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 sequencing in more than 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 costimulatory 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 with CD4⁺ T cells in the central memory (T_CM) and effector memory (T_EM) subsets. Simultaneous T cell antigen receptor (TCR) analysis in single cells and bulk subsets revealed that CD4⁻T_EMRA cells show marked clonal expansion compared with T_CM and T_EM cells and that most of CD4⁻T_EMRA were dengue virus (DENV)–specific in donors with previous DENV infection. The profile of CD4⁻T_EMRA was highly heterogeneous across donors, with four distinct clusters identified by the single-cell analysis. 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 used to generate durable and effective CD4-CTL immunity.

INTRODUCTION
After exposure to pathogens, naive CD4⁺ T helper (T_H) lymphocytes differentiate into memory and effector T_H cell subsets: tissue-resident memory 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 on the basis of their cytokine profile and functional properties into T_H1, T_H2, T_H17, T_H* (T_H1/T_H17), regulatory T cell, and follicular helper (T_FH) T cell subsets (3, 4). Although T lymphocytes with cytotoxic function (CTLs) are predominantly restricted to conventional major histocompatibility complex (MHC) class I–restricted CD8⁺ T lymphocytes, the existence of MHC class II–restricted T_H1 cells with cytotoxic potential (CD4-CTLs) in humans, nonhuman primates, and mice has been reported for many decades (5, 6). However, compared with the other T_H1 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), HIV, dengue virus (DENV), and hepatitis C virus (5, 7–15). CD4-CTLs have also been detected in mouse lungs as early as 1 week after acute influenza viral infection (16, 17). 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 donors carrying human leukocyte antigen (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). 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 antitumor 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 notable intra- and interdonor 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-7 receptor (IL-7R).
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 sequencing (RNA-seq) in more than 9000 cells isolated ex vivo from the TEMRA subset and, as a control, in the TEM and TCM subsets (Fig. 1A and table S1) (21). Using complementary methods of single-cell differential gene expression analysis [SCDE and MAST analysis, see Supplementary Materials and Methods and in (22, 23)], we compared the full-length transcriptome of CD4+ T lymphocytes present in TEMRA, TEM, and TCM subsets from three donors 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 with the T_{EM} and T_{CM} subsets (see 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, CT5W, FCRL6, SPON2, CX3CR1, SIPR5, NKG7, and CD244 (Fig. 1, B and C) (6, 25); 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 the T_{EMRA} subset (Fig. 1, E and F). Transcripts encoding transcription factors (TFs) related to CTL function such as ZNF683 (Hobit), and 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. Coexpression analysis of T_{EMRA}-enriched transcripts also revealed a number of genes (PFN1, PFN1P1, EFHD2, VCL, DIP2A, SYNE1, and PLEK) (29–32) whose expression was highly correlated with cytotoxicity signature genes, suggesting that the 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 after exposure to certain viruses such as DENV or CMV (5, 7–10), we expected to see a more restricted TCR repertoire, that is, greater clonal expansion in the CD4-T_{EMRA} subset compared with 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 to 89% of single cells, the TCRα chain in 31 to 81%, and both chains in 22 to 70% of cells across all memory subsets (table S3). As expected, a greater clonal expansion was observed in the T_{EMRA} subset compared with other subsets, as shown by highly interconnected clonotype network graphs for single cells from donor #6 (Fig. 2A). Furthermore, the analysis of single cells that shared TCRα or TCRβ chain clonotypes showed that more than 50% of cells in the T_{EMRA} subset were clonally expanded (Fig. 2B, left and middle). 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 with only ~5% and none of the cells in T_{EM} and T_{CM} subsets, respectively (Fig. 2B, right, and table S4). Together, these results suggested a highly restricted TCR repertoire in the T_{EMRA} subset (Fig. 2, A and B). The clonally expanded cells had a higher mean expression of cytotoxicity 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 donors with previous DENV infection (Fig. 2, E and F, fig. S1, and table S4). Next, we asked what fraction of these DENV-specific TCR clonotypes was present in the general pool of cells present in the T_{EMRA} subset from the same donor. 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 donors, one or both of the top two clonally expanded clonotypes in the T_{EMRA} population were always DENV-specific (Fig. 2, E and F, fig. S1, and table S4), which suggested that most of the clonally expanded cells in the T_{EMRA} population in these individuals were specific for DENV.

