

Cell Reports Methods, Volume 5

Supplemental information

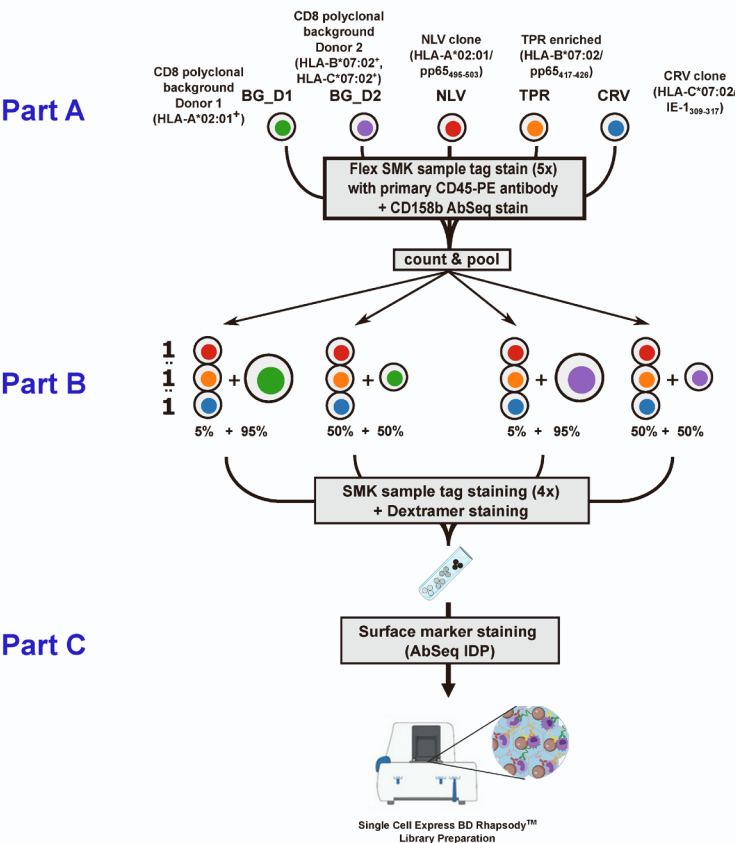
Characterization of human CMV-specific

CD8⁺ T cells using multi-layer single-cell omics

Ioanna Gemünd, Lorenzo Bonaguro, Matthias Becker, Sophie Müller, Clemens Joos, Elena De Domenico, Anna C. Aschenbrenner, Joachim L. Schultze, Andreas Moosmann, and Marc D. Beyer

