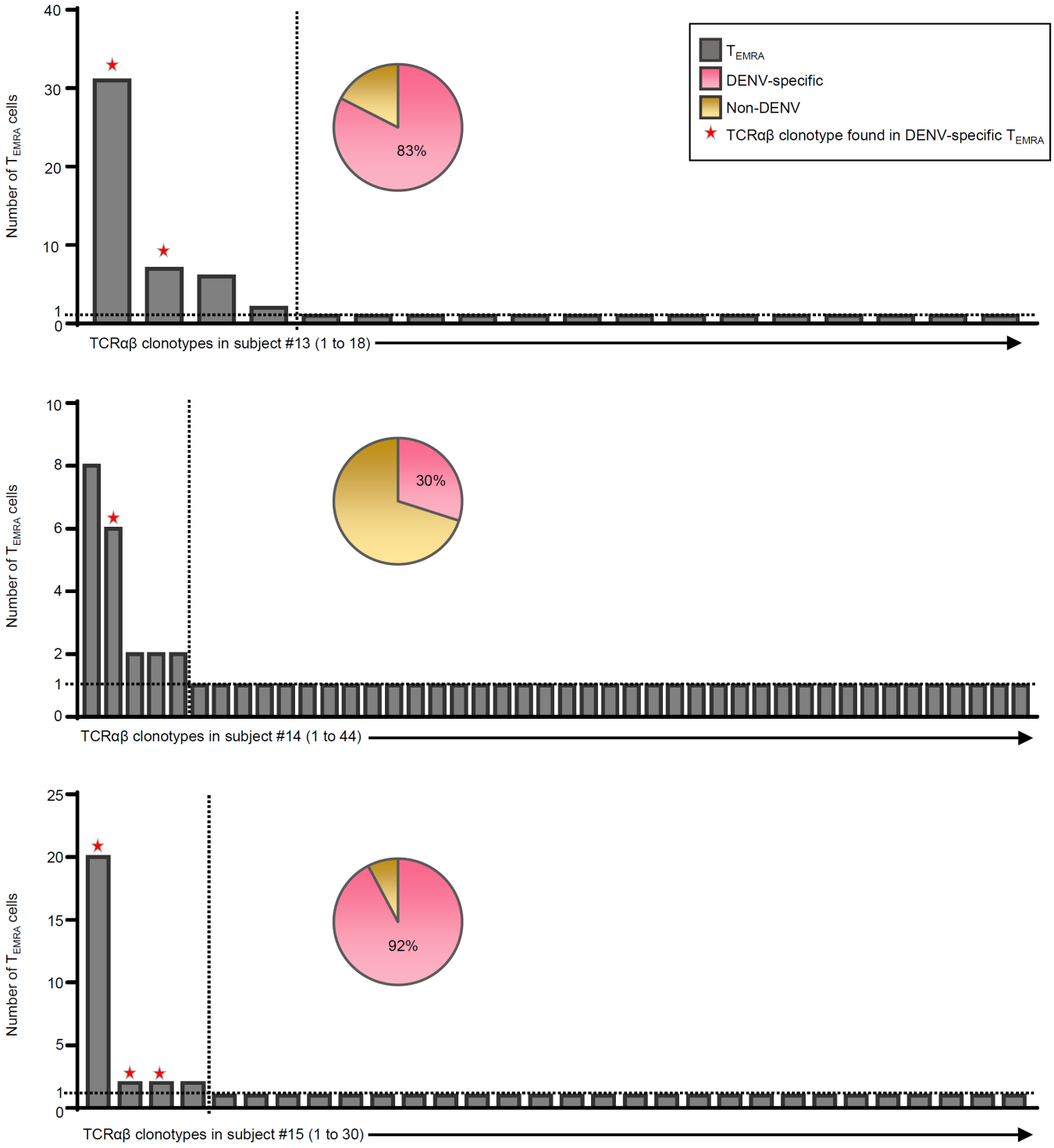


Fig. S1. The majority of the expanded TCR clonotypes are DENV-specific. Bar graphs show the number of T_{EMRA} cells for each unique TCRα and TCR β chain clonotype for subjects #13, 14 and 15. Stars in red color indicate the clonotypes shared by DENV-specific T_{EMRA}. Pie chart shows the percentage of expanded T_{EMRA} cells that share unique TCRα and TCRβ chain clonotypes of DENV-specific T_{EMRA} cells (pink).

