



High-throughput and high-dimensional single-cell analysis of antigen-specific CD8⁺ T cells

Ke-Yue Ma^{1,11}, Alexandra A. Schonnesen^{2,11}, Chenfeng He², Amanda Y. Xia³, Eric Sun², Eunise Chen⁴, Katherine R. Sebastian⁵, Yu-Wan Guo^{6,7}, Robert Balderas⁸, Mrinalini Kulkarni-Date⁵ and Ning Jiang^{1,2,7,9,10} ✉

Although critical to T cell function, antigen specificity is often omitted in high-throughput multiomics-based T cell profiling due to technical challenges. We describe a high-dimensional, tetramer-associated T cell antigen receptor (TCR) sequencing (TetTCR-SeqHD) method to simultaneously profile cognate antigen specificities, TCR sequences, targeted gene expression and surface-protein expression from tens of thousands of single cells. Using human polyclonal CD8⁺ T cells with known antigen specificity and TCR sequences, we demonstrate over 98% precision for detecting the correct antigen specificity. We also evaluate gene expression and phenotypic differences among antigen-specific CD8⁺ T cells and characterize phenotype signatures of influenza- and Epstein-Barr virus-specific CD8⁺ T cells that are unique to their pathogen targets. Moreover, with the high-throughput capacity of profiling hundreds of antigens simultaneously, we apply TetTCR-SeqHD to identify antigens that preferentially enrich cognate CD8⁺ T cells in patients with type 1 diabetes compared to healthy controls and discover a TCR that cross-reacts with diabetes-related and microbiome antigens. TetTCR-SeqHD is a powerful approach for profiling T cell responses in humans and mice.

Due to their multifaceted role in controlling infection, fighting cancer and responding to vaccines, T cells have been subjected to extensive analysis^{1,2}. Recently developed multiomics single-cell profiling methods have enabled multidimensional analysis in single T cells, such as combining assay for transposase-accessible chromatin using sequencing (ATAC-seq) with single-cell RNA sequencing (scRNA-seq)³, DNA-labeled antibody-based phenotyping with scRNA-seq (CITE-seq⁴ and REAP-seq⁵) and DNA-labeled antibody-based phenotyping with targeted single-cell gene expression⁶. Such approaches have greatly advanced our understanding of T cell immune responses in multiple disease settings^{7–9}.

T cell antigen specificity, although critical to T cell function and T cell-based immunotherapy development, has been challenging to analyze in a high-throughput manner until recently. Using T cell trogocytosis¹⁰ or reporter genes^{11–14}, a suite of technologies have been developed in this area enabling high-throughput screening for T cell antigens, such as SABR¹¹, MCR-TCRs¹², T-scan¹³ and granzyme B-based target cell tag¹⁴. These methods have provided much-needed T cell epitope information in the context of cancer^{11–14} and severe acute respiratory syndrome coronavirus 2¹⁵. However, because these methods use expanded T cells or TCR-transduced cell lines, they do not support the profiling of phenotype or gene expression in primary T cells and are thus unable to provide the endogenous activation and functional status of antigen-specific T cells that are important to disease diagnosis and treatment. Peptide-major histocompatibility complex (pMHC) tetramer-based methods can

be applied to primary T cells. When combining these methods with mass cytometry, it has been shown that over 100 antigens can be screened in parallel along with phenotype^{16,17}. Yet, the destructive nature of mass cytometry prevents the acquisition of TCR sequences, which is critical for T cell antigen validation.

We previously developed TetTCR-Seq¹⁸ to link the TCR sequence information to its cognate antigens by sequencing DNA-barcoded pMHC tetramers bound on individual T cells¹⁸. TetTCR-Seq took advantage of ultraviolet light (UV)-mediated peptide-exchange approach¹⁹. Combined with *in vitro* transcription and translation (IVTT) for rapid generation of peptides and pMHCs, TetTCR-Seq enables the screening of hundreds of antigens on primary T cells. To better understand functional profiles of antigen-specific CD8⁺ T cells, a method to simultaneously profile two other ‘dimensions’ of parameters, gene expression and surface-protein expression, is imperative. Such applications can help us to thoroughly understand the heterogeneity among different antigen-specific T cells in the settings of infection, cancer or autoimmune diseases, as well as identify possible biomarkers for disease diagnosis and prognosis. For example, type 1 diabetes (T1D) is a T cell-mediated autoimmune disease, where pancreatic insulin-secreting β cells are selectively destroyed. Autoreactive CD8⁺ T cells has a critical role in this process, and many autoantigens have been identified²⁰. However, due to technology limitations to simultaneously profile a large library of autoantigen-specific CD8⁺ T cells, there is still a lack of understanding of the functional and repertoire characteristics of these autoantigen-specific CD8⁺ T cells.

¹Interdisciplinary Life Sciences Graduate Programs, The University of Texas at Austin, Austin, TX, USA. ²Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX, USA. ³Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX, USA. ⁴Department of Nutritional Sciences, The University of Texas at Austin, Austin, TX, USA. ⁵Department of Internal Medicine, Dell Medical School, The University of Texas at Austin, Austin, TX, USA. ⁶McKetta Department of Chemical Engineering, The University of Texas at Austin, Austin, TX, USA. ⁷Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA. ⁸Becton Dickinson, San Jose, CA, USA. ⁹Department of Oncology, Dell Medical School, The University of Texas at Austin, Austin, TX, USA. ¹⁰Institute for Immunology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. ¹¹These authors contributed equally: Ke-Yue Ma, Alexandra A. Schonnesen. ✉e-mail: jinjiang@upenn.edu

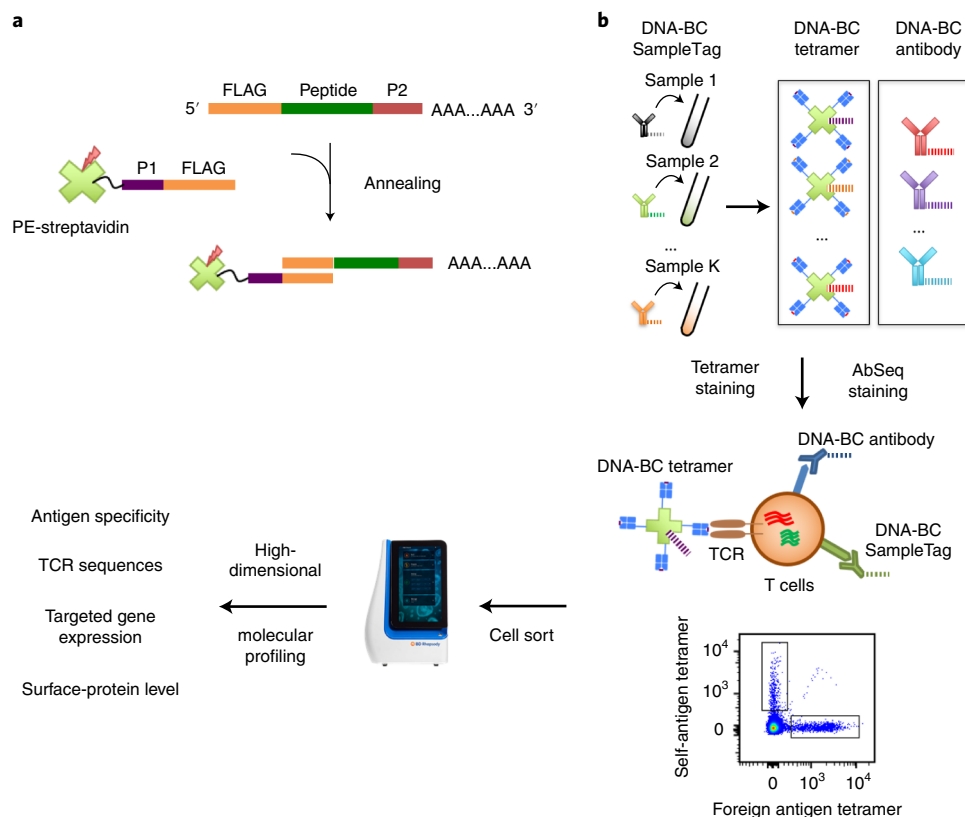


Fig. 1 | Schematics of the TetTCR-SeqHD workflow. **a**, DNA barcode for pMHC tetramer was synthesized with a 3' poly(A) tail. Fluorophore-labeled streptavidin conjugated with an oligonucleotide sequence complementary to the 5' end of tetramer DNA barcode was then used to anneal to each unique tetramer DNA barcode to generate barcoded streptavidin. Barcoded streptavidin was further used to form tetramers with peptide-loaded MHCs. **b**, Each human T cell sample was stained with a unique DNA-barcoded anti-CD50 antibody as a SampleTag, a panel of DNA-barcoded pMHC tetramers and a panel of 59 DNA-barcoded antibodies. Stained cells were sorted and then loaded on the BD Rhapsody single-cell analysis platform for high-throughput and high-dimensional molecular profiling, including of cognate antigen specificity, TCR sequences, targeted gene expression and surface protein level. BC, barcode.

In this study, we describe a high-dimensional TetTCR-Seq (TetTCR-SeqHD) method that enables us to simultaneously profile paired TCR sequences, cognate antigen specificities, targeted gene expression and selected surface-protein expression in tens of thousands of single cells from multiple biological samples. Using a panel of T1D-related antigens and foreign antigens, we show that human foreign pathogen-specific T cells exhibit infection-dependent states. We identify three peptides that have elevated antigen-specific CD8⁺ T cell frequencies in patients with T1D compared with healthy controls and identify a TCR that cross-reacts with a T1D-related peptide and a peptide derived from microbiome. TetTCR-SeqHD, together with the flexibility and speed of generating high-throughput antigen libraries through IVTT, creates a powerful technology to characterize the function and phenotype and track clonal lineage of antigen-specific T cells at the single-cell level in one assay.

Results

Accurate antigen-specificity detection by TetTCR-SeqHD. In TetTCR-SeqHD technology, each peptide encoding DNA oligonucleotide was individually in vitro transcribed/translated to generate a corresponding peptide, which was later loaded onto MHC molecules. Then, the pMHC tetramer was tagged with its corresponding peptide oligonucleotide bearing a 3' poly(A) overhang, which serves as the DNA barcode for that antigen specificity (Fig. 1a). This enables the tetramer barcodes to be captured by BD Rhapsody beads and reverse transcribed together with other mRNAs captured, including TCR transcripts (Fig. 1b). Concurrently, 59 DNA-labeled

antibodies²¹ were used to stain cells. Similar to the tetramer, the DNA barcodes ligated to the antibodies were captured by the same beads. Thus, TetTCR-SeqHD integrates TCR sequencing with TCR antigen specificity, gene expression and phenotyping in tens of thousands of single cells for hundreds of antigens simultaneously.

We first assessed the precision of TetTCR-SeqHD to detect correct antigen specificities using polyclonal CD8⁺ T cells sorted and stimulated with seven known antigens, including potentially cross-reactive epitopes (Supplementary Table 1). Phycoerythrin (PE)-labeled, DNA-barcoded tetramers were used to stain cultured T cells. Tetramer-positive CD8⁺ T cells were sorted (Extended Data Fig. 1a) and loaded to BD Rhapsody to perform reverse transcription and PCRs. A total of 4,533 single cells were recovered after sequencing (Supplementary Table 2). Further filtering of low-quality cells and putative multiplets led to 4,462 cells retained, among which, a median of 140 genes were detected, and TCR α and TCR β capture efficiencies were 89% and 94%, respectively. For each of these six polyclonal CD8⁺ T cell cultures, our previously developed molecular identifier (MID) clustering-based immune repertoire sequencing (MIDCIRS)²² was used to assess TCR β sequence diversity and distribution. These TCR β sequences were then set as internal references for identifying true antigen specificities (Supplementary Table 3). Although the tetramer-negative cells had a lower level of target gene expression, a similar level of gene expression was observed among different antigen-specific T cell clones (Extended Data Fig. 1b,c). An average of 17,249 reads per cell were sequenced for tetramer DNA barcodes.

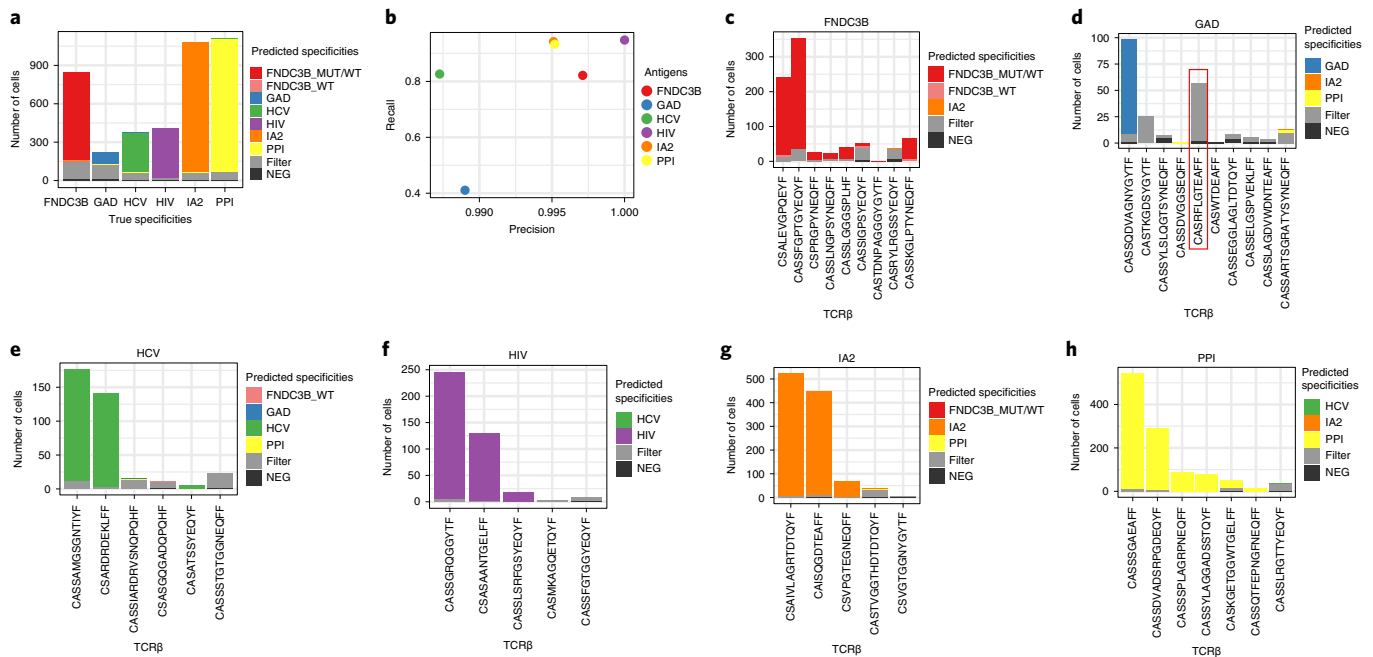


Fig. 2 | TetTCR-SeqHD validation in cultured human polyclonal CD8⁺ T cells. Seven pMHCs were used to sort six groups of polyclonal T cells that were then expanded *in vitro*. These included a group of cross-reactive T cells sorted by two similar antigens, FNDC3B_WT and FNDC3B_MUT. TCRβ sequencing was performed for each polyclonal T cell culture using MIDCIRS. These TCRs and their associated antigen specificity were used to assess the recall and precision rates (Methods) for TetTCR-SeqHD. **a**, Fraction of antigen specificities identified in different categories for each of the six polyclonal T cell cultures. True specificities were assigned based on the TCRβ sequence found in the TetTCR-SeqHD experiment that matches known TCR sequences from bulk TCRβ sequencing. Cells were classified into the ‘filter’ category based on the criteria described in Methods. **b**, The recall and precision rates for each polyclonal T cell culture shown in **a**. **c–h**, Distribution of predicted antigen specificities for each T cell clone within each polyclonal T cell culture. The x axis for each plot was ranked by TCRβ-associated transcript copy numbers from MIDCIRS assay (left to right, high to low). The red box denotes the contaminant TCR clone in GAD antigen-specific polyclonal culture. Filter category is defined the same as in **a**. HIV, human immunodeficiency virus; NEG, negative; PPI, preproinsulin.

We detected antigen-binding events based on MID count distribution of tetramer DNA barcodes in each cell, which helped us to define antigen specificity and possible cross-reactive binding antigens for individual T cells (Methods). Using the known TCR sequences from T cell clones, their known antigen specificities and detected antigen specificity by tetramer DNA barcode, we showed that the precision is over 98% and the recall is over 80%, except for glutamic acid decarboxylase (GAD)-specific clones (Fig. 2a,b). Additional analysis revealed that the lower recall rate for GAD-specific clones was due to one non-GAD-binding clone (TCRβ: CASRFLGTEAFF) that accounted for 26% of all GAD-specific T cells, which is likely to be a nonspecific contaminant in the polyclonal culture (Fig. 2c–h). Thus, TetTCR-SeqHD showed high precision in detecting antigen specificity.