Figure S1

A



B

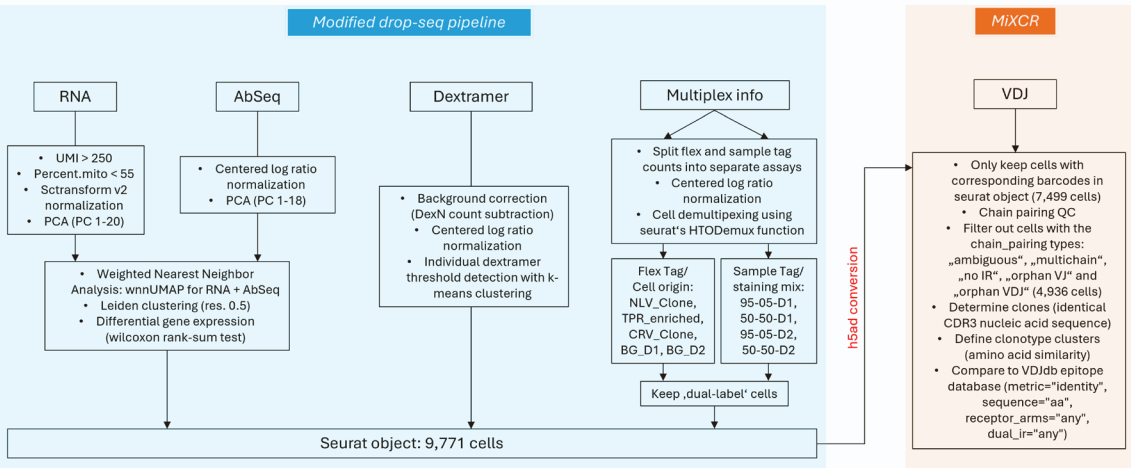
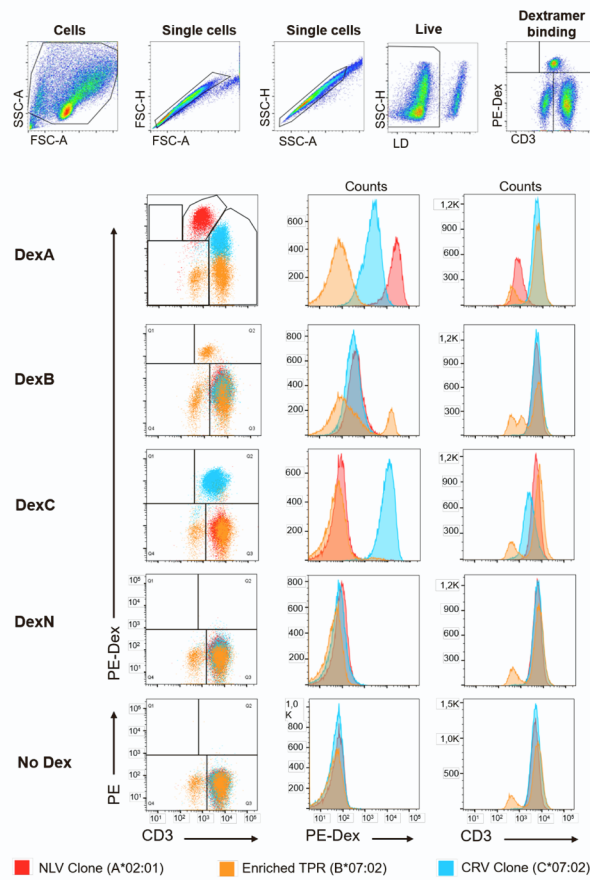


Figure S1. Single-cell multi-omics workflow allows delineation of spike-in and background samples integrating paired and dual-multiplexed RNA, AbSeq, dextramer, and VDJ sequencing data, related to Figure 1 and Methods.

