Select sequencing of clonally expanded CD8$^+$ T cells reveals limits to clonal expansion

Huang Huang[a,b,1], Michael J. Sikora[c,1], Saiful Islam[c,1], Roshni Roy Chowdhury[a], Yueh-hsiu Chien[a], Thomas J. Scriba[d], Mark M. Davis[a,b,e,2], and Lars M. Steinmetz[c,f,g,2]

[a]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305; [b]Institute for Immunity, Transplantation and Infection, Stanford University School of Medicine, Stanford, CA 94305; [c]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305; [d]South African Tuberculosis Vaccine Initiative, Institute of Infectious Disease and Molecular Medicine and Division of Immunology, Department of Pathology, University of Cape Town, Cape Town 7925, South Africa; [e]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305; [f]Stanford Genome Technology Center, Stanford University, Palo Alto, CA 94304; and [g]Genome Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany

Contributed by Mark M. Davis, March 10, 2019 (sent for review December 6, 2018; reviewed by Mitchell Kronenberg and Olivier Lantz)

To permit the recognition of antigens, T cells generate a vast diversity of T cell receptor (TCR) sequences. Upon binding of the TCR to an antigen–MHC complex, T cells clonally expand to establish an immune response. To study antigen-specific T cell clonality, we have developed a method that allows selection of rare cells, based on RNA expression, before in-depth scRNA-seq (named SELECT-seq). We applied SELECT-seq to collect both TCR sequences and then transcriptomes from single cells of peripheral blood lymphocytes activated by a *Mycobacterium tuberculosis* (*Mtb*) lysate. TCR sequence analysis allowed us to preferentially select expanded conventional CD8$^+$ T cells as well as invariant natural killer T (iNKT) cells and mucosal-associated invariant T (MAIT) cells. The iNKT and MAIT cells have a highly similar transcriptional pattern, indicating that they carry out similar immunological functions and differ considerably from conventional CD8$^+$ T cells. While there is no relationship between expression profiles and clonal expansion in iNKT or MAIT cells, highly expanded conventional CD8$^+$ T cells down-regulate the interleukin 2 (IL-2) receptor alpha (IL2RA, or CD25) protein and show signs of senescence. This suggests inherent limits to clonal expansion that act to diversify the T cell response repertoire.

single-cell transcriptomics | T cell | iNKT | MAIT | senescence

The immune system of higher organisms is responsible for detecting and neutralizing a broad range of pathogens (1, 2). The collective immune response is complex and often requires the cooperation of many cell types. T cells play numerous essential roles in this system, from enabling antibody secretion by B lineage cells (3, 4), to detecting and eliminating pathogen-infected or cancerous cells in the organism. T cells can be highly specific, which they achieve by recombining and expressing heterodimeric T cell receptors (TCRs) on their cell surface, theoretically producing in excess of ($\sim 10^{16}$) unique heterodimers (5). If a naïve αβ T cell's TCR binds a specific antigen–MHC complex with sufficient affinity, the T cell will clonally expand while retaining its initial TCR sequence. This significantly increases the abundance of T cells that recognize a specific pathogen and consequently, enable an effective immune response. After clearance of the pathogen, most of these expanded T cells undergo apoptosis and the remainder mature to a memory stage. These cells can remain in circulation for decades in a quiescent but activatable state for a more immediate and robust response when recognizing the pathogen again (6, 7). Reactivation will lead to renewed T cell clonal expansion. During both phases of clonal expansion, however, the immune system must maintain a diverse T cell repertoire to fight any other invading insults. Recent data have demonstrated that repertoire richness in both CD4$^+$ and CD8$^+$ T cell compartments remains highly diverse even in aged people (8). To maintain such richness, T cell clonal expansion is well controlled by a balance of signals that impact the coordinately regulated processes of quiescence, proliferation, and cell death (9). However, how this balance is achieved to keep the T cell repertoire diversified is not well defined.

The main challenge in defining the mechanism of T cell repertoire diversity lies in the rarity of these cells. Memory T cells against a specific antigen are very rare (1–100 in 10^5 T cells) (10) and only a limited number of cells (10~1,000) may be sorted from one individual. Furthermore, to characterize clonal expansion and function of this rare cell population, both transcriptome and paired TCR sequences are needed from a single cell. For this purpose, gene expression profiling with high-throughput methods like 10× Chromium or DropSeq (11, 12) are not applicable because of their requirement for a large number of input cells (usually many thousands). An alternative is to use well-based sequencing strategies like Smart-Seq2, which is more attractive because it produces higher sequence resolution data relative to droplet-based methods (13). However, these are very costly. And, because for memory T cells there is no marker for those that are clonally expanded, much of the experimental costs will go to singletons that are often not of primary interest. This restriction on cell input numbers and the high costs are a limitation in single-cell immunology studies.

Significance

T cells are a central component of our immune system. They express a T cell receptor (TCR) on their surface, which detects pathogens and stimulates the T cell to initiate an immune response. Upon exposure to a specific pathogen, some T cells are activated and clonally expand. These pathogen-specific clonally expanded T cells, however, are generally rare and difficult to isolate. We have developed a technology (SELECT-seq) to isolate this rare population and to analyze their transcriptome and TCR composition. Therefore, our work may be the key to achieving a better understanding of pathogen-specific T cell clonality and function.

Author contributions: H.H., M.J.S., S.I., M.M.D., and L.M.S. designed research; H.H., M.J.S., and S.I. performed research; R.R.C., Y.-h.C., and T.J.S. contributed new reagents/analytic tools; H.H., M.J.S., and S.I. analyzed data; H.H., M.J.S., S.I., M.M.D., and L.M.S. wrote the paper; and M.M.D. and L.M.S. provided mentorship.

Reviewers: M.K., La Jolla Institute for Immunology; and O.L., Institut Curie.

The authors declare no conflict of interest.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo (accession no. GSE107646).

[1]H.H., M.J.S., and S.I. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: mmdavis@stanford.edu or Lars.Steinmetz@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1902649116/-/DCSupplemental.

Published online April 16, 2019.

To circumvent these issues, we developed a direct approach that employs targeted PCR on Smart-Seq2–derived cDNA libraries, permitting a more precise selection of cells to analyze, avoiding the costly processing steps of tagmentation and in-depth sequencing in less relevant cells. We refer to this approach as SELECT-seq and applied it to the CDR3α and CDR3β TCR sequences, selecting duplicated TCR sequences, which indicates clonal expansion, to characterize the function of clonal expanded antigen-specific T cells in human.