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

We next asked whether the clonality and transcriptome of CD4-CTLs differed between donors with and without previous DENV infection or among donors across different geographical locations [Sri Lanka (Asia) and the Americas (Nicaragua and San Diego, California)]. We observed a wide range in the proportion (0 to 88%; median, 44%) of clonally expanded cells across the 12 donors with no major differences in their proportion when classifying donors based on previous DENV infection status or geographical location [Sri Lanka (Asia) versus the Americas; table S1]. When we compared the percentage of clonally expanded cells between donors with higher versus lower proportion of CD4+ T lymphocytes in the T_{EMRA} subset (classified as T_{EMRA}^high or T_{EMRA}^low donors), we observed greater clonal expansion in T_{EMRA}^high donors (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 donors with a larger T_{EMRA} pool have a greater degree of clonal expansion (Fig. 3B).

In most donors, we observed sharing of a single unique TCRα and TCRβ chain clonotype in a large fraction of CD4-T_{EMRA} cells, as exampled in donor #4 [58% (11 of 19)], donor #1 [50% (21 of 42)], donor #2 [41% (24 of 58)], donor #5 [37% (11 of 30)], donor #8 [22% (6 of 27)], donor #3 [16% (4 of 25)], and donor #6 [10% (7 of 67)] of cells (Fig. 3C). Considering some of these donors were previously infected with DENV (DENV+ donors) raised the hypothesis that selection and expansion of such TCR clonotypes may be linked to the DENV infection. Even in donors without previous DENV infection (DENV− donors), 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 donors (Fig. 3, D and E). For several cytotoxicity-related transcripts, the expression pattern was highly variable across the 12 donors (Fig. 3, D and E). CD4-T_{EMRA} cells from donors with a larger preexisting 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, and IFTT2 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 donors with larger preexisting T_{EMRA} pool (Fig. 3F).

Therefore, our combined transcriptomic, protein, and TCR analysis suggests that the CD4-T_{EMRA} subset exhibits quantitative and qualitative differences across different donors irrespective of their DENV
Fig. 2. CD4-TEMRA cells show marked clonal expansion. (A) Clonotype network graphs of cells from the TEMRA, TEM, and TCM subset and DENV-specific TEMRA and DENV-specific TEM cells from donor #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 dashed line indicates a precursor cell and an effector cell sharing the same unique TCRs and TCRβ chain clonotype. (B) Bar graph shows the percentage of cells with TCRα (left), TCRβ (middle), and TCRβ (right) chain clonotype frequency ≥2 (clonally expanded cells) in TEMRA, TEM, and TCM subsets. Error bars are mean ± SEM from three donors with the 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 TEMRA-enriched transcripts (SCDE and MAST analysis comparing TEMRA cells with clonotype frequency ≥2 or =1, Benjamini-Hochberg adjusted P < 0.05 and ≥2-fold change) in clonally expanded cells (unique TCRα and TCRβ chain clonotype frequency ≥2; magenta) and nonexpanded cells (unique TCRα and TCRβ chain clonotype frequency =1; green). The cells are arranged on the basis of 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 nonexpanded TEMRA cells. The shapes represent the distribution of cells based on their log2(TPM + 1) expression values (y axis). The color scale represents the mean expression. (E) Bar graph shows the number of TEMRA cells with a unique TCRα and TCRβ chain clonotype for donor #6. Stars in red color indicate the clonotypes shared by DENV-specific TEMRA. Pie chart shows the percentage of expanded TEMRA cells that share the unique TCRα and TCRβ chain clonotypes of DENV-specific TEMRA cells (pink). (F) Bar graph shows the percentage of the expanded TEMRA cells that share TCRβ clonotypes of DENV-specific TEMRA cells. Each dot represents a donor. Error bars are mean ± SEM from four donors.

infection status or geographical location and, in many instances, related to the CD4-CTL function.