Diverse T cell phenotypes revealed by TetTCR-SeqHD. To further demonstrate the advantages of TetTCR-SeqHD in characterizing antigen-specific CD8⁺ T cells, we curated a panel of 215 endogenous and 65 foreign antigens from the IEDB database (Supplementary Table 4 and Methods) covering HLA-A01:01, HLA-A02:01 and HLA-B08:01 alleles and applied TetTCR-SeqHD in ten healthy donors without T1D and eight patients with T1D (Supplementary Table 5). Each donor had at least one HLA allele amenable to UV peptide exchange. Endogenous and foreign peptides were UV exchanged¹⁸ onto PE and allophycocyanin (APC)-labeled tetramers, respectively. Enzyme-linked immunosorbent assay (ELISA) on all 280 pMHC species showed no difference in pMHC UV-exchange efficiency between detected and undetected species (Supplementary Note). CD8⁺ T cells were stained and sorted similarly to the polyclonal T cell cultures (Supplementary Note). A hepatitis C virus

(HCV) antigen-specific CD8⁺ T cell clone was spiked into primary CD8⁺ T cells for all HLA-A02:01-positive donors. A total of 35,168 cells were recovered across four experiments. An average of 50,000 reads per cell were sequenced, covering all six groups of attributes (Supplementary Table 2). After single-cell quality filtering and removing putative multiplets (Methods), 32,992 cells were retained with a median of 62 detected genes and 47 detected antibodies per cell. Among all primary cells, 45% and 68% of cells had TCRα and TCRβ captured, with a pairing efficiency of 34%. Because the primary CD8⁺ T cells were recovered from frozen samples, lower gene and TCR capture rates were seen compared with cultured clones.

We started by performing joint modeling of RNA expression and surface-protein expression using totalVI²³, followed by dimensionality reduction using uniform manifold approximation and projection (UMAP)²⁴ and single-cell clustering with the Leiden algorithm (Fig. 3a)²⁵. Minimum batch effects among chips were detected (Fig. 3b). A total of 13 clusters were identified, consisting of major conventional CD8⁺ T cell phenotypes including naive T cells (T_{naive}; clusters 1–4), central memory T cells (T_{CM}; cluster 6), effector memory T cells (T_{EM}; clusters 8–10), effector T cells (T_{eff}; clusters 11 and 12) and transitional T cells between effector and memory populations (T_{trans}; cluster 7) based on CCR7 and CD45RA/CD45RO protein expression, spike-in HCV-specific clone (cluster 13) and CD56⁺ T cells, which are likely to be natural killer-like T cells²⁶ (cluster 5) (Fig. 3c,d). The large number of primary CD8⁺ T cells processed and the combined analysis of target gene and surface-protein expression provided a superior resolution to identify subpopulations. While clusters 8, 9 and 11 represent early stages of T_{EM} and T_{eff} cells, clusters 10 and 12 represent late-stage T_{EM} and T_{eff} cells based on the gradual changes of gene/protein expression. Similarly, T_{naive}

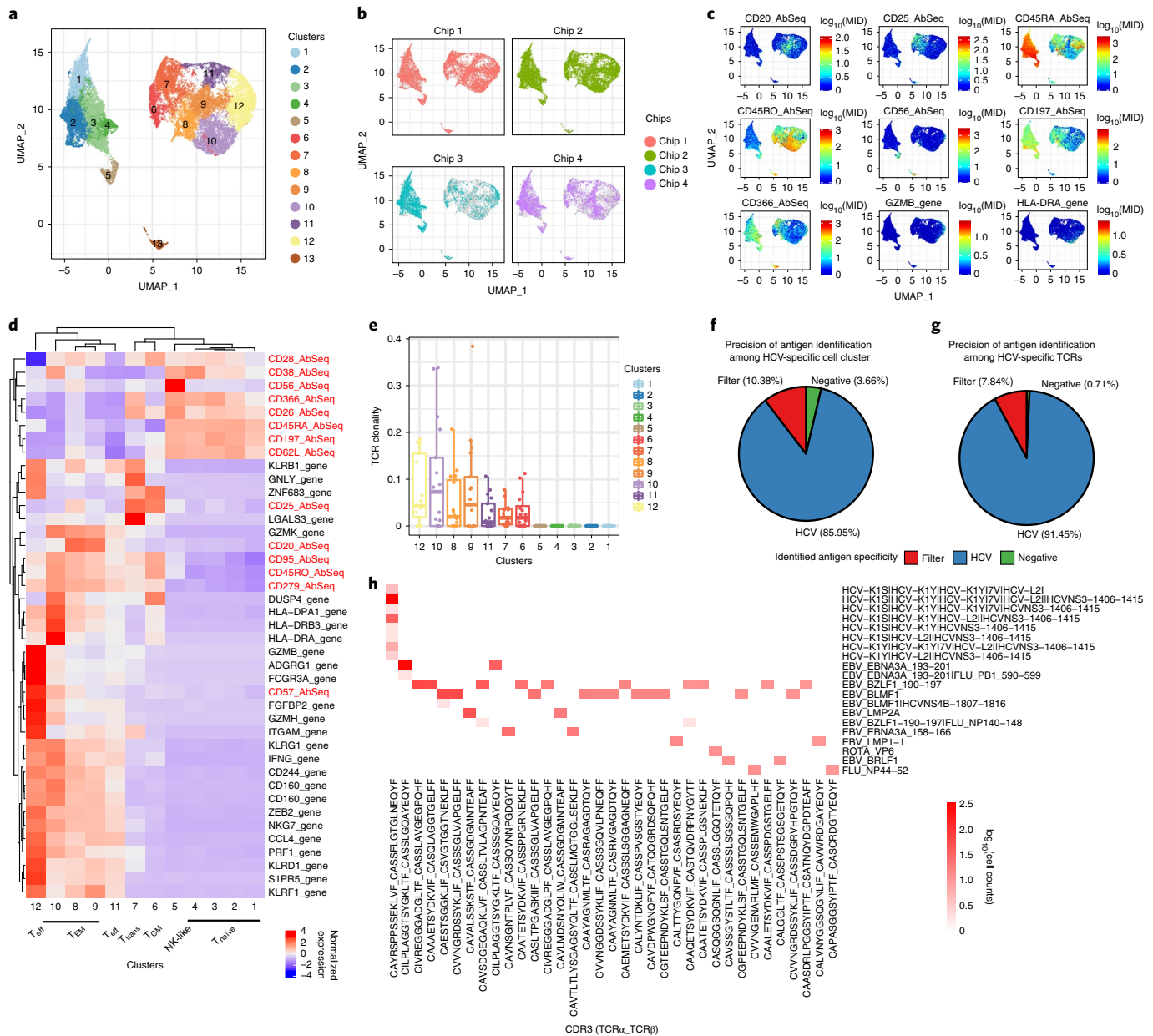


Fig. 3 | TetTCR-SeqHD enables combined gene expression, phenotype and TCR clonality comparison among antigen-specific CD8⁺ T cells. **a**, UMAP of 32,992 single cells sorted from healthy donors and donors with T1D. Thirteen clusters, including a cluster consisting of HCV-specific spike-in T cell clones, were identified. **b**, UMAP of single cells from different chips. Gray dots represent all cells, and colored dots are cells from different chips. **c**, Expression level of seven surface proteins (CD20, CD25, CD45RA, CD45RO, CD56, CD197 (CCR7) and CD366 (TIM3)) and two genes (*GZMB* and *HLA-DRA*) across single T cells illustrated in **a**. **d**, Z score normalized mean expression of differentially expressed genes and surface proteins (by antibodies) in each identified cluster. **e**, TCR clonality in 12 primary CD8⁺ T cell clusters among 18 donors. Horizontal lines represent median values, with whiskers extending to the farthest data point within a maximum of 1.5 × interquartile range. **f, g**, Precision of antigen identification among an HCV-specific T cell cluster (cluster 13) (**f**) and HCV-specific TCR-bearing cells (**g**). Cells were classified into a filter category based on the following criteria: (1) more than one antigen binds to a single cell, and these antigens are a distance of more than three amino acids away from each other; and (2) correlation of tetramer MID between a single cell and the median of all cells with the same TCR sequence is below 0.9, identified as described in Methods. **h**, Heatmap of the cognate antigen specificities of the top enriched TCRs (T cell clonality ≥10 cells). Top enriched TCRs are listed on the x axis, and the antigen specificities detected by TetTCR-SeqHD are listed on the y axis. Colored blocks indicate antigen binding to a particular TCR. White background represents no binding, which was true for most of the TCR-antigen combinations.

cells were also further separated into four clusters (1–4) (Fig. 3d and Extended Data Fig. 2a,b). Differentially expressed surface proteins showed a bimodal distribution among CD8⁺ T cells (Extended Data Fig. 2c,d). Of note, we found that cluster 5 (CD56⁺ T cells) is characterized by a fraction with low tetramer DNA-barcode

signal (Supplementary Note), and no enrichment of antigen-specific CD8⁺ T cells was identified. Among the 12 clusters of primary CD8⁺ T cells identified from all donors, the four T_{naive} clusters and the CD56⁺ T cell cluster have the lowest TCR clonality, which is ubiquitous in all donors. However, different activated T cell

subpopulations display various degrees of clonal expansion, and clusters 8–10 and 12 (T_{EM} and T_{eff}) have a relatively high TCR clonality in the majority of donors (Fig. 3e).

TetTCR-SeqHD was able to reveal diverse phenotypic classes represented by gradual changes of gene expression depicting CD8⁺ T cell activation status. In addition, we demonstrated that there was no cross-talk among different profiled dimensions, and adding AbSeq did not impact gene expression (Supplementary Note).

Distinct features of foreign antigen-specific T cells. Different donors show distinguishable phenotypic distributions on UMAP (Extended Data Fig. 3a), which prompted us to further examine the heterogeneous functional profiles of antigen-specific T cells among donors. Altogether, 12,518 viral antigen-specific, 3,626 non-T1D-related endogenous antigen-specific and 1,952 T1D-related endogenous antigen-specific T cells were detected, but the ratio varied in different individuals (Supplementary Table 6). Examples of tetramer MID count distribution among viral antigen-specific, T1D-related endogenous antigen-specific and cross-reactive T cells demonstrated clear antigen-specificity detection by TetTCR-SeqHD (Extended Data Fig. 4). Almost all the clonally expanded TCRs had unique antigen specificities identified, confirming the precision of TetTCR-SeqHD in primary CD8⁺ T cells from human peripheral blood mononuclear cells (PBMCs) (Fig. 3h). We further used the HCV-specific clone to characterize the precision and recall of TetTCR-SeqHD in primary CD8⁺ T cell experiments. In cluster 13, which was identified to harbor the HCV-specific spike-in clone, there were a total of 623 cells, 536 (86%) of which were accurately identified as binding to at least one HCV wild-type (WT) and associated variant antigen (Fig. 3f). Of these cells, a total of 421 cells were identified to have the same paired TCR α/β sequences as the HCV-specific clone, and 91% of them bound to at least one HCV WT and associated variant antigen (Fig. 3g).

A very small number of cells bound to mismatched HLA alleles, which further demonstrates the accuracy of antigen-specificity detection by TetTCR-SeqHD. There are two sources of mismatches: first, T cells binding more than one antigen displayed on multiple different HLA alleles were deemed HLA mismatch; second, any T cell bound to antigens derived from non-donor-specific HLA alleles was deemed HLA mismatch. Collectively, we showed that only a very small percentage (2.14%) of cells exhibited mismatched HLA binding (Extended Data Fig. 5a,b). Comparison of phenotypes among these cells demonstrated that mismatches occurred at random, without cells being in a particular phenotype (Extended Data Fig. 5c). In addition, we analyzed the distribution of tetramer MID signal fraction between the expanded clones and the rare cells (Supplementary Note). There was some enrichment of expanded T cell clones in the region with high tetramer MID signal fraction; however, the ranges of the two distributions were similar.

We compared the distribution of phenotypes between tetramer-positive and tetramer-negative populations. As expected, the tetramer-positive population had a relatively higher percentage of cells with non-naive phenotypes (Extended Data Fig. 3b). The majority of naive tetramer-positive CD8⁺ T cells were MART-1 and PGT-178 antigen specific (Extended Data Fig. 3c,d). Forty viral antigens that were detected in greater than five cells across all donors were selected for further analysis (Methods). As expected, different T cell phenotypic clusters were composed of distinct antigen specificities, with endogenous antigens occupying T_{naive} clusters and foreign antigens populating non-naive T cell clusters (Fig. 4a). In general, different donors, regardless of their T1D status, presented varying frequencies and phenotypic profiles of viral antigen-specific CD8⁺ T cells, possibly due to different infection or vaccination histories (Fig. 4b–e). However, we also found that some viral antigens induced distinct T cell phenotypes. Influenza

antigen-experienced T cells were mostly within cluster 7, where T cells display a Tim3⁺CD25⁺CD26⁺ phenotype^{27–29} (Fig. 4b and Supplementary Note). Epstein–Barr virus (EBV) antigens showed distinguishable phenotypes compared with influenza antigens (Fig. 4b and Supplementary Note). Two different categories of EBV antigens that originated from lytic and latent viral proteins also presented distinct phenotypes. Antigens from latent viral proteins, such as LMP1 and LMP2, preferentially induced T cells in central memory states (cluster 6), while lytic viral proteins, such as BRLF1 and BLMF1, displayed effector and effector memory phenotypes (clusters 8–10 and 12) (Fig. 4b), consistent with previous findings using mass cytometry by time-of-flight¹⁶. We also found that influenza matrix protein M1-specific CD8⁺ T cells displayed a more uniform phenotype distribution among donors compared with other antigens (Fig. 4d,e). In addition, increasing the tetramer MID threshold to be more stringent in antigen identification (Methods) did not change phenotypic features of viral antigen-specific CD8⁺ T cells (Extended Data Fig. 6).

Another advantage of TetTCR-SeqHD is its capacity to identify putative cross-reactive CD8⁺ T cells. Similar to TetTCR-Seq¹⁸, 85% of HCV-specific clones displayed binding to all five HCV antigens¹⁸ (Fig. 4f). We also examined the cross-reactivity detection in primary CD8⁺ T cells using MART1 antigens. MART1 WT antigens (MART1_{27–35} nonamer and MART1_{26–35} decamer) and their variants have been widely used as a model system of human cancer antigens. By changing one or two amino acids, such as MART1_{26–35} A27L and MART1_{26–35} E26A/A27L, it was noted that these variant peptides greatly improved the binding and stability of peptide–HLA-A*0201 complexes and enabled the otherwise-weak WT antigens to become potent immunogens^{30,31}. We thus used these peptides and studied the robustness of TetTCR-SeqHD in detecting both strong and weak pMHC ligands. Among cross-reactive cells, a total of 2,308 cells were identified to bind MART1 WT or variant antigens. Of these, 84% of cells bound to more than one MART1 WT or variant antigen (Fig. 4g). Interestingly, our method also detected previously noted cross-reactivities among the PGT-178 (LLAGIGTVPI) peptides and a MART1 variant antigen (ELAGIGILTV)³² and an additional MART1 variant cross-reactive antigen (ALAGIGILTV), despite differences of ≥ 5 amino acids in these peptides (Fig. 4h). TetTCR-SeqHD has the advantage of comparing phenotypes of both monospecific and cross-reactive CD8⁺ T cells simultaneously in high throughput.

Autoantigen differences between healthy individuals and individuals with T1D.