We used SELECT-seq to characterize clonally expanded $CD8^+$ memory T cells responding to a *Mycobacterium tuberculosis* (*Mtb*) lysate. Tuberculosis (TB) is one of the top 10 causes of death worldwide (14). While previous studies have shown an essential role of $CD4^+$ T cells in protection against *Mtb* (15, 16), the role of *Mtb*-specific $CD8^+$ T cells is unclear. *Mtb*-specific $CD8^+$ T cells may be protective, since it is known that murine and human $CD8^+$ T cells can recognize hundreds of *Mtb* epitopes (17–19). Besides conventional $CD8^+$ T cells, unconventional, non–MHC-restricted T cells, including mucosal associated invariant T (MAIT) and invariant natural killer T (iNKT) cells, are known to be involved in *Mtb* and other bacterial responses as well (20–23). In contrast to conventional cells, which have highly diverse alpha and beta TCR sequences, MAIT and iNKT cells have limited TCR diversity, typically using a semiinvariant or invariant alpha chain, containing a signature amino acid motif (21).

Thus, to study *Mtb*-reactive $CD8^+$ memory T cells, we stimulated peripheral blood mononuclear cells (PBMCs) from subjects with asymptomatic *Mtb* infection and sorted activated $CD8^+$ T cells for SELECT-seq. In total, we analyzed over 3,000 *Mtb*-responsive peripheral $CD8^+$ αβ T cells at the single-cell level. By analyzing TCR sequences, we found both conventional and unconventional (MAIT/iNKT) cells were activated by the *Mtb* lysate. We selected activated conventional and MAIT/iNKT cells for transcriptome sequencing. Analysis revealed a clear relationship between gene expression profiles and clone size in conventional $CD8^+$ T cells. However, we did not find the same association within unconventional cells, nor did we observe differences in expression profiles between MAIT and iNKT cells. Highly expanded conventional clones decrease interleukin 2 (IL-2) receptor alpha (IL2RA, or CD25) expression and increase indicators of senescence, suggesting that this is a mechanism for limiting clonal expansion, thereby increasing TCR diversity.

Results

Development of SELECT-seq for Clonally Expanded T Cells. The aim of SELECT-seq is to precisely select clonally expanded T cells for high-coverage sequencing within a single sequencing pipeline. This is achieved by screening single-cell cDNA libraries of large populations of cells with targeted PCR. We utilized a modified Smart-Seq2 protocol (24) to generate cDNA libraries (~15 μL) from single T cells, from which a small aliquot (1 μL) was taken for nested PCR to amplify and sequence the CDR3 regions of both TCRα and TCRβ chains (25). We then applied tagmentation and transcriptome sequencing of selected cell libraries based on their TCR sequence to obtain high-coverage in-depth sequence data (Fig. 1*A* and *SI Appendix*, Fig. S1). Notably, the cost of cDNA library generation and TCR sequencing (~$6) was much less than the combined cost of tagmentation and whole-transcriptome sequencing (~$32), which at scale, and when the cell type of interest is rare, this provides substantial cost savings for high-resolution methods (*SI Appendix*, Fig. S1*D*).

Applying this approach, we isolated PMBCs from donors who showed evidence of asymptomatic *Mtb* infection as gauged by a positive quantiferon-TB gold in-tube (QFT-GIT) test (26). PBMCs were activated with *Mtb* lysate ex vivo for 12 h, circumventing long-term culture after isolation from blood. This quick activation was sufficient to activate $CD8^+$ T cells as measured by activation markers, permitting the enrichment of *Mtb*-reactive cells.

We then used fluorescence-activated cell sorting (FACS) to select both conventional and unconventional $CD8^+$ T cells—MAIT and iNKT cells (*SI Appendix*, Fig. S2*A*). To define the surface markers to distinguish these cell types, we correlated $CD8^+$ T cell surface marker expression to T cell types as defined by their TCR sequences. We found that in addition to CD137, an activated $CD4^+$ T cell marker, CD154 (27–29) is up-regulated within $CD8^+$ T cells after lysate stimulation ($CD137^+$: 0.01~0.3% and $CD154^+$: 0.01~0.7% of $CD8^+$ T cells). While $CD8^+CD137^+$ T cells were enriched for conventional cells, $CD8^+CD154^+$ T cells were enriched for iNKT/MAIT cells (*SI Appendix*, Fig. S2 *A* and *B*). Further surface staining confirmed that only the $CD8^+CD154^+$ population is $CD161^+$, a marker for iNKT/MAIT $CD8^+$ T cells (*SI Appendix*, Fig. S2*C*) (30). In addition, both populations largely displayed memory T cell surface markers ($CCR7^-CD45RA^+$) and many contained identical TCR sequences, indicating clonal expansion (*SI Appendix*, Fig. S2*C*). Within the $CD8^+CD154^+$ population, invariant TCRα chains were ubiquitous, a property that defines both MAIT and iNKT $CD8^+$ T cells (Dataset S1) (21). Thus, while it had been reported that activated iNKT/MAIT cells up-regulate CD154 (31, 32), CD154 appears to be a precise and sortable activation marker for iNKT/MAIT $CD8^+$ T cells. We sorted, therefore, $CD8^+CD69^+CD137^+$ and $CD8^+CD69^+CD154^+$ populations for isolation of conventional and iNKT/MAIT cells, respectively. In addition, we sorted nonactivated cells for comparison (conventional: $CD8^+CD69^-CD137^-CD154^-$; iNKT/MAIT: $CD8^+CD161^+CD69^-CD137^-CD154^-$).

We then assigned each cell a clone size based on the number of cells with an identical alpha and beta TCR sequence. For nonactivated conventional cells, we did not find any clonally expanded sequences, nor did any TCR sequence match those of the activated population, indicating that our clonally expanded and activated cells are indeed pathogen specific. We found some expanded nonactivated iNKT/MAIT TCR sequences, and one-third of nonactivated sequences overlapped with activated sequences (Dataset S1). This is in line with a reported low level of TCR sequence diversity in iNKT/MAIT cells (21, 30, 33).

Together, we sequenced 3,321 TCRs from five different donors from both conventional and iNKT/MAIT $CD8^+$ T cells. Notably, only one in three activated T cells were clonally expanded. We detected 399 clonal (clone size >1) activated cells, with a wide range of clone sizes (Fig. 1*C* and see *Materials and Methods*). Many of these were doublets or very large clones (n = 51) so we deliberately picked out the most informative sequences. We selected over 1,000 cells for sequencing, 921 of which passed our quality-control thresholds and were used for downstream analysis (*SI Appendix*, Figs. S1*C* and S3 *A–C* and Dataset S2). Among them, the majority of the iNKT cells (152 of 169) are from donor 09/0769, while most of the MAIT cells are from four of five donors, and conventionally activated $CD8^+$ T cells are from three of the five donors. In two donors (01/0468 and 09/0769), we found almost all iNKT/MAIT $CD8^+$ T cells, with too few activated conventional $CD8^+$ T cells to analyze.

Cell Type Classification and Gene Expression Signatures. Each cell was assigned a cell type (conventional/iNKT/MAIT) based on its TCR sequence and activation marker, as well as its activation state (activated vs. nonactivated) (Fig. 1*B*). Applying t-distributed stochastic neighbor embedding (tSNE) to our transcriptome data, we detected distinct clusters that delineated conventional vs. unconventional and activated vs. nonactivated cell states (Fig. 1*D*).