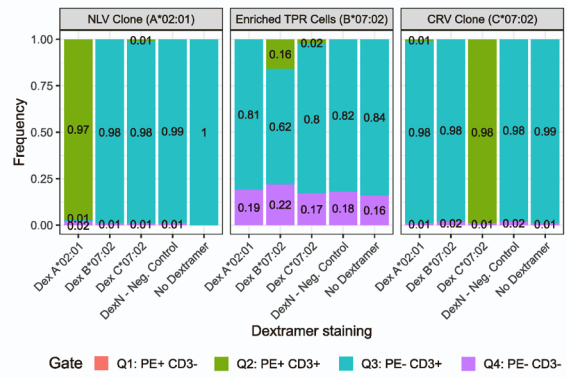
(A) Overview of stainings in experimental multi-omics workflow. In total, three staining steps were performed before single-cell capture. (a) First, clones and CD8 polyclonal backgrounds were labeled with flex sample tags and a custom CD158b antibody. (b) Next, clones and CD8 polyclonal backgrounds were counted and pooled in indicated ratios. Staining mixes consist of different combinations of CMV spike-in mix (1:1:1 ratio of NLV & CRV clone, as well as TPR enriched cells) and CD8⁺ polyclonal T cells from two background donors, either BG_D1 or BG_D2. These samples were stained with dextramers and standard sample tags. All four samples were stained with DexA, DexB, DexC and DexN. (b) Lastly, all four samples were pooled and stained with the BD AbSeq IDP panel before proceeding with single-cell capture. (B) The workflow outlines the analysis pipeline for paired RNA, AbSeq, dextramer, multiplex, and VDJ (primarily TCR) single-cell omics data. With the exception of VDJ data, which was preprocessed using MiXCR, all other data types were processed using a modified Drop-seq pipeline with downstream analysis in Seurat (R). For whole transcriptome ("RNA") data, only cells with more than 250 unique molecular identifiers (UMIs) and less than 55% mitochondrial gene content were retained. Gene expression was normalized using SCTransform v2, followed by principal component analysis (PCA). Data derived from oligo-conjugated antibodies ("AbSeq") for cell surface proteins was normalized using centered log-ratio (CLR) transformation, also followed by PCA. The RNA and AbSeq data were integrated using a weighted nearest neighbor (WNN) approach, generating a combined UMAP embedding using the top 20 principal components (PCs) from the RNA and the top 18 PCs from the AbSeq data. Leiden clustering was performed at a resolution of 0.52, and cluster markers for gene expression and surface proteins were identified using the Wilcoxon rank-sum test. For dextramer data, background correction was applied by subtracting negative control dextramer (DexN) counts from HCMV-specific dextramer counts. The corrected data was CLR-normalized, and thresholds for dextramer-positive cells were identified using k-means clustering. For sample demultiplexing ("Multiplex info"), flex and sample tag counts were separated into distinct Seurat assays, CLR-normalized, and demultiplexed using Seurat's HTODemux function. Only cells, which were categorized as "Singlet" in both assays were kept for further analysis. VDJ data and the h5ad-converted multimodal data from the Seurat workflow were analyzed using the scirpy toolkit (Python). Cells with matching barcodes in the Seurat object were retained for downstream VDJ analysis (7,499 cells). For chain pairing quality control (QC), cells with non-canonical receptor combinations, such as TRA+TRD or TRA+IGH combinations, as well as cells with more than two VJ or VDJ chains ("multichain"), no immune receptor chains ("no IR"), or single-chain cells ("orphan VJ/VDJ") were removed. Clonotypes were defined by CDR3 nucleotide identity, and clonotype networks were built based on CDR3 amino acid sequence similarity. Antigen specificity predictions were made by comparing CDR3 motifs to the VDJdb epitope database.

Figure S2

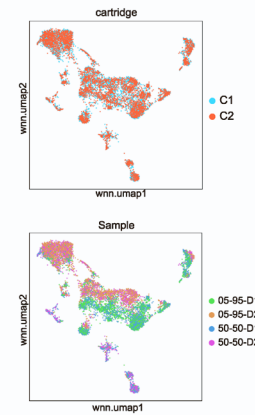
A



B



C



D

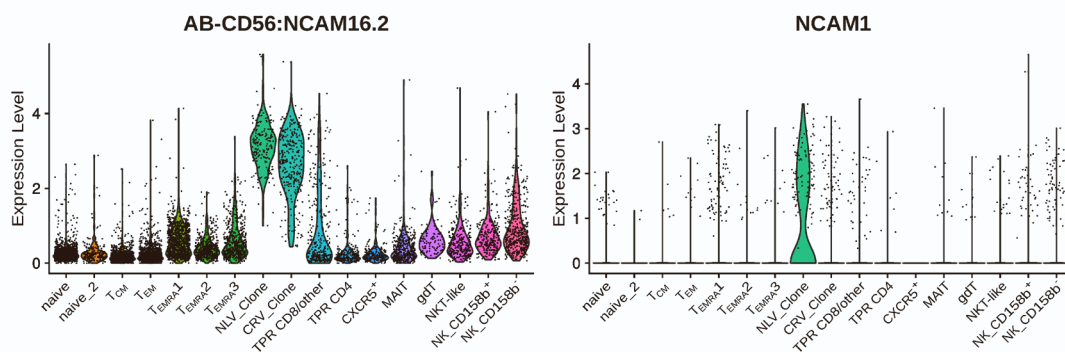
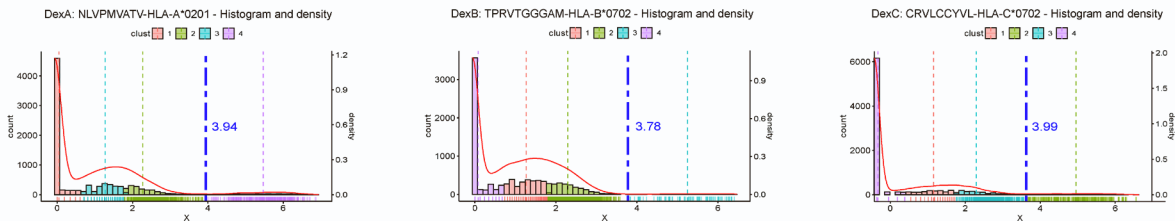


