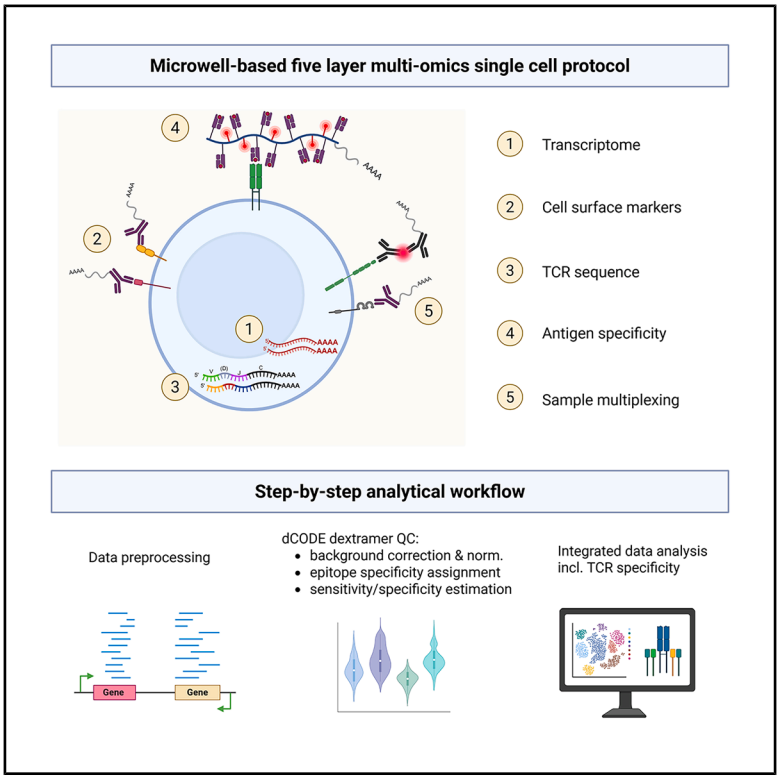
Cell Reports
Methods

Report

Characterization of human CMV-specific CD8+ T cells using multi-layer single-cell omics

Graphical abstract



Highlights

- Workflow for multi-omics single-cell data collection with the BD Rhapsody platform
- Integration of multiplexed transcriptome, TCR, antigen specificity, surface markers
- Two-tiered hashing enables flexible, high-throughput sample multiplexing
- Analysis pipeline for processing sequencing-based peptide-MHC multimer data

Authors

Ioanna Gemünd, Lorenzo Bonaguro, Matthias Becker, ..., Joachim L. Schultze, Andreas Moosmann, Marc D. Beyer

Correspondence

marc.beyer@dzne.de

In brief

Gemünd et al. present a multi-omics single-cell workflow that leverages sequencing-based dCODE dextramers to sensitively identify rare, antigen-specific T cells. This approach enables comprehensive characterization of human immune responses and provides a valuable resource for researchers studying T cell specificity and function in complex biological samples.

Gemünd et al., 2025, Cell Reports Methods 5, 101085
July 21, 2025
https://doi.org/10.1016/j.crmeth.2025.101085

CellPress

Cell Reports Methods

CellPress
OPEN ACCESS

Report

Characterization of human CMV-specific CD8+ T cells using multi-layer single-cell omics

Ioanna Gemünd,[1,2,3,4] Lorenzo Bonaguro,[1,5] Matthias Becker,[1,6] Sophie Müller,[1,2,3] Clemens Joos,[7,8] Elena De Domenico,[1,5] Anna C. Aschenbrenner,[1] Joachim L. Schultze,[1,2,5] Andreas Moosmann,[7,8,9] and Marc D. Beyer[1,5,10,11,*]

[1]Systems Medicine, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany
[2]Genomics and Immunoregulation, Life and Medical Sciences (LIMES) Institute, University of Bonn, Bonn, Germany
[3]Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, The University of Melbourne, Parkville, VIC, Australia
[4]Institute of Virology, Technical University of Munich School of Medicine/Helmholtz Munich, Munich, Germany
[5]PRECISE Platform for Single Cell Genomics and Epigenomics, DZNE, and University of Bonn and West German Genome Center, Bonn, Germany
[6]Modular High-Performance Computing and Artificial Intelligence, DZNE, Bonn, Germany
[7]Medizinische Klinik und Poliklinik III, LMU Klinikum, LMU München, Munich, Germany
[8]Helmholtz Munich, Munich, Germany
[9]German Center for Infection Research (DZIF), Munich Partner Site, Munich, Germany
[10]Immunogenomics & Neurodegeneration, Systems Medicine, DZNE, Bonn, Germany
[11]Lead contact
*Correspondence: marc.beyer@dzne.de
https://doi.org/10.1016/j.crmeth.2025.101085

MOTIVATION Single-cell multi-omics approaches are powerful tools to understand regulatory events in T cells; however, integrated use of the full portfolio of different modalities remains rare. Here, we report a highly standardizable workflow to generate a five-layer multi-omics dataset using the BD Rhapsody platform for the characterization of human antigen-specific T cells. Modalities include whole transcriptome, T cell receptor (TCR) sequences, T cell antigen specificity, surface marker proteins, and combinatorial sample multiplexing.

SUMMARY

In this study, we established a comprehensive workflow to collect multi-omics single-cell data using a commercially available micro-well-based platform. This included whole transcriptome, cell surface markers (targeted sequencing-based cell surface proteomics), T cell specificities, adaptive immune receptor repertoire (AIRR) profiles, and sample multiplexing. With this technique, we identified paired T cell receptor sequences for three prominent human CMV epitopes. In addition, we assessed the ability of dCODE dextramers to detect antigen-specific T cells at low frequencies by estimating sensitivities and specificities when used as reagents for single-cell multi-omics.

INTRODUCTION

Improving our understanding of immune responses requires sophisticated methods capable of capturing multiple layers of information simultaneously. The emergence of advanced single-cell multi-omics technologies has revolutionized our understanding of cellular heterogeneity and immunological processes.[1,2] These technologies are particularly valuable for the characterization of lymphocytes and their diverse adaptive immune receptor repertoires (AIRRs). Importantly, the integration of multiple layers not only facilitates the identification of antigen-specific lymphocyte clones but also sheds light on the intra- and interclonal phenotypic heterogeneity, enabling clonal tracking over time and across tissues.[3–5] This is important as T cell clones can exhibit varied functionalities based on their peptide specificities, major histocompatibility complex (MHC) restriction, tissue localization, and antigen exposure.[6–8]

Recent studies incorporating non-barcoded multimer strategies have not only revealed the extensive phenotypic diversity and clonal dynamics of antigen-specific T cells in infections such as influenza,[9,10] hepatitis C,[11,12] and COVID-19[13–16] but they have also set a benchmark for integrative immune profiling. By simultaneously linking transcriptional states, T cell receptor (TCR) sequence, and antigen specificity, these approaches

provide a comprehensive framework for understanding adaptive immune responses across diverse clinical settings. DNA barcode-labeled multimers were first used in bulk to screen >1,000 peptide specificities in parallel, enabling large-scale detection of antigen-specific $CD8^+$ T cells by sequencing-based readout.[17] Building on these advances, single-cell adaptations enabled the simultaneous capture of antigen specificity and paired TCR$\alpha\beta$ sequences,[18–21] with further developments allowing integration of transcriptome and surface protein expression profiling.[22] As a result, DNA-barcoded peptide-MHC (pMHC) multimer technologies provide a robust, high-throughput approach to study antigen-specific T cells in both research and clinical settings.

The identities and proportions of T cells specific for a certain antigen are hard to estimate. Factors such as the organ of origin as well as disease status and age of the donor add additional variability. In many cases, frequencies of antigen-specific T cells of interest will be low. Enrichment methods like fluorescence-activated cell sorting (FACS) with pMHC multimers are often required to improve the detection of low-abundance transcripts and proteins in rare antigen-specific cell populations and allow sequencing experiments to be conducted in a cost-effective manner. However, additional processing steps can be unfeasible and may even alter cellular states. More importantly, prior pMHC multimer enrichment might not fully capture the entire response to a given antigen.[23] This can result in the failure to notice responses to uncharacterized MHC-presented peptide epitopes and missing out on analyzing T cells in the context of their overall immune landscape. Therefore, barcoded pMHC multimer staining—either without subsequent FACS enrichment or with simultaneous analysis of both enriched and unenriched cells—enables comprehensive characterization of gene expression and TCR profiles for antigen-specific T cells with both known and unknown specificities for which potential reactivity can be computationally predicted.[24,25]

This raises the question of whether DNA-barcoded pMHC complexes, such as dCODE dextramers, can identify antigen-specific T cells at low frequencies. In this study we address this through a spike-in experiment using human cytomegalovirus (CMV)-specific T cells with known pMHC specificities and compare the performance of dCODE Dextramers when used as sequencing reagents against flow cytometry reagents. Furthermore, we present a workflow for collecting multi-omics data encompassing whole transcriptome, surface markers, AIRR, multiple antigen specificities, and combinatorial multiplexing using the BD Rhapsody platform.[26] Our approach integrates a two-tiered hashing strategy that enables flexible, high-throughput multiplexing. We utilize this technology to characterize TCR sequences of CMV-specific $CD8^+$ T cells. Additionally, we address technical aspects crucial for planning multi-omics experiments including DNA-barcoded pMHC multimers.

Taken together, he presented workflow offers significant potential for studying antigen-specific T cells in various biological contexts and provides a valuable resource for researchers using the BD Rhapsody platform or considering experiments with dCODE dextramers.

RESULTS

Single-cell multi-omics analysis allows for the delineation of antigen-specific T cells and background samples

T cells have diverse functionalities and phenotypes among different T cell clones, and single-cell multi-omics approaches enable comprehensive exploration of this diversity. Here, we provide a framework of how this can be technically achieved employing a five-layer BD Rhapsody multi-omics approach including whole-transcriptome analysis (WTA), TCR sequencing, antibody sequencing (AbSeq), dCODE dextramers, and dual sample multiplexing (Figures 1A and S1).

As a proof of concept to determine the ability to identify CMV-specific T cells at low frequency using dCODE dextramers, we conducted a spike-in experiment. As spike-in, a mixture of three samples enriched for CMV-specific $CD8^+$ T cells in a ratio of 1:1:1 was used. This included two previously described T cell clones[7,27] and one sample of peripheral blood mononuclear cells (PBMCs) 10 days after peptide stimulation. These samples are called "NLV clone," "CRV clone," and "TPR enriched" according to the CMV peptides they recognize: NLVPMVATV ($pp65_{495\text{-}503}$ on human leukocyte antigen [HLA]-A*02:01) for the NLV clone, CRVLCCYVL ($IE\text{-}1_{309\text{-}317}$ on HLA-C*07:02) for the CRV clone, and TPRVTGGGAM ($pp65_{417\text{-}426}$ on HLA-B*07:02) for the TPR-enriched sample, which was derived from PBMCs of a healthy donor (HLA-B*07:02$^+$) (Figure 1B; Table S1A). Correspondingly, we utilized three dCODE dextramers targeting the respective CMV epitopes: DexA (HLA-A*02:01/$pp65_{495\text{-}503}$), DexB (HLA-B*07:02/$pp65_{417\text{-}426}$), and DexC (HLA-C*07:02/$IE\text{-}1_{309\text{-}317}$) (Figure 1B). A negative control dextramer (DexN) carrying a nonsense peptide was also employed for background correction (HLA-B*08:01/AAKGRGAAL). Prior to the spike-in experiment, stainings with dCODE dextramers (phycoerythrin [PE] labeled) revealed that the NLV clone binds to DexA, with 97% of cells being DexA-PE$^+$ (Figures S2A and S2B). Similarly, 98% of the CRV clone cells were DexC-PE$^+$. The TPR-enriched sample contained about 16% DexB-PE$^+$ cells and 18% non-T cells (Figure S2B), reflecting that this sample originated from TPRVTGGGAM peptide-stimulated PBMCs.