Among 209 T1D-related autoantigens included in the antigen pool, 106 and 102 different autoantigens were detected more than three times in 1,109 and 814 T1D antigen-specific cells from donors with and without T1D, respectively. The total T1D autoantigen tetramer-positive CD8⁺ T cell frequency was comparable between patients with T1D and healthy donors (Extended Data Fig. 7). However, comparing the frequency of T1D autoantigen-specific CD8⁺ T cells individually, we found that INS-WMR-10-, PPI-29-38- and PTPRN-805-813-specific cells exhibited a significantly higher cell frequency in patients with T1D compared to healthy donors within this donor cohort (Fig. 5a). Among them, PTPRN-805-813 was reported before as a potential marker in PBMCs of patients with T1D³³, and PPI-29-38 was identified as an HLA-A02:01-low binder but is present in patients with T1D³⁴. To ensure the sensitivity of our analysis, we increased the tetramer-negative MID threshold to 15 and compared the frequency difference between T1D and healthy donors. Five antigens were identified, including previously identified INS-WMR-10 and PTPRN-805-813, further validating the potential of these two antigens to distinguish between individuals with T1D and healthy donors (Extended Data Fig. 8a). We also noticed varying degrees of clonal expansion in T1D autoantigen-specific T cells isolated from

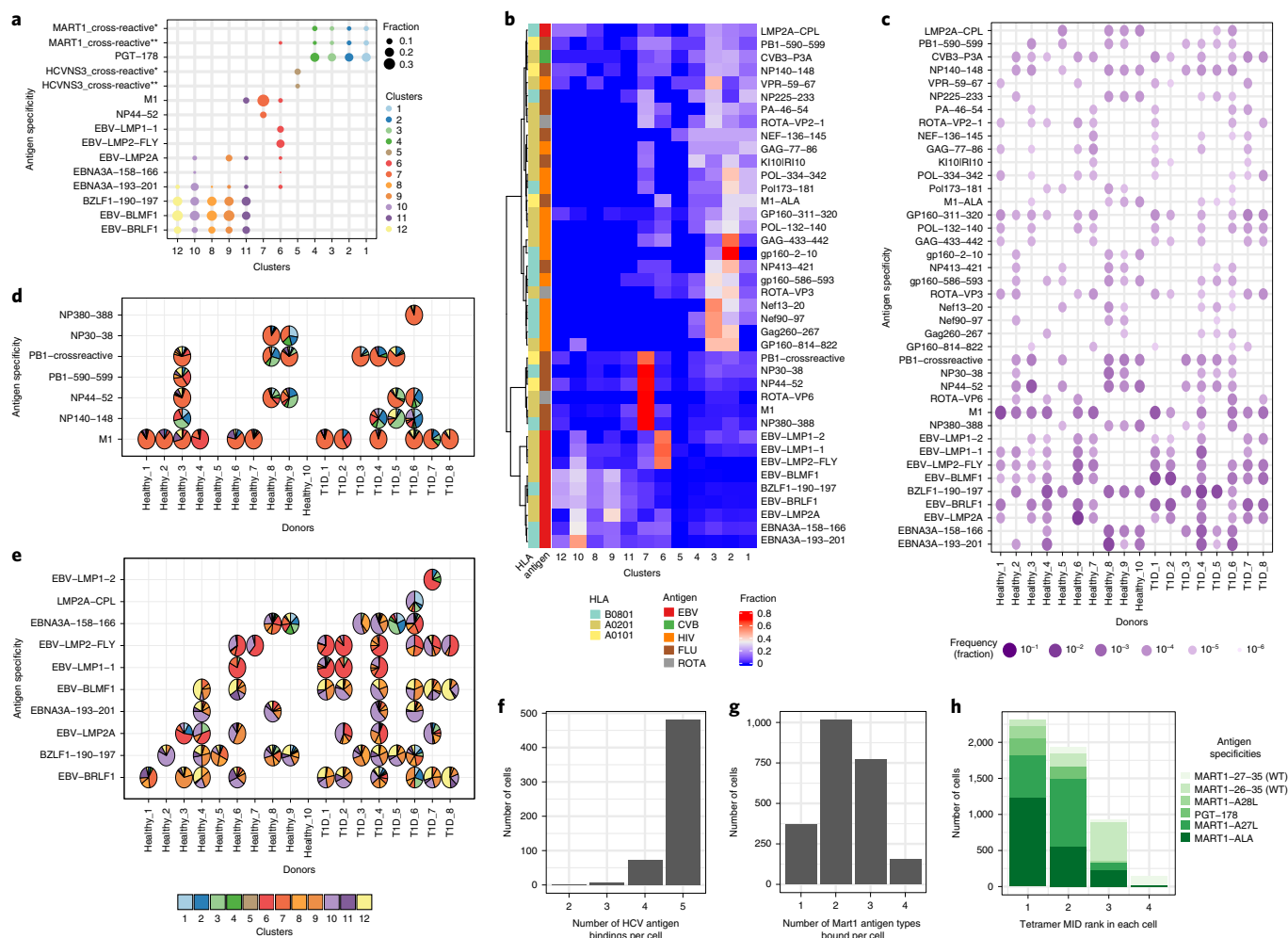


Fig. 4 | Gene expression and phenotype analysis of foreign antigen-specific T cells and cross-reactivity validation. **a**, Top representative antigen specificities in 12 primary CD8⁺ T cell clusters (MART1_crossreactive*: MART1-26-35 | MART1-A27L | MART1-ALA; MART1_crossreactive***: MART1-A27L | MART1-ALA; HCVNS3_crossreactive*: HCV-K15 | HCV-K1Y | HCV-K1YI7V | HCV-L2I | HCVNS3-1406-1415; HCVNS3_crossreactive***: HCV-K15 | HCV-K1Y | HCV-L2I | HCVNS3-1406-1415). **b**, The distribution of viral-antigen specific CD8⁺ T cells from all 18 donors among 12 primary CD8⁺ T cell clusters. Fraction was calculated as the fraction of each phenotype cluster among each antigen specificity. CVB, Coxsackie virus B; FLU, influenza virus; ROTA, rotavirus. **c**, Frequency comparison of viral antigen-specific T cells in 18 donors (in **b-d**; “PB1-crossreactive” represents cross-reactivity with PB1-590-599 and PB1-591-599). **d,e**, Phenotype distribution of influenza-specific (**d**) and EBV-specific (**e**) T cells in each individual. Each pie represents the T cell distribution in 12 phenotypes of primary CD8⁺ T cells for the corresponding donor (x axis) and antigen (y axis) combination. Empty spaces in **d** and **e** mean no antigens were detected above the threshold for the corresponding donor-antigen combination (antigen specificities with <10 cells were excluded in some donors). **f**, Histogram of the number of different types of HCV antigens, WT and variant antigens, bound per cell. **g**, Histogram showing the number of different types of Mart1 antigens, WT and variant antigens, bound per cell. **h**, Distributions of tetramer MID ranks of Mart1 WT and variant antigens and a cross-reactive antigen in each cell for four groups of binding patterns. MART1-27-35: AAGIGILTV (WT); MART1-26-35: EAAGIGILTV (WT); MART1-A28L: ALGIGILTV; MART1-A27L: ELAGIGILTV; MART1-ALA: ALAGIGILTV; PGT-178: LLAGIGTVPI; PGT-178: LLAGIGTVPI.

different patients with T1D, revealing the complexity of the antigen landscape in T1D (Extended Data Fig. 8b). This could also be caused by limited sampling from PBMCs. In addition, the phenotype distribution of T1D autoantigen tetramer-positive CD8⁺ T cell between healthy individuals and patients with T1D was comparable (Extended Data Fig. 9).

We further identified an expanded T cell clone cross-recognizing three different antigens (INSDRIP-1-9, DUF5119-124-133 and PTPRN-797-805) in a patient with type 2 diabetes (T2D). This led us to test the plasma banked from the same blood draw and showed the patient was positive for GAD-reactive autoantibodies. Further review of the medical record showed that the patient was diagnosed with latent autoimmune diabetes in adults after the sample was collected for this study. Interestingly, INSDRIP-1-9 is derived from an

alternative open reading frame within human insulin mRNA, and significantly higher levels of INSDRIP-1-9-specific CD8⁺ T cells were reported to be detected in patients with T1D³⁵. DUF5119-124-133 is derived from *Bacteroides fragilis* and *Bacteroides thetaiotaota*-micron, common bacteria found in human gut microbiota³⁶, and PTPRN-797-805 is derived from IA2 protein, a previously known T1D autoantigen³⁷. This is likely due to cross-reactivity of the three antigens by the same TCR. To confirm the result, we transduced this TCR together with other TCRs identified among patients with T1D and healthy donors to further validate the accuracy of TetTCR-SeqHD (Supplementary Table 7). Tetramer staining (Fig. 5b and Extended Data Fig. 10) and antigen-stimulation experiments (Fig. 5c) both confirmed that cognate TCRs identified in TetTCR-SeqHD can bind and be stimulated by respective antigens.

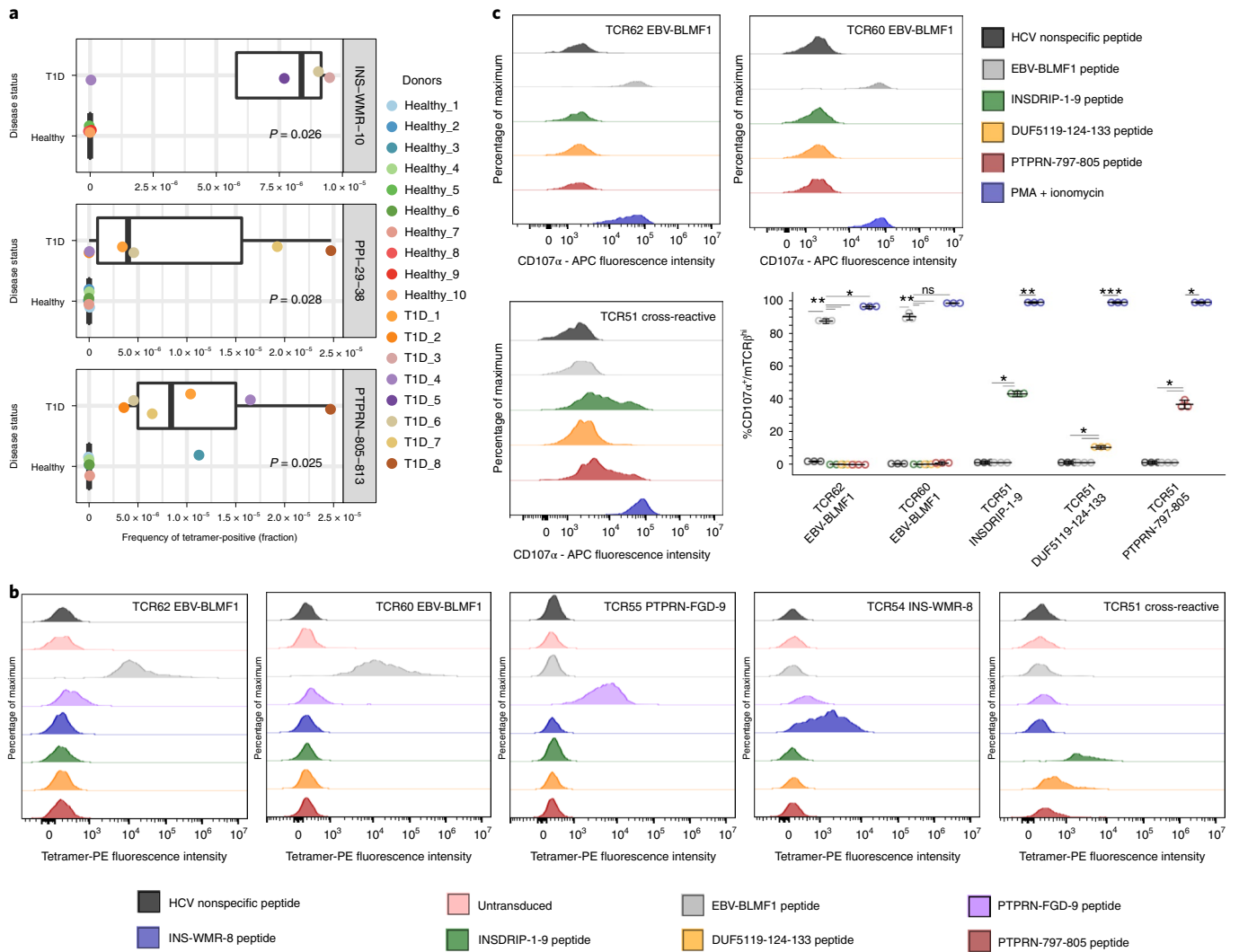


Fig. 5 | Identification of three T cell specificities selectively enriched in patients with T1D and TCR specificity and cross-reactivity validation. **a**, Three T1D-related antigens (INS-WMR-10, PPI-29-38 and PTPRN-805-813) were identified to have a significantly higher frequency of cognate T cells in the peripheral blood of patients with T1D compared to healthy donors. A two-sided Wilcoxon signed-rank test was performed. Horizontal lines represent 25th percentile, median and 75th percentile values, with whiskers extending to the farthest data point within a maximum of 1.5 × interquartile range. The number of donors with HLA-A01:01, HLA-A02:01 and HLA-B08:01 are 10, 12 and 10, respectively. **b**, TCR specificity and cross-reactivity validation by pMHC tetramer staining. Five TCRs that were identified to recognize six different antigens in complex with distinct HLA alleles, including TCR51 that recognized three unrelated antigens, were transduced into human primary CD8⁺ T cells and stained with respective cognate pMHC tetramers or other pMHC tetramers. The percentage of tetramer-positive cells was gated on the TCR β ^{hi} fraction of the cells. Statistical comparisons are listed in Extended Data Fig. 10. HLA-A02:01: EBV-BLMF1, INSDRIP-1-9, DUF5119-124-133, PTPRN-797-805; HLA-B08:01: INS-WMR-8; HLA-A01:01: PTPRN-FGD-9. **c**, TCR specificity and cross-reactivity validation by T cell functionality. The HLA-A02:01-restricted TCR-transduced cells generated in **b** were further stimulated with respective cognate pMHC tetramers or other pMHC tetramers. The percentage of CD107 α ⁺ cells was measured on the TCR β ^{hi} fraction of the cells. Experiments in **b** and **c** were performed in triplicate. PMA, phorbol myristate acetate. A two-tailed Student's *t* test was performed. Data are presented as mean values \pm s.d. EBV-BLMF1: GLCTLVAML; INSDRIP-1-9: MLYQHLLPL; DUF5119-124-133: MVWGPDPPLYV; PTPRN-797-805: MVWESGCTV; INS-WMR-8: WMRLPLL; PTPRN-FGD-9: FGDHPGHSY. ns, not significant; * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$.

In summary, we applied TetTCR-SeqHD to compare autoantigen-specific CD8⁺ T cells in healthy donors and those with T1D and identified putative autoantigens with differing frequencies in PBMCs from a large pooled autoantigen library.

Discussion

In this study, we developed a method to simultaneously profile TCR sequences, cognate antigen specificity, gene expression and surface-protein expression for single primary CD8⁺ T cells in a high-throughput manner. We addressed the precision of TetTCR-SeqHD, its ability to profile TCR cross-reactivity

and its application to study diverse phenotypes of foreign- and self-antigen-specific CD8⁺ T cells. By using in vitro-cultured polyclonal T cells with known antigen specificities and TCR sequences, we established that TetTCR-SeqHD has the precision to detect the correct antigen specificity of over 98% and a recall rate of over 80%, except for GAD-specific clones. The low recall rate was mainly caused by the nonspecific clone with TCR β sequence CASRFLGTEAFF. The in vitro-expanded polyclonal population was sorted for GAD tetramer-positive cells before applying bulk TCR sequencing²² and TetTCR-SeqHD. Although this nonspecific clone passed tetramer-based flow cytometry sorting and was

present in both bulk TCR repertoire and TetTCR-SeqHD data, TetTCR-SeqHD indeed assigned the majority of cells belonging to this clone to the filter, with nonspecific binding. This further demonstrates the superior specificity of TetTCR-SeqHD in identifying nonspecific clones that would otherwise be identified as tetramer positive by fluorescence-based detection.

Recently, a DNA-barcoded dextramer technique, dCODE dextramer, was adopted by the 10x Genomics platform to enable the profiling of antigen-specific CD8⁺ T cells. However, the dCODE dextramer suffers from the high cost of generation of dextramers and thus lacks the flexibility to screen large antigen panels. This prevents it from profiling antigens in a high-throughput manner. By combining IVTT with the UV-exchange technique, TetTCR-SeqHD enables the creation of a panel of antigens (on the scale of hundreds) affordably and quickly (within 1 week). Therefore, we created a large panel of antigens consisting of foreign antigens derived from various viruses and self-antigens derived from known T1D autoantigens and profiled CD8⁺ T cells to recognize these antigens from healthy individuals and those with T1D.

With the ability to profile targeted gene expression and surface-protein expression simultaneously using the BD Rhapsody platform, we resolved 12 clusters for primary CD8⁺ T cells plus one cluster for in vitro-cultured HCV-specific T cell clones. Most importantly, T cell phenotypic and functional subclasses, represented by gradual changes of gene expression, were revealed among these 12 clusters, from naive to early-stage effector and memory populations, to the transitional state between effector and memory, and to late-stage effector and memory populations. By investigating the composition of phenotypic clusters for each antigen, phenotype signatures of distinct antigens were assessed. We found that viral antigens from influenza and EBV display distinct phenotypes. Influenza-specific CD8⁺ T cells were mostly enriched in cluster 7, displaying a transitional phenotype between effector and memory populations, while EBV-specific CD8⁺ T cells were largely memory and effector populations. Similar phenotypic differences between EBV latent and lytic antigens were observed previously using mass cytometry¹⁶. This example further validates the robustness of TetTCR-SeqHD to capture the phenotypic profiles of antigen-specific CD8⁺ T cells. Moreover, studied donors also showed diverse phenotype signatures of influenza- and EBV-specific CD8⁺ T cells due to different viral infection (or vaccination) histories.

In addition to its high precision and high-throughput capacity, TetTCR-SeqHD also enables the detection of cross-reactive T cells. We examined cross-reactivity in both in vitro-cultured HCV-specific CD8⁺ T cell clones and primary CD8⁺ T cells. We not only detected cross-reactivity among HCV and MART1 WT and variant antigens but also found cross-reactivity among INSDRIP-1-9, DUF5119-124-133 and PTPRN-797-805 in a patient with T2D. The TCR sequences obtained simultaneously demonstrated their critical role in validating antigen specificity and cross-reactivity in high-throughput antigen screening and antigen-specific T cell profiling. Interestingly, these three antigens are more than three amino acids away from each other, underscoring the flexibility of TCR-antigen recognition between dissimilar peptides. Given that DUF5119-124-133 is derived from the human gut microbiota, the association between certain dysbiosis of gut microbiome and the role of T cells in the onset of T1D requires further investigation.