CD4-CTL effector cells revealed by single-cell analysis

We next asked whether the heterogeneity observed in the single-cell transcriptomes of CD4-TEMRA was due to the presence of multiple distinct subsets. Four clusters were revealed by unbiased clustering of CD4-TEMRA cells from the 12 study donors (917 cells) using the Seurat software (Fig. 4A) (38); other methods of clustering also revealed a similar pattern, and no major changes were introduced by technical and batch effects (see figs. S2 and S3, table S1, and Supplementary Materials and Methods) (39). The proportion of cells in each cluster varied greatly among donors, with marked differences observed between donors from Sri Lanka and the Americas (Fig. 4B, left), suggesting that the nature and type of infections may shape the molecular profiles of CD4-CTLs. Clonal expansion was observed more frequently in cells from clusters 1 and 2 (Fig. 4B, right, and table S4).
Single-cell differential gene expression analysis among the four clusters revealed notable differences in their molecular profiles (Fig. 4C, fig. S4, A to D, and table S5). IPA and differential expression analysis of the transcripts enriched in clusters 1 and 2 with comparison of clusters 3 and 4 showed significant overrepresentation of genes encoding products related to cytotoxicity, such as GZMB, GZMH, PRF1, GNLY, NKG7, FASLG, and CASP10, many were expressed in a cluster-specific manner (Fig. 4, D and E, fig. S4D, and table S5). On the basis of 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, and 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 to 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 costimulation molecules (Figs. 4E and 5, A and B) (44, 45). These results indicate that cells in clusters 1 and 2 have a terminal effector state and hence are CD4-CTL effectors with high cytotoxic potential. Genes linked to cell survival and CD4-CTL function such as PRSS23, SPON2, TCF7 (encodes for TCF1), CRTAM, and 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, and TCF7) (47, 48) in the survival and long-term persistence of these high cytotoxic terminal effector cells deserves further investigation (45).