The three-sample spike-in mix was compared against polyclonal $CD8^+$ T cells from the PBMCs of two healthy donors, in the following referred to as "background." Specifically, DexA was HLA matched with background donor 1 (BG_D1) (HLA-A02:01$^+$; HLA-B07:02$^-$ and HLA-C*07:02$^-$), while DexB and DexC were HLA matched with background donor 2 (BG_D2) (HLA-B07:02$^+$ and HLA-C*07:02$^+$; HLA-A02:01$^-$) (Figure 1B). To estimate the sensitivity and specificity of antigen-specific T cells captured with dCODE dextramers in our multi-omics approach, we needed a low-frequency spike-in condition. For this we chose to perform a 5% spike-in of the three-sample spike-in mix, meaning that each sample's actual spike-in frequency was about 1.67% (Table S1B). Additionally, we included a 50% spike-in plus 50% background condition to ensure sufficient cells for characterization of the spike-in mix (Figures 1B and S1A).

To reidentify all input populations (three CMV-enriched populations and two background donors), as well as four dCODE

CellPress
OPEN ACCESS

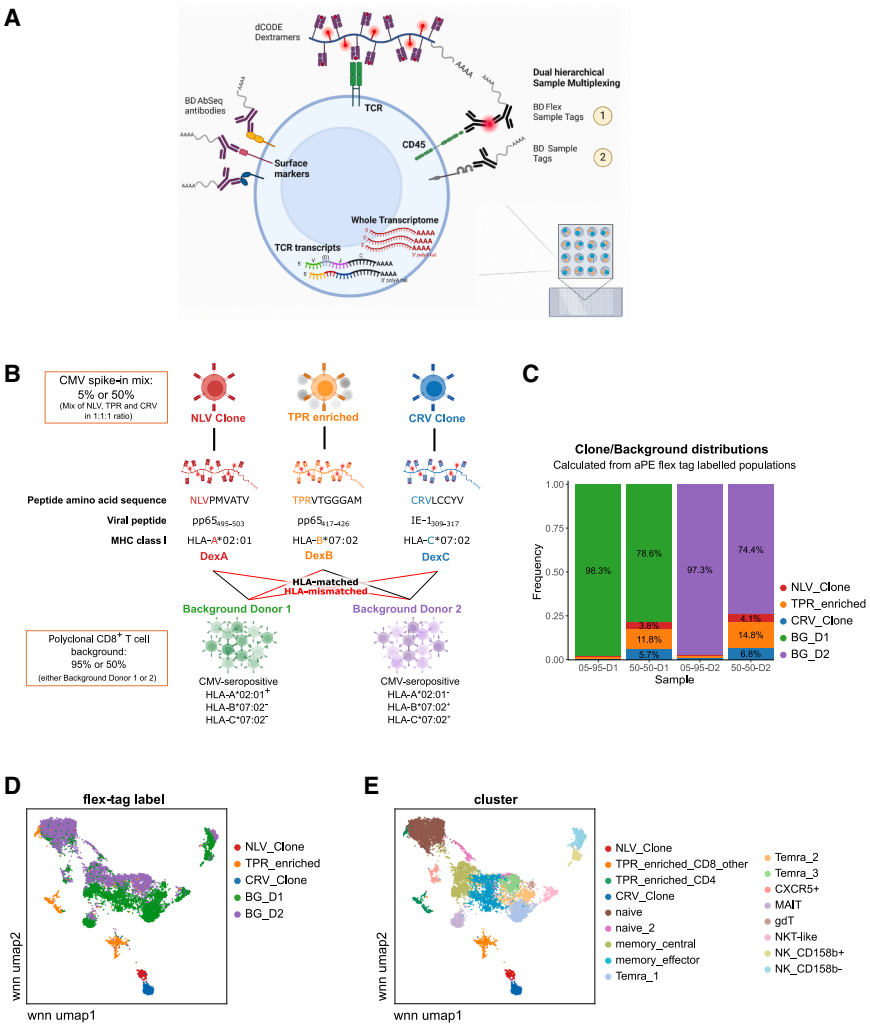


Figure 1. Single-cell multi-omics workflow allows delineation of spike-in and background samples

(A) Scheme of the different multi-omic modalities captured per cell.

(B) Schematic of the spike-in versus background experiment. The types of input material are depicted, and for the three CMV spike-in samples at the top, the corresponding CMV epitope as well as dCODE dextramer specificity are indicated by black lines. For polyclonal background samples, HLA-matched and -mismatched combinations with the dCODE dextramers are shown. HLA-matched and -mismatched combinations with dCODE dextramers are depicted by black and red lines, respectively.

(C) Distribution of spike-in and background cells in the four dCODE dextramer staining conditions.

(D and E) wnnUMAP representation of single-cell data with coloring according to flex-tag label (origin of five input populations) (D) and Leiden clustering (E).

See also Figures S1 and S2 and Table S1.

dextramer staining conditions (5% and 50% CMV spike-in mix against either BG_D1 or BG_D2), we employed a dual hierarchical indexing strategy. Individual input populations were labeled with anti-PE flex-tags binding PE-labeled anti-CD45 in the primary staining (Figure S1A). Subsequently, the four spike-in/donor mixes were labeled with regular Sample-tags (Figure S1A). This allowed for the simultaneous allocation of spike-in input samples and staining mixes and facilitated the estimation of dextramer performance within each setting.

We loaded 20,000 cells on 2 cartridges, performed the BD Rhapsody workflow, and proceeded to bioinformatic analysis after sequencing and preprocessing (see method details and Figure S1B). The first step of the analysis was assigning cells to their corresponding input material (flex-tag for BG_D1, BG_D2, NLV_Clone, TPR_enriched, or CRV_Clone) and staining mix (Sample-tag for 5:95 or 50:50 of spike-in/background with either background donor 1 or 2) (Table S1B). Cells were retained for analysis only if both tags were present, resulting in a dataset of 9,771 cells. After demultiplexing, the distribution of the 5 input samples showed a ratio of approximately 3%:30% spike-in and 97%:70% background cells (Figure 1C), indicating lower than expected spike-in frequencies likely due to lower viability of the clones after cell thawing.

We utilized uniform manifold approximation and projection (UMAP) to reduce data dimensionality.[28] A weighted nearest-neighbor (wnnUMAP) approach was chosen for multimodal UMAP embedding, integrating paired single-cell RNA sequencing (scRNA-seq and AbSeq information.[29] This approach allowed similar cells of the same subset or type to accumulate in specific UMAP clusters, indicating related transcriptome and surface marker expression. The CMV spike-in and background samples clustered apart in the wnnUMAP (Figure 1D), and the two background donors could also be visually distinguished. No technical batches from both cartridges were observed (Figure S2C). Cell clustering identified 17 clusters (Figure 1E), which we further characterized using transcriptome and surface protein expression (Figure 2). Background $CD8^+$ T cells clustered into four major states: naive, central memory (T_{CM}), effector memory (T_{EM}), and effector memory T cells re-expressing CD45RA (T_{EMRA}) (Figure 1E). Naive T cells were characterized by *CCR7* gene expression and surface markers CCR7, CD62L, and CD45RA (Figure 2). T_{CM} cells exhibited CD127, CD27, and CD28 expression, with intermediate levels of CCR7 and CD62L but no CD45RA (Figure 2A). T_{EM} cells showed high expression of PD-1 (CD279) and CXCR6 (Figure 2), while they lacked the surface markers CD62L, CCR7, and CD45RA. T_{EMRA} cells re-expressed CD45RA and lacked CD127 and CD27 (Figure 2A). Additional smaller clusters included MAIT cells, $\gamma\delta$ T cells, $CXCR5^+$ cells, natural killer (NK)T cells, and NK cell clusters (Figure 1E).

As outlined above, the NLV clone, TPR-enriched sample, and CRV clone clustered separately from the polyclonal background

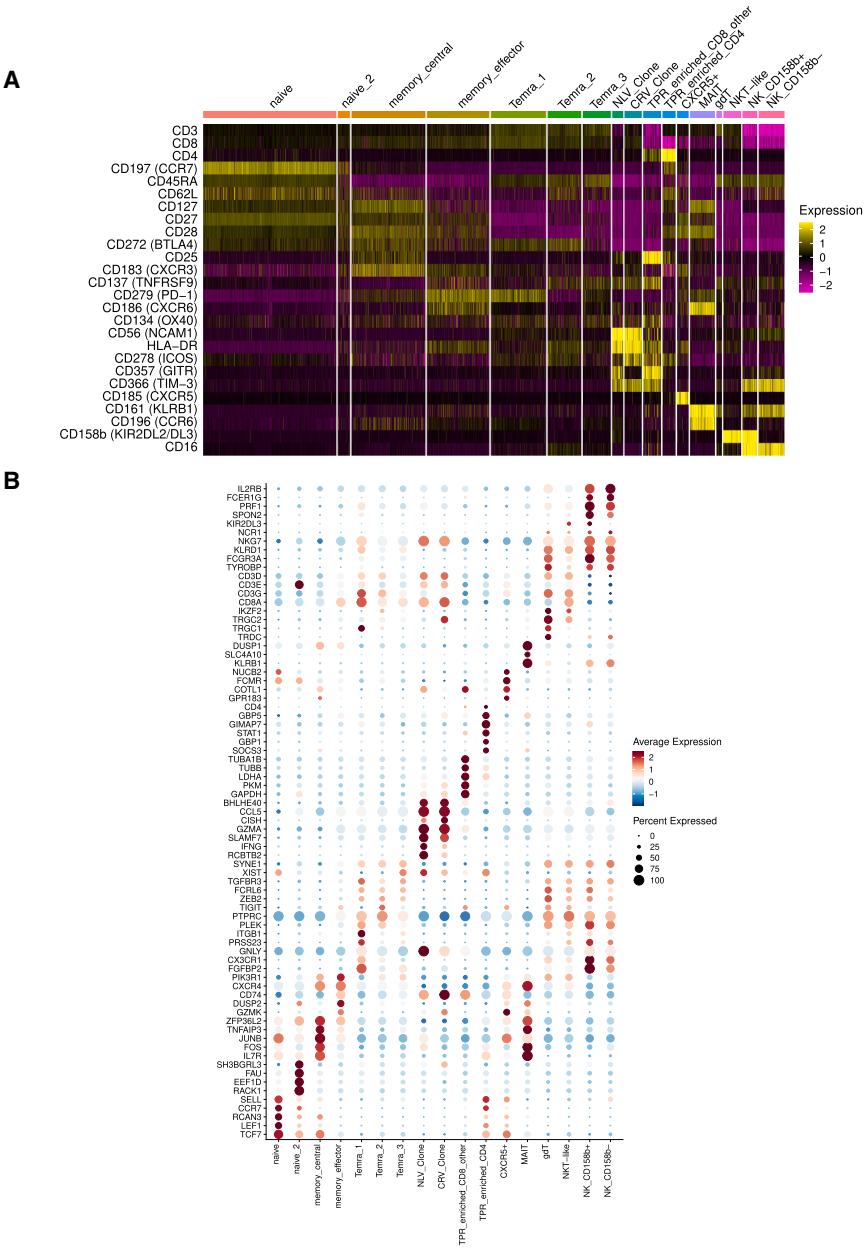


Figure 2. Cell cluster characterization by surface protein and marker gene expression

(A) Expression heatmap of 31 surface markers by cluster (colored by z-transformed expression).

(B) Dot plot displaying expression of the top five markers for each cluster, as well as additional selected T cell and NK cell gene markers. Dots are colored by z-transformed mean expression per cluster, dot size represents the frequency of expression within the cluster.