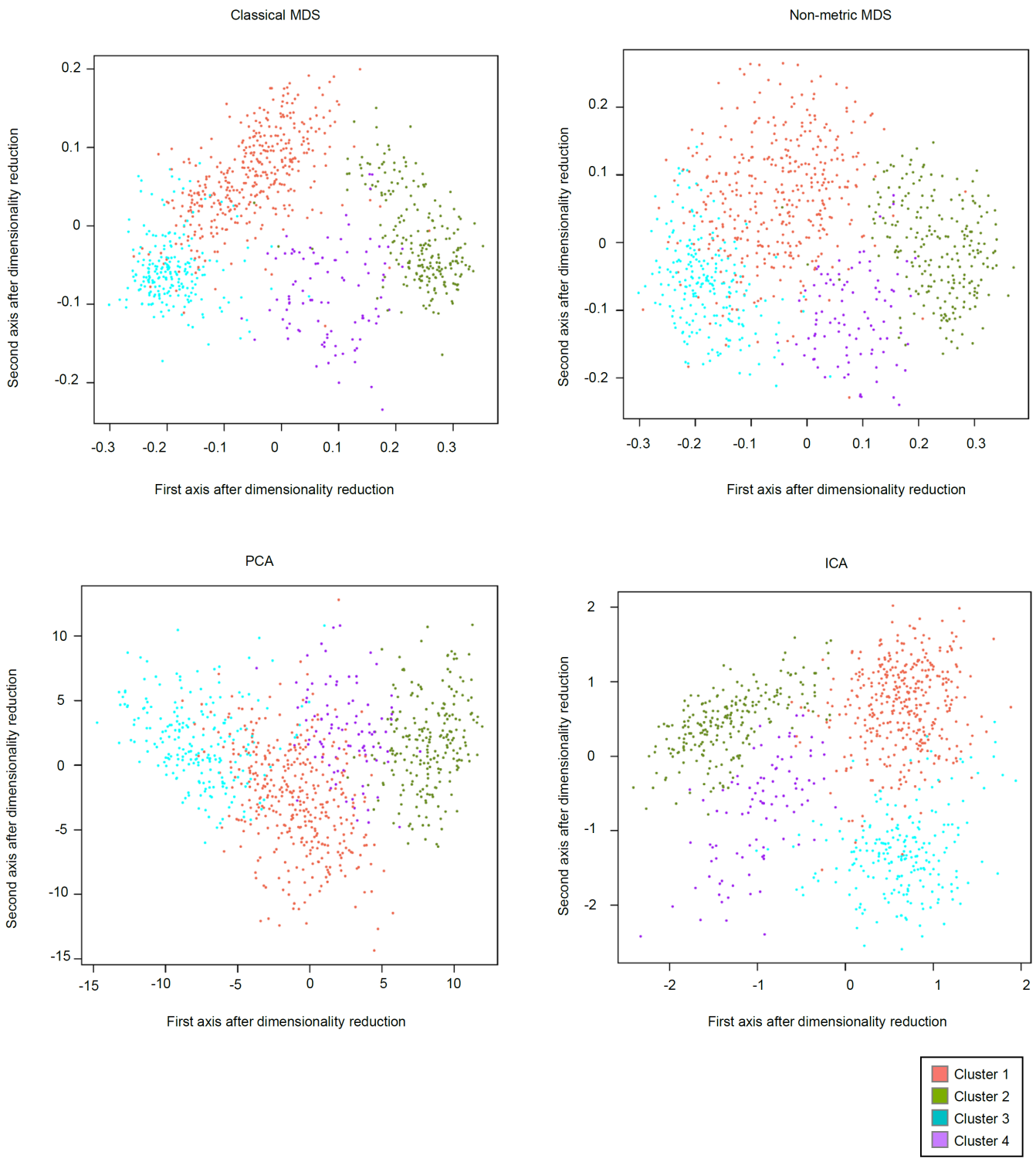


Fig. S2. T_{EMRA} cells cluster into four distinct clusters. T_{EMRA} cells were clustered based on the single-cell expression of the 1000 most variable transcripts by using SINCELL software package. The four different methods of clustering used were Classical MDS; classical multidimensional scaling, Non-metric MDS; non-metric multidimensional scaling, PCA; principal component analysis, ICA; independent component analysis.