Figure S2. Single-cell multi-omics workflow allows delineation of spike-in and background samples, related to Figure 1 and 2.

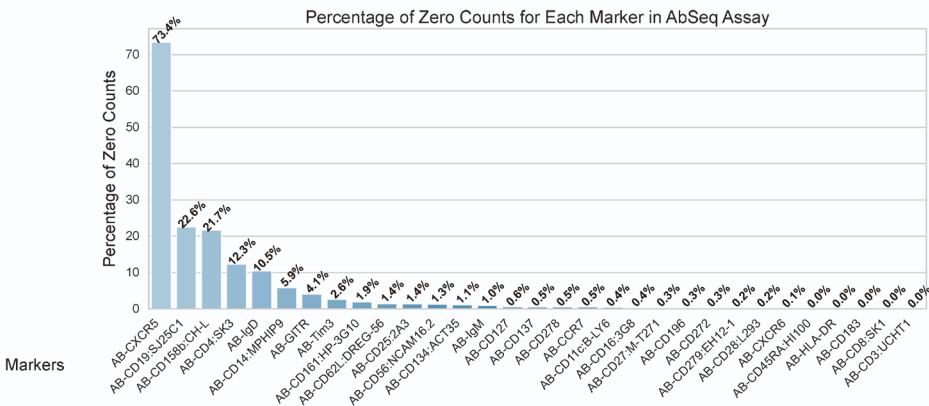
(A-B) Superimposition of flow cytometry data of individual single dextramer stainings (DexA, DexB, DexC, DexN and negative control without dextramer) on three CMV enriched input populations (NLV clone, TPR enriched and CRV clone). (A) Gating strategy for the quantification of live, single, CD3⁺ and dextramer-PE⁺ cells. (B) Gating for the identification of dextramer-PE⁺ cells for four dextramers (DexA, DexB, DexC and DexN) in three different samples with enriched CMV-specific cells (NLV clone, enriched TPR and CRV clone). (C-D) Plots originating from single-cell omics data. (C) wnnUMAP representation with cells colored by cartridge origin (technical replicates). (D) CD56 surface expression (left) and NCAM1 gene expression (right) are visualized across the annotated cell cluster populations.