See also Figure S2.

samples (Figure 1D). Interestingly, during quality control, we detected a cluster containing a mixture of clone-derived and T_{EM} cells; however, we observed that this cluster exhibited high mitochondrial transcripts, low unique molecular identifier (UMI) and gene counts, and elevated *MALAT1* expression (data not shown). Additionally, despite dextramer positivity, TCR information could not be recovered from these cells. These findings suggest that the cluster likely consisted of cells undergoing activation-induced cell death (AICD), and it was therefore excluded from further analysis. The TPR-enriched sample appeared more heterogeneous and contained $CD4^+$ T cells (Figure 2). Both NLV and CRV clones, along with the TPR fraction containing $CD8^+$ T cells, exhibited high surface expression of activation markers HLA-DR and inducible co-stimulatory molecule (ICOS) (Figure 2A). CD56 protein and neural cell adhesion molecule 1 (NCAM1) transcript quantification confirmed that both CMV-specific clones upregulated CD56 relative to other clusters, with *NCAM1* transcript expression markedly enriched in the NLV clone (Figures 2A and S2D). The NLV and CRV clones were further distinguished by increased gene expression of $CD8^+$ T cell activation markers *IFNG*, *SLAMF7*, *GZMA*, and *GNLY* (Figure 2B). These observations support the notion that the CMV-specific clones were of $CD8^+$ T cell origin but were transcriptionally different from *ex vivo* isolated $CD8^+$ T cells, likely due to their *in vitro* cultivation and possibly influenced by the distinct activation and differentiation history of CMV-specific $CD8^+$ T cells *in vivo*.

Characterization of CMV-specific $CD8^+$ T cells

To identify dextramer-positive (Dex^+) cells, we analyzed the distribution of background-corrected normalized expression for each dextramer to establish an appropriate cutoff (Figure S3; see method details). The data showed a clear separation between Dex^{high} and Dex^{low} populations (Figures 3A and S3A). Cutoffs were set at local minima of the distribution, which exhibited a multimodal pattern. Thresholds for Dex^+ cells were automatically defined using k-means clustering (Figure S3A).

We then investigated which clusters contained Dex^+ cells (Figure 3A). Expectedly, the majority of the NLV clone was defined as $DexA^+$ and the majority of the CRV clone as $DexC^+$. Additionally, a fraction of the "TPR_enriched_CD8/other" cluster was $DexB^+$. Background donor 1 (BG_D1), which was HLA matched to DexA, displayed a distinct fraction of $DexA^+$ cells, whereas background donor 2 (BG_D2), HLA matched to DexB and DexC, showed a small cluster of $DexB^+$ and some interspersed $DexC^+$ cells (Figure 3B). Dex^+ cells in the background

CellPress
OPEN ACCESS

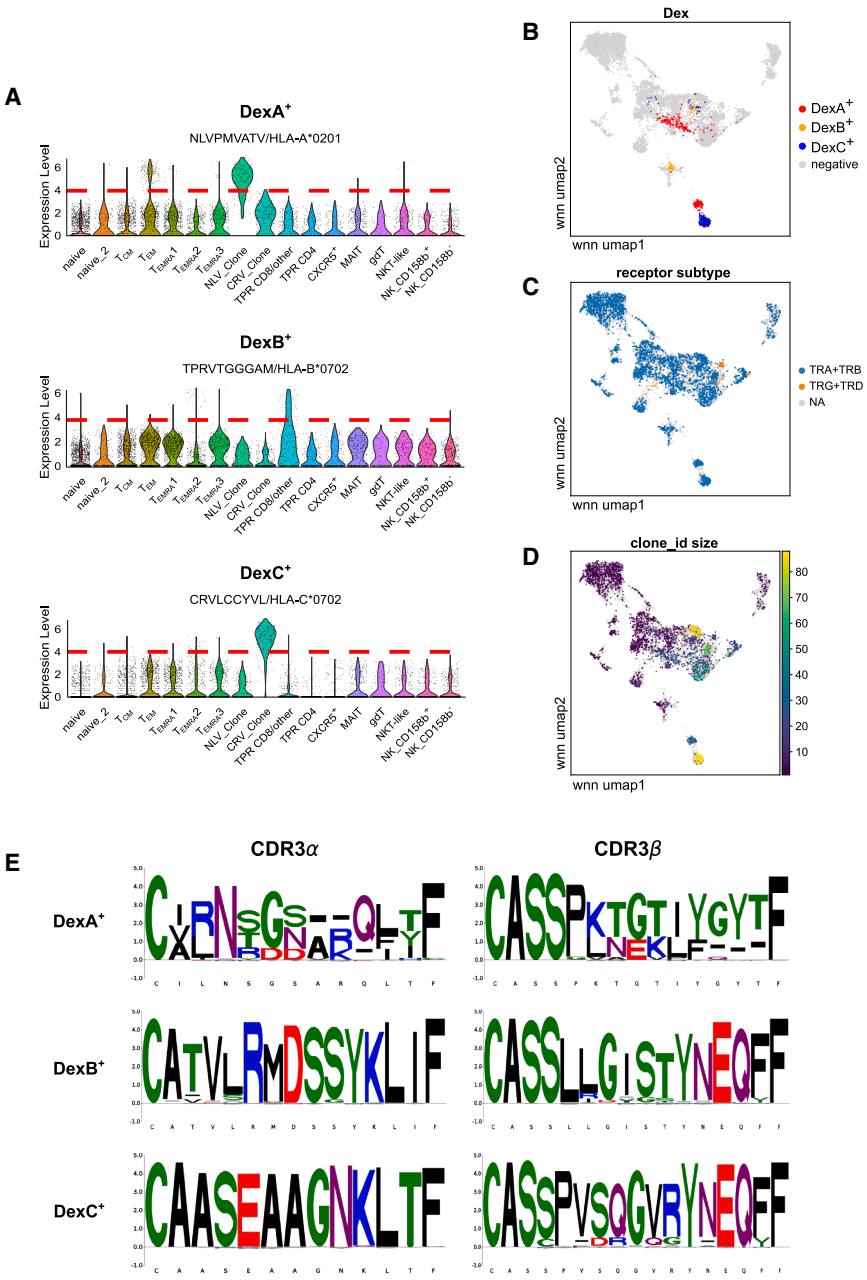


Figure 3. Characterization of CMV-specific CD8$^+$ T cells

(A) Distribution of background noise-corrected CLR-normalized DexA, DexB, and DexC counts per cluster. Red line indicates cutoff for Dex$^+$ label.

(B–D) wnnUMAP representation of dataset with highlighted DexA$^+$, DexB$^+$, and DexC$^+$ cells (B), colored by receptor type after filtering for "single pair," "extra VJ," and "extra VDJ" sequences (C) and showing the corresponding clone size for each cellular TCR sequence (D).

(E) CDR3 sequence logo plots depicting under- and overrepresented amino acids of TCR CDR3α and CDR3β amino acid sequences. Logo plots are shown for DexA$^+$, DexB$^+$, and DexC$^+$ cells. Colors indicate amino acid chemistry.

See also Figures S3 and S4 and Tables S2, S3, and S4.

samples mainly localized to T_{EMRA} and T_{EM} clusters (Figures 3B and S4A). These cells potentially correspond to endogenous CMV-specific memory T cells in these donors.

Subsequently, we examined the occurrence of CMV-specific CD8$^+$ T cells under different dCODE dextramer staining conditions. The frequencies of Dex$^+$ cells in the four spike-in/background staining conditions for each dextramer are shown in Figure S4B. Samples with HLA-matched backgrounds consistently contained higher frequencies of Dex$^+$ cells compared to HLA-mismatched samples. However, even in samples with low-frequency spike-ins, Dex$^+$ cells were consistently detectable (Figure S4B), indicating that the assay was sensitive enough to detect antigen-specific cells across all spike-in frequencies.

As it has been reported that HLA-C multimers like DexC can bind to CD158b (KIR2DL2/3) on CD8$^+$ T cells with low affinity in a CMV-epitope-independent manner,[30] we also investigated the co-expression of CD158b (KIR2DL2/3) and DexC. Co-expression analysis did not reveal a KIR2DL2/3$^+$ DexC^{high} population (Figure S4C), indicating that DexC specifically bound to TCRs targeting the IE-$1_{309-317}$ epitope.

To evaluate the performance of oligo-coupled dCODE dextramers we compared them to flow cytometry data from the same reagents and conditions in the three CMV spike-in populations. In the flow cytometry data, DexA and DexC molecules bound to nearly the entirety of the corresponding clones (Figure S2B). We used this as a reference to estimate the sensitivity and specificity of DexA and DexC in the single-cell omics protocol. Cells from the NLV or CRV clones stained with flex-tags were assumed capable of binding DexA and DexC, respectively, while DexB was excluded due to the heterogeneity of the TPR-enriched sample.

We assessed the overlap between DexA$^+$ and DexC$^+$ cells based on dextramer detection ("predicted class") and cells labeled as NLV or CRV clones by flex-tags ("actual class") and calculated sensitivity and specificity (Table S2). For DexA, the positive "actual class" included the NLV clone, with the negative class consisting of cells either HLA mismatched to DexA or cells from the CRV clone. For DexC, the positive actual class were cells from the CRV clone, and the negative class contained cells HLA mismatched to DexC or cells from the NLV clone. The predicted classes were previously defined based on detected centered

log-ratio (CLR)-normalized dextramer sequencing counts (Dex^+ and Dex^- labels). Using these definitions, we observed that DexA and DexC exhibited specificities of 100% and 99.9%, respectively, indicating a minimal false positive rate. Sensitivities were 75.8% for DexA and 84.8% for DexC (Table S2).

Taken together, this demonstrates that dCODE dextramers can be highly specific and sensitive when used for scRNA-seq, providing reliable identification of antigen-specific T cells. However, differences in sensitivity between scRNA-seq and flow cytometry exist, highlighting the need for careful consideration when interpreting results across modalities.

TCR repertoires of CMV-specific $CD8^+$ T cells

Analysis of the AIRR data identified 7,499 cells with paired immune receptor sequences and corresponding transcriptome data. After filtering, 7,281 cells contained TCR sequences, 2 were B cell receptor (BCR) expressing, and 213 exhibited multi-chain expression. Complete pairs of TCRα/β or TCRγ/δ chains could be reconstructed for approximately 25% of cells (Figure S4D). For the majority of cells, at least one TCRα or TCRβ chain was detected (6,561 cells), while the remaining cells either had TCRγ/δ chains (270 cells) or non-canonical receptor combinations (450 cells), such as TCRβ+γ. Notably, TCRγδ sequences co-localized with the γδ T cell cluster identified in the scRNA-seq dataset (based on *TRGC1*/*TRGC2* and *TRDC* gene expression) (Figures 3C and 2B). Subsequently, we analyzed clonotypes by first excluding BCR, VJ, or VDJ chain-only sequences, as well as non-canonical and multi-chain receptors. Clonotypes were then defined as cells with identical nucleotide TCR sequences (Table S3).

To explore antigen specificity beyond identical nucleotide sequences, we performed a secondary clonotype clustering based on the similarity of combined TCRα and TCRβ CDR3 amino acid sequences. Similarity was assessed using a metric based on the BLOSUM62 substitution matrix enabling the grouping of TCRs that likely recognize the same antigen. Network graphs (Figure 4) depict clonotypes as nodes and their inferred relationships as edges. Clone sizes defined by identical nucleotide TCR sequences were highest in T_{EM} and T_{EMRA} populations, as well as the NLV and CRV clones (Figures 3D, 4A, and 4B).

