**Supplementary Information**

# High-throughput determination of the antigen specificities of T cell receptors in single cells

Shu-Qi Zhang[1†], Ke-Yue Ma[2†], Alexandra A. Schonnesen[3†], Mingliang Zhang[4,5†], Chenfeng He[3], Eric Sun[3], Chad M. Williams[3], Weiping Jia[4,5*], Ning Jiang[2,3,6*]

[1]McKetta Department of Chemical Engineering, University of Texas at Austin, Austin, TX 78712, USA

[2]Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX 78712, USA

[3]Department of Biomedical Engineering, University of Texas at Austin, Austin, TX 78712, USA

[4]Department of Endocrinology and Metabolism, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai 200233, China

[5]Shanghai Key Laboratory of Diabetes Mellitus, Shanghai Clinical Center of Diabetes, Shanghai 200233, China

[6]LIVESTRONG Cancer Institutes, Dell Medical School, University of Texas at Austin, Austin, TX 78712, USA

[†]These Authors contributed equally

[*]Correspondence should be addressed to:

Ning Jiang, Ph.D.

Email: jiang@austin.utexas.edu

Phone: 512-471-4860

Weiping Jia, M.D., Ph.D.

Email: wpjia@sjtu.edu.cn

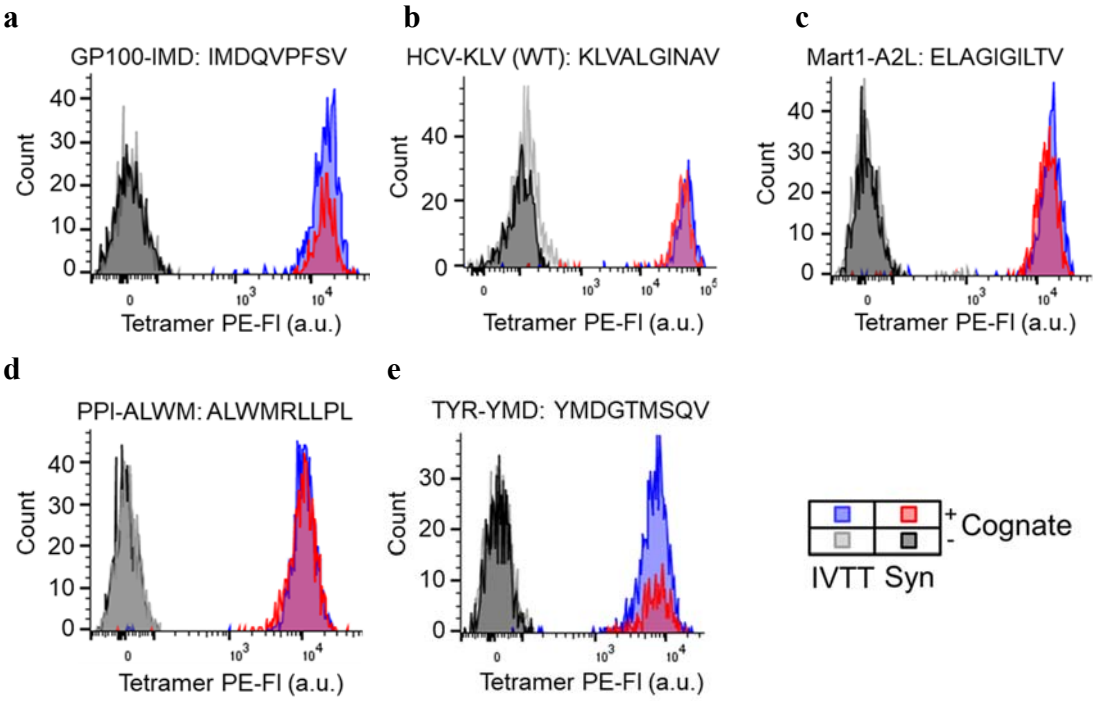
Phone: 86-21-64369181-8922

**Criteria for peptide classification: MID threshold and signal-to-noise ratio**

In order to characterize the non-specific binding level of DNA-BC peptides to T cells, we define a peptide to be positively binding if the fluorescence intensity of the corresponding pMHC tetramer is above background level, which is set using the flow through fraction after tetramer enrichment. To measure background, we sorted fluorescent tetramer negative (Tetramer$^-$) single CD8+ T cells from the tetramer enriched fraction and measured the number of MIDs associated with each of the non-specifically bound peptides. Results show that these non-specific bound DNA-BCs from Tetramer$^-$ single cells have low MID counts associated with each peptide (Fig. 1d and Supplementary Fig. 11a, 13a, 16a, 16e).