Notably however, between $CD8^+$ iNKT and MAIT cells, tSNE analysis found no observable differences in the overall gene expression profiles (Fig. 1*E*). In fact, we were only able to detect seven significantly differentially expressed genes between cells of these types ($P < 0.05$ and fold change >2, Dataset S3). Thus, we concluded that at least within $CD8^+$ T cell lineages (a minority of iNKT cells), the iNKT/MAIT cells defined by TCR sequences, respond very similarly to *Mtb* lysate at a transcriptional level.

Additionally, we characterized the genes that were specific to each major cell type, which allowed for the first comprehensive

comparison of genes specific to conventional and iNKT/MAIT cells in humans (Fig. 1 *F* and *G* and *SI Appendix*, Fig. S4). Both conventional and iNKT/MAIT subtypes showed upregulation of cytotoxic genes GZMB, GZMH, and GNLY upon activation, consistent with their ability to kill *Mtb*-infected cells or *Mtb* directly (34, 35). While iNKT/MAIT cells showed more robust and ubiquitous IFN-gamma and TNF-alpha production, our data support only limited and sparse IFN-gamma gene expression from activated conventional cells. In terms of chemokine activity, conventional cells specifically expressed CCL3, CCL4, and CCL5, while iNKT/MAIT cells expressed mainly CCL20 (*SI Appendix*, Fig. S4). Elevated levels of CCL3, CCL4, and CCL5 in $CD8^+$ T cells had been reported in viral infections (36) and also in liver-related infections (37). However, an elevated level of these cytokines in lung-related infections, where *Mtb* mainly resides, had not been characterized. Furthermore, we observe a set of NK cell receptors in conventional $CD8^+$ T cells after activation, such as KLRK1, KIR3, ITM2A, KLRC, and HLA class II genes HLA-DRB1 and HLA-DRA (*SI Appendix*, Fig. S4). Furthermore, SLAMF1, SLC4A10, and DPP4 were expressed uniquely in iNKT/MAIT cells (Fig. 1 *F* and *G*). This was consistent with recent reports that MAIT cells preferentially express greater levels of SLAMF1 and DPP4 (38, 39). Interestingly, DPP4 is involved in the costimulatory signal essential for T cell receptor (TCR)-mediated T cell activation (Fig. 1 *F* and *G*).

In addition to well-defined clusters based on cell type and states, we find subclustering by donor (Fig. 1 *D* and *F*). This was driven by a set of genes differentially expressed between donors, including IFIT3, IFI44L, IFI6, and ISG15, which are part of the type I IFN signaling pathway. Although we could not associate a significance to these expression differences because