In addition, CDR3 amino acid logo plots were generated for both Dex^+ cells (DexA, DexC, DexB; Figure 3E) and flex-tag-labeled samples (NLV, CRV, TPR; Figure S4E). For the CRV clone and $DexC^+$ populations, cells exhibited nearly identical CDR3 sequences across both TCRα and TCRβ chains (Figures 3E and S4E). CRV clone cells shared a common TCRα CDR3 amino acid sequence CAASEAAGNKLTF (*TRAV29/DV5*, *TRAJ17*), with the most prominent TCRβ CDR3 amino acid sequence being CASSPVSQGVRYNEQFF (*TRBV28*, *TRBD1*, *TRBJ2-1*) (Table S4).

Interestingly, while the clone size of the NLV clone appeared smaller compared to the CRV clone (Figure 3D), the NLV clone contained two distinct clonotype clusters in the network graph (Figure 4A). This was because NLV clones shared a common TCRβ chain (CDR3: CASSPKTGTIYGYTF from *TRBV6-5*, *TRBD1*, *TRBJ1-2*), but diverged in their TCRα chains—one with a TCRα CDR3 CVRNRDDKIIF (*TRAV38-1*, *TRAJ30*) and the other with CARNTGNQFYF (*TRAV24*, *TRAJ49*) (Table S4). While $DexA^+$ cells and those labeled with a flex-tag for the NLV clone co-localized to clonotype clusters '1' and '15' (Figures 4A and 4C), two additional $DexA^+$ clusters ('105' and '164') originated from the polyclonal background sample of donor 1 (BG_D1) and displayed distinct TCR sequences (Table S4).

Lastly, clonotype cluster 21 contained $DexB^+$ cells from both the TPR enriched and BG_D2 samples, consistent with their common donor origin (Figures 4A and 4C). Despite the heterogeneity of the TPR-enriched sample, DexB staining effectively isolated TCR sequences specific for the $pp65_{417-426}$ epitope on HLA-B*07:02, with the most frequent sequences being CDR3α CATVLRMDSSYKLIF (*TRAV17*, *TRAJ12*) and CDR3β CASSLLGISTYNEQFF (*TRBV7-9*, *TRBJ2-1*) (Figures 3E and S4E).

The specificities of TCRs were assessed by comparing sequences to a curated database of TCR sequences with known antigen specificities (VDJdb). This comparison revealed that the CDR3 sequences of the NLV and CRV clones were correctly predicted to be specific for pp65 and IE-1 CMV epitopes, respectively (Figure S4F). In summary, the analysis demonstrated that large clones originated from the three spike-in samples, or, alternatively, from clusters of differentiated T cell populations in the background samples (Figure 4).

DISCUSSION

In this study, we generated a comprehensive single-cell multi-omics dataset using the BD Rhapsody system, combining flex-tags with standard Sample-tags in a dual-layer hashing approach alongside whole transcriptome, 31 surface markers, AIRR, and antigen specificities. Unlike conventional combinatorial barcoding methods applied to fixed cells such as single-cell combinatorial indexing RNA sequencing (sci-RNA-seq) or split pool ligation-based transcriptome sequencing (SPLiT-seq),[31,32] our method multiplexes viable cells using cell surface markers via two sequential antibody stainings on distinct sample pools with differing cell compositions. This enables a significant increase in sample throughput and potential scaling for screening large sample volumes.

The dataset was designed to address both technical and biological questions, notably assessing dCODE dextramers for rare antigen-specific T cell detection and characterizing CMV-specific TCR sequences. Five samples were analyzed: three enriched for CMV-specific $CD8^+$ T cells and two $CD8^+$ polyclonal T cell backgrounds from CMV-seropositive individuals with matching HLA allotypes. Despite lower-than-expected spike-in frequencies, dual hashing recovered 9,771 cells with both Sample-tag information. Notably, negative selection for enrichment of $CD8^+$ T cells introduced minor non-T cell contamination in polyclonal background samples.

To detect antigen-specific T cells using dCODE dextramers as sequencing reagents, we established an alternative workflow that employs k-means clustering to automatically define Dex^+ cells based on dextramer count distributions without the need to pre-filter cells on TCRα/β expression, as required by the dextramer analysis framework ICON.[33] With this workflow, we demonstrated that dextramers can effectively detect antigen-specific T cells at low frequencies. Although the detection limit for antigen-specific T cells was not reached in our data, we

CellPress
OPEN ACCESS

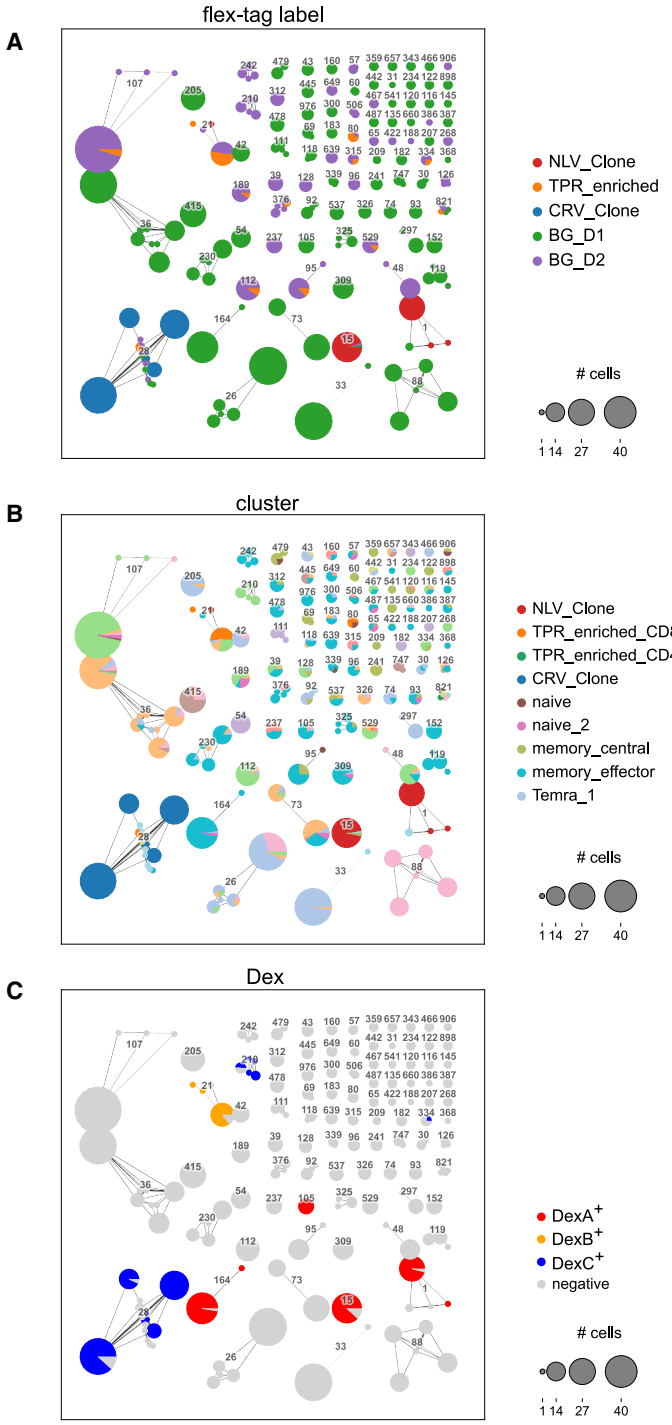


Figure 4. TCR repertoires of CMV-specific CD8⁺ T cells

(A–C) Clonotype network graphs for filtered TCR sequences are colored by flex-tag label (A), annotated cluster from scRNA-seq analysis (B), and dextramer label (C). Each circle depicts a clonotype. Connecting lines between clones indicate a "clonotype cluster," thus a high degree of CDR3 amino acid similarity between the VJ and VDJ chains of the connected clones. The numeric label indicates the clonotype cluster number. Clonotype clusters in the graph layout are arranged according to their size.

See also Tables S3 and S4.

observed differences in dextramer sensitivity when used as a sequencing versus flow cytometry reagent. Flow cytometry indicated that 97% of NLV clone cells were DexA-PE$^+$ and 98% of CRV clone cells were DexC-PE$^+$. Cells detected in the sequencing dataset labeled with flex-tags for the NLV and CRV clone could therefore be assumed to bind to DexA and DexC, respectively. The sequencing data confirmed high specificity of DexA and DexC for the NLV and CRV clones, respectively, with negligible false positives and no HLA-mismatched binding. However, the sensitivity was lower than flow cytometry, with estimated sensitivities of 75.8% for DexA and 84.8% for DexC. This discrepancy may arise from signal loss during multiple processing steps in the sequencing workflow and needs to be further monitored and investigated in future studies. In addition, AICD may pose a challenge in experiments with low-frequency antigen-specific T cells as prior cell stimulation and dextramer-labeling might contribute to this. Furthermore, increasing sequencing depth of the dextramer library might be necessary to reliably capture rare cells or counter lower binding efficacies. Moreover, systematic benchmarking of different oligo-barcoded pMHC multimer and single-cell multi-omics platforms, using matched clones, reagents, and protocols, will be critical to establish standardized performances.

TCR analysis of CMV-specific T cell clones revealed two known pp65-binding CDR3α motifs: CARNTGNQFYF (*TRAV24-TRAJ49*, clonotype cluster 1/DexA$^+$) binding pp65$_{495\text{-}503}$ on HLA-A*02:01[34] and CATVLRMDSSYKLIF (*TRAV17-TRAJ12*, clonotype cluster 2/DexB$^+$) binding pp65$_{417\text{-}426}$ on HLA-B*07:02.[35] Known CDR3β sequences included CASSLLGISTYNEQFF (*TRBV7-9*, *TRBJ2-1*) and CASSPVSQGVRYNEQFF (*TRBV28*, *TRBJ2-1*) specific for CMV epitopes pp65$_{417\text{-}426}$ and IE-1$_{309\text{-}317}$, respectively.[36] Additionally, the CDR3β sequence CASSPKTGTIYGYTF (*TRBV6-5*, *TRBJ1-2*) was previously linked to a pp65 mini-lymphoblastoid cell line (LCL) without epitope determination.[36]

We observed that dCODE dextramers can identify antigen-specific T cells at low frequencies and may be used on

samples without further enrichment. Flow cytometry can help to estimate antigen-specific T cell frequencies and guide sequencing experiments. In cases where antigen-specific T cells are rare, sorting before sequencing may be advantageous to ensure comprehensive gene expression analysis.

By presenting a step-by-step protocol, a dual-layer hashing strategy, and a detailed description of key quality-control steps, our workflow offers practical guidance that may support wider adoption of oligo-barcoded pMHC multimer single-cell multi-omics, an area where standardized best practices have only recently begun to emerge. This approach enhances our understanding of immune responses at the single-cell level, particularly in the context of TCR specificity and function and holds significant potential for improving vaccine trials and developing targeted immunotherapies.

Limitations of the study

Various approaches exist for identifying Dex^+ cells, each of which has its own merits. However, the lack of standardized best practices in the community hampers the comparability of pMHC multimer-omics data between studies. Establishing community standards would improve the reliability and reproducibility. In our study, we used k-means clustering to identify Dex^+ cells, which proved effective due to the presence of distinct Dex^{high} and Dex^{low} populations. However, this method may encounter challenges in datasets featuring populations with intermediate affinity or high variability between donors.