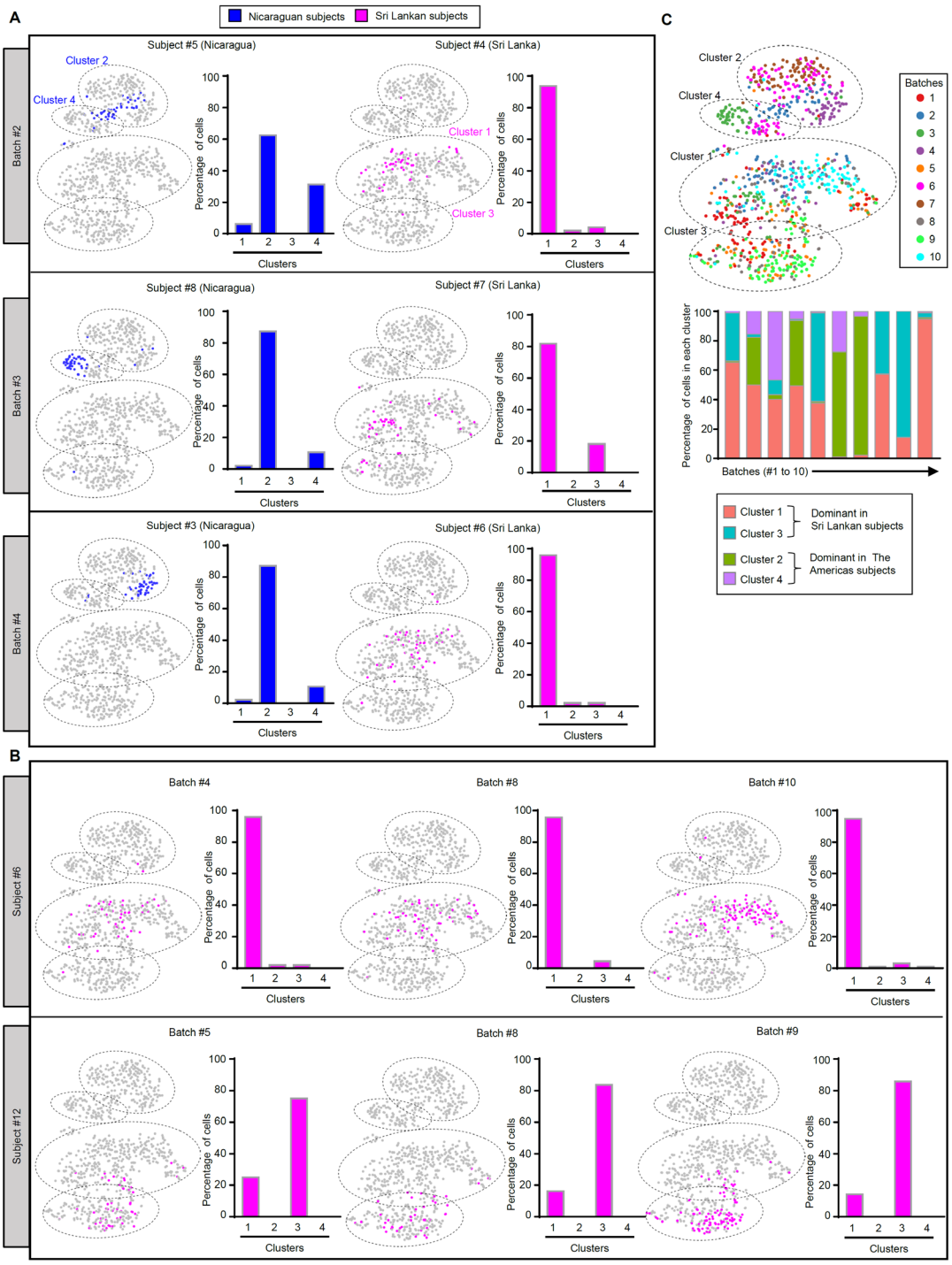


Fig. S3. Batch analysis of T_{EMRA} cells. (**A**) tSNE 2D plots of single-cell RNA-seq data (left) and the bar graphs (right, percentages) show the distribution of cells across four T_{EMRA} clusters from the indicated subjects in each batch. Each batch of ~96 cells, included ~48 cells from a Nicaraguan subject (blue, left panel) and a Sri Lankan subject (magenta, right panel). (**B**) tSNE 2D plots of single-cell RNA-seq data (left) and bar graphs (right, percentages) show the distribution of cells (48 to 96 cells) across four T_{EMRA} clusters in the indicated batch and the subject. Cells were processed in three independent batches from the same subject (subject #6, top panel; subject #12, bottom panel). (**C**) tSNE 2D plot of single-cell RNA-seq data (top) and the bar graph (bottom) show the distribution of cells, processed in 10 different batches of ~96 cells each, across four T_{EMRA} clusters. Cells in tSNE 2D plot are colored based on the batch number. The batch and the subject numbers are shown in table S1.

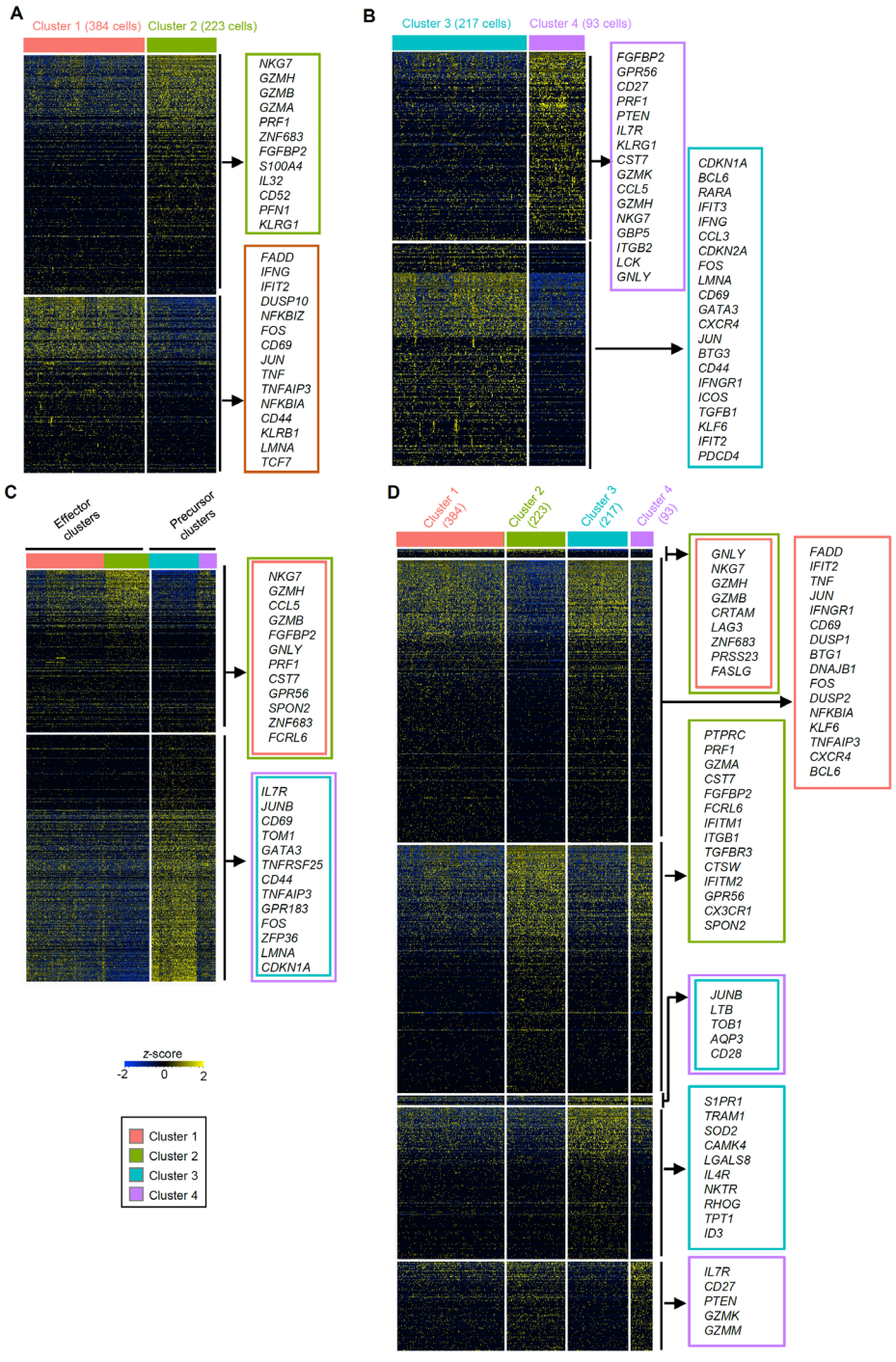
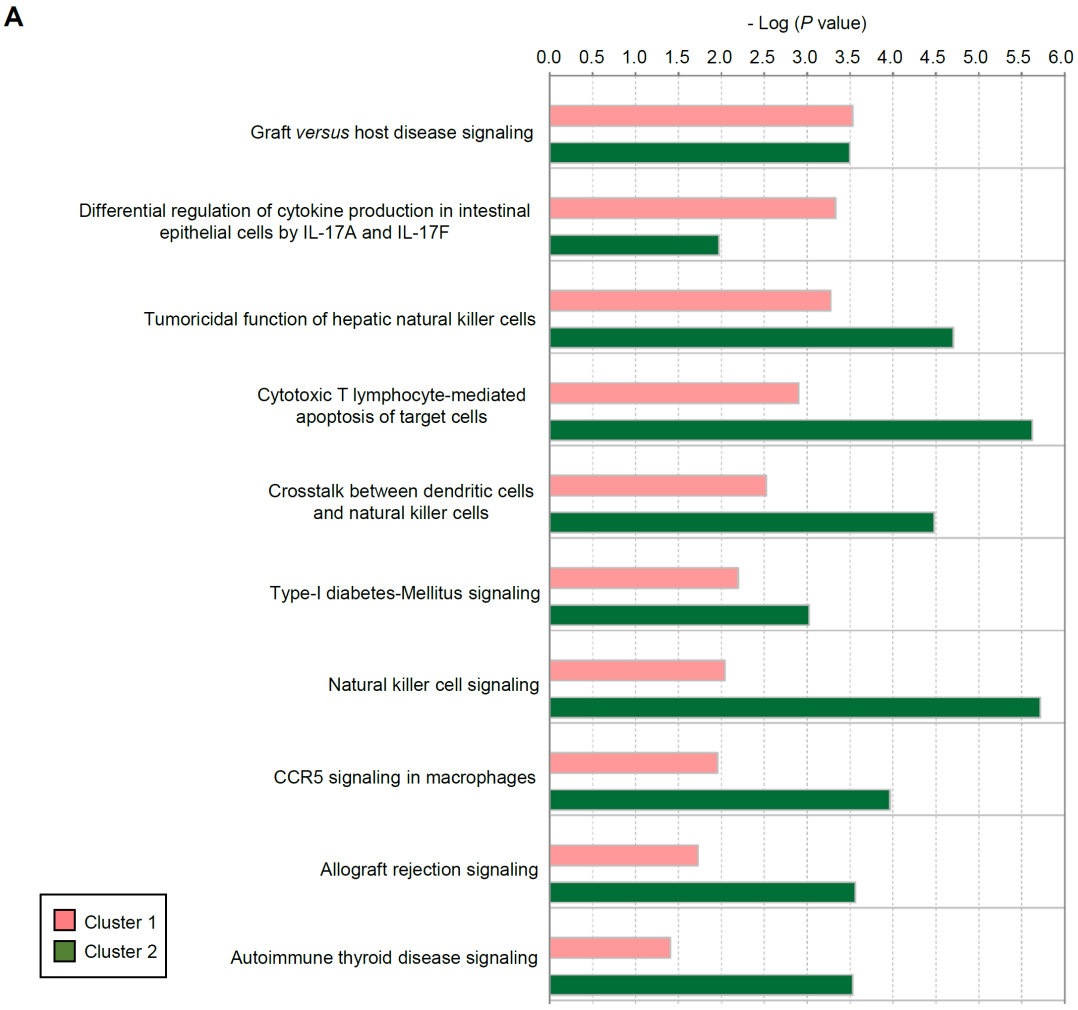