Lastly, with the panel of T1D autoantigens, we investigated the differences in autoantigen-specific CD8⁺ T cells between healthy individuals and those with T1D. Although we did not identify any phenotypic differences, we found three antigens (INS-WMR-10, PPI-29-38 and PTPRN-805-813) that exhibited a significantly higher antigen-specific CD8⁺ T cell frequency in patients with T1D within this donor cohort. Of note, with simultaneously screening of 280 antigens, the false discovery rate-corrected

P value is 0.99 for all identified antigens. However, due to the large number of peptide panels and the limited number of blood samples, this false discovery rate-corrected *P* value could be falsely inflated. Instead, we performed a sensitivity analysis by increasing the tetramer MID threshold to be more stringent in antigen identification. Five antigens were identified, including previously identified INS-WMR-10 and PTPRN-805-813, further providing evidence for including these antigens in future larger-cohort studies. Wiedeman et al. recently found that activated islet-specific CD8⁺ memory T cells were prevalent in donors with T1D who experienced a rapid loss of C-peptide, whereas slow disease progression was associated with an exhaustion-like profile³⁸. By contrast, Culina et al. reported a predominantly naive phenotype for circulating islet-specific CD8⁺ T cells in T1D³⁹, similar to our results. These contradicting results are likely due to different patient cohorts with different T1D onset timing as well as the choice of T1D antigens. Given that a similar attempt using a much smaller panel of T1D-related autoantigens failed to identify any antigens within PBMCs that would separate healthy individuals from those with T1D³⁹, our results provide a premise that warrants further tests in a much larger cohort, which could be very useful in T1D early diagnosis.

Due to the advantage of multidimensional profiling of single cells, TetTCR-SeqHD enables one to identify phenotypic differences of antigen-specific CD8⁺ T cells, distinguish disease status, screen antigens with a high throughput and identify TCRs with therapeutic potential. TetTCR-SeqHD is likely to be a game changer in basic and translational research focusing on T cells.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgments, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41590-021-01073-2>.

Received: 18 December 2020; Accepted: 15 October 2021;

Published online: 22 November 2021

References

1. Davis, M. M. & Boyd, S. D. Recent progress in the analysis of alphaT cell and B cell receptor repertoires. *Curr. Opin. Immunol.* **59**, 109–114 (2019).
2. Pulendran, B. & Davis, M. M. The science and medicine of human immunology. *Science* **369**, 1582–1593 (2020).
3. Satpathy, A. T. et al. Transcript-indexed ATAC-seq for precision immune profiling. *Nat. Med.* **24**, 580–590 (2018).
4. Stoekius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
5. Peterson, V. M. et al. Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* **35**, 936–939 (2017).
6. Mair, F. et al. A targeted multi-omic analysis approach measures protein expression and low-abundance transcripts on the single-cell level. *Cell Rep.* **31**, 107499 (2020).
7. Granja, J. M. et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat. Biotechnol.* **37**, 1458–1465 (2019).
8. Fernandez, D. M. et al. Single-cell immune landscape of human atherosclerotic plaques. *Nat. Med.* **25**, 1576–1588 (2019).
9. Su, Y. et al. Multi-omics resolves a sharp disease-state shift between mild and moderate COVID-19. *Cell* **183**, 1479–1495 (2020).
10. Li, G. et al. T cell antigen discovery via trogocytosis. *Nat. Methods* **16**, 183–190 (2019).
11. Joglekar, A. V. et al. T cell antigen discovery via signaling and antigen-presenting bifunctional receptors. *Nat. Methods* **16**, 191–198 (2019).
12. Kisielow, J., Obermair, F. J. & Kopf, M. Deciphering CD4⁺ T cell specificity using novel MHC-TCR chimeric receptors. *Nat. Immunol.* **20**, 652–662 (2019).
13. Kula, T. et al. T-scan: a genome-wide method for the systematic discovery of T cell epitopes. *Cell* **178**, 1016–1028 (2019).

14. Sharma, G., Rive, C. M. & Holt, R. A. Rapid selection and identification of functional CD8⁺ T cell epitopes from large peptide-coding libraries. *Nat. Commun.* **10**, 4553 (2019).
15. Ferretti, A. P. et al. Unbiased screens show CD8⁺ T cells of COVID-19 patients recognize shared epitopes in SARS-CoV-2 that largely reside outside the spike protein. *Immunity* **53**, 1095–1107 (2020).
16. Newell, E. W. et al. Combinatorial tetramer staining and mass cytometry analysis facilitate T-cell epitope mapping and characterization. *Nat. Biotechnol.* **31**, 623–629 (2013).
17. Simoni, Y. et al. Bystander CD8⁺ T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature* **557**, 575–579 (2018).
18. Zhang, S. Q. et al. High-throughput determination of the antigen specificities of T cell receptors in single cells. *Nat. Biotechnol.* **36**, 1156–1159 (2018).
19. Rodenko, B. et al. Generation of peptide-MHC class I complexes through UV-mediated ligand exchange. *Nat. Protoc.* **1**, 1120–1132 (2006).
20. Bender, C., Rajendran, S. & von Herrath, M. G. New insights into the role of autoreactive CD8 T cells and cytokines in human type 1 diabetes. *Front. Endocrinol. (Lausanne)* **11**, 606434 (2020).
21. Shahi, P., Kim, S. C., Haliburton, J. R., Gartner, Z. J. & Abate, A. R. Abseq: Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. *Sci. Rep.* **7**, 44447 (2017).
22. Ma, K. Y. et al. Immune repertoire sequencing using molecular identifiers enables accurate clonality discovery and clone size quantification. *Front. Immunol.* **9**, 33 (2018).
23. Gayoso, A. et al. Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nat. Methods* **18**, 272–282 (2021).
24. McInnes, L., Healy, J. & Melville, J. Umap: uniform manifold approximation and projection for dimension reduction. Preprint at <https://arxiv.org/abs/1802.03426> (2018).
25. Levine, J. H. et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* **162**, 184–197 (2015).
26. Pita-Lopez, M. L., Pera, A. & Solana, R. Adaptive memory of human NK-like CD8⁺ T-cells to aging, and viral and tumor antigens. *Front. Immunol.* **7**, 616 (2016).
27. Keating, R. et al. Potential killers exposed: tracking endogenous influenza-specific CD8⁺ T cells. *Immunol. Cell Biol.* **96**, 1104–1119 (2018).
28. Sharma, S. et al. T cell immunoglobulin and mucin protein-3 (Tim-3)/galactin-9 interaction regulates influenza A virus-specific humoral and CD8 T-cell responses. *Proc. Natl Acad. Sci. USA* **108**, 19001–19006 (2011).
29. Ibegbu, C. C. et al. Differential expression of CD26 on virus-specific CD8⁺ T cells during active, latent and resolved infection. *Immunology* **126**, 346–353 (2009).
30. Men, Y. et al. Assessment of immunogenicity of human Melan-A peptide analogues in HLA-A*0201/Kb transgenic mice. *J. Immunol.* **162**, 3566–3573 (1999).
31. Derre, L. et al. A novel population of human melanoma-specific CD8 T cells recognizes Melan-AMART-1 immunodominant nonapeptide but not the corresponding decapeptide. *J. Immunol.* **179**, 7635–7645 (2007).
32. Dutoit, V. et al. Degeneracy of antigen recognition as the molecular basis for the high frequency of naive A2/Melan-a peptide multimer⁺ CD8⁺ T cells in humans. *J. Exp. Med.* **196**, 207–216 (2002).
33. Blancou, P. et al. Immunization of HLA class I transgenic mice identifies autoantigenic epitopes eliciting dominant responses in type 1 diabetes patients. *J. Immunol.* **178**, 7458–7466 (2007).
34. Abreu, J. R. et al. CD8 T cell autoreactivity to preproinsulin epitopes with very low human leucocyte antigen class I binding affinity. *Clin. Exp. Immunol.* **170**, 57–65 (2012).
35. Kracht, M. J. et al. Autoimmunity against a defective ribosomal insulin gene product in type 1 diabetes. *Nat. Med.* **23**, 501–507 (2017).
36. Cole, D. K. et al. Hotspot autoimmune T cell receptor binding underlies pathogen and insulin peptide cross-reactivity. *J. Clin. Invest.* **126**, 3626 (2016).
37. Velthuis, J. H. et al. Simultaneous detection of circulating autoreactive CD8⁺ T-cells specific for different islet cell-associated epitopes using combinatorial MHC multimers. *Diabetes* **59**, 1721–1730 (2010).
38. Wiedeman, A. E. et al. Autoreactive CD8⁺ T cell exhaustion distinguishes subjects with slow type 1 diabetes progression. *J. Clin. Invest.* **130**, 480–490 (2020).
39. Culina, S. et al. Islet-reactive CD8⁺ T cell frequencies in the pancreas, but not in blood, distinguish type 1 diabetic patients from healthy donors. *Sci. Immunol.* **3**, 20 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Samples and material. Human whole blood from patients diagnosed with T1D and T2D was obtained at Seton Family of Hospitals at Austin with informed consent. The use of whole blood from these patients was approved by the institutional review board of the Ascension Seton University Physicians Group under institutional review board number 2013-10-0140 and is compliant with all relevant ethical regulations. Human PBMCs from healthy donors were purchased from ePBMC.

Generation of DNA-barcoded fluorescent streptavidin. The conjugation of DNA linker (Supplementary Table 8) to PE- or APC-labeled streptavidin was performed as previously described, with slight modifications¹⁸. During S-HyNic modification of PE- or APC-labeled streptavidin, 2 mol equivalent of S-HyNic was used. Following the conjugation of DNA linker, peptide-encoding DNA barcodes (Supplementary Table 4) were annealed to the complementary DNA linker on the DNA-linker PE or APC streptavidin conjugate in the presence of 1× NEBbuffer2 (NEB) with the following program: 60 °C for 30 s, then –1 °C/cycle for 35 cycles. The final DNA-barcoded fluorescent streptavidin conjugate was stored at 4 °C.

IVTT. Peptide-encoding DNA oligonucleotides were purchased from Sigma-Aldrich. DNA templates (50 nM) were first amplified by PCR as described previously with modifications¹⁸. IVTT_r and IVTT_f primers (1 μM; Supplementary Table 8) were used in the following reaction conditions: 95 °C for 3 min; then 22 cycles of 95 °C for 20 s, 59 °C for 30 s and 72 °C for 30 s; then 72 °C for 5 min. The PCR product was then diluted with 50 μl nuclease-free water before proceeding to the IVTT reaction.

Generation of the pMHC tetramer library. IVTT-generated peptides were mixed with biotinylated pMHC monomers containing a UV-labile peptide. The UV-labile peptide-loaded pMHC monomers were provided by the National Institutes of Health tetramer core. The final concentration of biotinylated pMHC is 0.2 mg ml⁻¹. Individual pMHC was formed through UV exchange as described previously¹⁹. Confirmation of the quality and concentration of UV-exchanged pMHC monomer was assessed by an ELISA assay as described previously¹⁹. Individual pMHC tetramers and the tetramer library pool were generated and tested as described previously¹⁸. pMHC tetramer library should only be pooled together immediately before cell staining.

Customization of CD2 SampleTag, custom AbSeq and custom CD50

SampleTag. Anti-CD2 antibody was purchased from Biolegend (clone RPA-2.10, Biolegend). Amine-modified oligonucleotide was purchased from Sigma-Aldrich (Supplementary Table 8). The conjugation between the oligonucleotide and anti-CD2 antibody followed the CITE-seq protocol⁴.

Corresponding antibodies and used oligonucleotides are listed in Supplementary Table 9.

Twelve CD50 antibody SampleTags¹⁰ were customized by BD Biosciences using the commercial SampleTag oligonucleotides.

Sorting and culture of antigen-specific CD8⁺ T cell polyclones. Seven types of tetramers with peptides chemically synthesized and UV-exchanged to MHC were used to raise antigen-specific polyclonal T cells (Supplementary Table 1). For each antigen specificity, 20 tetramer-positive CD8⁺ single T cells were sorted into each well of the 96-well plate and cultured for 3 weeks. Polyclonal T cell expansion and culture were performed according to a previously published protocol⁴¹.

pMHC tetramer staining and sorting of primary human CD8⁺ T cells. PBMCs from T1D whole blood were isolated using Ficoll-Paque density-gradient centrifugation (GE Healthcare). CD8⁺ T cells were then enriched from PBMCs of T1DM and healthy donors using the EasySep Human CD8⁺ T cell isolation kit (Stemcell Technologies).

CD8⁺ T cells were resuspended in FACS buffer containing 0.05% sodium azide and 50 nM dasatinib. CD8⁺ T cells were then incubated at 37 °C for 30–60 min. Approximately 10,000 cells from an HCV peptide-binding clone used previously¹⁸ were prestained with BV510 anti-CD8a antibody (clone RPA-T8, Biolegend) and spiked into the primary CD8⁺ T cells. Following dasatinib treatment, the tetramer pool, together with BV421 anti-CD8a antibody (clone RPA-T8, Biolegend), was directly added into the cells. Cells were incubated at 4 °C for 1 h with continuous rotation. After washing, cells were further stained at 4 °C for 20 min with the presence of 5 μg ml⁻¹ mouse anti-PE (clone PE001, Biolegend) and/or mouse anti-APC (clone APC003, Biolegend). AbSeq staining mastermix was prepared by pooling 1 μl of each AbSeq together (Supplementary Table 9). Cells were washed in FACS buffer once and stained with the AbSeq mastermix. Additional dump-channel antibodies (AF488-anti-CD4, AF488-anti-CD14 and AF488-anti-CD19), 7-aminoactinomycin D and 2 μl anti-CD50 SampleTag were mixed in cells. Cells were incubated at 4 °C for 40 min prior to washing in FACS buffer twice and then sorted.

During cell sorting, approximately 50,000 tetramer-negative CD8⁺ T cells were also sorted and then later spiked into tetramer-positive T cells. FlowJo V10 was used to process FACS data.

BD Rhapsody sequencing library preparation and sequencing. Prior to BD Rhapsody processing, tetramer-negative CD8⁺ T cells were first stained with 2 μl CD2 SampleTag at 4 °C for 30 min. Cells were washed in FACS buffer three times and resuspended in 100 μl BD Sample Buffer. Sorted tetramer-positive CD8⁺ T cells and tetramer-negative CD8⁺ T cells were counted using BD Rhapsody. Tetramer-positive and tetramer-negative CD8⁺ T cells were pooled and processed on a BD Rhapsody cartridge following the user's manual. Single-cell mRNA, AbSeq barcodes, tetramer barcodes and SampleTag barcodes were all captured by BD Rhapsody beads coated with poly(T) oligonucleotide, with a unique cell barcode and molecular barcode on each bead. Single-cell cDNA synthesis and library amplification were performed following the manufacturer's protocol, with some modifications. Briefly, in PCR1, 1.2 μl tetramer PCR1 primer was added to the PCR reaction in addition to primers for gene expression panel, AbSeq, SampleTag and universal oligonucleotides (Supplementary Table 9). Nine and ten PCR cycles were used for 5,000–10,000 and 10,001–20,000 cells, respectively. Double-sided size selection with AMPure beads was performed to purify short amplicons (AbSeq, SampleTag and tetramer DNA-barcodes) and long amplicons (target genes and TCRα/β) separately. In PCR2, five separate PCR reactions with 15 reaction cycles were carried out to amplify gene panel, SampleTag, TCRα, TCRβ and tetramer DNA barcodes. AbSeq, tetramer and TCRα/β libraries were gel extracted for the desired band before proceeding to PCR3. Finally, eight cycles of PCR reactions were performed for all six elements following the manufacturer's instructions. All PCR libraries were quantified using Bioanalyzer 2100 and pooled. Fifteen percent PhiX was used in all sequencing runs. Pooled libraries were sequenced on HiSeq X with PE150.

BD Rhapsody sequencing preprocessing. Sequencing reads from target gene expression, AbSeq, SampleTag, TCRα/β and tetramer DNA barcodes were processed as described below (Supplementary Note).

For target gene expression and AbSeq sequencing, reads were processed with BD Targeted Multiplex Rhapsody Analysis Pipeline Version 1.5 on the Seven Bridges platform following the manufacturer's instructions. For tetramer and SampleTag sequencing, reads were processed with custom codes and are available in GitHub. True cell barcodes were converted to oligonucleotide sequences according to BD cell barcode indexing rules. Then, sequencing data of tetramer, TCRα and TCRβ were processed using umitools⁴² to extract a cellular barcode and unique MID for each read. Reads that are mapped to true cell barcodes were obtained.

For tetramer DNA barcodes, only reads that were an exact match for the tetramer DNA-barcode reference were retained. The number of reads of the same MID-tagged tetramer DNA barcode (unique tetramer DNA barcode) was counted for each cell. The distribution of the reads of unique tetramer DNA barcode follows a bimodal distribution as reported previously¹⁸. The first peak corresponds to PCR and sequencing errors, and thus, reads falling under the first peak were filtered out. Further, the number of MIDs aligned to each tetramer DNA barcode in each cell was determined to construct a tetramer DNA-barcode count matrix.

For the SampleTag DNA barcodes, reads were mapped to SampleTag DNA-barcode reference using bowtie2 with –norc and –local mode⁴³. Aligned reads were then processed using umitools to count the number of MIDs for each SampleTag DNA barcode in each cell. The distribution of MID counts for each SampleTag was fitted by a bimodal distribution, and the cutoff between two distributions was set as the negative threshold for the corresponding SampleTag. In addition, to recover false-negative SampleTag signals, SampleTags whose MID counts accounted for more than 50% of total SampleTag MID counts were also classified as a positive event. Cells containing CD2 SampleTag were tetramer-negative cells, whereas cells with more than two regular SampleTags were multiplets and were removed from further analysis.