The existing literature illustrates the diversity in methodologies for defining pMHC-binding T cells. The TetTCR-SeqHD method employs a bimodal distribution fitting and knee point analysis to set thresholds for defining tetramer-binding events.[22] Other studies classify cells as Dex^+ as those with UMI counts surpassing negative controls multifold.[23,37] Yet another study defined cells as Dex^+ if more than 30% of dextramer-derived UMI counts were derived from a specific dextramer barcode.[38] Lastly, ICON, a tool for the normalization of high-throughput TCR-pMHC binding data and the identification of TCR-pMHC interactions incorporated negative control dextramers as well as TCRα/β information to determine dextramer binders.[33] Our study provides a flexible approach that circumvents the limitation of pre-selecting TCRα/β-expressing cells, reducing the risk of excluding relevant antigen-specific cells due to incomplete TCR recovery, although incorporating paired clonotype information remains essential for high-fidelity pMHC specificity assignment. Instead, we define cutoffs based on the distribution of CLR-normalized dextramer data. Similarly, ensuring downstream analysis tools can natively import data in AIRR-compliant TCR rearrangement format[39] will be critical for reproducible repertoire analyses.

In conclusion, while our study demonstrates the potential of combining dCODE dextramers and single-cell workflows like the BD Rhapsody platform for the identification of antigen-specific T cells, it underscores the necessity for standardized methodologies and highlights the challenges inherent in multi-modal single-cell analysis.

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to and will be fulfilled by the lead contact, Marc D. Beyer (marc.beyer@dzne.de).

Materials availability

This study did not generate unique reagents.

Data and code availability

- Raw single-cell sequencing data derived from human samples have been deposited on the European Genome-Phenome Archive (EGA: EGA50000000594) and are available upon approval of the Data Access Committee.
- All the original code to reproduce the analyses reported in this paper has been deposited at the following link: https://gitlab.dzne.de/ag-beyer/gemuend_cmv_2025. An archival DOI to access scripts and h5Seurat and MuData objects to reproduce the figures is listed in the key resources table.
- Any other information required to reproduce the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank PRECISE for technical assistance and Tal Pecht and Vadir Lopez-Salmeron (BD Biosciences) for critical discussions and technical support. The graphical abstract, Figure 1A, and parts of Figure 1B were created with BioRender.com. L.B., A.C.A., J.L.S., and M.D.B. are members of the excellence cluster ImmunoSensation2 (EXC 2151, no. 390873048) funded by the German Research Foundation (DFG). A.C.A., J.L.S., and M.D.B. are supported by the DFG via the SFB 1454 (grant no. 432325352) and IRTG2168 (grant no. 272482170), and the European Union (EU)-funded project NEUROCOV received funding from RIA HORIZON Research and Innovation under GA no. 101057775. A.C.A. is supported by the DFG (grant nos. 458854699, 466168337, and 466168626) and the BMBF-funded project IMMME (01EJ2204D). J.L.S. is supported by the EU-funded projects discovAIR (grant no. 874656) and SYSCID (grant no. 733100); the BMBF-founded project DietBody-Brain (DietBB, 01EA1809A) and iTREAT (01ZX1902B). M.D.B. is supported by the Else Kroner-Fresenius Foundation (grant no. 2018_A158). L.B. is supported by the DFG-funded project ImmuDiet (grant no. 513977171) and the European Research Council under the European Union's Horizon 2020 research and innovation program (project no. 101163024, POLIS).

AUTHOR CONTRIBUTIONS

Conceptualization: L.B., A.C.A., J.L.S., A.M., and M.D.B.; methodology, visualization, and writing – original draft: I.G., L.B., and M.D.B.; software: I.G., L.B., and M.B.; formal analysis: I.G.; investigation: I.G., S.M., C.J., and E.D.D.; resources: A.M.; funding acquisition: L.B., A.C.A., J.L.S., and M.D.B.; supervision: L.B., J.L.S., and M.D.B.; writing – review & editing: all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

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

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Generation of CMV-specific T cells
  - Generation of polyclonal T cell background
  - Thawing of human Cell material
  - Single Dextramer stainings and flow cytometry

- Multi-omics staining and single-cell capture
- First layer sample multiplexing and CD158b staining
- dCODE dextramer staining and second layer sample multiplexing
- AbSeq surface marker staining
- Multi-omics library preparation
- Sequencing and raw data demultiplexing
- Data pre-processing of scRNA-seq data
- Bioinformatic analysis of multi-omics data
- Bioinformatic analysis of AIRR data

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2025.101085.

Received: October 14, 2024
Revised: March 25, 2025
Accepted: May 28, 2025
Published: June 23, 2025

REFERENCES

1. Rajewsky, N., Almouzni, G., Gorski, S.A., Aerts, S., Amit, I., Bertero, M. G., Bock, C., Bredenoord, A.L., Cavalli, G., Chiocca, S., et al. (2020). LifeTime and improving European healthcare through cell-based interceptive medicine. Nature *587*, 377–386. https://doi.org/10.1038/s41586-020-2715-9.
2. Baysoy, A., Bai, Z., Satija, R., and Fan, R. (2023). The technological landscape and applications of single-cell multi-omics. Nat. Rev. Mol. Cell Biol. *24*, 695–713. https://doi.org/10.1038/s41580-023-00615-w.
3. Efremova, M., Vento-Tormo, R., Park, J.-E., Teichmann, S.A., and James, K.R. (2020). Immunology in the Era of Single-Cell Technologies. Annu. Rev. Immunol. *38*, 727–757. https://doi.org/10.1146/annurev-immunol-090419-020340.
4. Valkiers, S., de Vrij, N., Gielis, S., Verbandt, S., Ogunjimi, B., Laukens, K., and Meysman, P. (2022). Recent advances in T-cell receptor repertoire analysis: bridging the gap with multimodal single-cell RNA sequencing. ImmunoInformatics *5*, 100009. https://doi.org/10.1016/j.immuno.2022.100009.
5. Hoehn, K.B., and Kleinstein, S.H. (2024). B cell phylogenetics in the single cell era. Trends Immunol. *45*, 62–74. https://doi.org/10.1016/j.it.2023.11.004.
6. Hislop, A.D., Annels, N.E., Gudgeon, N.H., Leese, A.M., and Rickinson, A. B. (2002). Epitope-specific evolution of human CD8(+) T cell responses from primary to persistent phases of Epstein-Barr virus infection. J. Exp. Med. *195*, 893–905. https://doi.org/10.1084/jem.20011692.
7. Ameres, S., Mautner, J., Schlott, F., Neuenhahn, M., Busch, D.H., Plachter, B., and Moosmann, A. (2013). Presentation of an immunodominant immediate-early CD8+ T cell epitope resists human cytomegalovirus immunoevasion. PLoS Pathog. *9*, e1003383. https://doi.org/10.1371/journal.ppat.1003383.
8. Daniel, B., Yost, K.E., Hsiung, S., Sandor, K., Xia, Y., Qi, Y., Hiam-Galvez, K.J., Black, M., J Raposo, C., Shi, Q., et al. (2022). Divergent clonal differentiation trajectories of T cell exhaustion. Nat. Immunol. *23*, 1614–1627. https://doi.org/10.1038/s41590-022-01337-5.
9. Koutsakos, M., Illing, P.T., Nguyen, T.H.O., Mifsud, N.A., Crawford, J.C., Rizzetto, S., Eltahla, A.A., Clemens, E.B., Sant, S., Chua, B.Y., et al. (2019). Human CD8+ T cell cross-reactivity across influenza A, B and C viruses. Nat. Immunol. *20*, 613–625. https://doi.org/10.1038/s41590-019-0320-6.
10. van de Sandt, C.E., Nguyen, T.H.O., Gherardin, N.A., Crawford, J.C., Samir, J., Minervina, A.A., Pogorelyy, M.V., Rizzetto, S., Szeto, C., Kaur, J., et al. (2023). Newborn and child-like molecular signatures in older adults stem from TCR shifts across human lifespan. Nat. Immunol. *24*, 1890–1907. https://doi.org/10.1038/s41590-023-01633-8.
11. Cai, C., Samir, J., Pirozyan, M.R., Adikari, T.N., Gupta, M., Leung, P., Hughes, B., Van der Byl, W., Rizzetto, S., Elthala, A., et al. (2022). Identification of human progenitors of exhausted CD8+ T cells associated with elevated IFN-γ response in early phase of viral infection. Nat. Commun. *13*, 7543. https://doi.org/10.1038/s41467-022-35281-7.
12. Hensel, N., Gu, Z., Emmerich, F., Wieland, D., Wieland, D., Jechow, K., Kemming, J., Llewellyn-Lacey, S., Gostick, E., Sogukpinar, O., et al. (2021). Memory-like HCV-specific CD8+ T cells retain a molecular scar after cure of chronic HCV infection. Nat. Immunol. *22*, 229–239. https://doi.org/10.1038/s41590-020-00817-w.
13. Agerer, B., Koblischke, M., Gudipati, V., Montaño-Gutierrez, L.F., Smyth, M., Popa, A., Genger, J.-W., Endler, L., Florian, D.M., Mühlgrabner, V., et al. (2021). SARS-CoV-2 mutations in MHC-I-restricted epitopes evade CD8+ T cell responses. Sci. Immunol. *6*, eabg6461. https://doi.org/10.1126/sciimmunol.abg6461.
14. Ogura, H., Gohda, J., Lu, X., Yamamoto, M., Takesue, Y., Son, A., Doi, S., Matsushita, K., Isobe, F., Fukuda, Y., et al. (2022). Dysfunctional Sars-CoV-2-M protein-specific cytotoxic T lymphocytes in patients recovering from severe COVID-19. Nat. Commun. *13*, 7063. https://doi.org/10.1038/s41467-022-34655-1.
15. Choy, C., Chen, J., Li, J., Gallagher, D.T., Lu, J., Wu, D., Zou, A., Hemani, H., Baptiste, B.A., Wichmann, E., et al. (2023). SARS-CoV-2 infection establishes a stable and age-independent CD8+ T cell response against a dominant nucleocapsid epitope using restricted T cell receptors. Nat. Commun. *14*, 6725. https://doi.org/10.1038/s41467-023-42430-z.
16. Aoki, H., Kitabatake, M., Abe, H., Xu, P., Tsunoda, M., Shichino, S., Hara, A., Ouji-Sageshima, N., Motozono, C., Ito, T., et al. (2024). CD8+ T cell memory induced by successive SARS-CoV-2 mRNA vaccinations is characterized by shifts in clonal dominance. Cell Rep. *43*, 113887. https://doi.org/10.1016/j.celrep.2024.113887.
17. Bentzen, A.K., Marquard, A.M., Lyngaa, R., Saini, S.K., Ramskov, S., Donia, M., Such, L., Furness, A.J.S., McGranahan, N., Rosenthal, R., et al. (2016). Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. Nat. Biotechnol. *34*, 1037–1045. https://doi.org/10.1038/nbt.3662.
18. Zhang, S.-Q., Ma, K.-Y., Schonnesen, A.A., Zhang, M., He, C., Sun, E., Williams, C.M., Jia, W., and Jiang, N. (2018). High-throughput determination of the antigen specificities of T cell receptors in single cells. Nat. Biotechnol. *36*, 1156–1159. https://doi.org/10.1038/nbt.4282.
19. Ng, A.H.C., Peng, S., Xu, A.M., Noh, W.J., Guo, K., Bethune, M.T., Chour, W., Choi, J., Yang, S., Baltimore, D., and Heath, J.R. (2019). MATE-Seq: microfluidic antigen-TCR engagement sequencing. Lab Chip *19*, 3011–3021. https://doi.org/10.1039/c9lc00538b.
20. Overall, S.A., Toor, J.S., Hao, S., Yarmarkovich, M., Sara, M O'Rourke, Morozov, G.I., Nguyen, S., Japp, A.S., Gonzalez, N., Moschidi, D., et al. (2020). High throughput pMHC-I tetramer library production using chaperone-mediated peptide exchange. Nat. Commun. *11*, 1909. https://doi.org/10.1038/s41467-020-15710-1.
21. Pai, J.A., and Satpathy, A.T. (2021). High-throughput and single-cell T cell receptor sequencing technologies. Nat. Methods *18*, 881–892. https://doi.org/10.1038/s41592-021-01201-8.
22. Ma, K.-Y., Schonnesen, A.A., He, C., Xia, A.Y., Sun, E., Chen, E., Sebastian, K.R., Guo, Y.-W., Balderas, R., Kulkarni-Date, M., and Jiang, N. (2021). High-throughput and high-dimensional single-cell analysis of antigen-specific CD8+ T cells. Nat. Immunol. *22*, 1590–1598. https://doi.org/10.1038/s41590-021-01073-2.
23. Zhang, B., Upadhyay, R., Hao, Y., Samanovic, M.I., Herati, R.S., Blair, J. D., Axelrad, J., Mulligan, M.J., Littman, D.R., and Satija, R. (2023). Multimodal single-cell datasets characterize antigen-specific CD8+ T cells

across SARS-CoV-2 vaccination and infection. Nat. Immunol. *24*, 1725–1734. https://doi.org/10.1038/s41590-023-01608-9.