Fig. S4. T_{EMRA} clusters express distinct set of transcripts. (**A, B, C** and **D**) Single-cell RNA-seq analysis showing row-wise *z*-scores of normalized TPM (transcripts per million) counts of cells in each cluster for each differentially expressed transcripts (rows) obtained by pairwise comparison of cluster 1 *versus* cluster 2 (A), cluster 3 *versus* cluster 4 (B), cluster 1 and 2 (effector clusters) *versus* cluster 3 and 4 (precursor clusters) (C). (**D**) Cluster-specific transcript list was obtained by comparing pairwise comparisons of clusters and taking the union of all the transcripts upregulated in a particular cluster *versus* each cluster separately (example: cluster 1 specific; union of transcripts upregulated in cluster 1 *versus* cluster 2, cluster 1 *versus* cluster 3, cluster 1 *versus* cluster 4). Transcripts that are commonly expressed by either the effector clusters (1 and 2) or the precursor clusters (3 and 4) was obtained by pairwise comparison of those clusters individually against the precursor clusters (example: for effector clusters, the transcript list common to cluster 1 *versus* cluster 3 or cluster 4 and cluster 2 *versus* cluster 3 or cluster 4). (SCDE and MAST analysis, Benjamini-Hochberg adjusted $P < 0.05$ and ≥2-fold change).



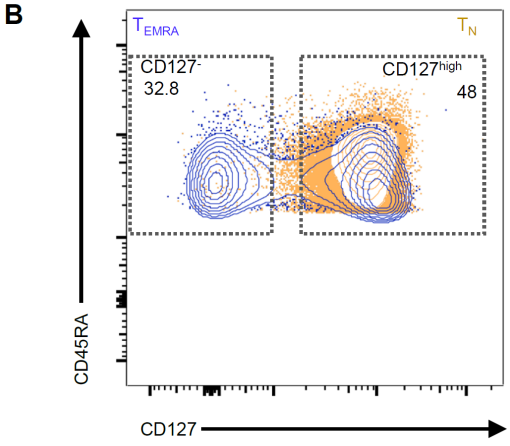


Fig. S5. Clusters 1 and 2 are enriched for transcripts involved in cytotoxicity-related pathways. (**A**) Comparative ingenuity pathway analysis of the transcripts enriched in cluster 1 *versus* cluster 3 and cluster 4 and cluster 2 *versus* cluster 3 and cluster 4. (**B**) Contour plot shows the co-expression of CD45RA and CD127 in T_{EMRA} and T_N subsets in live and singlet-gated $CD3^+CD4^+$ T cells obtained from PBMCs (peripheral blood mononuclear cells). The gating strategy shows $CD127^-$ and $CD127^{high}$ proportion for T_{EMRA} subset.