For the TCR sequencing reads, we adapted a subclustering algorithm as previously described⁴⁴ to remove PCR and/or sequencing errors and identify VDJ and CDR3, with some changes. Reads were first aligned to TCR J and C region reference. Only reads that are more than 62.5% identical were retained. Reads with the same cellular barcodes and MID were grouped together. Under each group, reads within a Levenshtein distance of 15% were further clustered into a subgroup. For each subgroup, a consensus sequence was built based on the average nucleotide at each position, weighted by quality score. After ranking the consensus sequences by their abundance, the most abundant consensus sequence was selected, and other sequences with an edited distance of less than three were removed. In cases where the most abundant consensus sequence was nonproductive, the next most abundant productive sequence (if it existed) was selected as the unique consensus sequence for that cell. The second TCR chain was retained when its MID count accounted for more than 20% of total TCRα or TCRβ MID counts.

Dimensionality reduction, clustering and differential expression of single cells.

All single cells were first filtered to exclude low-quality cells whose total gene and AbSeq expression MID counts were in the last 1% quantile. Then, cells identified as multiplets with SampleTag and cells with two productive TCRβ chains were also removed. Additionally, genes or AbSeqs whose expression was detected in fewer than 50 cells were filtered out. Gene expression and AbSeq data from different Rhapsody chips were pooled together and used to perform joint probabilistic modeling of RNA expression and surface-protein measurement with totalVI²³.

Each donor was treated as an independent batch factor and 200 epochs were used to train the model. Other parameters were set as default in totalVI. The posterior dataset was then used for dimensionality reduction (UMAP algorithm) and clustering (Leiden algorithm), both with Scanpy⁴⁵.

Calling tetramer specificity for each cell. First, for each tetramer fluorescent color, the distribution of total tetramer DNA-barcode counts per cell was fitted to a bimodal distribution. The cutoff points were set as the negative threshold to capture positive tetramer-binding events. Tetramer DNA-barcode counts were then ranked for each cell, and the knee point on the count-rank plot was selected. Antigens that ranked higher than the inflection point were included as putative binding antigens, and antigens that ranked below the inflection point but showed a difference of three amino acids or fewer compared with higher-ranking antigens were also included as putative cross-reactive binding antigens. For each cell, the tetramer MID signal fraction was defined as the fraction of the cumulative MID count from putative binding antigens divided by the cumulative MID count from all bound antigens:

$$\text{tetramer MID signal fraction} = \frac{\sum \text{MID}_{\text{putative binding antigens}}}{\sum \text{MID}_{\text{all}}}$$

Cells with a tetramer MID signal fraction below 0.4 were prefiltered in the preprocessing step to identify antigen specificities. Further, cells with the same TCR α/β were pooled together. The correlation coefficient of antigen binding for each cell in the pool was calculated between detected tetramer DNA-barcode counts and the corresponding median tetramer DNA-barcode counts within the pool. This correlation coefficient for each cell was used as the tetramer-binding noise. The knee point of the distribution of correlation coefficients was set as the threshold below which cells were removed due to high tetramer-binding noise.

For analysis of viral antigens, we selected antigens detected in more than five cells to ensure the capture of low-frequency antigen-specific CD8⁺ T cells while limiting nonspecific binding.

For sensitivity analysis to demonstrate the robustness of TetTCR-SeqHD, we set the negative threshold of tetramer MID to 15 to capture positive binding events. This threshold was then used for all experiments.

Precision and recall rate calculation for TetTCR-SeqHD. In the TetTCR-SeqHD clone experiment, true positive is defined as antigen-matched TCRs between MIDCIRS and TetTCR-SeqHD. Predicted condition positive is defined as antigen-specific TCRs identified by pMHC DNA barcodes. The condition positive is defined as antigen-specific TCRs identified by MIDCIRS. Precision and recall are then calculated as follows:

$$\text{precision} = \frac{\sum \text{true positive}}{\sum \text{predicted condition positive}}$$

$$\text{recall} = \frac{\sum \text{true positive}}{\sum \text{condition positive}}$$

Prediction of pMHC class I binding. HLA-A02:01-bound T1D autoantigens were curated from the IEDB (www.iedb.org) database, while HLA-A01:01- and HLA-B08:01-bound T1D autoantigens were predicted using NetMHCpan 4.0 (ref. ⁴⁶). The half-maximum inhibitory concentration cutoff for HLA-A01:01 and HLA-B08:01 was 950 nM and 500 nM, respectively.

TCR clonality calculation. TCRs that have productive paired α and β chains were used to calculate TCR clonality, which is a score to characterize T cell expansion. Higher TCR clonality indicates that corresponding TCRs are more clonally expanded. If there is a singleton TCR, we define the TCR clonality as 0, while single TCR species with multiple copies have a TCR clonality of 1. For all other situations, the TCR clonality is defined using the following formula:

$$\text{TCR clonality} = 1 - \left(- \sum p_i \log_e p_i / \log_e N \right) \quad (i = 1 \text{ to } N)$$

Calculation of antigen-specific T cell frequency. The absolute frequency of antigen-specific T cells for antigen a_i in each donor was calculated as follows:

$$\text{Freq}(a_i) = \left(\frac{\text{number of } a_i \text{ specific CD8}^+ \text{ T cells}}{\text{total sorted CD8}^+ \text{ T cells}} \right) \times \left(\frac{\text{number of loaded cells on Rhapsody Chip}}{\text{number of recovered cells on Rhapsody Chip}} \right)$$

For cross-reactive cells, especially when cells are cross-reactive with more than two antigens in the antigen panel, one cell can be identified to bind a combination of antigen specificities by TetTCR-SeqHD. Each combination is a binding pattern. The frequency for cross-reactive antigen-specific T cells was calculated for each binding pattern.

TCR transduction. We generated TCR constructs as previously described¹⁸ and cloned them into an empty pCDH (System Biosciences) vector driven by the MSCV promoter. Lentivirus was generated using the Virapower (ThermoFisher Scientific) system and concentrated 10 times using an Amicon Ultra column. Freshly thawed CD8⁺ T cells from an HLA-A2-, HLA-B8 and HLA-A1-negative donor were stimulated with Immunocult (Stemcell Technologies) and incubated with the concentrated virus for 2–3 days. The cells were expanded for a minimum of 10 days and then assessed for murine TCR β chain expression.

Flow cytometry on transduced cells. Tetramer staining was performed as previously described¹⁸ with tetrameric MHC loaded with chemically synthesized peptides (Genscript). Briefly, the transduced cells and negative controls were stained with an anti-CD8a antibody (clone RPA-T8, Biolegend) before the addition of tetramer for 1 h on ice. Negative controls were established using nonspecific tetramer (HLA-A*02:01:HCVns3:1406-1415 – KLVALGINAV) and untransduced T cells from the same donor. Cross-TCR and cross-HLA negative controls were also included to assess the degree of nonspecific activity. After washing, the cells were stained with an anti-murine TCR β antibody (Biolegend) and 7-aminoactinomycin D before analysis on a BD Accuri.

T2 cells (generously provided by the Mark Davis lab) were pulsed with a chemically synthesized peptide (10 μM) for 2 h at 37 °C. The cells were then washed and incubated 1:1 with the transduced cells for 4 h at 37 °C. Negative controls were performed using nonspecific peptide (HCVns3:1406-1415) and cross-TCR nonspecific peptides (for example, EBV-BLMF1 peptide was used as a negative control for T1D antigen cross-reactive TCR, TCR51), while positive control was performed using PMA/ionomycin (Cell Stimulation Cocktail, Biolegend). During incubation, anti-CD107 α (Biolegend) antibody and monensin were added to detect and stabilize degranulation events. The assay was stopped via the addition of cold PBS and subsequent staining for CD107 α , CD8 α and murine TCR β (Biolegend). Cells were analyzed via a BD Accuri.

Detection of autoantibodies. The presence of anti-GAD, anti-IA2 and anti-Znt8 antibodies was determined via ELISA obtained from Kronus and performed according to the manufacturer's instructions. Whole, undiluted plasma was used in this assay. Absorbance was measured using a SpectraMax M3 plate reader, and analysis of the standard curve was performed in R using a cubic-spline fit. The antibody concentration for each sample was then interpolated, with all positive controls falling within the reported concentrations. Patients were reported as positive if the detectable antibody levels were in excess of 5 IU ml⁻¹, 7.5 U ml⁻¹ and 15 U ml⁻¹ for the anti-GAD, anti-IA2 and anti-Znt8 antibodies, respectively, according to the manufacturer's instructions.

Statistics and reproducibility. The relevant statistical test, sample size, replicate type and *P* values for each figure are found in the figure and/or corresponding figure legend.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All TCR and peptide information is in the supplementary tables. The accession number for raw sequencing data is [phs002441.v1.p1](https://dbGaP.ncbi.nlm.nih.gov/study/GEO/geoquery.cgi?acc=GSE113177) on dbGaP. Source data are provided with this paper.

Code availability

Custom analysis code is available on GitHub (<https://github.com/JiangLabSysImmune>).

References

- Shum, E. Y., Walczak, E. M., Chang, C. & Fan, H. C. in *Single Molecule and Single Cell Sequencing* (ed. Suzuki, Y.) 63–79 (Springer Singapore, 2019).
- Yu, W. et al. Clonal deletion prunes but does not eliminate self-specific alpha beta CD8(+) T lymphocytes. *Immunity* **42**, 929–941 (2015).
- Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**, 491–499 (2017).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
- Wendel, B. S. et al. Accurate immune repertoire sequencing reveals malaria infection driven antibody lineage diversification in young children. *Nat. Commun.* **8**, 531 (2017).
- Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
- Jurtz, V. et al. NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.* **199**, 3360–3368 (2017).

Acknowledgments

We thank the patients with T1D for donating blood samples to our study. We also thank anonymous blood donors and staff members at We Are Blood for sample collection. We thank P. Parker for assistance with blood sample purification. We thank S. Chizari for assistance with uploading sequencing data to dbGaP and source code to GitHub. This work was supported by National Institutes of Health grants S10OD020072 (N.J.), R33CA225539 (N.J.) and R56AG064801 (N.J.); National Science Foundation CAREER award 1653866 (N.J.); Welch Foundation grant F1785 (N.J.); the Robert J. Kleberg, Jr. and Helen C. Kleberg Foundation (N.J.); and the Chan Zuckerberg Initiative Neurodegeneration Challenge Network Ben Barres Early Career Acceleration Awards 191856 (N.J.). We would also like to acknowledge funding from the University of Texas at Austin Cockrell School of Engineering Fellowship (A.A.S.), Mario E. Ramirez Endowed Graduate Fellowship (A.A.S.) and the Harry and Rubye Gaston Graduate Scholarship (A.A.S.).

Author contributions

K.-Y.M. and N.J. conceived and designed the study. K.-Y.M. designed and developed the technology platform; K.-Y.M. and A.A.S. performed and analyzed data for the majority of experiments; K.-Y.M. developed the pipeline to analyze tetramer DNA-barcode data; C.H. and K.-Y.M. developed the script for analyzing TCR sequence data; A.A.S., A.X., E.C., and Y.W.G. performed TCR transduction experiments; E.S. performed in vitro cell culture; A.A.S. and Y.W.G. performed tetramer staining and CD107 α experiments.

K.R.S. and M.K.-D. recruited patients with T1D and collected blood samples from them. R.B. provided help with Rhapsody-related experiments; K.-Y.M. and N.J. wrote the manuscript with help from all co-authors.

Competing interests

N.J. is a scientific advisor and holds equity interest in ImmuDX and Immune Arch, companies that are developing products related to the research reported. R.B. is an employee of Becton Dickinson, which provided some of the equipment and reagents used in the study. The remaining authors declare no competing interests.

Additional information

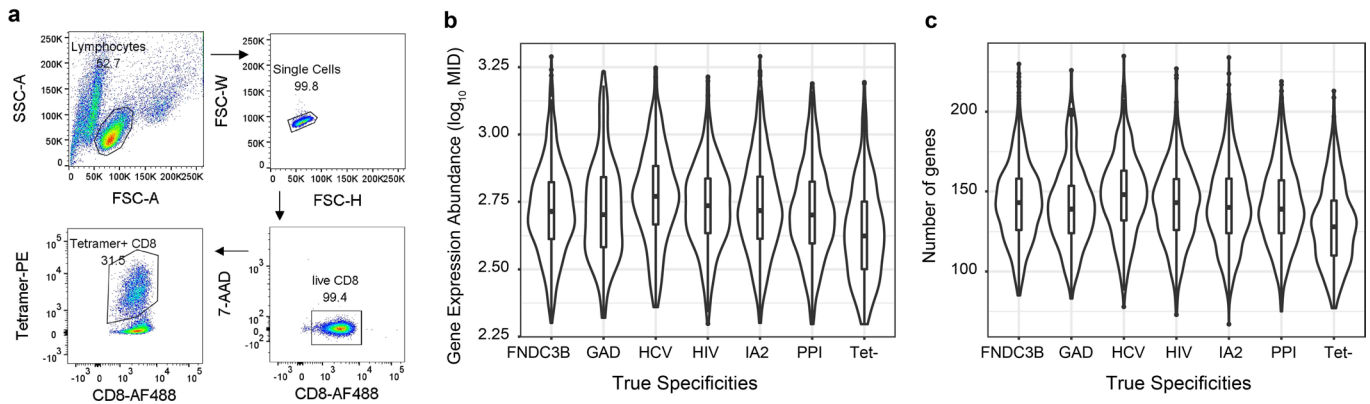
Extended data is available for this paper at <https://doi.org/10.1038/s41590-021-01073-2>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41590-021-01073-2>.

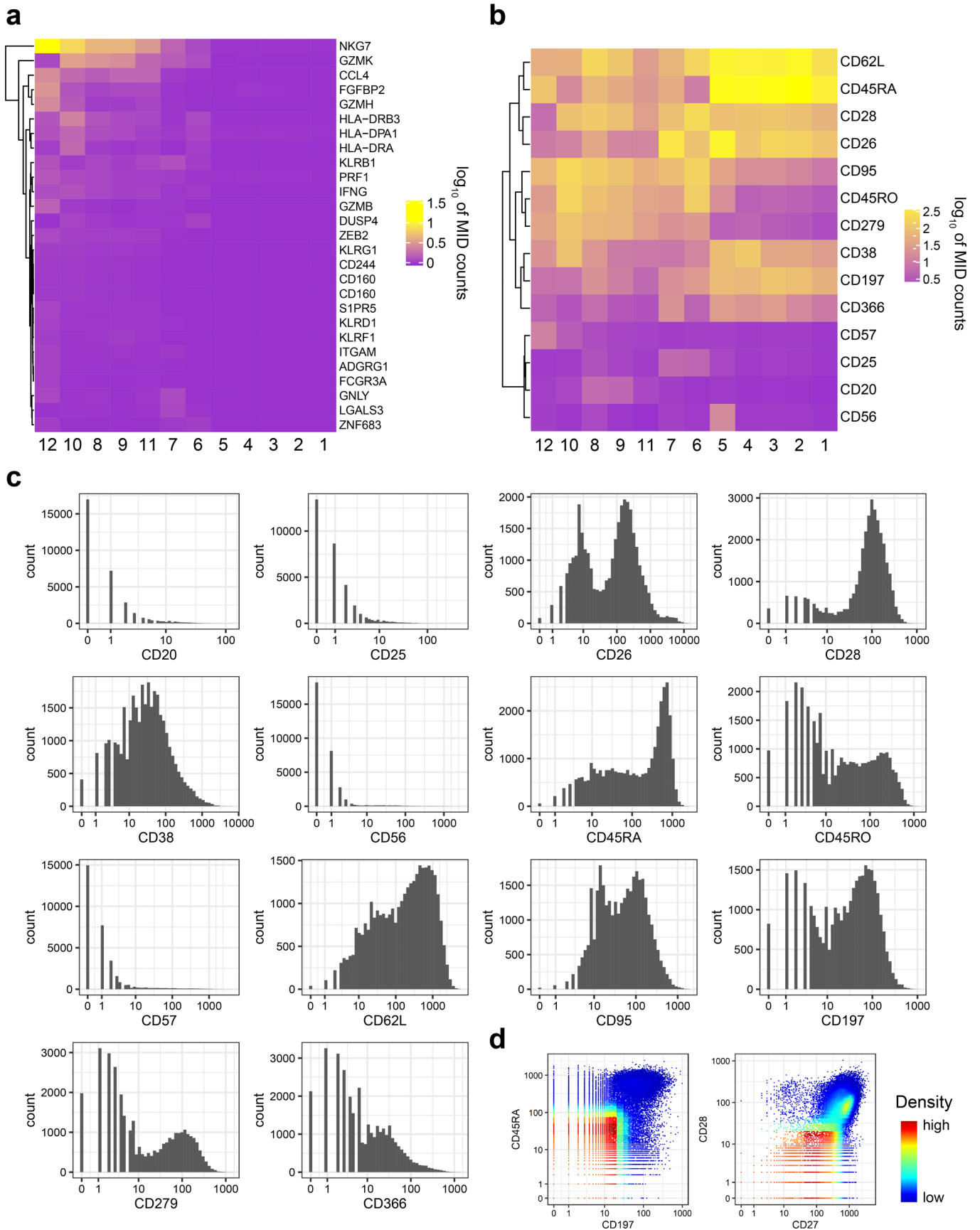
Correspondence and requests for materials should be addressed to Ning Jiang.