24. Greiff, V., Yaari, G., and Cowell, L.G. (2020). Mining adaptive immune receptor repertoires for biological and clinical information using machine learning. Curr. Opin. Syst. Biol. *24*, 109–119. https://doi.org/10.1016/j.coisb.2020.10.010.
25. Hudson, D., Fernandes, R.A., Basham, M., Ogg, G., and Koohy, H. (2023). Can we predict T cell specificity with digital biology and machine learning? Nat. Rev. Immunol. *23*, 511–521. https://doi.org/10.1038/s41577-023-00835-3.
26. Ulbrich, J., Lopez-Salmeron, V., and Gerrard, I. (2023). BD Rhapsody™ Single-Cell Analysis System Workflow: From Sample to Multimodal Single-Cell Sequencing Data. Methods Mol. Biol. *2584*, 29–56. https://doi.org/10.1007/978-1-0716-2756-3_2.
27. Moosmann, A., Khan, N., Cobbold, M., Zentz, C., Delecluse, H.-J., Hollweck, G., Hislop, A.D., Blake, N.W., Croom-Carter, D., Wollenberg, B., et al. (2002). B cells immortalized by a mini-Epstein-Barr virus encoding a foreign antigen efficiently reactivate specific cytotoxic T cells. Blood *100*, 1755–1764. https://doi.org/10.1182/blood.V100.5.1755.h81702001755_1755_1764.
28. McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: uniform manifold approximation and projection. JOSS *3*, 861. https://doi.org/10.21105/joss.00861.
29. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587. https://doi.org/10.1016/j.cell.2021.04.048.
30. Schlott, F., Steubl, D., Ameres, S., Moosmann, A., Dreher, S., Heemann, U., Hösel, V., Busch, D.H., and Neuenhahn, M. (2018). Characterization and clinical enrichment of HLA-C*07:02-restricted Cytomegalovirus-specific CD8+ T cells. PLoS One *13*, e0193554. https://doi.org/10.1371/journal.pone.0193554.
31. Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L.T., Peeler, D.J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Sci. Technol. Humanit. *360*, 176–182. https://doi.org/10.1126/science.aam8999.
32. Martin, B.K., Qiu, C., Nichols, E., Phung, M., Green-Gladden, R., Srivatsan, S., Blecher-Gonen, R., Beliveau, B.J., Trapnell, C., Cao, J., and Shendure, J. (2023). Optimized single-nucleus transcriptional profiling by combinatorial indexing. Nat. Protoc. *18*, 188–207. https://doi.org/10.1038/s41596-022-00752-0.
33. Zhang, W., Hawkins, P.G., He, J., Gupta, N.T., Liu, J., Choonoo, G., Jeong, S.W., Chen, C.R., Dhanik, A., Dillon, M., et al. (2021). A framework for highly multiplexed dextramer mapping and prediction of T cell receptor sequences to antigen specificity. Sci. Adv. *7*, eabf5835. https://doi.org/10.1126/sciadv.abf5835.
34. Link, C.S., Eugster, A., Heidenreich, F., Rücker-Braun, E., Schmiedgen, M., Oelschlägel, U., Kühn, D., Dietz, S., Fuchs, Y., Dahl, A., et al. (2016). Abundant cytomegalovirus (CMV) reactive clonotypes in the CD8(+) T cell receptor alpha repertoire following allogeneic transplantation. Clin. Exp. Immunol. *184*, 389–402. https://doi.org/10.1111/cei.12770.
35. Dössinger, G. (2014). MHC multimer purification of rare antigen specific T cells and direct T cell receptor isolation by single cell PCR. PhD thesis. https://mediatum.ub.tum.de/doc/1136748/1136748.pdf.
36. Huth, A., Liang, X., Krebs, S., Blum, H., and Moosmann, A. (2019). Antigen-Specific TCR Signatures of Cytomegalovirus Infection. J. Immunol. *202*, 979–990. https://doi.org/10.4049/jimmunol.1801401.
37. Adamo, S., Michler, J., Zurbuchen, Y., Cervia, C., Taeschler, P., Raeber, M.E., Baghai Sain, S., Nilsson, J., Moor, A.E., and Boyman, O. (2022). Signature of long-lived memory CD8+ T cells in acute SARS-CoV-2 infection. Nature *602*, 148–155. https://doi.org/10.1038/s41586-021-04280-x.
38. Minervina, A.A., Pogorelyy, M.V., Kirk, A.M., Crawford, J.C., Allen, E.K., Chou, C.-H., Mettelman, R.C., Allison, K.J., Lin, C.-Y., Brice, D.C., et al. (2022). SARS-CoV-2 antigen exposure history shapes phenotypes and specificity of memory CD8+ T cells. Nat. Immunol. *23*, 781–790. https://doi.org/10.1038/s41590-022-01184-4.
39. Vander Heiden, J.A., Marquez, S., Marthandan, N., Bukhari, S.A.C., Busse, C.E., Corrie, B., Hershberg, U., Kleinstein, S.H., Matsen Iv, F.A., Ralph, D.K., et al. (2018). AIRR community standardized representations for annotated immune repertoires. Front. Immunol. *9*, 2206. https://doi.org/10.3389/fimmu.2018.02206.
40. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.
41. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. j. *17*, 10. https://doi.org/10.14806/ej.17.1.200.
42. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420. https://doi.org/10.1038/nbt.4096.
43. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. *20*, 296. https://doi.org/10.1186/s13059-019-1874-1.
44. Wickham, H. (2016). ggplot2 - Elegant Graphics for Data Analysis (Springer-Verlag). https://doi.org/10.1007/978-0-387-98141-3.
45. Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. J. Stat. Softw. *28*, 1–26. https://doi.org/10.18637/jss.v028.i05.
46. Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233. https://doi.org/10.1038/s41598-019-41695-z.
47. Bolotin, D.A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I.Z., Putintseva, E.V., and Chudakov, D.M. (2015). MiXCR: software for comprehensive adaptive immunity profiling. Nat. Methods *12*, 380–381. https://doi.org/10.1038/nmeth.3364.
48. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.
49. Sturm, G., Szabo, T., Fotakis, G., Haider, M., Rieder, D., Trajanoski, Z., and Finotello, F. (2020). Scirpy: a Scanpy extension for analyzing single-cell T-cell receptor-sequencing data. Bioinformatics *36*, 4817–4818. https://doi.org/10.1093/bioinformatics/btaa611.
50. Virshup, I., Rybakov, S., Theis, F.J., Angerer, P., and Wolf, F.A. (2021). anndata: Annotated data. Preprint at bioRxiv. https://doi.org/10.1101/2021.12.16.473007.
51. Bredikhin, D., Kats, I., and Stegle, O. (2022). MUON: multimodal omics analysis framework. Genome Biol. *23*, 42. https://doi.org/10.1186/s13059-021-02577-8.
52. Longmate, J., York, J., La Rosa, C., Krishnan, R., Zhang, M., Senitzer, D., and Diamond, D.J. (2001). Population coverage by HLA class-I restricted cytotoxic T-lymphocyte epitopes. Immunogenetics (N. Y.) *52*, 165–173. https://doi.org/10.1007/s002510000271.
53. Weekes, M.P., Wills, M.R., Mynard, K., Carmichael, A.J., and Sissons, J.G. (1999). The memory cytotoxic T-lymphocyte (CTL) response to human cytomegalovirus infection contains individual peptide-specific CTL clones that have undergone extensive expansion in vivo. J. Virol. *73*, 2099–2108. https://doi.org/10.1128/JVI.73.3.2099-2108.1999.
54. Wiesner, M., Zentz, C., Hammer, M.H., Cobbold, M., Kern, F., Kolb, H.-J., Hammerschmidt, W., Zeidler, R., and Moosmann, A. (2005). Selection of CMV-specific CD8+ and CD4+ T cells by mini-EBV-transformed B cell

lines. Eur. J. Immunol. *35*, 2110–2121. https://doi.org/10.1002/eji.200425936.

55. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774. https://doi.org/10.1101/gr.135350.111.

56. De Domenico, E., Bonaguro, L., Schulte-Schrepping, J., Becker, M., Händler, K., and Schultze, J.L. (2020). Optimized workflow for single-cell transcriptomics on infectious diseases including COVID-19. STAR Protoc. *1*, 100233. https://doi.org/10.1016/j.xpro.2020.100233.