The TEMRA 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 with 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 IL-7R...
(IL7R or CD127), which is known to play an important role in mediating the long-term homeostatic survival of naive and memory T lymphocytes (Fig. 5A) (52–55). The mean expression of IL7R transcripts in cluster 4 cells was even higher than that in TCM 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 (IL-7R$^{high}$ and KLRG1$^{−}$) described in murine models of resolving acute infections (52, 56). Transcripts encoding CD27 and CD28 costimulation molecules, lymphotoxin B (LTB), and JUNB TF were also expressed at higher levels in clusters 3 and 4, suggesting that CD4-TEMRA cells are poised to acquire effector function as indicated by the expression of the CD4-TEMRA subset in the CD4$^{+}$ T cell population across 104 samples from 89 individuals. Contour plots show the surface expression of CD45RA and CCR7 (left) or CD127 (right) in live and singlet-gated CD3$^{+}$CD4$^{+}$ T cells from 15 donors with two longitudinal samples (V1 and V2). Donor #16, marked in red color, shows marked increase in the proportion of CD4$^{+}$TEMRA and CD4$^{+}$TEMRA (effectors) cells from V1 to V2. Contour plots show the surface expression of CD45RA and CCR7 (top) or CD127 (bottom) in live and singlet-gated CD3$^{+}$CD4$^{+}$ T cells obtained from PBMCs (left). Contour plots show coexpression of CD45RA and CCR7 in TEMRA and TCM subsets (right). The gating strategy shows CD127$^{−}$ and CD127$^{high}$ proportion. (F) Proportion of total CD4$^{+}$TEMRA (left), CD127$^{−}$CD4$^{+}$TEMRA (middle), and CD127$^{high}$CD4$^{+}$TEMRA cells (right) in CD4$^{+}$ T cells from 15 donors with two longitudinal samples (V1 and V2). Donor #16, marked in red color, shows marked increase in the proportion of CD4$^{+}$TEMRA and CD4$^{+}$TEMRA (effectors) cells from V1 to V2. Contour plots show the surface expression of CD45RA and CCR7 (top) or CD127 (bottom) in live and singlet-gated CD3$^{+}$CD4$^{+}$ T cells obtained from PBMCs of donor #16 at two time points (168 days apart) [V1 (left) and V2 (right)]. Bottom panels show the expression of CD127 in the CD4$^{+}$TEMRA subset for the same donor. n.s, not significant; from Student’s two-tailed paired t test. Numbers inside the contour plots show the proportion of the indicated cell type. (H) Bulk RNA-seq analysis of the IL-7R$^{high}$ TEMRA subset (CD4-CTL precursors), IL-7R$^{−}$ TEMRA (CD4-CTL effectors), and TCM cells shows the row-wise normalized TPM counts of top 200 (100 up-regulated and down-regulated, based on fold change) differentially expressed transcripts obtained by pairwise comparison of IL-7R$^{−}$ TEMRA (CD4-CTL effectors) versus TCM from the five donors (DESeq2 analysis, Benjamini-Hochberg adjusted P < 0.05 and ≥2-fold change). The panel above the heat map identifies the cell type, donor, and visit. (I) Contour plots show the coexpression of CD244 or GPR56 with CD28 in TCM (red), IL-7R$^{high}$ TEMRA (CD4-CTL precursors) (magenta), and IL-7R$^{−}$ TEMRA (CD4-CTL effectors) (blue) in singlet-gated CD3$^{+}$CD4$^{+}$ T lymphocytes.
cells in that they expressed several transcripts that are specifically enriched in TTEMRA cells (KLRC1, TBX21, S1PR5, FGFBP2, CCL4, PRF1, GZMH, GNLY, NKG7, ZEB2, and GPR56), albeit at lower levels compared with CD4-CTL effectors (cells in clusters 1 and 2) (Figs. 4E and 5C). On the basis of these results, we hypothesize that cells in clusters 3 and 4 likely represent memory precursor cells for the cells present in cluster 1 and 2 that have a terminal CD4-CTL effector phenotype.

Given that surface expression of IL-7R (CD127) can be readily determined by flow cytometry, we looked for the presence of potential CD4-CTL precursors within the CD4-TTEMRA subset from an independent cohort of healthy donors. For these studies, we capitalized on the La Jolla Institute cohort of 89 healthy donors, of which 15 donors had longitudinal samples (two time points) (104 samples in total). As expected, there was a large variation (0.12 to 15.2%) in the proportion of the CD4-TTEMRA subset across the study donors (Fig. 5D). On the basis of the surface expression pattern of IL-7R, cells in the CD4-TTEMRA subset were classified into IL-7Rhigh and IL-7R− (Fig. 5E). The expression level of IL-7R on IL-7Rhigh TTEMRA cells was similar to that observed in TCM and Tn (naïve T cell) CD4+ T lymphocytes (Fig. 5E and fig. S5B). These cells (IL-7Rhigh TTEMRA cells) may represent the CD4-CTL precursors defined by our single-cell transcriptome analysis. In longitudinal samples obtained at 3- to 6-month intervals, the proportion of cells in the CD4-TTEMRA subset was quite stable (Fig. 5F); however, in one donor, we noted a marked (>8-fold) expansion of the CD4-TTEMRA subset; this expansion was mainly confined to the IL-7R TTEMRA cells (CD4-CTL effectors), and not the IL-7Rhigh TTEMRA (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 the IL-7Rhigh TTEMRA subset (CD4-CTL precursors) shared the molecular profile of both TCM (long-term memory cells) and cytotoxic cells, we isolated IL-7Rhigh TTEMRA and IL-7R TTEMRA (CD4-CTL effectors) along with TCM and TEM subsets and performed RNA-seq in two longitudinal samples from five donors (Fig. 5H). Consistent with our single-cell cluster analysis (Fig. 5C), the IL-7Rhigh TTEMRA subset shared molecular features of both TCM (memory precursor cells) and cytotoxic cells (IL-7R TTEMRA, CD4-CTL effectors) that were stable over two longitudinal visits (Fig. 5H and table S6). The IL-7Rhigh TTEMRA subset (CD4-CTL precursors) expressed both cytotoxicity-related molecules (GPR56 and CD244) and a co-stimulatory molecule (CD28, TEM-enriched) at the protein level (Fig. 5I). Together, these results show that cells in the IL-7Rhigh TTEMRA subset have properties of memory precursor cells and cytotoxic cells.