Peer review information *Nature Immunology* thanks Iwijn De Vlamincq, Angela Wu, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. L. A. Dempsey was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Tetramer-positive CD8⁺ T cells from a mixture of pMHC tetramer-sorted polyclonal T cells cultured in vitro and quality check of gene expression. a, Gating strategy for sorting tetramer-positive CD8⁺ T cells from a mixture of pMHC tetramer sorted polyclonal T cells cultured in vitro. **b,c**, Distribution of mRNA counts (\log_{10}) (**b**) and number of detected genes per cell (**c**) among different antigen-specific T cell populations. Horizontal lines represent 25th percentile, median and 75th percentile values, with whiskers extending to the farthest data point within a maximum of 1.5 × interquartile range.

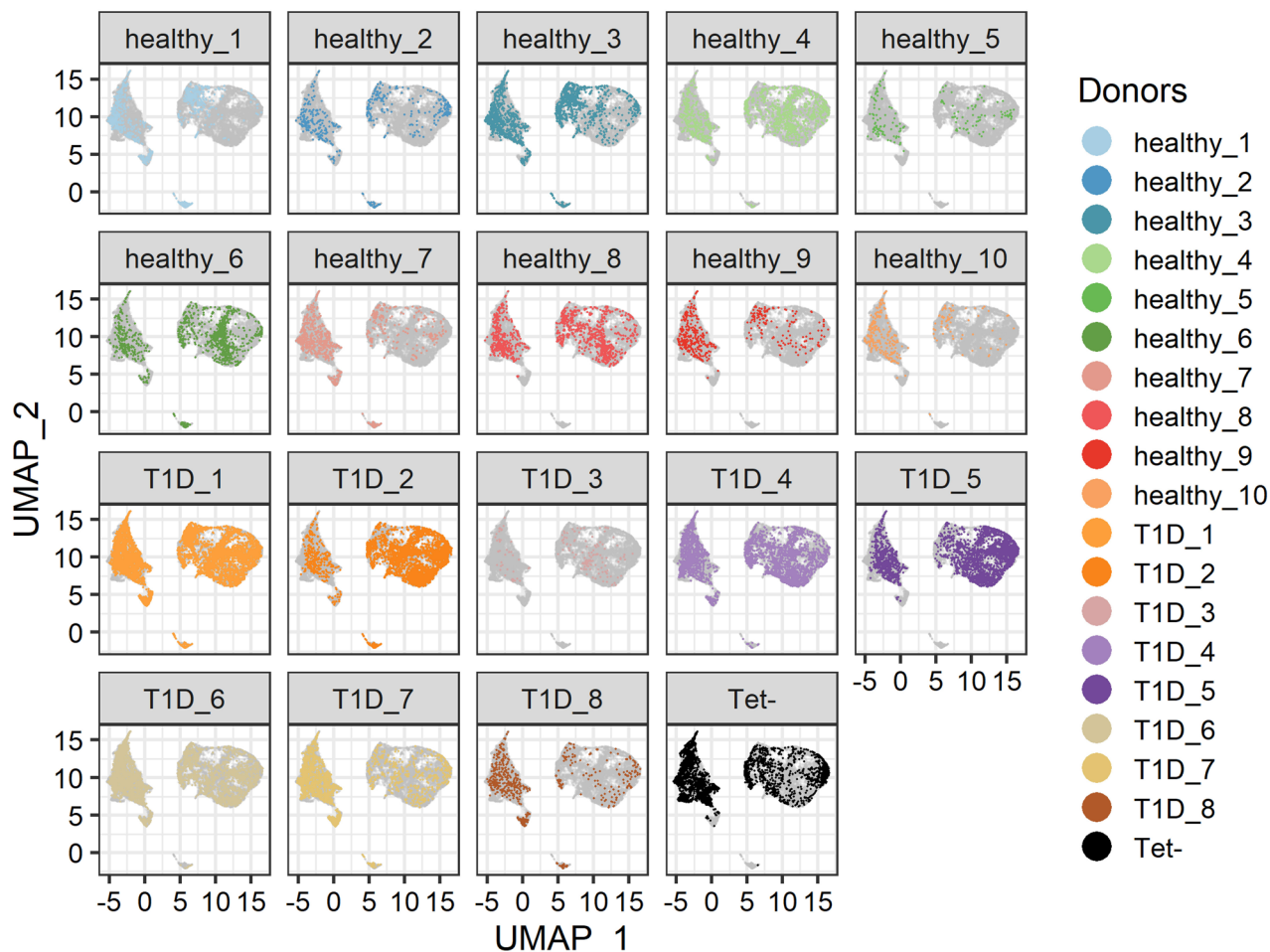


Extended Data Fig. 2 | See next page for caption.

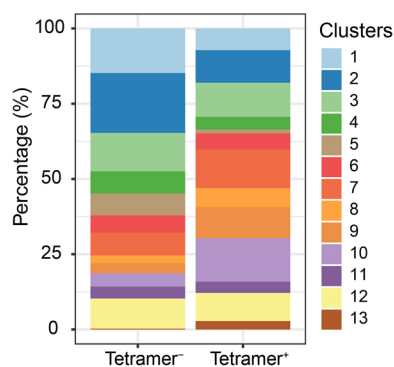
Extended Data Fig. 2 | Absolute expression (\log_{10} of MID counts) of differentially expressed genes and surface proteins among different clusters.

a,b. Genes and surface protein plotted here are the same set as in Fig. 3d. **c.** Distributions of AbSeq MID counts for differentially expressed surface proteins shown in Fig. 3d. **d.** Density plot showing of all CD8 T cells by AbSeq MID counts of CD45RA vs. CD197, and CD28 vs. CD27, respectively. Colors represent the local density of cells on the two-dimensional space.

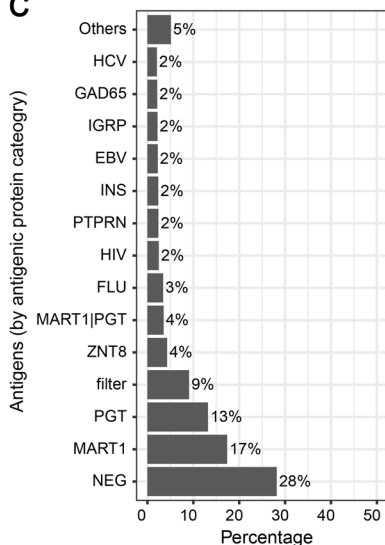
a



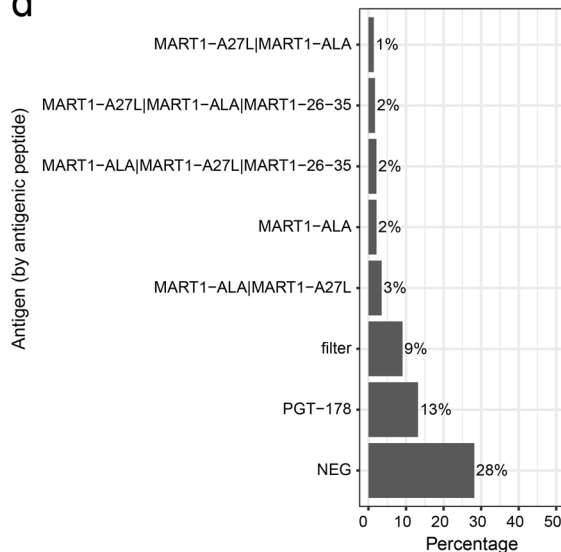
b



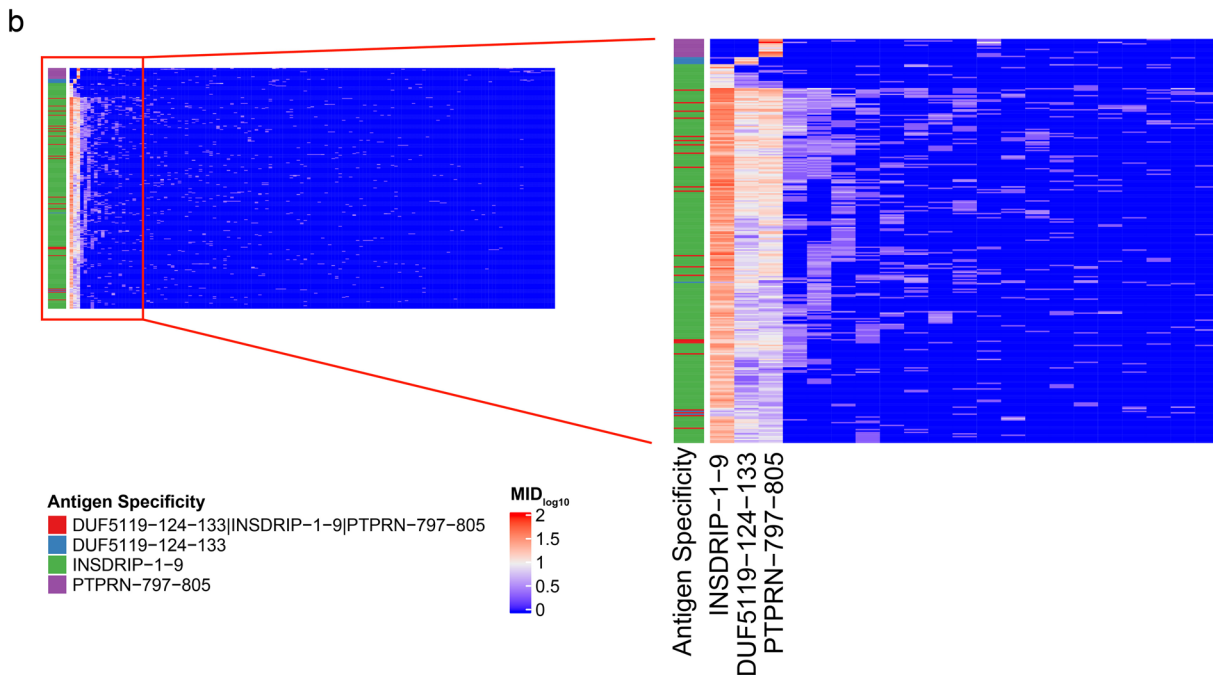
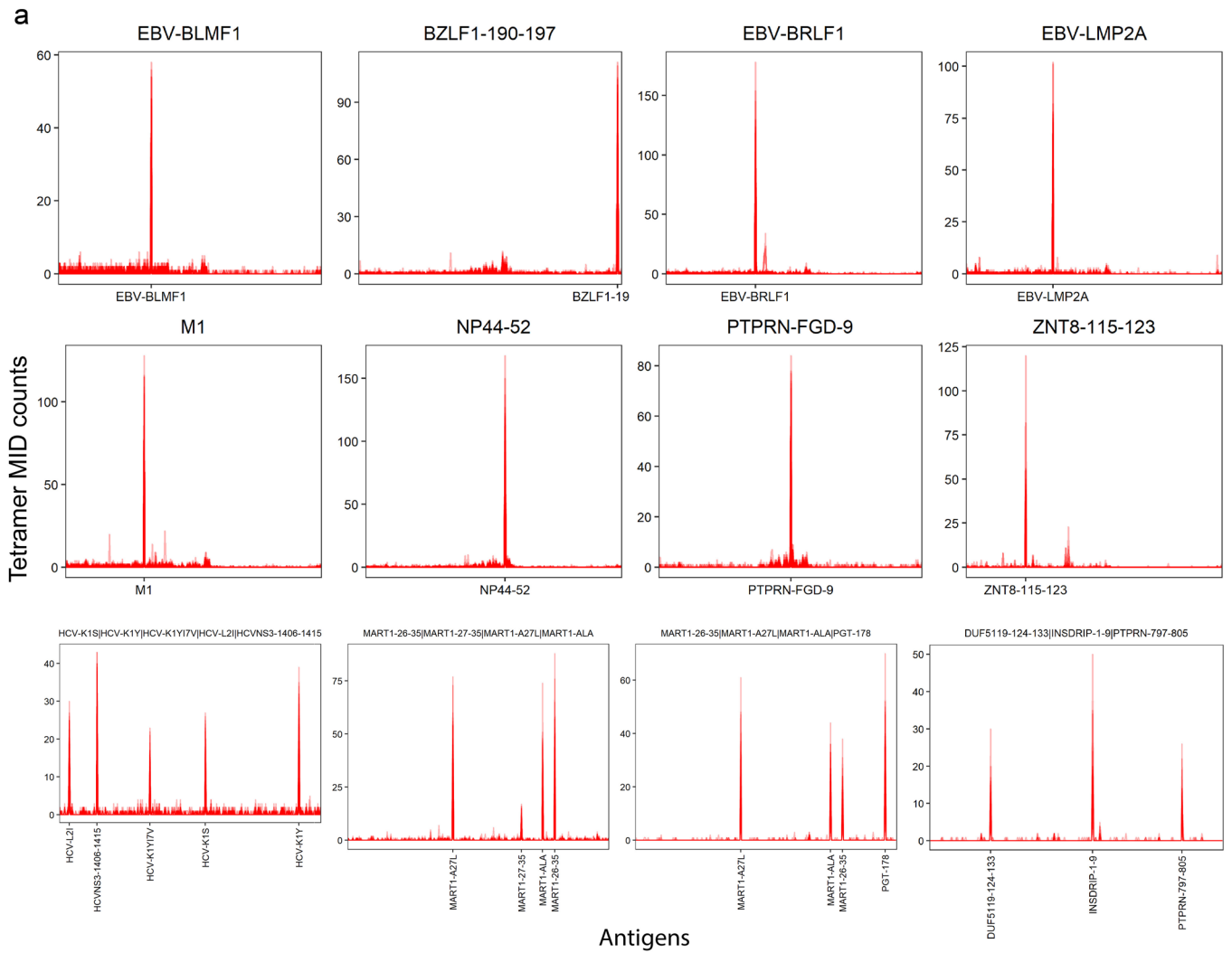
c



d

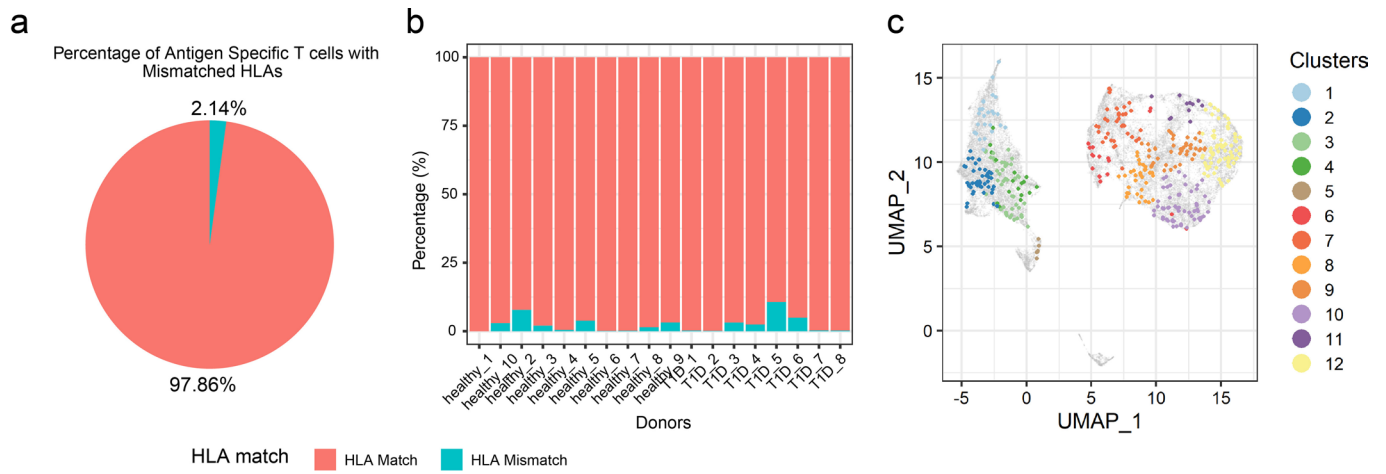


Extended Data Fig. 3 | TetTCR-SeqHD enables combined gene expression, phenotype and TCR clonality comparison among antigen-specific CD8⁺ T cells. **a**, UMAP of single cells among different donors. Grey dots represent all cells and colored dots are cells from different chips. **b**, Comparison of the distribution of phenotypes between tetramer-positive and tetramer-negative CD8⁺ T cells. **c**, Percentage of naive population in each antigen-specific CD8⁺ T cell group. **d**, Percentage breakdown of naive CD8⁺ T cells among all major antigen specificities. Only antigen specificities with a percentage greater than 1% within the naive population are shown. 'NEG' cells are tetramer-negative CD8⁺ T cells identified by tetramer MID counts. Cells were classified into filter category based on the following criteria: (1) more than one antigen binds to a single cell, and these antigens are a distance of more than 3 amino acids away from each other; (2) correlation of tetramer MID between single cells and the median of all cells with same TCR sequence is below 0.9, identified as described in Methods.