57. Stoeckius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B.Z., Mauck, W.M., Smibert, P., and Satija, R. (2018). Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biol. *19*, 224. https://doi.org/10.1186/s13059-018-1603-1.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-human CD158b (clone CH-L), custom AbSeq DNA barcode	BD Biosciences	Cat# 559783, RRID:AB_397324
PE mouse anti-human CD45 (clone HI30)	BioLegend	Cat# 304039, RRID:AB_2562057
BV421 mouse anti-human CD3 (clone UCHT1)	BD Biosciences	Cat# 562426, RRID:AB_11152082
Chemicals, peptides, and recombinant proteins		
FcR Block Reagent, human	Miltenyi	Cat# 130-059-901
Invitrogen UltraPure Herring Sperm DNA Solution (sheared)	Thermo Fischer	Cat# 15634017
D-Biotin	Thermo Fischer	Cat# B20656
AMPure XP Bead-Based Reagent	Beckman Coulter	Cat# A63881
BD Pharmingen Stain Buffer (FBS)	Becton Dickinson	Cat# 554656
Fetal bovine serum	PAN Biotec	Cat# 3302
RPMI 1640 medium	Gibo	Cat# 21875034
PBS, pH 7.4	Gibco	Cat# 15374875
Critical commercial assays		
Human $CD8^+$ T cell isolation kit	Miltenyi	Cat# 130-096-495
BD Rhapsody cDNA Kit	Becton Dickinson	Cat# 633773
BD Rhapsody WTA Amplification Kit	Becton Dickinson	Cat# 633801
BD Rhapsody TCR/BCR Amplification Kit	Becton Dickinson	Cat# 665345
BD Rhapsody Enhanced Cartridge Reagent Kit	Becton Dickinson	Cat# 664887
BD Rhapsody Cartridge Kit	Becton Dickinson	Cat# 633733
BD AbSeq Immune Discovery Panel	Becton Dickinson	Cat# 625970
BD Single-Cell Multiplexing Kit, Human Sample Tag (12) Component	Becton Dickinson	Cat# 633781
BD Flex Single-Cell Multiplexing Kit C, Flex Sample Tag 13-18	Becton Dickinson	Cat# 633851
LIVE/DEAD Fixable Near-IR Dead Cell Stain Kit, for 633 or 635 nm excitation, dilution 1:1000	Thermo Fischer	Cat# L34975
NovaSeq 6000 S4 Reagent Kit v1.5 (300 cycles)	Illumina	Cat# 20028312
Deposited data		
Single-cell multi-omics raw data	This paper	EGA: EGA50000000594
Software and algorithms		
FlowJo	BD Biosciences	https://flowjo.com, RRID:SCR_008520
Bcl2fastq2 v2.20	Illumina	https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html, RRID:SCR_015058
STAR v2.6.1b	Dobin et al.[40]	https://github.com/alexdobin/STAR/, RRID:SCR_004463
Cutadapt v1.16	Martin[41]	https://github.com/marcelm/cutadapt, RRID:SCR_011841
Dropseq-tools	Broad Institute	https://github.com/broadinstitute/Drop-seq/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
R v4.1.2 (scRNA-seq/AbSeq analysis)	R Foundation	http://www.r-project.org/, RRID: SCR_001905
Python v3.10.13 (scTCR-seq analysis)	Python Software Foundation	https://www.python.org/, RRID: SCR_008394
Seurat v4.1.0 (R package)	Butler et al.[42]	http://seurat.r-forge.r-project.org/, RRID: SCR_007322
sctransform v0.3.3 (R package)	Hafemeister and Satija[43]	https://github.com/satijalab/sctransform, RRID:SCR_022146
ggplot2 v3.3.5 (R package)	Wickham[44]	https://cran.r-project.org/web/packages/ggplot2/index.html, RRID:SCR_014601
stats v4.1.2 (R package)	N/A	https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html, RRID: SCR_025968
caret v6.0-94 (R package)	Kuhn[45]	https://github.com/topepo/caret, RRID: SCR_022524
SeuratDisk v0.0.0.9019 (R package)	N/A	https://github.com/mojaveazure/seurat-disk
igraph - Leiden clustering	Traag et al.[46]	https://igraph.org/r/doc/cluster_leiden.html
MiXCR v4.2.0	Bolotin et al.[47]	https://mixcr.com/mixcr/about/, RRID:SCR_018725
Scanpy v1.9.8	Wolf et al.[48]	https://github.com/theislab/scanpy, RRID: SCR_018139
Scirpy v0.17.2	Sturm et al.[49]	https://scirpy.scverse.org/en/latest/
anndata v0.10.8	Virshup et al.[50]	https://github.com/theislab/anndata, RRID: SCR_018209
muon v0.1.6	Bredikhin et al.[51]	https://gtca.github.io/muon/, RRID: SCR_022804
mudata v0.2.3	N/A	https://github.com/scverse/mudata
MuDataSeurat v0.0.0.9000 (R package)	N/A	https://pmbio.github.io/MuDataSeurat/
numpy v1.26.4	N/A	http://www.numpy.org/, RRID:SCR_008633
scipy v1.13.0	N/A	http://www.scipy.org/, RRID:SCR_008058
pandas v2.2.2	N/A	https://pandas.pydata.org/, RRID: SCR_018214
matplotlib v3.7.2	N/A	http://matplotlib.sourceforge.net/, RRID: SCR_008624
Custom code	This paper	Zenodo: https://doi.org/10.5281/zenodo.15423937
Other		
dCODE Dextramer (RiO) - Gold, HLA-A*0201/NLVPMVATV/PE	Immudex	Cat# WB02132DRG
dCODE Dextramer (RiO) - Gold, HLA-B*0702/TPRVTGGGAM/PE	Immudex	Cat# WH02136DRG
Custom reagent dCODE Dextramer (RiO), HLA-C*07:02/CRVLCCYVL/PE	Immudex	N/A
dCODE Dextramer (RiO)-Gold, HLA-B*0801/Neg. Control/PE	Immudex	Cat# WI03233DRG

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

PBMCs were prepared from anonymous peripheral blood buffy coats purchased from the Institut für Transfusionsmedizin, Ulm, Germany. Alternatively, PBMCs were isolated from peripheral blood donations provided by healthy adult CMV-positive donors with informed consent. PBMCs were used to establish T cell lines and clones and generate mini-EBV-transformed B cell lines. The institutional review board (Ethikkommission bei der Medizinischen Fakultät der Ludwig-Maximilians-Universität München, Project no. 17–455, 16.10.2017) has approved these procedures. HLA typing was performed by MVZ Labor Dr. Klein and Dr. Rost, Martinsried, Germany. Demographic information for human material, along with details on the Donor origin of each sample ('NLV clone', 'TPR enriched', 'CRV Clone', 'Background Donor 1′, and 'Background Donor 2′), can be found in Table S1.

The small number of human subjects involved in this study limits its generalizability to any population group defined by ethnicity, geography etc. However, each of the donors was carrier of one or more HLA alleles (A*02:01) or haplotypes (B*07:02-C*07:02) known to be overrepresented in populations of European origin compared to the worldwide average. Information about self-ascribed ethnicity or origin was not obtained from donors as per the study protocol reviewed by the institutional review board.

METHOD DETAILS

Generation of CMV-specific T cells

The T cells used in this study were a $CD8^+$ T cell clone specific for HLA-A*02:01-restricted epitope NLVPMVATV from tegument phosphoprotein pp65 (abbreviated NLV, amino acids 495–503), referred to as 'NLV clone'; a $CD8^+$ T cell clone specific for HLA-C*07:02-restricted epitope CRVLCCYVL from immediate-early protein IE-1 (abbreviated CRV, amino acids 309–317), referred to as 'CRV clone'; and a polyclonal T cell culture enriched for $CD8^+$ T cells specific for HLA-B*07:02-restricted epitope TPRVTGGGAM from pp65 (abbreviated TPR, amino acids 417–426), referred to as 'TPR enriched'. These $CD8^+$ T cell epitopes were previously identified as follows: NLV,[52] CRV,[7] TPR.[53] For TPR, a polyclonal T cell culture was used, because of technical limitations in the analysis of a TPR-specific T cell clone. Cell culture medium was RPMI 1640 (Invitrogen) with 10% fetal bovine serum (FBS; Anprotec). Recombinant human interleukin-2 (Proleukin, Novartis) was added as indicated.

To establish T cell clones, peripheral blood mononuclear cells (PBMCs) from CMV-positive carriers of the required HLA allele were repeatedly stimulated,[54] in the presence of interleukin-2 (50–100 U/ml), with irradiated cells from an autologous mini-EBV-transformed B-lymphoblastoid cell line (mini-LCL) that expressed the antigen in question (pp65 or IE-1) encoded by its mini-EBV genome under control of an SV40 promoter.[27] This procedure results in polyclonal T cell cultures strongly enriched for CMV antigen-specific T cells after 4–5 stimulations (31–38 days). In a second step, T cell clones were produced from these polyclonal T cell cultures by limiting dilution[7,54] into 96-well plate microcultures at 0.5 or 0.7 T cells per well and cultivation with irradiated feeder/stimulator cells (mixed allogeneic PBMCs at 1–1.5x10^6 cells/ml and autologous pp65-or IE-1-expressing mini-LCL at 5–10x10^4 cells/ml) and 1000 U/ml Interleukin-2. Microcultures were checked for outgrowth and first screened for specificity at 2–3 weeks after limiting dilution, and outgrowing cultures were expanded under the same conditions. Screens for specificity included testing for recognition of CMV antigen-expressing autologous mini-LCLs over control mini-LCLs, demonstrating recognition of intracellularly processed antigen, testing for recognition of specific peptide, and staining with HLA/peptide multimers.

Polyclonal T cell cultures enriched for TPR-specific T cells were generated by stimulation of PBMCs with peptide.[36] PBMCs from a CMV-positive, HLA-B*07:02-positive donor were loaded with peptide (1 μg/mL) for 1 h at 37°C, washed three times with phosphate-buffered saline (PBS), and plated at 2.5x10^6 cells per mL in medium with interleukin-2 (50 U/ml). Cells were cultivated for 10 days, with partial medium renewal at 5 days. Cells were cryoconserved in 50% FBS, 40% RPMI 1640, and 10% DMSO, and stored over liquid nitrogen until further use.

T cell clones, polyclonal T cells, and control PBMCs were derived from different donors for reasons of availability, except for one donor from whom both the TPR-enriched T cells and PBMCs for one of the polyclonal $CD8^+$ T cell background (BG_D2) originated.

Generation of polyclonal T cell background

'Background' samples were $CD8^+$ polyclonal T cells enriched from PBMCs of two different donors (referred to as 'BG_D1' and 'BG_D2'). $CD8^+$ T cells were isolated by negative selection using MS columns with a Miltenyi human $CD8^+$ T Cell Isolation Kit (Cat No. 130-096-495) according to manufacturer's instructions. Sterile filtered PBS with 0.5% BSA and 2 mM EDTA was used as a buffer for magnetic cell separation.

Thawing of human Cell material

Cryovials were removed from liquid nitrogen storage and briefly placed on dry ice before thawing in a 37°C water bath for 2–3 min. Thawed cells were gently transferred to a 50 mL Falcon tube using a 1 mL pipette tip. The cryovial was rinsed with 1 mL of pre-warmed complete growth medium (RPMI +10% FCS), which was gradually added dropwise to the Falcon tube. Cells were serially diluted 1:1 with complete growth medium five times, reaching a final volume of 32 mL. The suspension was centrifuged at 300 × g for 5 min at room temperature. After aspirating the supernatant to 1–2 mL, cells were resuspended and stored on ice until further use.

Single Dextramer stainings and flow cytometry

First, cell labeling, wash, and FACS buffers were prepared. Cell labeling buffer consisted of BD Pharmingen Stain Buffer (FBS) supplemented with 0.1 g/L sheared herring sperm DNA. Wash buffer was PBS (pH 7.4) with 2.5% fetal bovine serum albumin, while FACS buffer contained PBS (pH 7.4) with 0.5% bovine serum albumin.

Aliquots of the NLV and CRV T cell clones, as well as the TPR-enriched sample, were thawed and counted. Each cell type was individually stained with one of the following dCODE Dextramers: HLA-A*02:01/pp65$_{495\text{-}503}$ (“DexA”), HLA-B*07:02/pp65$_{417\text{-}426}$ (“DexB”), HLA-C*07:02/IE-1$_{309\text{-}317}$ (“DexC”), HLA-B*08:01/AAKGRGAAL (“DexN”), or left unstained as a control. For each condition, 300,000 cells were resuspended in 50 μL of cell labeling buffer and stained with 0.2 μL of 100 μM d-Biotin (diluted in PBS) and 2 μL of dCODE Dextramer (DexA, DexB, DexN) or 10 μL of dCODE Dextramer (DexC). Staining was performed for 10 min at room temperature, protected from light.

During incubation, a 2x concentrated antibody master mix was prepared, containing CD3 (BV421) antibody (1:200) and a viability dye (1:1000) in BD stain buffer. After incubation, the cell suspension volume was adjusted to 100 μL with cell labeling buffer, followed by the addition of 100 μL of the 2× antibody master mix, yielding a final staining volume of 200 μL. Samples were incubated for 30 min at 4°C in the dark. Cells were then washed twice by centrifugation at 400*g* for five minutes, discarding the supernatant and resuspending in 1 mL of wash buffer. Finally, cells were resuspended in 200 μL of FACS buffer and immediately acquired on a FACSymphony A5 Cell Analyzer. Data were analyzed using FlowJo software.

Multi-omics staining and single-cell capture

The three samples ‘NLV clone’, ‘CRV clone’ and ‘TPR enriched’ are referred to as "spike-in” samples. CD8$^+$ polyclonal T cells enriched from PBMCs of two different donors are referred to as “background”.

First layer sample multiplexing and CD158b staining

Each of the five samples were individually labeled with different BD Flex SMK antibodies. To do so, three to four million cells per sample were pelleted (400 × g, 5 min), resuspended in 100 μL blocking buffer (95 μL BD Stain Buffer (FBS) + 5 μL FcR human blocking reagent), and incubated on ice for 10 min. Cells were then first stained with 7.5 μL primary CD45-PE antibody (BioLegend, Cat No. 304008) for 20 min on ice, then washed three times with 2 mL BD stain buffer (400 × g, 5 min), and resuspended in 80 μL BD stain buffer.