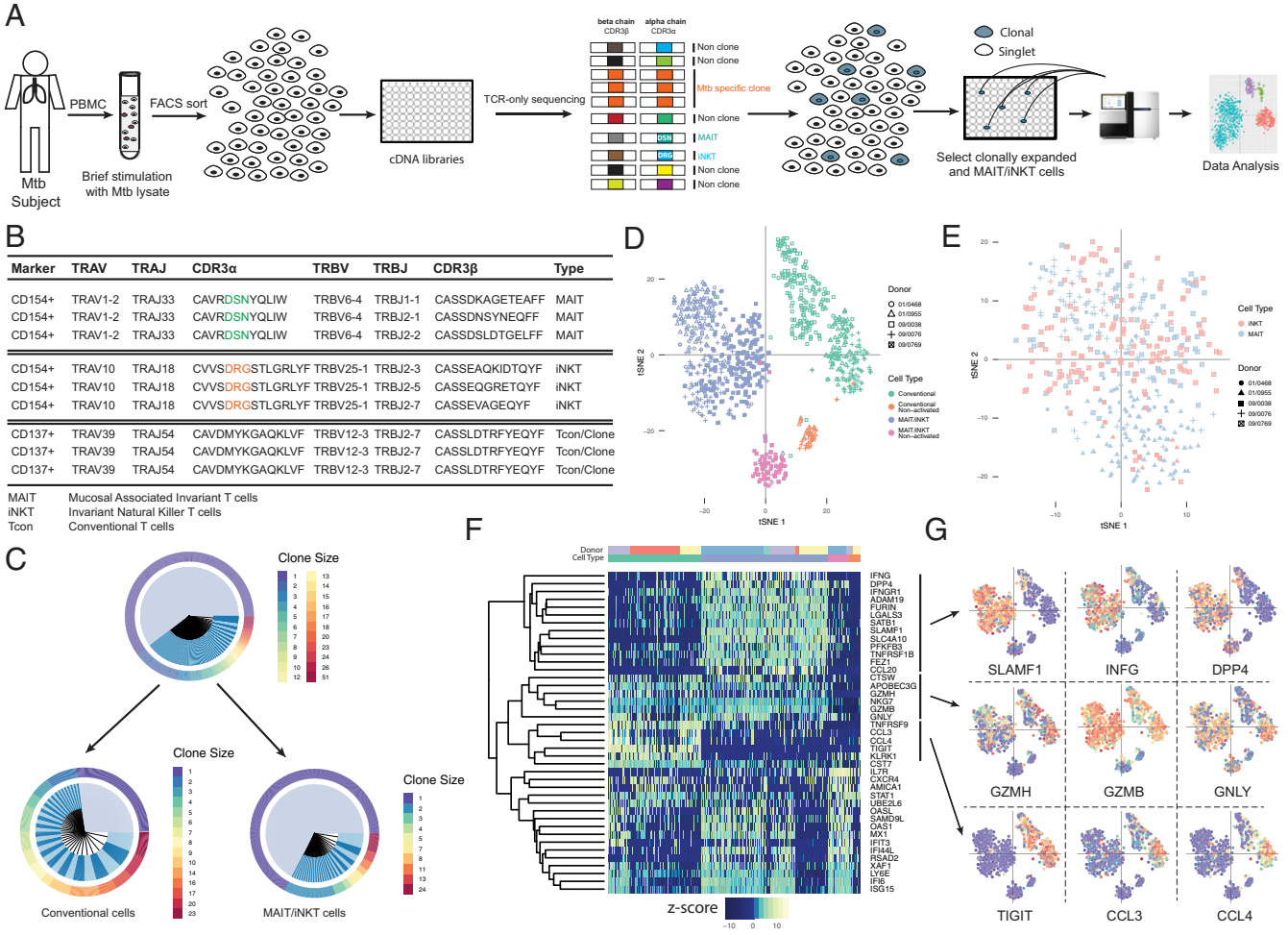

Fig. 1. Method for preferential selection of clonally expanded *Mtb*-reactive $CD8^+$ T cells via TCR sequencing, for subsequent single-cell RNA sequencing. (*A*) Schematic view of the experimental workflow for determining TCR clonotypes and transcriptomes from single *Mtb*-reactive $CD8^+$ T cells. (*B*) Table showing representative CDR3α and CDR3β sequences in different cell types derived from single T cells sorted based on CD137 and CD154 staining. Green text highlights "DSN" motif for MAIT cells, and red text highlights "DRG" motif for iNKT cells. (*C*) Circular phylogenetic trees representing the distribution of clone sizes for all activated TCR-sequenced cells, and the conventional and iNKT/MAIT cells selected for scRNA-seq. Each "leaf" (tip) represents one cell, and is grouped by identical TCR sequences (beta only for conventional, alpha and beta for iNKT/MAIT). The tree root is situated in the center of the circle, with branches for each clone, and colors distinguishing separate clones. The gray, rooted branches represent singlets. The inner ring color (blue and light blue) distinguishes neighboring clones but have no other significance. The outer ring color (rainbow) represents the size of the clone. (*D*) t-SNE analysis of all single cells (921 cells). Each point represents an individual cell. Colors indicate cell type and activation status; symbols indicate donor origin. (*E*) t-SNE analysis of all activated iNKT/MAIT cells (n = 169 and 276 for iNKT and MAIT, respectively). Each point represents an individual cell. Color indicates cell type. Symbol indicates donor origin. (*F*) Heatmap of the top 50 genes ranked by their contribution to PCA components 1–5 from all of our sequenced cells. The columns are individual cells ordered first by cell type (coloring as in *D*) then by donor, as shown in the color bars. The rows are genes clustered by expression similarity using hierarchical clustering. The z score is calculated by subtracting the gene's expression in each cell by the mean expression across all cells. (*G*) t-SNE plot (from *D*) colored by the expression of the corresponding gene.

we did not have the requisite patient phenotype information, we noted that it is consistent across multiple donors, including within the nonactivated cells. Therefore, these genes might reflect expected differences between environment and genetic backgrounds of donors, and our ability to detect them confirmed the sensitivity of our SELECT-seq approach.

Dynamic Gene Expression Profiles of Clonally Expanded T Cells. To describe the dynamics of expression profiles based on clonal information, we performed principal component analysis (PCA) to determine if clustering was informed by clone size. Two observations stand out from this analysis: first, there is a striking dependence of conventional, but not iNKT/MAIT cell, gene expression profiles on clone size; second, the clone size distribution is the dominant layer of transcriptomic heterogeneity, i.e., it is associated with the first principal component (Fig. 2*A*). Thus, the transcriptional variation between large and small clones in conventional cells is greater than the difference between conventional and unconventional (iNKT/MAIT) cells. Indeed, we find that these two principal component dimensions contribute a much greater share of expression variance than the subsequent dimensions, demonstrating the importance of these features to their cell-state definitions (Fig. 2*B*). Together, this provides additional validation of our approach to specifically select clonally expanded cells because these cells are clearly unique relative to the singlet population.

Although we do not observe an association between clone size and expression profiles in iNKT/MAIT cells, we explored these data further. We checked the first four principal components from PCA of activated iNKT/MAIT cells and found no association between clone size and expression profiles, but rather a primarily donor-specific contribution as mentioned above (Fig. 2 *D* and *E* and *SI Appendix*, Fig. S5*B*). In contrast, the first principal component of conventional cells revealed a clear correlation to clone size (Fig. 2*C*). We further analyzed this trend within each donor, and indeed found that conventional cells but not iNKT/MAIT cells cluster by clone size in aggregate, as well as within each donor (*SI Appendix*, Fig. S5 *A* and *C*).

To compare our results in additional contexts, we included similar analyses utilizing published datasets with single-cell RNA-seq (scRNA-seq) of clonally expanded $CD8^+$ T cells. Surprisingly, none of these contexts yielded similar associations, or any discernible pattern, between clone size and gene expression profiles (*SI Appendix*, Fig. S6). These assays included T cells from blood, tumor, and adjacent tissue in cancer patients (40), as well as dextramer-isolated T cells from donors after receiving a yellow fever vaccine (41). Furthermore, a recently published study from patients hospitalized with severe influenza A virus also reported no difference in the transcriptomes of clonally expanded and nonclonally expanded T cells (42). A notable difference with our experimental design is our brief lysate activation. This may suggest that clonal-dependent gene expression profiles are only apparent immediately following activation, which in vivo studies may not capture. In Reinius et al. (41), the authors did sequence some T cells stimulated ex vivo; however, they allowed the cells to proliferate for 18 d, which increases the potential for artifacts. In addition, they included few clones, all of which were large (~30–50), and thus the lack of an association may have been due to their limited clone size range (*SI Appendix*, Fig. S6).

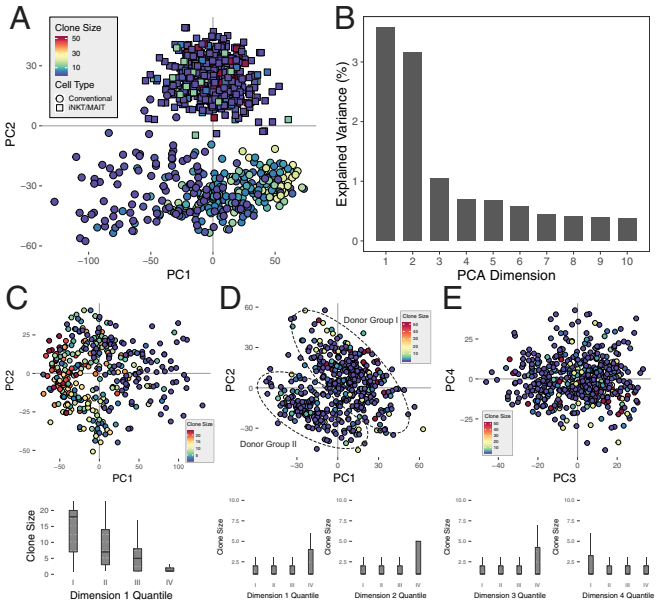

Fig. 2. Dynamic gene expression patterns among clonotypes and cell types. (*A*) PCA from all *Mtb*-reactive $CD8^+$ T cells using highly variable genes. The size of each clonotype is represented by color; each dot represents one cell. Symbol represents cell type. (*B*) PCA, as in *A*, with the explained percent variances plotted for the top 10 PCA dimensions. (*C*) PCA of all *Mtb*-reactive $CD8^+$ conventional cells using highly variable genes. The size of each clonotype is represented by color. (*Lower*) Cells are ranked according to their PC1 coordinate value and split into four equal-sized groups; a boxplot displays the distribution of clone sizes for these groups. Each boxplot represents the median (center line), the first quartiles and the third quartiles (box limits), and 1.5 times the interquartile range (whiskers). (*D* and *E*) As in *C* but for iNKT/MAIT cells, and includes PCA dimensions 1–4. Dotted-lined circles in *D* represent manually observed donor-dependent grouping (*SI Appendix*, Fig. S5*B*).