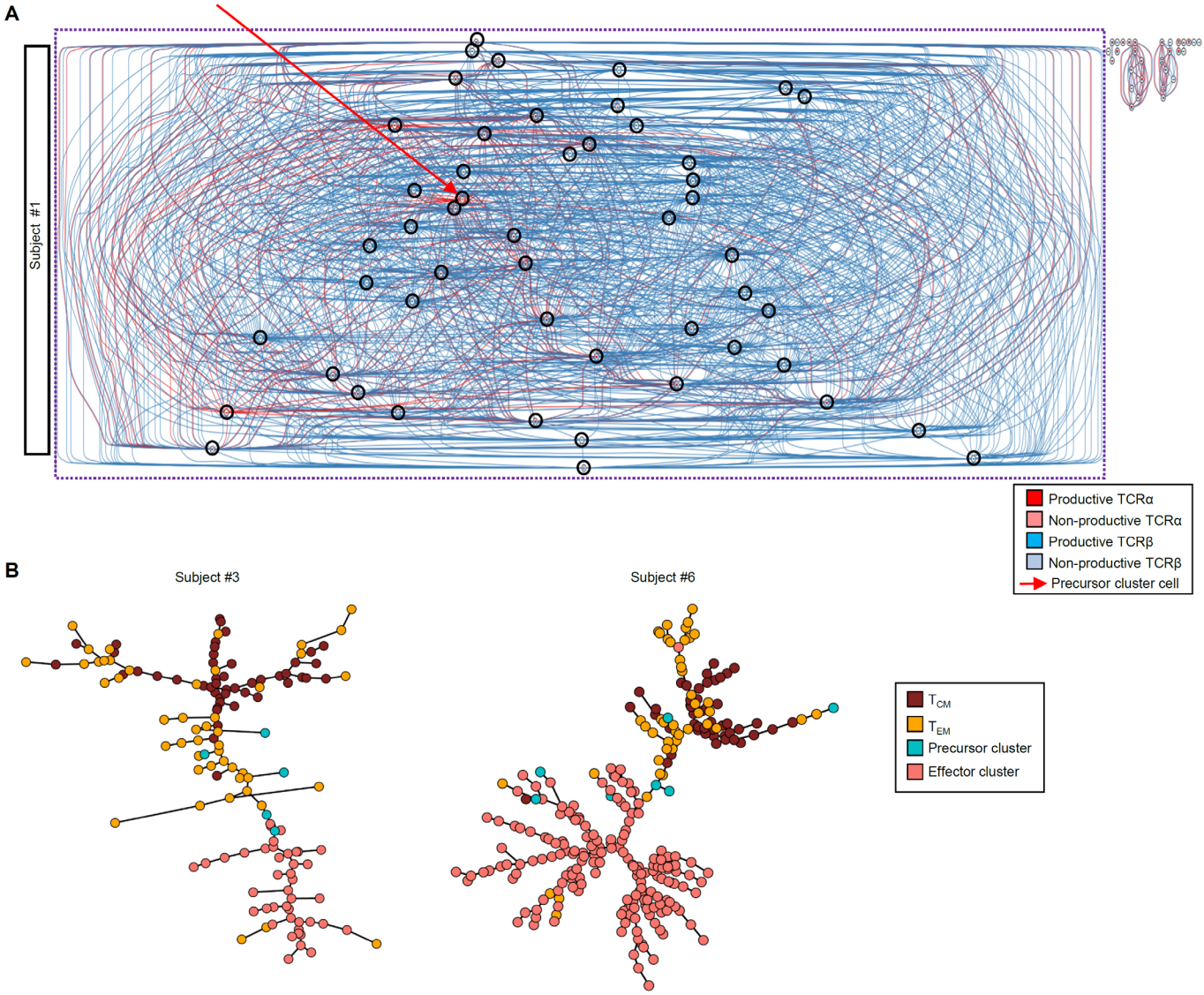


Fig. S6. Precursor cells in T_{EMRA} subset. (**A**) Clonotype network graph of T_{EMRA} cells from subject #1. Each circle represents a single cell; the reconstructed TCRα and TCRβ sequences for each cell are depicted as red and blue colored bars, respectively, inside each circle. Dark and light colored bars represent productive, and nonproductive TCRs, respectively. Red connecting lines indicate shared TCRα sequences, blue lines indicate shared TCRβ sequences. Red arrow indicates a precursor cell. Purple dotted line indicates a cluster of cells with intermixed precursor and effector cells sharing the same clonotype. Black circles encircle the cells. (**B**) Cell state hierarchy maps based on single-cell transcriptome data is shown for the indicated subjects. Precursor cluster-T_{EMRA} (light green), effector cluster-T_{EMRA} (orange), T_{CM} (brown) and T_{EM} (yellow). Each circle is a cell and black line the connection between cells.