Figure S3

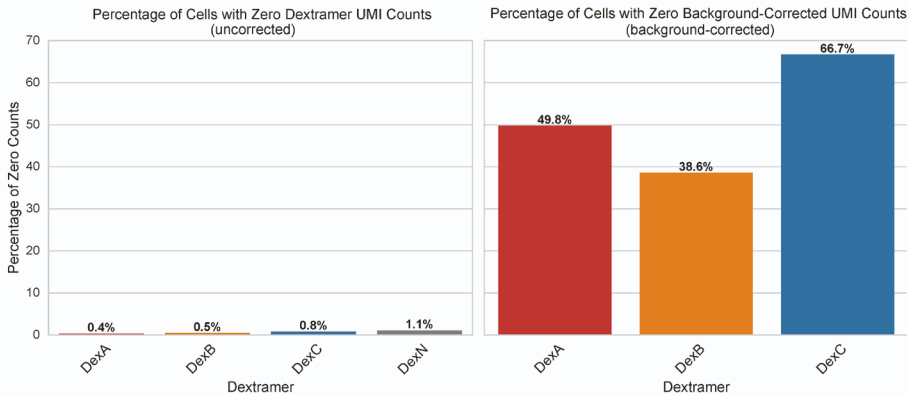
A



B



C



D

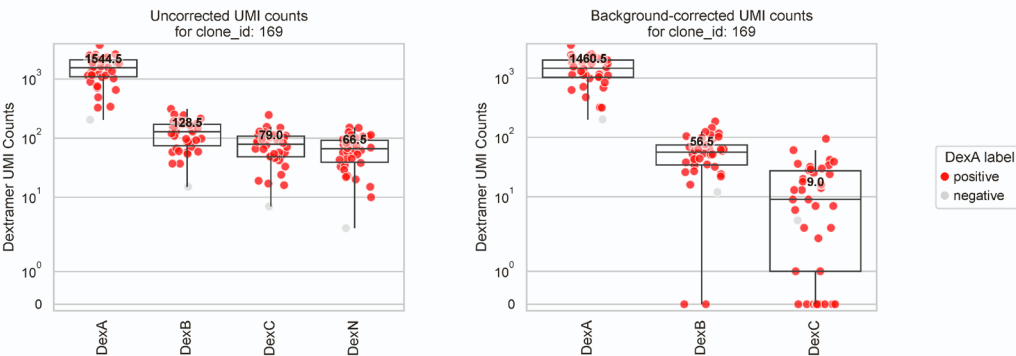


Figure S3. Assignment of dextramer-positive cells and comparison of background in AbSeq and dCODE dextramer sequencing data, related to STAR methods.

(A) Histogram and density plots for distribution of noise-corrected CLR-normalized DexA, DexB and DexC counts. Blue line indicates cutoff for Dex⁺ label (3.94 for DexA, 3.78 for DexB and 3.99 for DexC), as determined by k-means clustering (k=4). The cutoff was defined as the lowest value of the cluster with the highest mean expression. (B) Barplot depicting for each of the 31 surface markers the percentage of cells with zero UMI counts within the full dataset. (C) Barplots showing for each dextramer the percentage of cells with zero UMI counts within the full dataset. Percentages for raw, uncorrected dextramer UMI counts are depicted on the left, and for background-corrected UMI counts on the right. On the left, the negative control dextramer DexN is included. The category does not exist on the right side because DexN UMI counts have been subtracted from the other dextramer counts on a per-cell basis. (D) Boxplots of raw uncorrected (left) and background-corrected UMI counts (right) of each dextramer for all cells of one exemplary clonotype (clone_id 169). The y-axis is displayed on a symlog scale. The boxplots are in the style of Tukey, where the center of the box represents the median of values, hinges represent the 25th and 75th percentile, and the whiskers are extended not further than the 1.53IQR (inter-quartile range). The median is highlighted on top of each boxplot. The dots are colored by whether the corresponding cell was previously classified as DexA-positive or not.