Underlying Expression Patterns of Activated Conventional Cells. Given the importance of clone size toward informing expression profiles in conventional cells only, we next focused on these cells by performing PCA with genes up-regulated upon T cell activation. We ranked all genes by their contribution to the first principal component, since PC1 reflects clone size (Fig. 3*A*). The highest ranked gene is IL2RA (CD25), which is induced by stimulation via the T cell receptor (43), and is critical in mediating T cell growth and proliferation (Fig. 3*B* and *SI Appendix*, Fig. S7 *A* and *B*). Two major cytotoxic cytokines LTA and IFNG (44) are also important contributors to PC1 and are generally expressed in small clones (Fig. 3 *B* and *F*).

Most of the genes that contribute the greatest level to PC1 variance are anticorrelated with clone size (Fig. 3*B*). Therefore, we next focused on genes specifically expressed in large clones (Fig. 3*D*). We found that these genes are part of the TCR signaling pathway (Fig. 3*E*), and are relevant to cytotoxic activity (PRF1 and GNLY), inflammatory chemokines (CCL4 and CCL5), in addition to, notably, senescence (TIGIT and KLRG1) (Fig. 3*G*).

We found similar indications in gene sets: chemokines are up-regulated in clonal cells, while cytokine and proliferation genes decrease expression with increasing clone size (*SI Appendix*, Fig. S7 *E*–*G*). Similarly, we looked at coreceptor genes, which colocalize with TCR molecules to inhibit or activate the TCR signaling pathway (45). We observed a decrease in expression of apoptotic coreceptor genes as clone size decreased, but a drop in costimulatory genes concordant with an increase of coinhibitory expression (*SI Appendix*, Fig. S7 *H*–*J*). Together, these data suggest that although large clones express genes important for cytotoxic and inflammatory activity, they also have lower expression levels of cytokines, costimulatory genes and proliferation related genes, suggesting that these large clones have a diminished cellular proliferation capacity and increased features of cellular senescence (46, 47).

Highly Expanded Clones Show Signatures of Senescence. Next, we confirmed that the RNA-seq expression changes of key genes associated with a change in FACS profiling of their concomitant proteins. In addition to individuals with *Mtb* infection, we also assayed subjects with latent cytomegalovirus (CMV) infection using the same activation procedure, but with a CMV-pp65 derived peptide pool rather than a lysate. In two QFN+ (Quantiferon test positive) donors and one CMV+ donor, we observed a bimodal

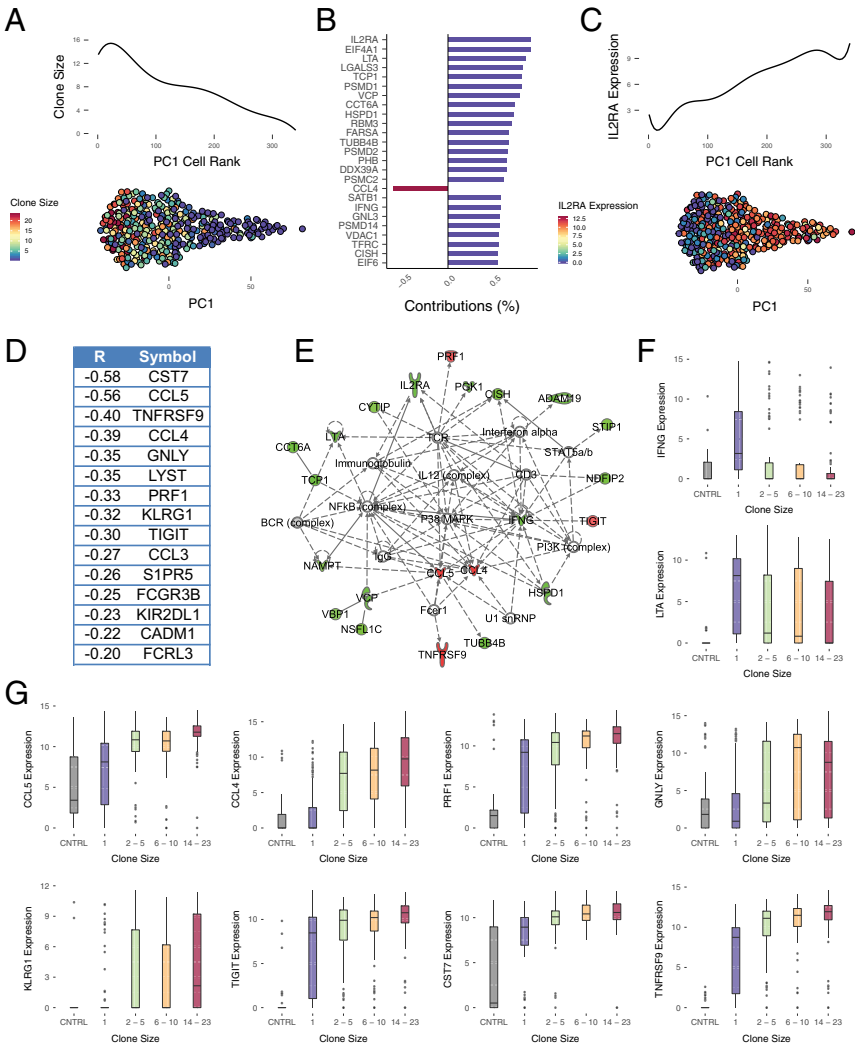

Fig. 3. Individual gene expression patterns associated with clone size in conventional cells. (*A*, *Lower*) Beeswarm plot of the PC1 coordinate values from PCA of activated conventional cells using the highly variable genes that are also up-regulated ($P < 0.1$ and fold change >0) in conventional cells (activated vs. nonactivated). The *y* axis has no significance. Color represents clone size. (*A*, *Upper*) Smoothed polynomial regression line of each cell's clone size ordered by its PC1 coordinate. (*B*) Top 25 genes contributing to principal component 1, ranked by their percent contribution (*x* axis) calculated with "factoextra" (60). Negative values indicate a negative correlation between gene expression and PC1. Red means high expression in large clones, and purple reflects high expression in singlets and small clones. (*C*) As in *A* but for IL2RA expression. (*D*) Top 15 genes negatively correlated (Spearman's rank) with PC1 coordinate values but positively correlated with clone size. (*E*) The highest scoring IPA network generated from genes that are highly correlated with clone size, as in *B* and *D*. (*F* and *G*) Boxplots showing individual gene expression among different clone sizes, and nonactivated conventional cells (CNTRL). Each boxplot represents the median (center line), the first quartiles, and the third quartiles (box limits), and 1.5 times the interquartile range (whiskers). Points are outliers.

distribution of IL2RA (CD25) on antigen-specific $CD8^+$ T cells (Fig. 4*A*). We compared these positive and negative IL2RA (CD25) populations by their expression of TIGIT and KLRG1 (Fig. 4*B*). We observed that the IL2RA (CD25) negative population is enriched in TIGIT and KLRG1 double positive cells while the IL2RA (CD25) positive population shows the opposite. Furthermore, we utilized index sorting to assay the expression of senescent markers, while retaining cell information to compare with TCR sequence data. Indeed, these data confirmed the scRNA-seq findings that IL2RA (CD25), and that costimulatory markers CD27 and CD28 (45), are expressed more frequently in singlets and small clones than in large clones. In contrast, the senescent markers, TIGIT and KLRG1, are expressed more frequently in large clones (Fig. 4*C* and *SI Appendix*, Fig. S8 *A–E*).