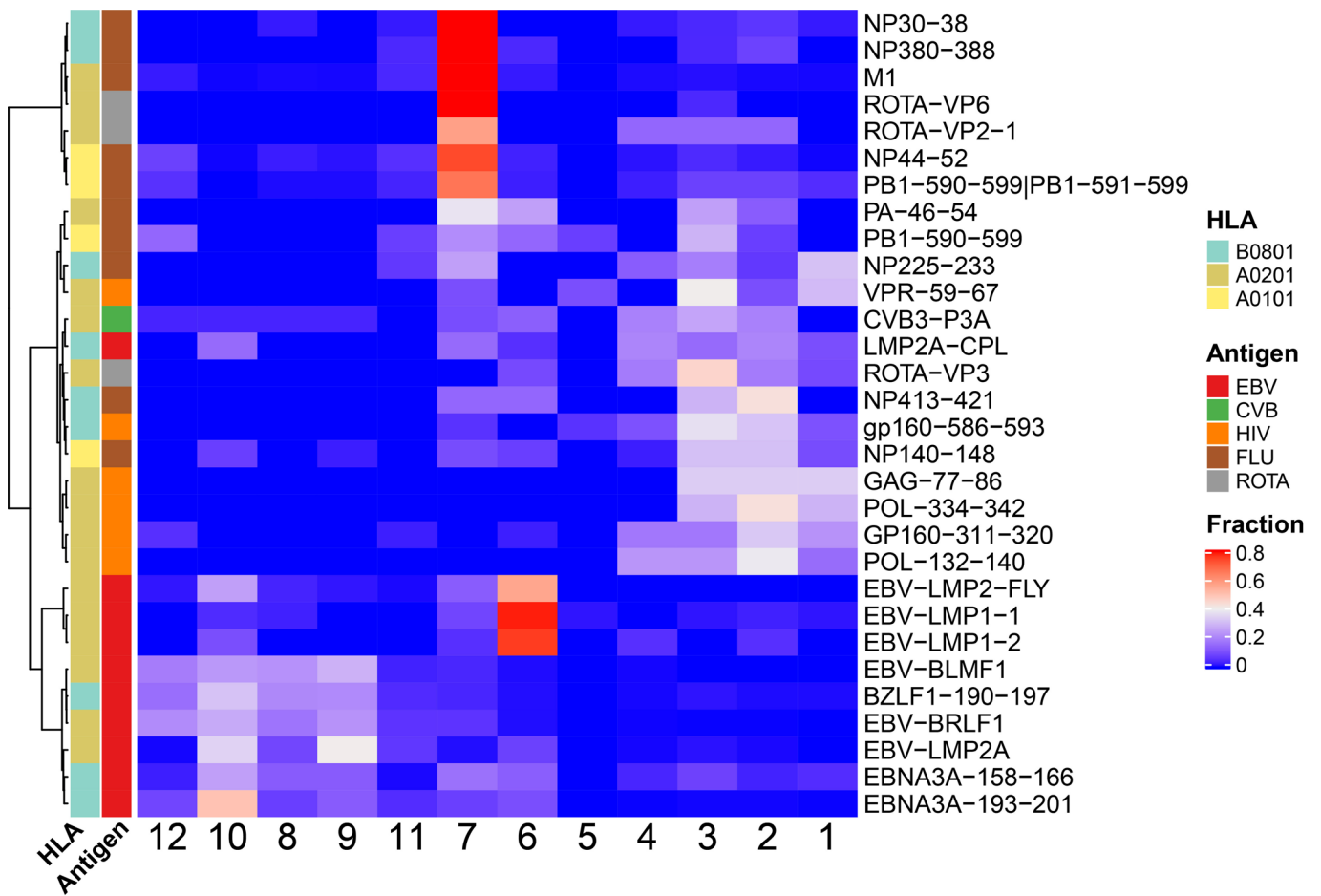


Extended Data Fig. 4 | See next page for caption.

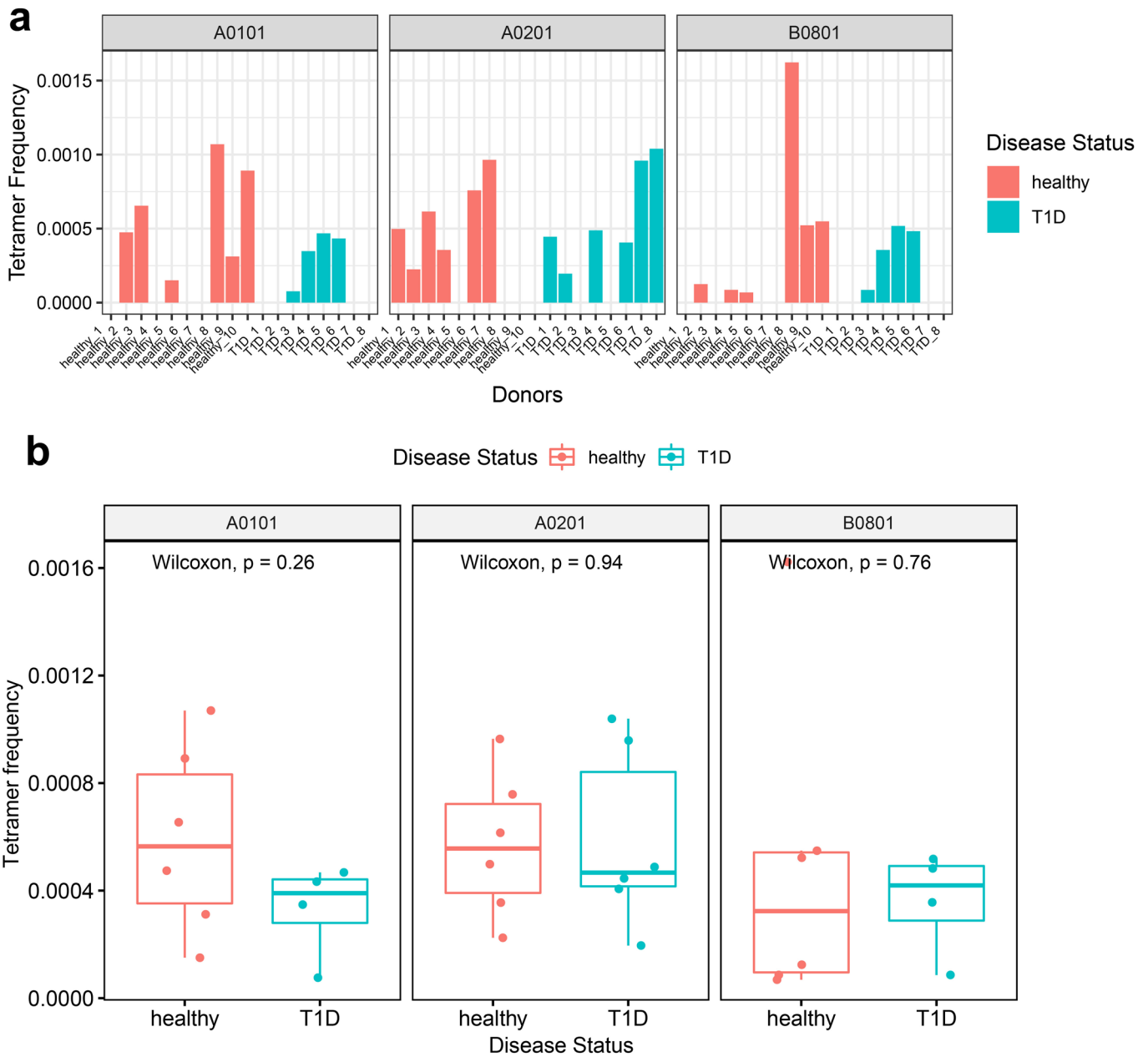
Extended Data Fig. 4 | Distribution of representative tetramer MID counts. **a**, Distribution of tetramer MID counts for eight antigen specificities, including EBV antigens (EBV-BLMF1, BZLF1-190-197, EBV-BRLF1 and EBV-LMP2A), influenza viral antigens (M1 and NP44-52), T1D-associated antigens (PTPRN-FGD-9 and ZNT8-115-123) or cross-reactive antigens (HCV, Mart1 and DUF5119-124-133/INSDRIP-1-9/PTPRN-797-805). For each cell in the group, the MID counts for each of the 280 antigens used in the experiment were tallied and then overlaid in the same order of the 280 antigens. Only the antigens that emerge after the filter are labeled on the x axis, and their position in the 280 antigen list is indicated by a tick on the x axis. Each panel with a sharp single peak indicated single antigen specificity, while panels with multiple sharp peaks indicated cross-reactive antigens. **b**, Comparison of tetramer MID counts among DUF5119-124-133/INSDRIP-1-9/PTPRN-797-805 cross-reactive and single antigen-specific cells.



Extended Data Fig. 5 | Analysis of T cells with bound antigen specificity being mismatched HLA alleles. **a**, Summary of percentage of antigen-specific T cells with mismatched HLA alleles in all donors. Combined percentages from two sources are presented (Discussion). **b**, Percentage of antigen-specific T cells with mismatched HLA alleles in each donor. Combined percentages of two sources are presented (Discussion). **c**, Comparison of phenotypes of cells with mismatched HLA alleles with the overall population. Gray dots represent all CD8⁺ T cells.

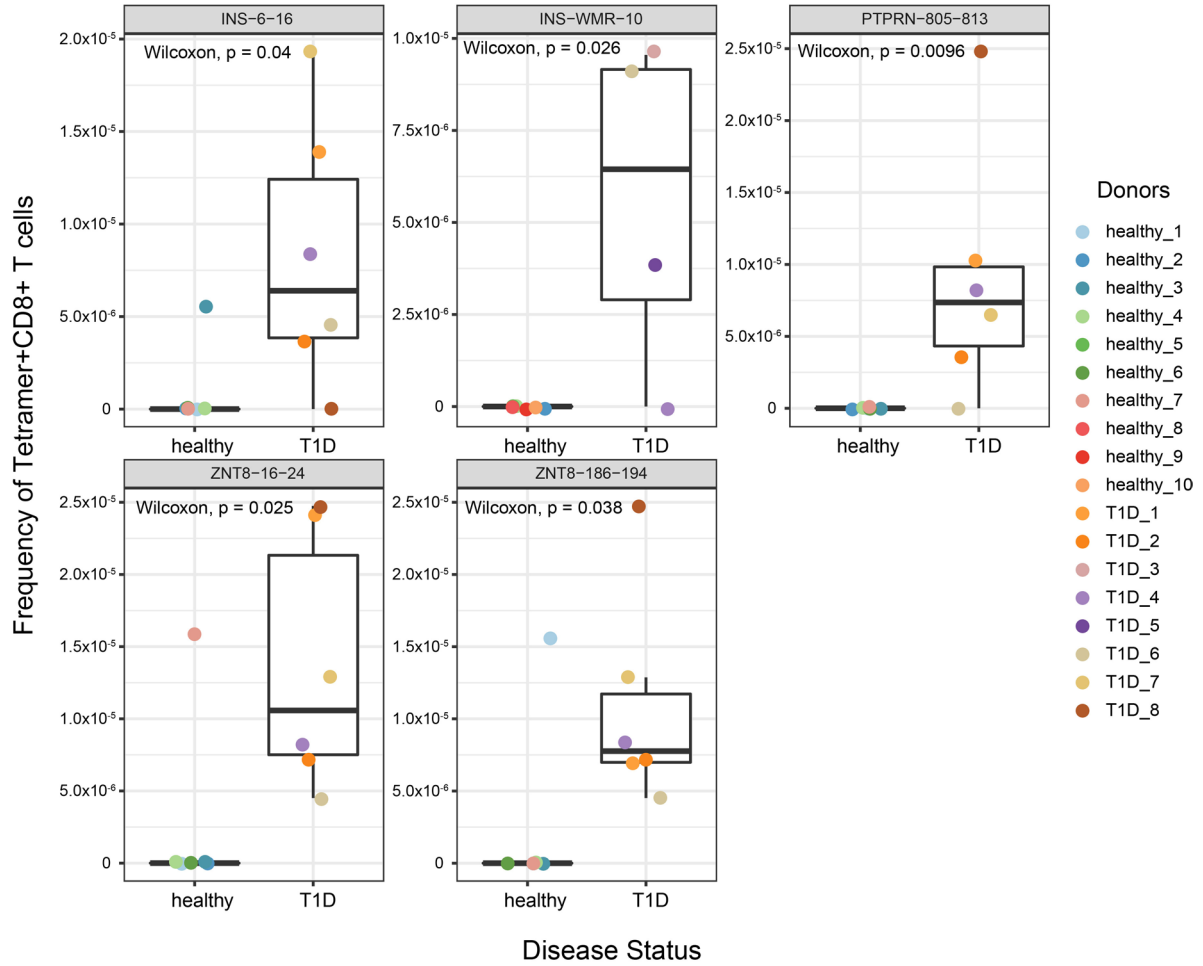


Extended Data Fig. 6 | Distribution of viral antigen-specific CD8⁺ T cells among 12 primary CD8⁺ T cells clusters in all 18 donors when the tetramer-negative MID threshold was set to 15. Distribution of viral antigen-specific CD8⁺ T cells among 12 primary CD8⁺ T cells clusters in all 18 donors when the tetramer-negative MID threshold was set to 15 (Methods).

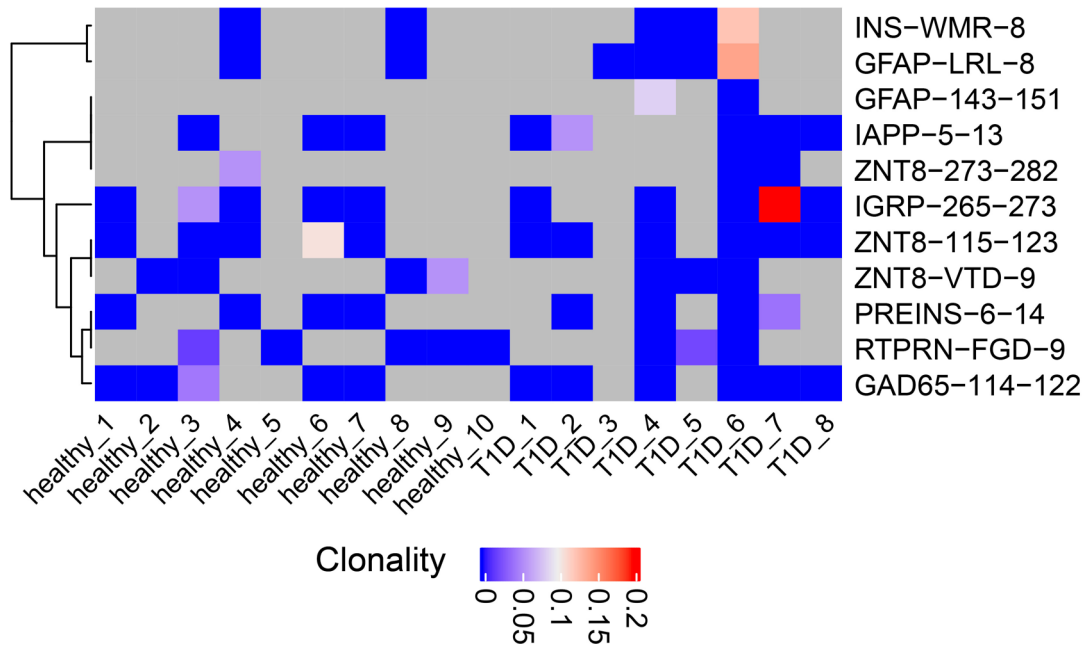


Extended Data Fig. 7 | Frequency of total T1D autoantigen-specific CD8⁺ T cells in healthy subjects and T1D patients. a, Frequency of T1D autoantigen tetramer-positive CD8⁺ T cells in different donors for various HLA alleles. **b**, Comparison of total T1D autoantigen tetramer-positive CD8⁺ T cells between healthy and T1D donors for various HLA alleles. A two-sided Wilcoxon nonparametric test was performed. Horizontal lines represent 25th percentile, median and 75th percentile values, with whiskers extending to the farthest data point within a maximum of 1.5× interquartile range. The number of subjects with HLA-A01:01, A02:01 and B08:01 are 10, 12 and 10, respectively.