**CD4-CTL precursors share TCRs with CD4-CTL effectors**

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 to 12 donors, TCR clonotypes were shared between CD4-CTL precursors and effectors (Figs. 2A and 6A, fig. S6A, and table S4). As examples, in donor #1, 23 of 80 TTEMRA cells had the same TCRα chain clonotype (fig. S6A) as a single cell from precursor cluster (cluster 3); in donor #12, four TCRβ chain clonotypes were shared between CD4-CTL precursors and effectors (Fig. 6A, clones highlighted in dashed lines with the precursor cells indicated with red arrows). A similar pattern was observed in donors #3, #5, and #6 (Figs. 2A and 6A and table S4). To further support our inference, we constructed cell-state hierarchy maps using the Sincell software (39) for cells in the TCM, TEM, and TTEMRA 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 TEM or TCM 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 sequencing (TCR-seq) assays (57) in TCM, TEM, IL-7Rhigh TTEMRA (CD4-CTL precursors), and IL-7R− TTEMRA (CD4-CTL effectors) subsets from 14 donors (table S7). Consistent with single-cell TCR analysis (Fig. 2, A and B), the CD4-CTL precursor and effector cells in the TTEMRA subset had highly restricted TCR repertoire, compared with cells in the TCM and TEM subsets, 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, and fig. S7, left). A larger fraction of the clonotypes in the CD4-CTL precursor cells relative to TEM and TCM was composed of the expanded clonotypes (frequency, ≥3) present in the CD4-CTL effector cells (Fig. 6E, middle, and fig. S7, middle). 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 TEM and TCM (Fig. 6E, right graphs, and fig. S7, right graphs).

We also performed TCR-seq in CD4+ T cell memory subsets from five donors who provided longitudinal samples obtained at 3- to 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 effector clonotypes from second visit (V2) that were present in the different CD4+ T cell memory subsets from the first visit (V1) (table S8). In four of the five donors analyzed, a greater fraction of the TCR clonotypes detected in CD4-CTL effectors at V2 was shared with cells in the precursor population (IL-7Rhigh TTEMRA subset) relative to the TEM or TCM subset at V1 (Fig. 6F, middle, fig. S7, and table S8). 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 TEM, TCM, 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 the IL-7Rhigh TTEMRA subset is the predominant precursor/progenitor for the CD4-CTL effector cells.