To identify functional differences between large and small clones, we used an IFNγ capture assay, which showed decreased IFNγ secretion in the IL2RA (CD25) negative populations after pp65 peptide pool stimulation (Fig. 4*D*). Then, we assayed the proliferation potential of both IL2RA (CD25) positive and negative populations by a CFSE (carboxyfluorescein succinimidyl ester) assay (Fig. 4*E* and *SI Appendix*, Fig. S8*F*), by treating each population with the pp65 peptide pool and varying concentrations of IL-2. We found that the IL2RA (CD25) positive population proliferates robustly with both high and low IL-2 concentrations, while the negative population only proliferates well at high IL-2 concentrations. Together, these assays corroborate our hypothesis that large clones have a diminished capacity for cytokine secretion and proliferation, both features of cellular senescence.

Discussion

In this report, we develop a method, SELECT-seq, to provide high-quality scRNA-seq data of clonally expanded *Mtb*-responsive T cells. This approach enabled us to discover a strong but heterogenous activation response in $CD8^+$ T cells upon *M. tuberculosis* lysate stimulation in infected individuals. In addition, we discovered an expression signature that limits further expansion

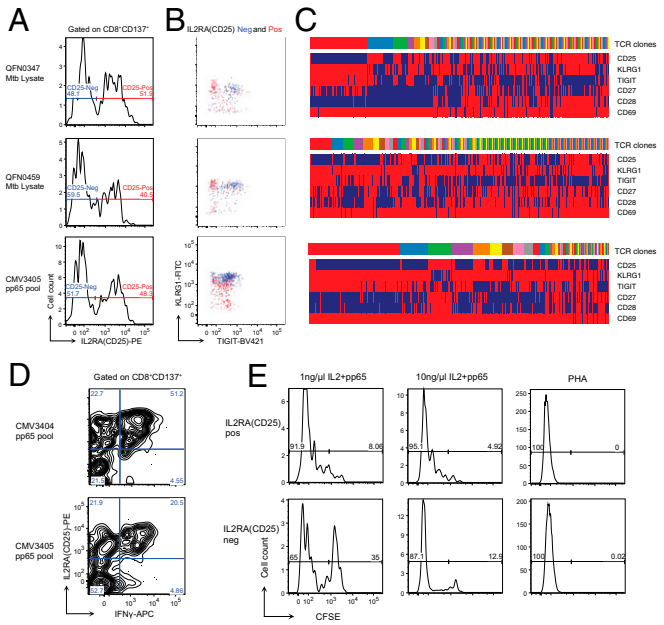

Fig. 4. Functional characterization of TB and CMV-reactive CD8$^+$ conventional T cell clones in relation to IL2RA (CD25). (*A*) Density plots showing the distribution of IL2RA (CD25) expression among *Mtb*-reactive (two donors) and CMV-reactive (one donor) CD8$^+$CD137$^+$ T cells after antigen-specific stimulation. Blue line indicates CD25$^-$ population and the red line indicates CD25$^+$. (*B*) Overlaid flow cytometry profiles of TIGIT and KLRG1 expression on CD25 negative (blue) and CD25 positive (red) cells, gated as in *A*. (*C*) Cells from *A* were single-cell index sorted and analyzed by single-cell TCR sequencing (25) six-parameter (*Right* of figure) phenotypic analysis of *Mtb* and CMV-specific CD8+ T cells. Individual T cells are grouped by TCR sequence; each color on the bar above the heat maps represents a distinct and clonal expanded TCR sequence. Red and blue represent positive and negative expression levels, respectively, of the protein acquired through index sorting. (*D*) Flow cytometry profiles showing combinations of CD25 expression with IFNγ-secreting CD8$^+$CD137$^+$ conventional T cells after stimulation with CMV-specific pp65 peptide pool. (*E*) Flow cytometry profiles showing CFSE dilution. CMV-specific CD25 positive and negative populations from donor CMV3405 were sorted, stained with CFSE, and stimulated for 7 d with pp65 peptide pool in different concentrations of IL-2 (*Left* and *Middle*) along with a positive control activated with phytohemagglutinin (PHA, *Right*).

within CD8$^+$ clonally expanded conventional, but not in unconventional (iNKT/MAIT) T cells.

SELECT-seq is a straightforward approach to circumvent the problem of analyzing large numbers of cells, low sequencing resolution, high costs, and sorting by cell-surface proteins. It utilizes Smart-Seq2 to sequence rare cells and provide high-quality libraries. And, due to the targeted PCR step before expensive whole-transcriptome scRNA-seq, SELECT-seq significantly reduces the expense of untargeted single-cell sequencing. We applied this approach to select clonally expanded CD8$^+$ T cells. This was achieved by identifying cells that had matching TCR sequences. Approximately two in five cells were clonally expanded, meaning the cost of tagmentation and sequencing all singlets would be $57,056, which can be avoided with SELECT-seq. Because it utilizes PCR to screen cDNA libraries, SELECT-seq may target any gene, or combination of genes, relevant to a specific cell population, like B cell receptors of B cells. Further, SELECT-seq is not limited to protein coding genes; it can also be applied to polyadenylated noncoding RNA or any expressed RNA if utilizing a total scRNA method (48, 49).

Applying SELECT-seq to major CD8$^+$ subpopulations, we were able to distinguish MAIT and iNKT cells from conventional CD8$^+$ T cells. Importantly, we found that gene expression profiles of CD8$^+$ iNKT and MAIT cells are largely indistinguishable from each other in response to *Mtb* lysate. It is surprising that we see such a robust iNKT response to the *Mtb* lysate, since there are no known iNKT ligands from *Mtb*, versus MAIT cells which are known to recognize riboflavin derivatives from *Mtb* (23, 50). This indicates that even though iNKT and MAIT cells recognize different ligands, they carry out similar immunological functions, at least in our experimental context. In addition, it has been reported that both iNKT and MAIT cells could be activated by cytokines independent of TCR engagement (23, 50), thus an alternative explanation could be that iNKT and MAIT cells are activated by cytokines IL-12/IL-18/IL-15 in *Mtb* lysate stimulated PBMC culture medium. Thus, future studies comparing cognate antigen and cytokine activation of iNKT and MAIT cells with transcriptome analysis will be valuable. In this study, as we focused on the response of CD8$^+$ T cells to *Mtb*, such responding iNKT and MAIT cells may only specify a particular subset of unconventional T cells, a similar comparison could be applied to iNKT/MAIT cells from additional lineages, such as CD4$^+$ or CD4$^-$CD8$^-$ T cells, to determine if this pattern is consistent.