Figure S4

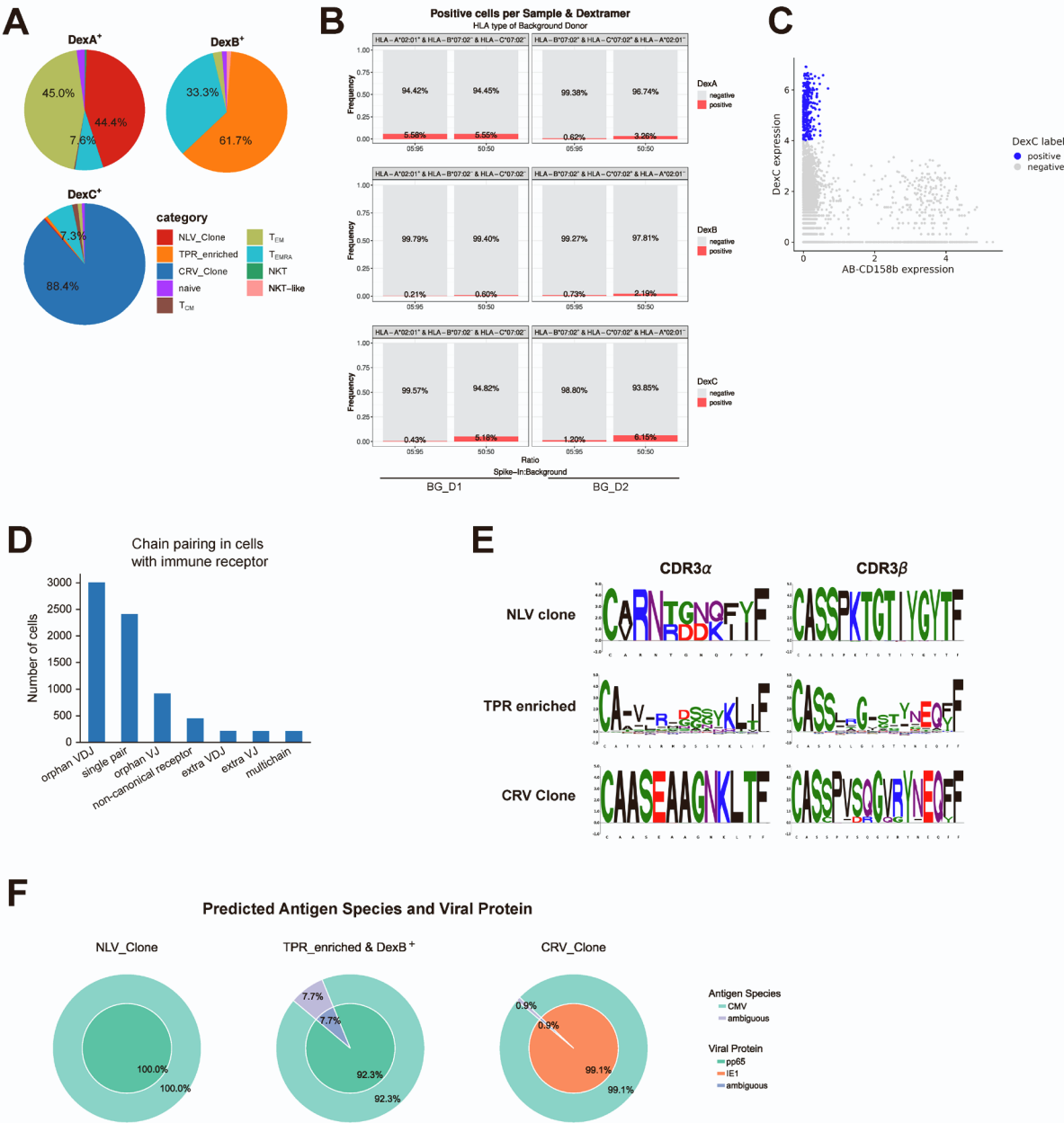


Figure S4. Characterization and TCR repertoires of CMV-specific CD8⁺ T cells, related to Figure 3 and 4.

(A) Dextramer-positive cells in each staining condition. Barplots indicating frequency of Dex⁺ cells (for DexA, DexB and DexC) in each of the four staining conditions: Spike-In to Background Ratio of 05:95 and 50:50, with background originating from Donor 1 (BG_D1, HLA-matched to DexA) or Donor 2 (BG_D2, HLA-matched to DexB and DexC). (B) Pie charts showing cluster category origin of Dex⁺ cells. (C) Co-expression of DexC and CD158b (AbSeq) per cell. CLR-normalized expression values are plotted and cells defined as DexC⁺ are highlighted. (D) Immune receptor sequence types in the dataset. Barplot depicting cell numbers categorized by type of chain pairing. Applies only to cells, which had both a recorded immune receptor sequence and paired RNA-seq information. (E) CDR3 sequence logos depicting under- and over-represented amino acids of CDR3 α and CDR3 β amino acid sequences. Logo plots are shown for samples carrying the flex-tag labels for the NLV clone, TPR enriched and CRV clone input material. Colours indicate amino acid chemistry. (F) Antigen specificity predictions for the NLV and CRV clones, as well as DexB⁺ cells from the TPR-enriched sample, were made by matching CDR3 amino acid VJ and VDJ sequences to the VDJdb reference database. A match was identified if similarity was found in either the VJ or VDJ chain.