**Variable number of CD4-CTL precursors across donors**

The proportion of CD4-CTL precursors varied greatly among donors, ranging from 3 to 92% of cells in the TTEMRA subset (fig. S8A). To determine whether additional CD4-CTL subsets exist and to definitively assess the presence of CD4-CTL precursors, we analyzed single-cell transcriptomes of 10 times as many cells (>6000 cells) from two donors using the high-throughput 10x genomics platform. The sensitivity of the 10x genomics platform, which uses beads to capture transcripts and also sequences only their 3′ ends (59), is twofold lower than the Smart-seq2 protocol used to study full-length transcriptomes of the same donors (fig. S8, B and C). In donor #2, where 24 of 87 cells (28%) (fig. S8A) were previously assessed as CD4-CTL precursors, unbiased clustering of more than 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). These two clusters differentially expressed transcripts characteristic of CD4-CTL effectors (GZMB and NKG7) and precursors (IL7R, JUNB, and LTB), respectively (Fig. 7B, left, and table S9). Even in
Fig. 6. CD4-CTL effectors share TCR clonotypes with CD4-CTL precursors. (A) Clonotype network graphs of single cells from T EMRA subsets from donors #12 (top) and #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 dashed 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 TEMRA (light green), effector cluster TEM (orange), TCM (brown), and TEM (yellow) cells from donor #12. The network shows connection (black line) between cells (circle). (C and D) TCR-seq assays in TCM, TEM, IL-7R hi TEMRA (CD4-CTL precursors), and the IL-7R– TEMRA (CD4-CTL effectors) subset from 14 donors; 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 to 14 donors. *P < 0.05, **P < 0.005, and ***P < 0.001 from Student’s paired two-tailed t test; ns, not significant. (E) Pie charts (left) 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) as well as the rest of the expanded (≥3) (all other expanded clones; yellow) and nonexpanded clones (frequency ≤2; gray). The pie charts (middle) 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) 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), as well as the rest of the expanded (≥3) (all other expanded clones; yellow) and nonexpanded clones (frequency ≤2; gray). The pie charts (middle) show the distribution of shared (overlapping) expanded V2 effector clonotypes (frequency ≥3) within other subsets from V1 (V2 samples). The graphs (right) show the percentage overlap of the most expanded TCRβ clonotype from the effector subset at V2 with other subsets at V1.

Donor #1, where only 3 of 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 of 3664 cells (3.2%)] when analyzing more cells by the 10x genomics platform (Fig. 7, A and B, right, and table S9). At the single-cell level, CD4-CTL precursors in the CD4-TEMRA 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 IL-7R hi and IL-7R–TEMRA cells (Fig. 7C and table S1).
DISCUSSION

CD4-CTLs have long been considered to be terminal effector cells derived from T\textsubscript{EM} cells after 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\textsubscript{EMRA} subset did identify cells that have features of terminal CD4-CTL effector cells (KLRG1\textsuperscript{high} and IL-7R\textsuperscript{low}; clusters 1 and 2). However, we observed another population of cells in the T\textsubscript{EMRA} subset that displayed a molecular program (KLRG1\textsuperscript{low} and IL-7R\textsuperscript{high}; clusters 3 and 4) indicative of memory.
precursor cells, intermixed with several features of CD4-CTLs (KLRG1, TBX21, 1P15R, FGFBP2, CCL4, PRF1, GZMH, GNLY, NK7G, ZEB2, and GPR56), albeit less prominently expressed than in the CD4-CTL effectors. Bulk and single-cell transcriptome analysis of the purified IL-7R	extsuperscript{high} and IL-7R	extsuperscript{T_EⅅRA} subset further confirmed our observation from single-cell transcriptome analysis of total T_EⅅRA subset. In several donors, 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. Although the kinetic studies support the development of effectors from precursor cells, the nature of our studies in humans 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 previously unknown molecules potentially important for their differentiation and function and represent attractive targets for further validation studies.