a

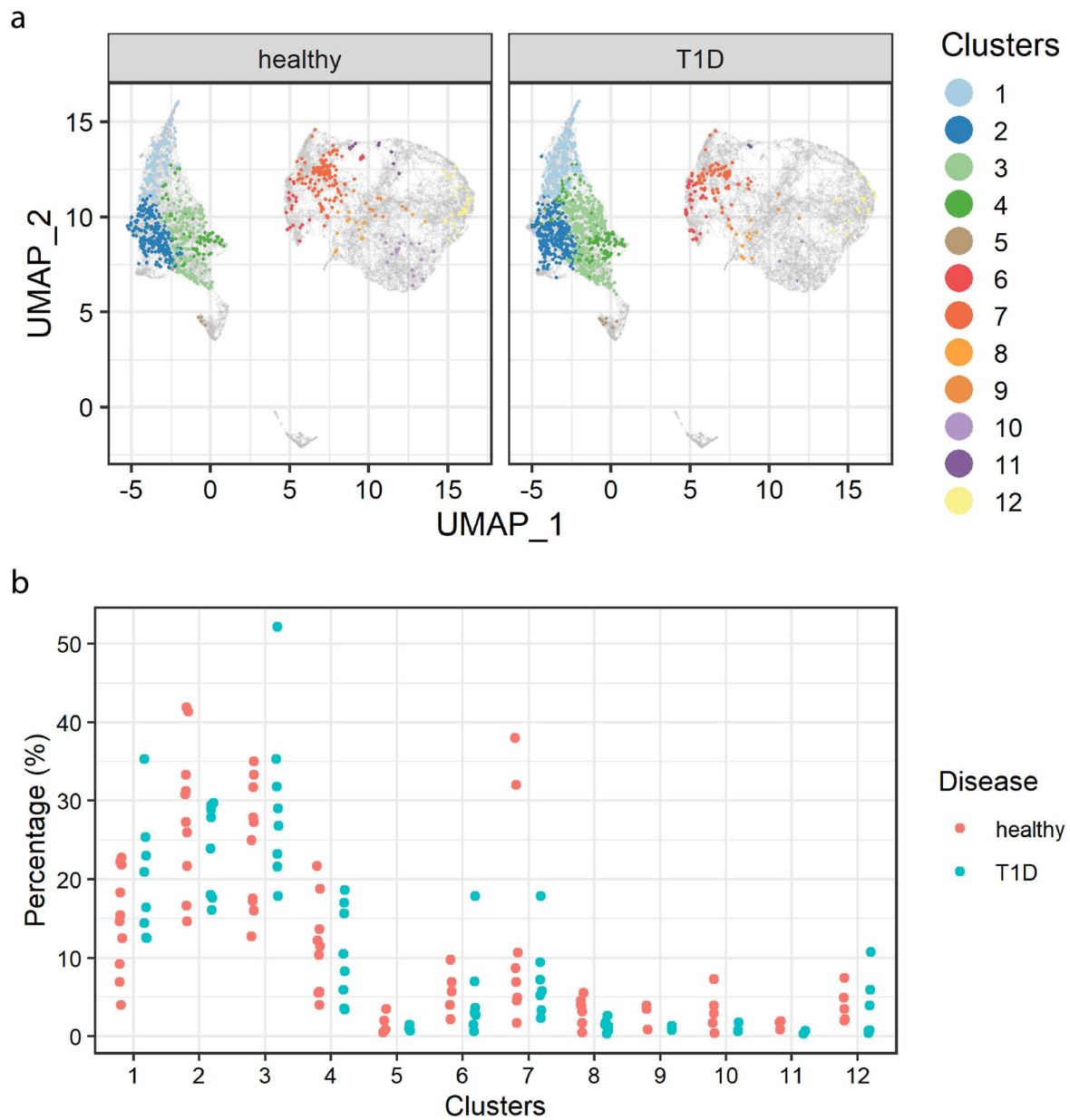


b

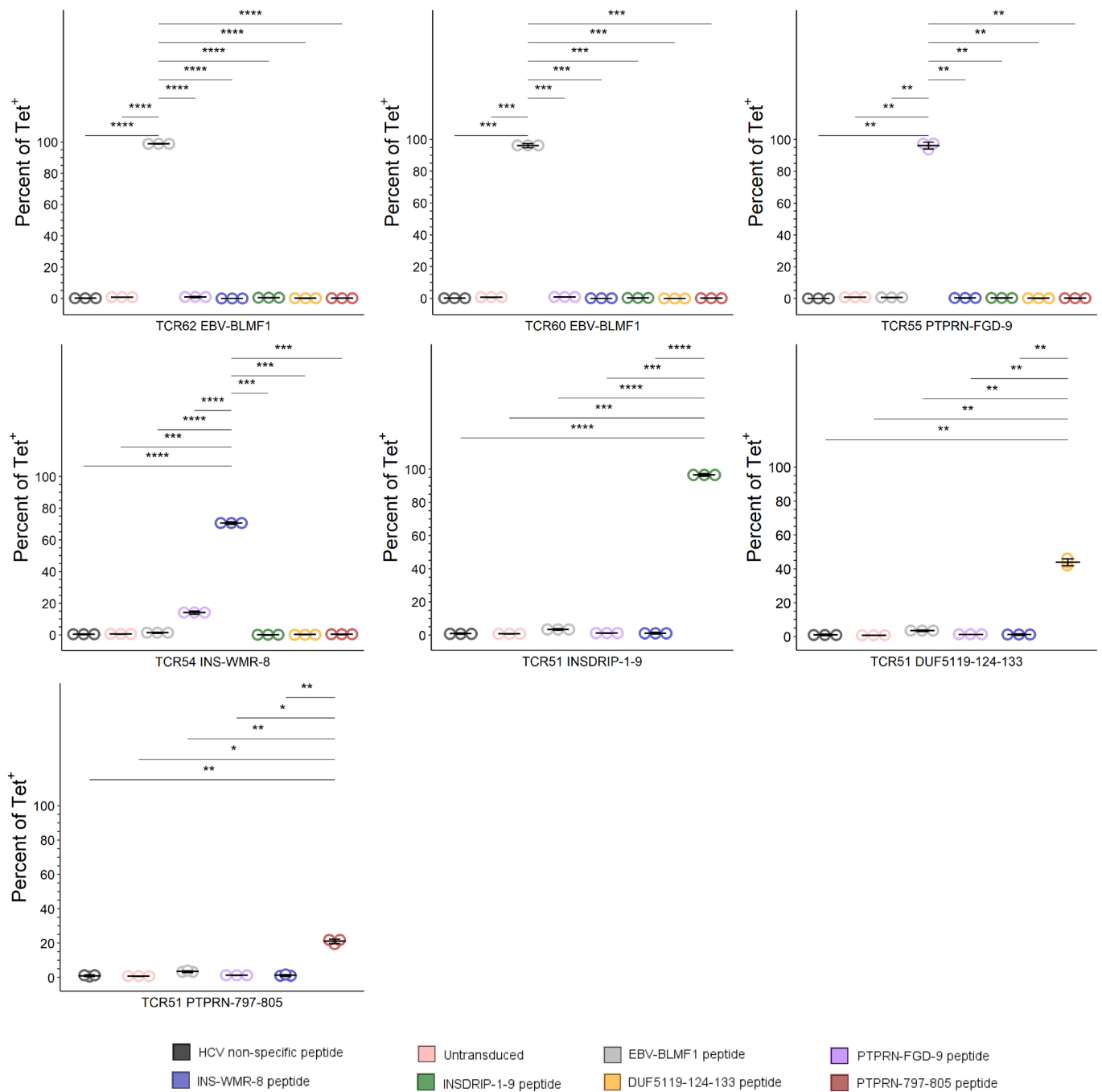


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | T1D autoantigens with different antigen-specific CD8⁺ T cell frequencies and clonality between healthy subjects and T1D patients. **a**, Five T1D autoantigens were identified to have a significantly higher frequency of antigen-specific T cells in peripheral blood when the MID-negative threshold was set to 15. A two-sided Wilcoxon nonparametric test was performed. Horizontal lines represent 25th percentile, median and 75th percentile values, with whiskers extending to the farthest data point within a maximum of 1.5 × interquartile range. The number of subjects with HLA-A01:01, A02:01 and B08:01 are 10, 12 and 10, respectively. **b**, TCR clonality heatmap of T1D antigenic-specific T cells for each antigen/donor combination. Grey, no T cells were detected.



Extended Data Fig. 9 | Comparison of T1D antigen-specific CD8⁺ T cells between T1D patients and healthy subjects. a, UMAP of T1D antigen-specific CD8⁺ T cells in T1D patients and healthy subjects respectively. Colored dots are T1D antigen-specific CD8⁺ T cells, and gray dots are other cells. **b**, Comparison of the distribution of phenotypes among T1D antigen-specific CD8⁺ T cells in each donor. Wilcoxon test was performed, with no significance between T1D and healthy subjects in any cluster.



Extended Data Fig. 10 | TCR specificity and cross-reactivity validation by tetramer staining. Bar plot showing the percentage of tetramer-positive cells gated on TCR β^{hi} fraction of the cells, corresponding to Fig. 5b. Tetramer staining experiments were performed in triplicate. A two-tailed Student's *t* test was performed between cognate tetramer and each negative control for all TCRs. EBV-BLMF1: GLCTLVAML; INSDRIP-1-9: MLYQHLLPL; DUF5119-124-133: MVWGPDPPLYV; PTPRN-797-805: MVWESGCTV; PTPRN-FGD-9: FGDHPGHSY; INS-WMR-8: WMRLPLL. ns, not significant; **P* ≤ 0.05; ***P* ≤ 0.01; ****P* ≤ 0.001; *****P* ≤ 0.0001.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Data was obtained through sequencing with HiSeq X PE150. HLA-A02:01 bound T1D autoantigens were curated from the IEDB (www.iedb.org) database. HLA-A01:01 and HLA-B08:01 bound T1D autoantigens were predicted using NetMHCpan 4.0.

Data analysis For target gene expression and AbSeq sequencing, reads were processed with BD Targeted Multiplex Rhapsody Analysis Pipeline Version 1.5 on Seven Bridges platform following manufacturer's instructions. Sequencing data of Tetramer, TCR α and TCR β were processed using umitools to extract cellular barcode and unique molecular identifier (MID) for each read. For the SampleTag DNA-barcodes, reads were mapped to SampleTag DNA-barcode reference using bowtie2 (version 2.3.4.2) with --norc and --local mode. Aligned reads were then processed using umi_tools (version 1.0.0) to count the number of MIDs for each SampleTag DNA-barcode in each cell. RNA expression and surface protein measurement were clustered with totalVI (version 0.6.6) and scanpy (version 1.5). R 3.6 is used for post-processing analysis. Flowjo V10 were used for flow cytometry analysis. Custom analysis code is available on Github (<https://github.com/JiangLabSysImmune>).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All TCR and peptide information are in the Supplementary Tables. The accession number for raw sequencing data is phs002441.v1.p1 on dbGaP. HLA-A02:01 bound T1D autoantigens were curated from the IEDB (www.iedb.org) database.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We typed the HLA alleles for all the T1D/T2D patient samples that were available, and then chose the donors with HLA alleles being HLA-A02:01, HLA-A01:01 and HLA-B08:01. Healthy were chosen based on HLA alleles being HLA-A02:01, HLA-01:01 and HLA-B08:01 and matching age with T1D/T2D patients.
Data exclusions	All single cells were first filtered to exclude low quality cells whose total gene and AbSeq expression MID counts were in the last 1% quantile. Then cells identified as multiplets with SampleTag and cells with two productive TCR β chains were also removed. Additionally, genes or AbSeqs whose expression were detected in less than 50 cells were filtered. Details are described in the Method section. This exclusion criteria were applied for all experiments and predefined.
Replication	The tetramer staining and functional assay were replicated three times and all attempts at replication were successful
Randomization	Not relevant to our study because subjects were not allocated into experimental groups
Blinding	There was no blinding applied in this experiment. We specifically aim to recruit T1D patients, and compare with healthy individual

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

BV510 anti-human CD8a antibody (clone: RPA-T8, Biolegend Cat# 301048, 5uL per million cells in 100uL)
 mouse anti-PE (clone: PE001, Biolegend Cat# 408101, 1uL per 100uL staining)
 mouse anti-APC (clone: APC003, Biolegend Cat# 408001, 1uL per 100uL staining)
 AF488-anti-human CD4 (clone: RPA-T4, Biolegend Cat# 300519, 2uL per million cells in 100uL)
 AF488-anti-human CD14 (clone: 63D3, Biolegend Cat# 367130, 2uL per million cells in 100uL)
 AF488-anti-human CD19 (clone: HIB19, Biolegend Cat# 363038, 2uL per million cells in 100uL)
 BV421 anti-CD8a antibody (clone: RPA-T8, Biolegend Cat# 344748, 5uL per million cells in 100uL)
 APC anti-murine TCR β antibody (clone: H57-597, Biolegend Cat# 109212, 1uL per million cells in 100uL)

PE/Cyanine5 anti-murine TCR β antibody (clone: H57-597, Biolegend Cat# 109209, 1ul per million cells in 100uL)
 APC anti-human CD107 α (clone: H4A3, Biolegend Cat# 328620, 5ul per million cells in 100uL)
 Anti-human CD2 (Clone RPA-2.10, Biolegend Cat # 300202, after customized conjugation according to method section, 1ul per million cells in 100uL)
 Antibodies used for AbSeq were provided by BD and the list of AbSeq used is listed in supplementary Table 9.

Validation

All antibodies used for staining were supplied by Biolegend or BD. Each lot of the antibody has been quality control tested by immunofluorescent staining with flow cytometric analysis, which can be obtained from Technical Data Sheet for each antibody on the manufacturer's website. For AbSeq supplied from BD, the production process underwent stringent testing and validation to assure that it generates a high-quality conjugate with consistent performance and specific binding activity.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

T2 were generously provided by Mark Davis's lab from Stanford
 Other T cell clones used in the clone experiment are in vitro expanded primary T cells from healthy de-identified human blood donors.

Authentication

All T cell clones and lines were authenticated by antibody staining with anti-CD8. T2 cell lines were verified by anti-HLA-A2 antibody staining.

Mycoplasma contamination

cells lines were not tested for mycoplasma contamination

Commonly misidentified lines
(See [ICLAC](#) register)

No commonly misidentified cell lines were used in this study

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

T1D_1, T1D, Male, 23yrs, HLA-A0201
 T1D_2, T1D, Male, 40yrs, HLA-A0201
 T1D_3, T1D, Male, 27yrs, HLA-A0101, HLA-B0801
 healthy_1, healthy, Male, 26yrs, HLA-A0201
 healthy_2, healthy, Female, 34yrs, HLA-A0101, HLA-A0201, HLA-B0801
 T1D_4, T1D, Female, 27yrs, HLA-A0101, HLA-A0201, HLA-B0801
 T1D_5, T1D, Female, 21yrs, HLA-A0101, HLA-B0801
 T1D_6, T1D, Male, 48yrs, HLA-A0101, HLA-A0201, HLA-B0801
 healthy_3, healthy, Female, 46yrs, HLA-A0101, HLA-A0201
 healthy_4, healthy, Male, 59yrs, HLA-A0201, HLA-B0801
 healthy_5, healthy, Male, 27yrs, HLA-A0101, HLA-B0801
 T1D_7, T1D, Male, 38yrs, HLA-A0201
 T1D_8, T1D, Male, 21yrs, HLA-A0201
 healthy_6, healthy, Male, 37yrs, HLA-A0201
 healthy_7, healthy, Female, 27yrs, HLA-A0201
 healthy_8, healthy, Female, 25yrs, HLA-A0101, HLA-B0801
 healthy_9, healthy, Female, 19yrs, HLA-A0101, HLA-B0801
 healthy_10, healthy, Female, 44yrs, HLA-A0101, HLA-B0801
 T2D_1, T2D and later diagnosed with Latent Autoimmune Diabetes in Adults, Male, 50yrs, HLA-A0201
 Information of these donors are also available in Supplementary Table 5.

Recruitment

At the time, a set of T1D and T2D patients seeking medical care through Seton Family of Hospitals at Austin were recruited into this study with informed consent. There was no other selection bias. Human whole blood from diagnosed T1D and T2D patients were obtained at Seton Family of Hospitals at Austin with informed consent. Human peripheral blood mononuclear cell (PBMC) from healthy donors were purchased from ePBMC.

Ethics oversight

The use of whole blood from these patients was approved by the Institutional Review Board of the Ascension Seton University Physicians Group and is compliant with all relevant ethical regulations.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

CD8+ T cells were then enriched from PBMC of T1DM and healthy donors using EasySep™ Human CD8+ T cell isolation kit (STEMCELL). CD8+ T cells were resuspended in FACS buffer containing 0.05% sodium azide and 50nM of Dasatinib. CD8+ T cells were then incubated at 37C for 30min-60min. Following the Dasatinib treatment, tetramer pool together with anti-CD8a antibody (clone: RPA-T8, Biolegend) was directly added into the cells. Cells were incubated at 4C for 1hr with continuous rotation. After washing, cells were further stained at 4C for 20min with the presence of 5 µg/ml mouse anti-PE (clone: PE001, Biolegend) and/or mouse anti-APC (clone: APC003, Biolegend). AbSeq staining mastermix was prepared by pooling 1µl of each AbSeq together. Cells were washed in FACS buffer once and stained with the AbSeq mastermix. Additional dump-channel antibodies (AF488-anti-CD4, AF488-anti-CD14 and AF488-anti-CD19), 7-AAD and 2µl of anti-CD50 SampleTag were mixed in cells. Cells were incubated at 4C for 40mins, prior to washing in FACS buffer twice and proceeded for sorting.

Instrument

FACSAria II Cell sorter and FACSAria Fusion SORP Cell Sorter

Software

flowjo V10 were used for flow cytometry analysis

Cell population abundance

Abundance of tetramer+ CD8 T cells sorted:
T1D_1: 14603; T1D_2: 9289; T1D_3: 522; healthy_1: 3361; healthy_2: 1677; T1D_4: 8293; T1D_5: 9715; T1D_6: 8615; healthy_3: 5911; healthy_4: 4160; healthy_5: 673; T1D_7: 5219; T1D_8: 1547; healthy_6: 4029; healthy_7: 2295; healthy_8: 2941; healthy_9: 1398; healthy_10: 759; T2D_1: 2566

Gating strategy

Cells were first gated on lymphocyte based on FSC and SSC. Then singlets were gated based on FSC-H and FSC-W. CD8+ T cells were gated by dump channel and anti-CD8 antibody. Lastly, tetramer+ cells were gated by fluorescence-labeled tetramer (PE and APC).

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.