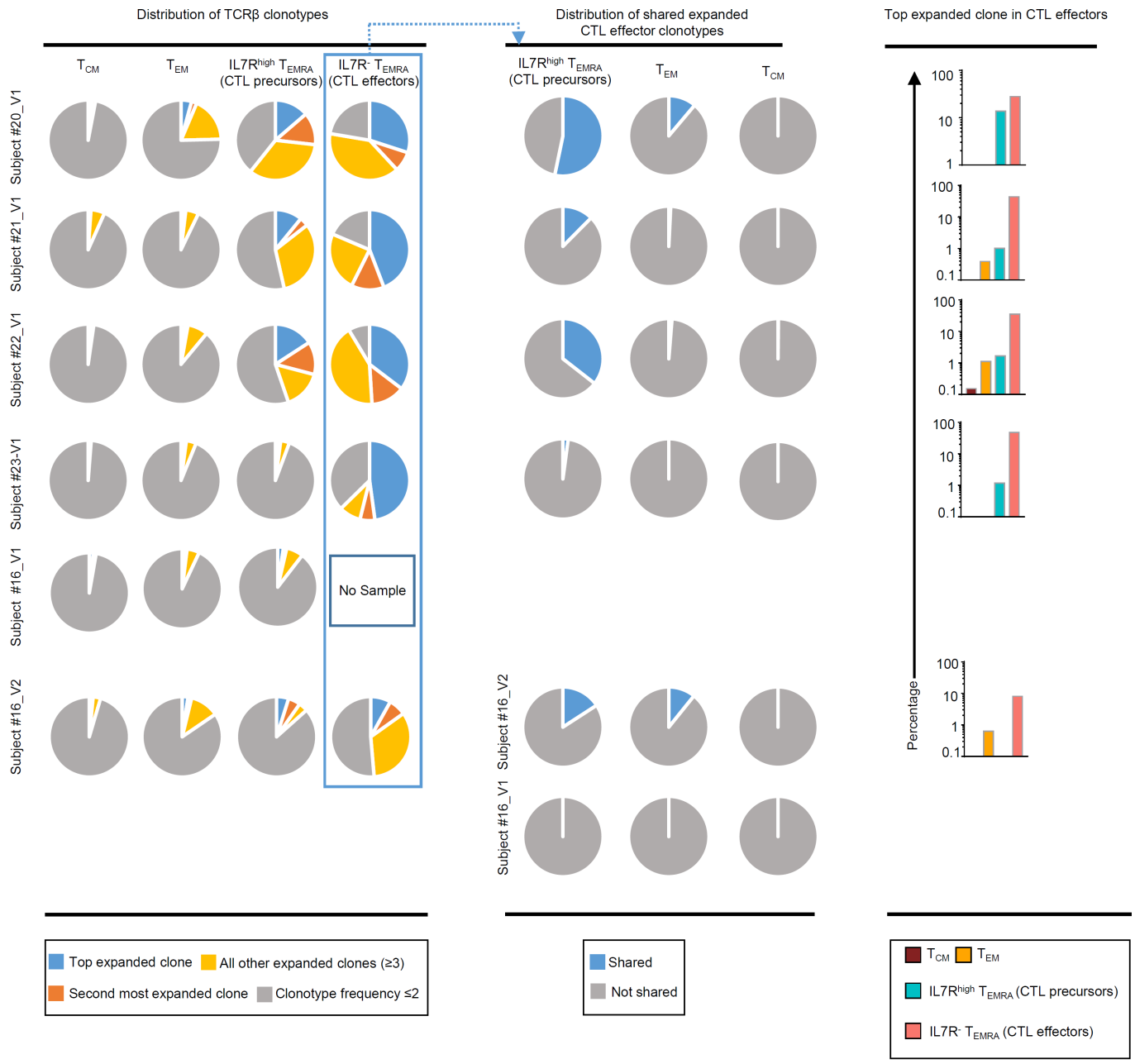


Fig. S7. CD4-CTL precursors share clonotypes with CD4-CTL effectors. Pie charts (left panel) show the distribution of TCRβ clonotypes based on clonal frequency, the most clonally expanded (top expanded clone, blue) and the next most clonally expanded (second most expanded clone, orange), rest of the expanded clonotypes (frequency ≥3) (all other expanded clones, yellow) and not expanded (frequency ≤2, grey). The pie charts (middle panel) show the distribution of shared (overlapping) expanded effector TCRβ clonotypes (frequency ≥3) within the other indicated subsets. The graphs (right) show the percentage overlap of the most expanded TCRβ clonotype from effector subset with the other indicated subsets.

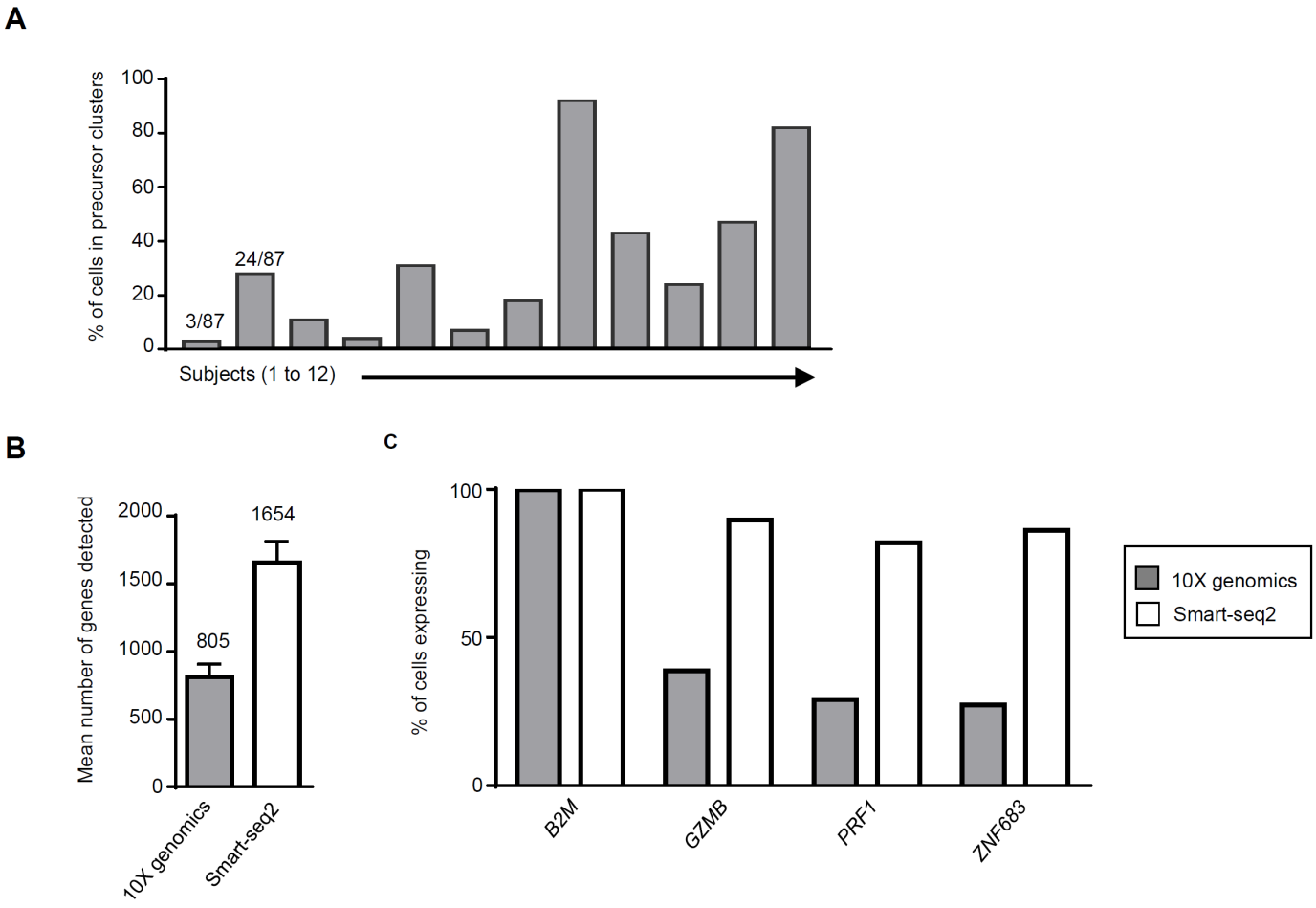


Fig. S8. Smart-seq2 is more sensitive compared to droplet based approach (10X genomics). (**A**) Bar graph shows the percentage of cells in precursor clusters across 12 subjects. The number on the top of the bars for subjects #1 and 2 indicates precursor cells out of total number of cells analyzed. (**B** and **C**) Comparison of the sensitivity of the single-cell RNA-seq methods, 10X genomics and modified smart-seq2 using cells from subjects #1 and 2. Bar graph shows the average number of genes detected by 10x genomics and smart-seq2 methods (B) and the percentage of cells expressing the indicated transcripts (C). *B2M* is shown as a control house-keeping gene.