We showed that surface expression of the IL-7R (CD127) in T_EⅅRA cells defines a subset of cells with the expression of IL-7R as high as that observed in the T_CM subset. Given the established role of the IL-7R signaling pathway in homeostatic T cell proliferation and survival, such IL-7R	extsuperscript{high} T_EⅅRA cells are likely to represent the precursors of CD4-T_EⅅRA. The isolation of a CD4-CTL precursor subset based on the surface expression of IL-7R 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 previously unknown players (PFN1, PENIP1, EFHD2, VCL, DIP2A, PLEK, and SYNE1) whose single-cell coexpression 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 because they express low levels of the costimulatory 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 donors, suggesting that, compared with other conventional T1 effector cells, different molecular mechanisms may operate in CD4-CTL effectors to promote their long-term survival. IL-7R	extsuperscript{T_EⅅRA} CD4-CTL effectors highly expressed several molecules that are linked to cell survival such as CRTAM, ZNF683 (Hobit), PRSS23, SPON2, and TCF7 (encodes 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_EⅅRA 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 donors. We identified a T_EⅅRA 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 originating from T_N 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_EⅅRA in response to viral infections such as DENV and CMV, it is possible that IL-7R	extsuperscript{high} T_EⅅRA 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 after 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 use single-cell RNA-seq assay to capture the transcriptome of individual cells in CD4+ T cell memory subsets in human peripheral blood mononuclear cells (PBMCs). 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 were performed using R, Qucoore 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**

immunology.sciencemag.org/cgi/content/full/3/19/eaan8664/DC1

Materials and Methods

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

Fig. S2. TEMRA cells cluster into four distinct clusters.

Fig. S3. Batch analysis of TEMRA cells.

Fig. S4. TEMRA 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 TEMRA subset.

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

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

Table S1. Summary of study donors.

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

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

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

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 chain sequences derived from TCR-seq analysis of 14 donors [for data in Fig. 6 (E and F) and fig. S7].
– chain clonotype sharing derived from TCR-seq analysis of α
Table S8. TCR β
SCIENCE IMMUNOLOGY

15.
17.
18.
10.
20.
19.

12.
8.
2.
6.
H.

N.
K.

virus infection.
86 expressing gamma interferon and perforin mediate protection against lethal influenza

M.
T.

K.

E.
Kuhlmann, T.

Zaunders, W.

Brown, S.

R.

Yurdaydin, J.

Leeuwen, E.

Dyer, B.

27.

M.

J.

G.

Cooper, S.

D.

Kolla, A.

M.

Crotty,

Kelleher, J.

Kellchen, A.

Beck, A.

Prlic, P.

Krawczyk, A.

S.

Shen, N.

Yang, T.

A.

D.

M.

H.

J.

Kreuzeck, H.

W.

M.

K.

Protzer, R.

Lugli, The who’

Eickhoff,

Giesen, D.

Engel, S.

Flecken, D.

G.

Meissner, Z.

Böttcher, M.

Bauer, T.

Fleckenstein, D.

Bauer, T.

M.

K.

s who of

Kastenmüller, P.

Roederer, E.

Lugli, The who’

Kastenmüller, P.

L.

Kastenmüller, P.

L.

Peters, U.

Takeuchi, M.

Slichter, H.

Miller,

S.

P.

Lindsten,

Lindsten,

Halloran, The transcriptome of human

A.

J.

J.

D.

M.

Peters, A.

Peters, A.

Taylor,

A.

D.

M.

D.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.

M.
Precursors of human CD4+ cytotoxic T lymphocytes identified by single-cell transcriptome analysis

Veena S. Patil, Ariel Madrigal, Benjamin J. Schmiedel, James Clarke, Patrick O'Rourke, Aruna D. de Silva, Eva Harris, Bjoern Peters, Gregory Seumois, Daniela Weiskopf, Alessandro Sette and Pandurangan Vijayanand

Sci. Immunol. 3, eaan8664. DOI: 10.1126/sciimmunol.aan8664

Committing to cytotoxicity

CD4+ cytotoxic T lymphocytes (CD4-CTLs) were initially identified in patients with chronic viral infections, including dengue virus (DENV) infection, and these cells have been associated with protection in the context of severe DENV infection. Here, Patil et al. have carried out single RNA-seq and sequenced the T cell receptors (TCRs) of CD4+ T cells from human blood to identify precursors that give rise to CD4-CTLs. They report that CD4-CTLs 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.