Another important finding was the significant gene expression changes along the continuum of clonally expanded conventional CD8$^+$ T cells. Within the clonally expanded conventional cells we found an overall decrease in the expression of IL2RA (CD25), costimulatory receptors CD27 and CD28, and the number of genes expressed. Likewise, we found that the coinhibitory receptor TIGIT, senescent marker KLRG1, and cytotoxic and proinflammatory genes were up-regulated. Together, these findings suggested that these highly expanded T cells have become senescent. We confirmed our findings in multiple donors exposed to either a bacterial (*Mtb*) or viral infection (CMV) by showing that the most clonally expanded cells do not secrete IFNγ and are limited in their capacity for further proliferation.

The senescent phenotype observed in our study could be the outcome of low and consistent antigen dose stimulation, or homeostatic T cell proliferation considering the donors' asymptomatic *Mtb* infection. In any case, an interesting consequence of this finding is that it would place an upper limit on the extent to which a given T cell could expand in response to antigens. This would allow "room" for other T cell clones to expand, diversifying the response considerably. This can be seen in the recent report by Stern and coworkers (51), who sequenced thousands of unique TCRs specific for one influenza peptide MHC and found an average of 516 unique TCRα clonotypes and 432 unique TCRβ clonotypes in each individual. In another report (52), the number of unique TCRs specific for NLV–HLA-A2 and GIL–HLA-A2 ranged from 21 to 784 for TCRα and from 13 to 1,030 for TCRβ per person. Similarly, in our studies (53) we found there are multiple TCR sequence solutions to particular viral or bacterial antigens in a given subject, even though they recognized the same ligand. An intrinsic limit to T cell clonal expansion could explain these TCR sequence results, generating a more diverse T cell response. This enforced diversity would limit the effectiveness of pathogen evolution to evade T cell recognition.

Given that we studied CD8$^+$ T cells specific for *Mtb* antigens, the senescence signature we found is significant. While tuberculosis is highly prevalent and the immune systems response to *Mtb* infection is well studied (15), the role of CD8$^+$ T cells in this disease is controversial. Their role is especially uncertain in human *Mtb* infection as there is no definitive evidence for a requirement of CD8$^+$ T cells in mitigating this infection (54–56). Rather, CD4$^+$ T cells have been implicated as the major functional responder to *Mtb* infection. However, we show that in asymptomatic *Mtb*-infected teenagers, not only is there extensive clonal expansion, but CD8$^+$ T cells exhibit a senescent phenotype when reactivated, which suggests they are continually activated in vivo. Furthermore, these highly expanded clones show increased perforin and granulysin expression, which has been reported in terminally differentiated CD8$^+$ T cells that are effective in killing *Mtb* (57). Unique to these expanded *Mtb*-specific CD8$^+$ T cells, however, is that they express elevated levels of chemokines CCL4 and CCL5, ligands for CCR5, which attract white blood cells to sites of infection. Thus, while there are some unique characteristics of these senescent *Mtb*-specific CD8$^+$ T cells, they are highly similar to CD8$^+$ T cells in chronic diseases like HIV and CMV (58, 59), which are widely accepted as functional responders. Thus, the differences in CD8$^+$ T cells responding to *Mtb* infection are more subtle than previously thought.

In conclusion, we developed SELECT-seq to permit economical, in-depth analysis of any specific cell type of interest. Without the limitation of cell-surface protein sorting, it provides significantly increased flexibility not associated with any current scRNA-seq method.

Materials and Methods

Detailed information on study cohort, antigen preparation, cell stimulation, flow cytometry, single-cell RNA-seq, TCR sequencing, and data analysis is provided in *SI Appendix*.

ACKNOWLEDGMENTS. We thank Shin-Heng Chiou for study conception discussions; Michelle Nguyen for help preparing single-cell samples; Lars Velten and Daniel Schraivogel for their comments on the manuscript; Hassan Mahomed, Willem Hanekom, and members of the Adolescent Cohort Study group for enrollment and follow-up of the *Mtb*-infected adolescents; the Stanford Human Immune Monitoring Center for high-throughput sequencing support; and the Bill and Melinda Gates Foundation, the National Institutes of Health (Grant 2U19 AI057229), and the Howard Hughes Medical Institute for financial support. Sorting was partially performed in the Stanford FACS Facility.