The first criteria we applied to detect positively bound peptides from background level of non-specific binding is a MID count threshold. We define this threshold to be the maximum MID count-per-peptide from the Tetramer$^-$ population with an added 25% buffer, rounded to the nearest tens digit (dashed lines in Fig. 1d and Supplementary Fig. 11a, 13a, 16a, 16e). This value was determined for each TetTCR-Seq experiment.

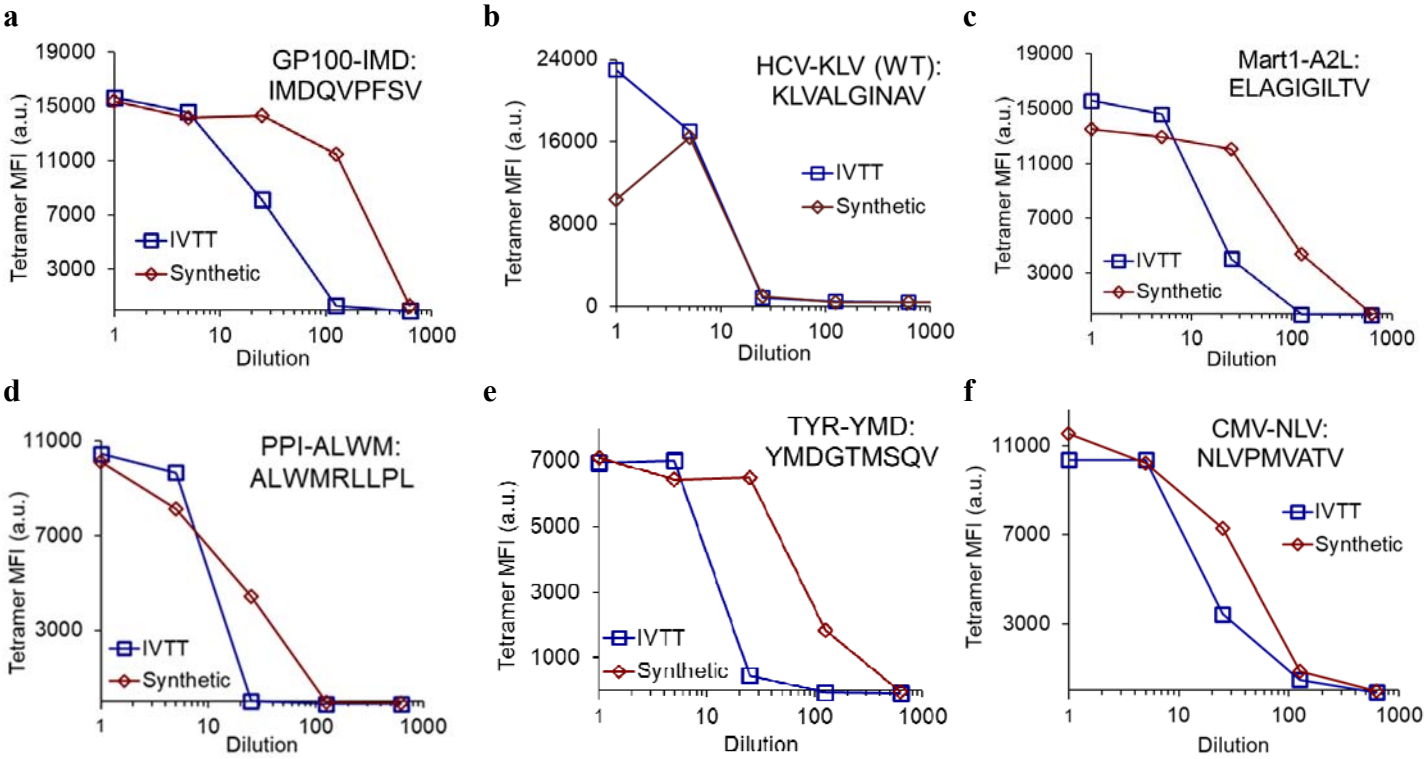
The second criteria we used for each cell is a signal-to-noise ratio between two borderline peptides, which is defined to be the ratio of the peptide with the lowest MID count above the MID threshold to the peptide with the highest MID count below the MID threshold. We used the spike-in clone from Experiment 1 as the positive control for the MID counts associated with positive and negatively binding peptides, which we validated using traditional tetramer staining (Fig. 1f and Supplementary Fig. 8). By aggregating all cells from this spike-in clone, the signal-to-noise ratio ranged from 3.6:1 to 61:1. Using this as a guide, we set the signal-to-noise ratio to be greater than 2:1; Cells with a signal-to-noise ratio below this threshold was removed from analysis because the segregation in MID counts between positive and negative binding peptides was too low.