DENV status	HLA	Geographical location	Subject number	T_{EMRA}	$IL7R^{high}$ T_{EMRA} (CTL precursors)	$IL7R^{-}$ T_{EMRA} (CTL effectors)	Batch number	T_{EM}	T_{CM}	DENV-specific T_{EM}	DENV-specific T_{EMRA}	Total/donor
Single-cell RNA-seq												
				Number of cells								
				Smart-seq2 method								
DENV-positive ($DENV^{+}$)	DRB1*0401 (Protective allele)	Sri Lanka	12	182 [189]			5, 8, 9	46 [48]	46 [48]			274 [285]
		Nicaragua	3	47 [48]			4	48 [48]	47 [47]			142 [143]
		Sri Lanka	6	189 [192]			4, 8, 10	45 [48]	48 [48]	70 [89]	62 [73]	414 [450]
	DRB1*0802 (Susceptible allele)	Sri Lanka	11	45 [48]			5					45 [48]
		Nicaragua	5	48 [48]			2					48 [48]
		Nicaragua	8	48 [48]			3					48 [48]
DENV-positive ($DENV^{+}$)		Sri Lanka	13	94							55	149
		Sri Lanka	14	94							58	152
		Sri Lanka	15	92							72	164
DENV-negative ($DENV^{-}$)		Sri Lanka	9	47 [47]			1					47 [47]
		Sri Lanka	10	45 [48]			1					45 [48]
		Sri Lanka	4	48 [48]			2					48 [48]
		Sri Lanka	7	44 [48]			3					44 [48]
		San Diego	2	87 [95]			7					87 [95]
		San Diego	1	87 [95]			6					87 [95]
		San Diego	16		188 [188]	182 [188]						370 [376]
	Total (Smart-seq2)		16	1197 [1234]	188 [188]	182 [188]		139 [144]	141 [143]	70 [89]	247 [258]	2164 [2244]
				10x genomics platform								
DENV-negative ($DENV^{-}$)		San Diego	2	3204								3204
		San Diego	1	3780								3780
	Total (10X genomics)		2	6984								6984
	Final total (both methods)		16	7938								9156 [9232]
Bulk RNA-seq (V1=Visit1, V2=Visit2)												
DENV-negative ($DENV^{-}$)		San Diego	16_V1		1	0		1	1			
		San Diego	16_V2		1	1		1	1			
		San Diego	20_V1		1	1		1	1			
		San Diego	20_V2		1	1		1	1			
		San Diego	21_V1		1	1		1	1			
		San Diego	21_V2		1	1		1	1			
		San Diego	22_V1		1	1		1	1			
		San Diego	22_V2		1	1		1	1			
		San Diego	23_V1		1	1		1	1			
		San Diego	23_V2		1	1		1	1			
TCR-seq (V1=Visit1, V2=Visit2)												
DENV-positive ($DENV^{+}$)		Sri Lanka	15		1	1		1	1			
DENV-negative ($DENV^{-}$)		San Diego	17		1	1		1	1			
		San Diego	18		1	1		1	1			
		San Diego	19		1	1		1	1			
		San Diego	16_V1		1	0		1	1			
		San Diego	16_V2		1	1		1	1			
		San Diego	20_V1		1	1		1	1			
		San Diego	20_V2		1	1		1	1			
		San Diego	21_V1		1	1		1	1			
		San Diego	21_V2		1	1		1	1			
		San Diego	22_V1		1	1		1	1			
		San Diego	22_V2		1	1		1	1			
		San Diego	23_V1		1	1		1	1			
		San Diego	23_V2		1	1		1	1			

Table S1. Summary of study subjects. For single-cell RNA-seq the number under each cell type indicates the cells that passed the quality control pipeline (Materials and Methods) out of the cells sequenced (number in the brackets). Batch number indicates the Nextera-XT library preparation batch.

Subject ID	Reconstructed TCRs	T_{EMRA}	T_{EM}	T_{CM}	DENV specific T_{EM}	DENV specific T_{EMRA}
12	TCRα	73 / 182 (40.1%)	21 / 46 (45.7%)	26 / 46 (56.5%)		
	TCRβ	111 / 182 (61.0%)	29 / 46 (63.0%)	30 / 46 (65.2%)		
	TCRαβ	57 / 182 (31.3%)	15 / 46 (32.6%)	17 / 46 (37.0%)		
3	TCRα	31 / 47 (66.0%)	39 / 48 (81.3%)	36 / 47 (76.6%)		
	TCRβ	36 / 47 (76.6%)	40 / 48 (83.3%)	42 / 47 (89.4%)		
	TCRαβ	25 / 47 (53.2%)	33 / 48 (68.8%)	33 / 47 (70.2%)		
6	TCRα	93 / 189 (49.2%)	21 / 45 (46.7%)	19 / 48 (39.6%)	31 / 72 (43.1%)	23 / 60 (38.3%)
	TCRβ	122 / 189 (64.6%)	31 / 45 (68.9%)	32 / 48 (66.7%)	33 / 72 (45.8%)	25 / 60 (41.7%)
	TCRαβ	62 / 189 (32.8%)	12 / 45 (26.7%)	17 / 48 (35.4%)	24 / 72 (33.3%)	17 / 60 (28.3%)
11	TCRα	16 / 45 (35.6%)				
	TCRβ	31 / 45 (68.9%)				
	TCRαβ	12 / 45 (26.7%)				
5	TCRα	35 / 48 (72.9%)				
	TCRβ	42 / 48 (87.5%)				
	TCRαβ	30 / 48 (62.5%)				
8	TCRα	30 / 47 (63.8%)				
	TCRβ	37 / 47 (78.7%)				
	TCRαβ	24 / 47 (51.1%)				
9	TCRα	16 / 47 (34.0%)				
	TCRβ	38 / 47 (80.9%)				
	TCRαβ	15 / 47 (31.9%)				
10	TCRα	17 / 45 (37.8%)				
	TCRβ	25 / 45 (55.6%)				
	TCRαβ	10 / 45 (22.2%)				
4	TCRα	19 / 48 (39.6%)				
	TCRβ	42 / 48 (87.5%)				
	TCRαβ	19 / 48 (39.6%)				
7	TCRα	14 / 44 (31.8%)				
	TCRβ	34 / 44 (77.3%)				
	TCRαβ	11 / 44 (25.0%)				
2	TCRα	63 / 87 (72.4%)				
	TCRβ	74 / 87 (85.1%)				
	TCRαβ	53 / 87 (60.9%)				
1	TCRα	42 / 87 (48.3%)				
	TCRβ	77 / 87 (88.5%)				
	TCRαβ	39 / 87 (44.8%)				
15	TCRα	59/92 (64.1%)				43/72 (59.7)
	TCRβ	79/92 (85.9%)				49/72 (68.1%)
	TCRαβ	52/92 (56.5%)				42/72 (58.3%)
13	TCRα	68/94 (72.3%)				31/55 (56.4)
	TCRβ	78/94 (83%)				31/55 (56.4%)
	TCRαβ	60/94 (63.8%)				20/55 (36.4%)
14	TCRα	64/94 (68.1%)				34/58 (58.6%)
	TCRβ	80/94 (85.1%)				37/58 (63.8)
	TCRαβ	59/94 (62.8%)				29/58 (50%)