Next, secondary BD flex SMK antibodies, each with a unique oligo barcode per sample, were added alongside CD158b AbSeq antibody (BD, Cat. No. 559784). Samples were incubated with 98 μL BD stain buffer, 20 μL BD flex SMK antibody, and 2 μL CD158b antibody for 30 min on ice. After three washes (400 × g, 5 min), cell pellets were resuspended in 500 μL stain buffer and stored on ice.

dCODE dextramer staining and second layer sample multiplexing

Cells were counted and pooled into four staining mixes (400,000 cells per staining mix), each containing either 95% or 50% background and 5% or 50% spike-in mix, respectively. Background samples consisted of either Background Donor 1 or 2, while the spike-in mix included NLV Clone, TPR enriched, and CRV Clone samples in a 1:1:1 ratio.

A dextramer labeling master mix was prepared for the four samples using 3.2 μL of 100 μM d-Biotin and 8 μL of each dextramer, except for HLA-C*07:02 (CMV IE-1), which was added at 40 μL. A total of 16.8 μL master mix was added per sample, followed by incubation at room temperature for 10 min. Then, 113.2 μL BD stain buffer and 20 μL standard BD SMK antibody were added to each reaction, using a unique DNA barcode tag per staining mix. After a 20 min incubation, samples were washed three times (400 × g, 5 min) and resuspended in 500 μL stain buffer, counted, and stored on ice.

AbSeq surface marker staining

BD AbSeq Immune Discovery Panel (IDP) antibodies (BD, Cat No. 625970) were reconstituted in nuclease-free water and stored on ice. The four staining mixes (100,000 cells each) were pooled in a 1:1:1:1 ratio, centrifuged (300 × g, 5 min, 4°C), and resuspended in 100 μL blocking buffer. The antibody mix was prepared by adding 65 μL BD stain buffer to the reconstituted antibodies, yielding 100 μL of 2× AbSeq labeling master mix, which was added to the cells and incubated for 40 min on ice.

After two washes with BD stain buffer (300 × g, 5 min, 4°C), cells were resuspended in 300 μL cold sample buffer. Finally, 20,000 cells were processed on each of two BD Rhapsody cartridges (replicates), using enhanced cell capture beads.

Multi-omics library preparation

From each cartridge, six libraries were prepared. Reverse transcription and template switching was performed following manufacturer’s instructions (Becton Dickinson, Doc ID: 23–24020(02)). During the subsequent denaturation, 75 μL of elution buffer was used. The denatured supernatant was saved and utilized for dCODE Dextramer (RiO), AbSeq and Sample Tag library production. Generation of these three libraries was done following the manufacturer’s protocol (Immudex, Doc ID: TF1196.07) with 11 instead of 13 cycles in the dCODE RiO PCR2. By contrast, indexed WTA, TCR and BCR sequencing libraries were generated directly from the cell capture beads by following the corresponding steps of the manufacturer’s protocol (Becton Dickinson, Doc ID: 23–24020(02)).

Sequencing and raw data demultiplexing

Libraries were pooled taking into account the lower clustering efficiency of BCR and TCR libraries due to the increased average length following manufacturer's instructions (Becton Dickinson, Doc ID: 23–24020(02)). dCODE dextramer libraries were sequenced with a sequencing depth of an average of 5750 reads per cell and dextramer. Abseq libraries were sequenced with a sequencing depth of an average of 2900 reads per cell and antibody marker. Sequencing was performed on a NovaSeq6000 instrument with an S4 300 cycles kit v1.5. Sequencing was performed paired end with read 1 measuring 86 cycles and read 2 216 cycles for optimal TCR/BCR assembly.

Data pre-processing of scRNA-seq data

After demultiplexing of bcl files using Bcl2fastq2 V2.20 from Illumina and quality control, paired-end scRNA-seq reads were filtered for valid cell barcodes using the barcode whitelist provided by BD. Cutadapt 1.16 was then used to trim NexteraPE-PE adaptor sequences where needed and to filter reads for a PHRED score of 20 or above.[41] Then, STAR 2.6.1b was used for alignment against the human GENCODE reference genome and transcriptome hg38 release 33.[40,55] Dropseq-tools 2.0.0 were used to quantify gene expression and collapse to UMI count data (https://github.com/broadinstitute/Drop-seq/). To quantify AbSeq and dCODE Dextramer counts and facilitate dual-layer hashtag-oligo-based demultiplexing of single-cell data for downstream assignment of cell barcodes to their sample of origin (CMV spike-in/background) or staining mix, the corresponding DNA tag sequences were added to the reference genome and quantified.

Bioinformatic analysis of multi-omics data

For the bioinformatic analysis of the multi-omics data (Figure S1B), UMI count matrices were imported into R (version 4.1.2) as previously described[56] and further analyzed with the R toolkit Seurat (version 4.1.0).[29] RNA-seq analysis using Seurat was performed with the predefined software environment of the docker image jsschrepping/r_docker:jss_R412_S41 (https://hub.docker.com/r/jsschrepping/r_docker). The WTA count matrix was filtered to only contain cells with more than 250 UMI counts and with a percentage of mitochondrial genes of less than 55%. Sample tag count matrices were split into separate assays for flex and regular sample tags. Dextramer count matrices were obtained for four dextramers – three HCMV-specific ones (DexA, DexB and DexC) and one negative control dextramer (DexN). Background noise correction was performed by subtracting counts of the DexN Dextramer from DexA, DexB and DexC counts on a per-cell basis.

Count matrices for whole transcriptome, AbSeq markers, flex sample tags, regular sample tags and background corrected Dextramers were merged into one Seurat object with distinct assays using the CreateAssayObject function. The count matrix of the whole transcriptome assay was normalized using sctransform v2 regularization (R release 0.3.3).[43] The count data in the assays for AbSeq, sample tags and background-corrected Dextramers were normalized using centered log-ratio (CLR) transformation followed by scaling. For WTA and AbSeq assays, principal components (PCs) were calculated with Seurat's RunPCA function.

Demultiplexing of cells was performed for both the flex and regular sample tag assay using the HTODemux function implemented in Seurat.[57] Only cells that were labeled as "singlets" during demultiplexing were kept. Cells were further annotated according to their sample tag labels (Table S1B).

Seurat's FindMultiModalNeighbors function was used to create an embedding that represented a weighted combination of the two modalities 'whole transcriptome' and 'AbSeq'.[29] The first 20 dimensions of the WTA PCA and the first 18 dimensions from the AbSeq PCA were used to calculate multimodal neighbors as well as a weighted shared nearest neighbor (WSNN) graph. The resulting multimodal neighbor object was used to compute a joint weighted-nearest neighbors Uniform Manifold Approximation and Projection (wnnUMAP) embedding with Seurat's RunUMAP function. The WSNN graph was used to identify clusters with the leiden algorithm (resolution 0.52). Gene and surface expression markers for each cluster were determined with a Wilcoxon rank-sum test using the functions FindAllMarkers function (with the following parameters: only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25). In case of the sctransform-normalized WTA assay, the function PrepSCTFindMarkers was applied to the Seurat object prior to marker identification. Data visualization was performed with Seurat's VlnPlot, DotPlot, DoHeatmap, FeaturePlot and DimPlot function, as well as with R package ggplot2 (version 3.3.5).

Dextramer count matrices were obtained for four dextramers – three HCMV-specific ones (DexA, DexB and DexC) and one negative control dextramer (DexN). To assess and correct unspecific background signal, we first computed the percentages of cells with zero UMI counts in the dataset across all 31 AbSeq markers (Figure S3B) and across raw dextramer counts (Figure S3C, left). AbSeq markers displayed low zero-count rates even for proteins not expressed by $CD8^+$ T cells (e.g., IgM, ~1% of cells with zero UMI counts), and raw dextramer counts likewise exhibited low zero-count rates despite low true-positive frequencies, indicating notable unspecific background signal. We then subtracted DexN counts from DexA, DexB, and DexC on a per-cell basis, which markedly increased the proportion of zero counts for all three dextramers (Figure S3C, right), and yielded a clearer separation of true dextramer binders, as shown for an exemplary clonotype (clone_id 169) which binds to DexA (Figure S3D). Negative corrected values were floored to zero.

Dextramer-positive cells were defined using k-means clustering applied to CLR-normalized and background-corrected counts for each Dextramer individually. This was done using the kmeans function (centers = 4) from the stats package (version 4.1.2). The lower boundary of the k-means cluster with the highest mean value was used as a threshold to separate dextramer-positive from dextramer-negative cells (DexA = 3.94, DexB = 3.78, DexC = 3.99; Figure S3A). Dextramer sensitivities and specificities for the

DexA & the NLV clone, as well as DexC & the CRV clone, were calculated with the confusionMatrix function of the R package caret (version 6.0–94).

Bioinformatic analysis of AIRR data

The raw fastq files of the TCR and BCR libraries were preprocessed with MiXCR (version 4.2.0) using the mixcr analyze function with the 'bd-rhapsody-human-tcr-full-length' or 'bd-rhapsody-human-bcr-full-length' preset, respectively. AIRR compliant tables were exported with MiXCR's exportAirr function.

The subsequent downstream analysis of the paired TCR/BCR and WTA data (Figure S1B) was performed in Python (version 3.10.13) with the docker image rnakato/shortcake_light_3.0.0 (https://hub.docker.com/r/rnakato/shortcake_light). The Scirpy toolkit (version 0.17.2) was used for the subsequent analysis and visualization of scAIRR-seq data as well as the integration with the corresponding scRNA-seq data.[49] Additionally, the following python packages were used: Scanpy (1.9.8), anndata (0.10.8), muon (0.1.6), mudata (0.2.3), numpy (1.26.4), scipy (1.13.0), pandas (2.2.2) and matplotlib (3.7.2).

The AIRR-compliant output tables were imported into Scirpy using scirpy.io.read_airr. Chain quality metrics were calculated with ir.tl.chain_qc. To enable downstream analysis with Scirpy and Scanpy, the filtered, normalized, and annotated Seurat object was converted into an multimodal data object (MuData) object via the MuDataSeurat R package (version 0.0.0.9000) and imported with scanpy.read_h5ad. Transcriptome data were merged with TCR annotations into a MuData object.

For clonotype construction, the immune receptor repertoire was filtered to exclude sequences Scirpy's chain_qc function classified as ''ambiguous'' (non-canonical receptor combinations), ''multichain'', ''no IR'', ''orphan VJ'', or ''orphan VDJ'', as well as any entries containing BCR sequence annotation. Clonotypes were defined based on CDR3 nucleotide sequence identity of the primary VJ and VDJ chains using scirpy.pp.ir_dist and scirpy.tl.define_clonotypes. Clonotype clustering was performed using amino acid sequence similarity with scirpy.pp.ir_dist(metric = "alignment", sequence = "aa", cutoff = 15), which computes pairwise sequence alignments based on a BLOSUM62 substitution matrix. Using scirpy.tl.define_clonotype_clusters, the clonotype clustering considered both alpha and beta chains (receptor_arms = "all") and included dual TCR-expressing cells (dual_ir = "any"). Variations in CDR3 length were accounted for by BLOSUM62's alignment gap penalties.

TCR CDR3 sequence logo plots were generated with palmotif (version 0.4). TCR specificity predictions were performed using scirpy.tl.ir_query with VDJdb (metric = "identity", sequence = "aa", receptor_arms = "any", dual_ir = "any"), and the multimodal data object was annotated using scirpy.tl.ir_query_annotate (strategy = "most-frequent").

QUANTIFICATION AND STATISTICAL ANALYSIS

The analysis was performed on R and Python: the specific packages used for the analysis, their version and relevant parameters used are explained in the Method details sections.