1. Chaplin DD (2010) Overview of the immune response. *J Allergy Clin Immunol* 125 (Suppl 2):S3–S23.
2. Finlay BB, McFadden G (2006) Anti-immunology: Evasion of the host immune system by bacterial and viral pathogens. *Cell* 124:767–782.
3. Claman HN, Chaperon EA, Triplett RF (1966) Thymus-marrow cell combinations. Synergism in antibody production. *Proc Soc Exp Biol Med* 122:1167–1171.
4. Crotty S (2015) A brief history of T cell help to B cells. *Nat Rev Immunol* 15:185–189.
5. Davis MM, Bjorkman PJ (1988) T-cell antigen receptor genes and T-cell recognition. *Nature* 334:395–402.
6. Farber DL, Yudanin NA, Restifo NP (2014) Human memory T cells: Generation, compartmentalization and homeostasis. *Nat Rev Immunol* 14:24–35.
7. Kaech SM, Cui W (2012) Transcriptional control of effector and memory CD8+ T cell differentiation. *Nat Rev Immunol* 12:749–761.
8. Qi Q, et al. (2014) Diversity and clonal selection in the human T-cell repertoire. *Proc Natl Acad Sci USA* 111:13139–13144.
9. Ch'en IL, et al. (2008) Antigen-mediated T cell expansion regulated by parallel pathways of death. *Proc Natl Acad Sci USA* 105:17463–17468.
10. Bacher P, Scheffold A (2013) Flow-cytometric analysis of rare antigen-specific T cells. *Cytometry A* 83:692–701.
11. Macosko EZ, et al. (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161:1202–1214.
12. Zheng GXY, et al. (2017) Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 8:14049.
13. Svensson V, et al. (2017) Power analysis of single-cell RNA-sequencing experiments. *Nat Methods* 14:381–387.
14. Dye C, Williams BG (2010) The population dynamics and control of tuberculosis. *Science* 328:856–861.
15. O'Garra A, et al. (2013) The immune response in tuberculosis. *Annu Rev Immunol* 31: 475–527.
16. Walzl G, Ronacher K, Hanekom W, Scriba TJ, Zumla A (2011) Immunological biomarkers of tuberculosis. *Nat Rev Immunol* 11:343–354.
17. Blythe MJ, et al. (2007) An analysis of the epitope knowledge related to Mycobacteria. *Immunome Res* 3:10.
18. Lindestam Arlehamn CS, et al. (2016) A quantitative analysis of complexity of human pathogen-specific CD4 T cell responses in healthy M. tuberculosis infected South Africans. *PLoS Pathog* 12:e1005760.
19. Lewinsohn DM, et al. (2013) Human Mycobacterium tuberculosis CD8 T cell antigens/epitopes identified by a proteomic peptide library. *PLoS One* 8:e67016.
20. Bendelac A, Savage PB, Teyton L (2007) The biology of NKT cells. *Annu Rev Immunol* 25:297–336.
21. Godfrey DI, Uldrich AP, McCluskey J, Rossjohn J, Moody DB (2015) The burgeoning family of unconventional T cells. *Nat Immunol* 16:1114–1123.
22. Kronenberg M (2005) Toward an understanding of NKT cell biology: Progress and paradoxes. *Annu Rev Immunol* 23:877–900.
23. Salou M, Franciszkiewicz K, Lantz O (2017) MAIT cells in infectious diseases. *Curr Opin Immunol* 48:7–14.
24. Picelli S, et al. (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* 10:1096–1098.
25. Han A, Glanville J, Hansmann L, Davis MM (2014) Linking T-cell receptor sequence to functional phenotype at the single-cell level. *Nat Biotechnol* 32:684–692.
26. Moyo S, et al. (2011) Tuberculin skin test and QuantiFERON® assay in young children investigated for tuberculosis in South Africa. *Int J Tuberc Lung Dis* 15:1176–1181, i.
27. Chattopadhyay PK, Yu J, Roederer M (2005) A live-cell assay to detect antigen-specific $CD4^+$ T cells with diverse cytokine profiles. *Nat Med* 11:1113–1117.
28. Frentsch M, et al. (2005) Direct access to $CD4^+$ T cells specific for defined antigens according to CD154 expression. *Nat Med* 11:1118–1124.
29. Wölfl M, Kuball J, Eyrich M, Schlegel PG, Greenberg PD (2008) Use of CD137 to study the full repertoire of CD8+ T cells without the need to know epitope specificities. *Cytometry A* 73:1043–1049.
30. Godfrey DI, MacDonald HR, Kronenberg M, Smyth MJ, Van Kaer L (2004) NKT cells: What's in a name? *Nat Rev Immunol* 4:231–237.
31. Salio M, et al. (2017) Activation of human mucosal-associated invariant T cells induces CD40L-dependent maturation of monocyte-derived and primary dendritic cells. *J Immunol* 199:2631–2638.
32. Tomura M, et al. (1999) A novel function of Valpha14+CD4+NKT cells: Stimulation of IL-12 production by antigen-presenting cells in the innate immune system. *J Immunol* 163:93–101.
33. Rossjohn J, Pellicci DG, Patel O, Gapin L, Godfrey DI (2012) Recognition of CD1d-restricted antigens by natural killer T cells. *Nat Rev Immunol* 12:845–857.
34. Semple PL, et al. (2011) Induction of granulysin and perforin cytolytic mediator expression in 10-week-old infants vaccinated with BCG at birth. *Clin Dev Immunol* 2011: 438463.
35. Stenger S, et al. (1998) An antimicrobial activity of cytolytic T cells mediated by granulysin. *Science* 282:121–125.
36. Cocchi F, et al. (1995) Identification of RANTES, MIP-1 α, and MIP-1 β as the major HIV-suppressive factors produced by CD8+ T cells. *Science* 270:1811–1815.
37. Sadeghi M, et al. (2015) Serum levels of chemokines CCL4 and CCL5 in cirrhotic patients indicate the presence of hepatocellular carcinoma. *Br J Cancer* 113:756–762.
38. Sharma PK, et al. (2015) High expression of CD26 accurately identifies human bacteria-reactive MR1-restricted MAIT cells. *Immunology* 145:443–453.
39. Dusseaux M, et al. (2011) Human MAIT cells are xenobiotic-resistant, tissue-targeted, CD161hi IL-17-secreting T cells. *Blood* 117:1250–1259.
40. Zheng C, et al. (2017) Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell* 169:1342–1356.e16.
41. Reinius B, et al. (2016) Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. *Nat Genet* 48:1430–1435.
42. Wang Z, et al. (2018) Clonally diverse $CD38^+HLA\text{-}DR^+CD8^+$ T cells persist during fatal H7N9 disease. *Nat Commun* 9:824.
43. Leonard WJ, Krönke M, Peffer NJ, Depper JM, Greene WC (1985) Interleukin 2 receptor gene expression in normal human T lymphocytes. *Proc Natl Acad Sci USA* 82: 6281–6285.
44. Browning JL, et al. (1996) Signaling through the lymphotoxin beta receptor induces the death of some adenocarcinoma tumor lines. *J Exp Med* 183:867–878.
45. Chen L, Flies DB (2013) Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nat Rev Immunol* 13:227–242.
46. Akbar AN, Henson SM, Lanna A (2016) Senescence of T lymphocytes: Implications for enhancing human immunity. *Trends Immunol* 37:866–876.
47. Effros RB, Dagarag M, Spaulding C, Man J (2005) The role of CD8+ T-cell replicative senescence in human aging. *Immunol Rev* 205:147–157.
48. Hayashi T, et al. (2018) Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. *Nat Commun* 9:619.
49. Sheng K, Cao W, Niu Y, Deng Q, Zong C (2017) Effective detection of variation in single-cell transcriptomes using MATQ-seq. *Nat Methods* 14:267–270.
50. Crosby CM, Kronenberg M (2018) Tissue-specific functions of invariant natural killer T cells. *Nat Rev Immunol* 18:559–574.
51. Song I, Gil A, Mishra R, Selin LK, Stern LJ (2017) Broad TCR repertoire and diverse structural solutions for recognition of an immunodominant CD8+ T cell epitope. *Nat Struct Mol Biol* 24:395–406.
52. Chen G, et al. (2017) Sequence and structural analyses reveal distinct and highly diverse human $CD8^+$ TCR repertoires to immunodominant viral antigens. *Cell Rep* 19: 569–583.
53. Glanville J, et al. (2017) Identifying specificity groups in the T cell receptor repertoire. *Nature* 547:94–98.
54. Lin PL, Flynn JL (2015) CD8 T cells and Mycobacterium tuberculosis infection. *Semin Immunopathol* 37:239–249.
55. Prezzemolo T, et al. (2014) Functional signatures of human CD4 and CD8 T cell responses to Mycobacterium tuberculosis. *Front Immunol* 5:180.
56. Jasenosky LD, Scriba TJ, Hanekom WA, Goldfeld AE (2015) T cells and adaptive immunity to Mycobacterium tuberculosis in humans. *Immunol Rev* 264:74–87.
57. Bruns H, et al. (2009) Anti-TNF immunotherapy reduces CD8+ T cell-mediated antimicrobial activity against Mycobacterium tuberculosis in humans. *J Clin Invest* 119: 1167–1177.
58. Xu W, Larbi A (2017) Markers of T cell senescence in humans. *Int J Mol Sci* 18:E1742.
59. McElhaney JE, Effros RB (2009) Immunosenescence: What does it mean to health outcomes in older adults? *Curr Opin Immunol* 21:418–424.
60. Le S, Julie J, Husson F (2008) FactoMineR: An R package for multivariate analysis. *J Stat Softw* 25:1–18.