Table S3. Summary of TCRα and TCRβ chains recovered from full-length single-cell transcriptomes of all single cells in the 15 subjects. The values indicate the identity of the complete CDR3 and VDJ genes.

Antigen	Conjugate	Clone	Supplier
CCR7	APC	G043H7	Biolegend
CCR7	PE	G043H7	Biolegend
CCR7	PerCP-Cy5.5	G043H7	Biolegend
CD127 (IL7R)	APC	eBioRDR5	eBioscience
CD127 (IL7R)	BV421	A019D5	Biolegend
CD14	BV510	M5E2	Biolegend
CD14	BV421	M5E2	Biolegend
CD14	V500	M5E2	BD Biosciences
CD16	BV605	3G8	Biolegend
CD19	BV510	HIB19	Biolegend
CD19	V500	HIB19	BD Biosciences
CD244	PE-Cy7	C1.7	Biolegend
CD244	PE-Dazzle	C1.7	Biolegend
CD28	BV785	CD28.2	Biolegend
CD3	PE-Cy7	UCHT1	Biolegend
CD3	BV605	UCHT1	Biolegend
CD3	AlexaFluor700	UCHT1	BD Pharmingen
CD314	FITC	1D11	Biolegend
CD4	APC-Cy7	RPA-T4	Biolegend
CD4	APC-eFluor780	RPA-T4	eBioscience
CD45RA	PE	JS-83	eBioscience
CD45RA	APC	JS-83	eBioscience
CD45RA	Alexa Fluor 700	HI100	Biolegend
CD45RA	eFluor450	HI100	eBioscience
CD56	PE	CMSSB	eBioscience
CD8a	BV510	RPA-T8	Biolegend
CD8α	FITC	RPA-T8	Biolegend
CD8α	BV650	RPA-T8	Biolegend
CX3CR1	PerCP-Cy5.5	2A9-1	Biolegend
GPR56	PE-Cy7	CG4	Biolegend
GZMB	PE	REA226	Milteny Biotech
IFNγ	FITC	4S.B3	eBioscience
KLRG1	PE	13A2	eBioscience
Live/Dead	ef506		eBioscience
PRF1	PE	B-D48	Biolegend
Aqua			Biolegend
DNA	DAPI		Life Technologies
DNA	PI		Sigma Aldrich

Table S10. List of antibodies used in the study.

Name of the primer	Sequence
Reverse transcription (RT)	
TCR-TSO	AAGCAGUGGTAUCAACGCAGAGUNNNNUNNNNUNNNNUCTTrGrGrGrG
ACR_RT	GTCTAGCACAGTTTTGTC
BCR_RT	GTATCTGGAGTCATTGA
1st PCR	
M1ss	AAGCAGTGGTATCAACGCA
ACR_Amp	GTCACTGGATTTAGAGTC
BCR_Amp	TGCTTCTGATGGCTCAAACAC
2nd PCR	
M1s_701	NNNNTAAGGCGACAGTGGTATCAACGCAGAG
M1s_702	NNNNCGTACTAGCAGTGGTATCAACGCAGAG
M1s_703	NNNNAGGCAGAACAGTGGTATCAACGCAGAG
M1s_704	NNNNTCCTGAGCCAGTGGTATCAACGCAGAG
M1s_705	NNNNGGACTCCTCAGTGGTATCAACGCAGAG
M1s_706	NNNNTAGGCATGCAGTGGTATCAACGCAGAG
M1s_707	NNNNCTCTCTACCAGTGGTATCAACGCAGAG
M1s_710	NNNNCGAGGCTGCAGTGGTATCAACGCAGAG
M1s_711	NNNNAAGAGGCACAGTGGTATCAACGCAGAG
M1s_712	NNNNGTAGAGGACAGTGGTATCAACGCAGAG
M1s_714	NNNNGCTCATGACAGTGGTATCAACGCAGAG
M1s_715	NNNNATCTCAGGCAGTGGTATCAACGCAGAG
TCR_A_502	NNNNCTCTCTATGGGTCAGGGTTCTGGATAT
TCR_A_503	NNNNTATCCTCTGGGTCAGGGTTCTGGATAT
TCR_A_505	NNNNGTAAGGAGGGGTCAGGGTTCTGGATAT
TCR_A_506	NNNNACTGCATAGGGTCAGGGTTCTGGATAT
TCR_A_507	NNNNAAGGAGTAGGGTCAGGGTTCTGGATAT
TCR_A_508	NNNNCTAAGCCTGGGTCAGGGTTCTGGATAT
TCR_A_510	NNNNCGTCTAATGGGTCAGGGTTCTGGATAT
TCR_A_511	NNNNTCTCTCCGGGGTCAGGGTTCTGGATAT
TCR_B_513	NNNNTCGACTAGACACSTTKTTCAGGTCCTC
TCR_B_515	NNNNTTCTAGCTACACSTTKTTCAGGTCCTC
TCR_B_516	NNNNCCTAGAGTACACSTTKTTCAGGTCCTC

Table S11. List of primers used for TCR-seq assays