Supplementary Figure 1

pMHC tetramers produced by IVTT has similar staining performance as the conventional method using chemically synthesized peptide (Syn)

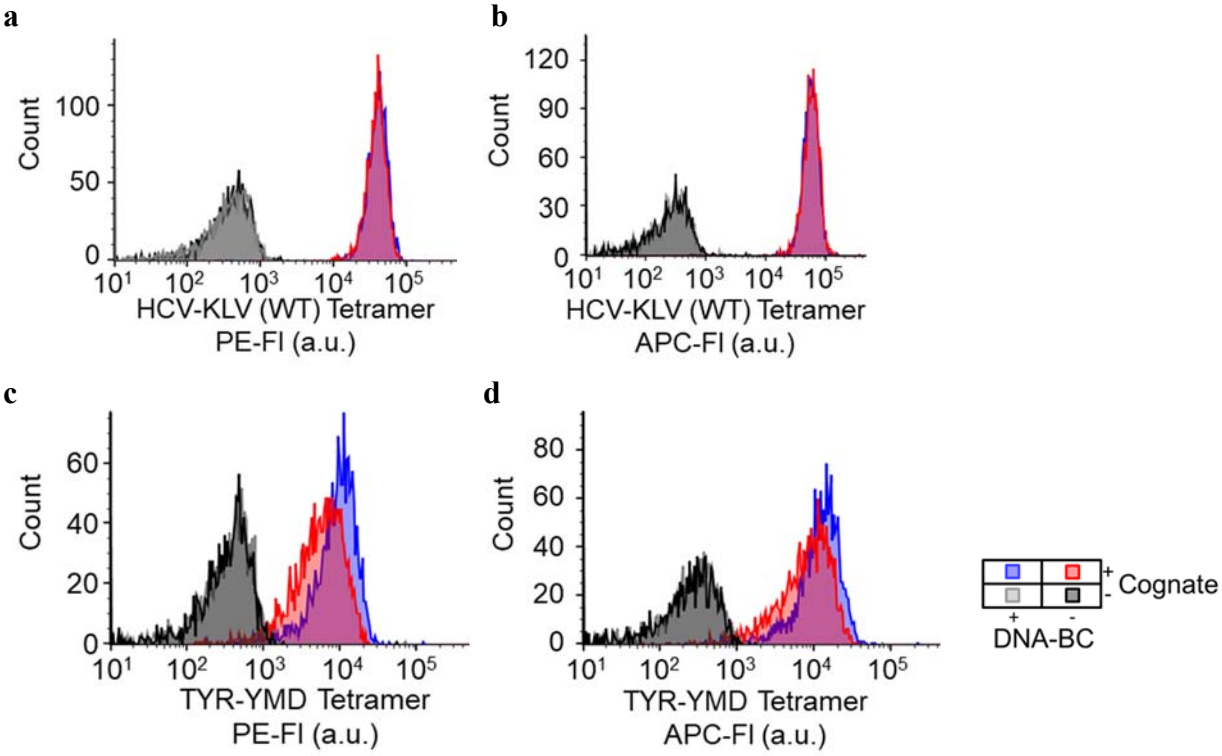
**(a-e)** Same experiment as Fig. 1c for 5 additional peptides. pMHC tetramers, containing the indicated peptide, were generated using IVTT or Syn method and used to stain a cognate and non-cognate T cell clone. Anti-CD8a (RPA-T8) was present throughout the staining. Experiments were repeated once independently with similar results.




Supplementary Figure 2

IVTT can generate 20-100 μM of the desired peptide

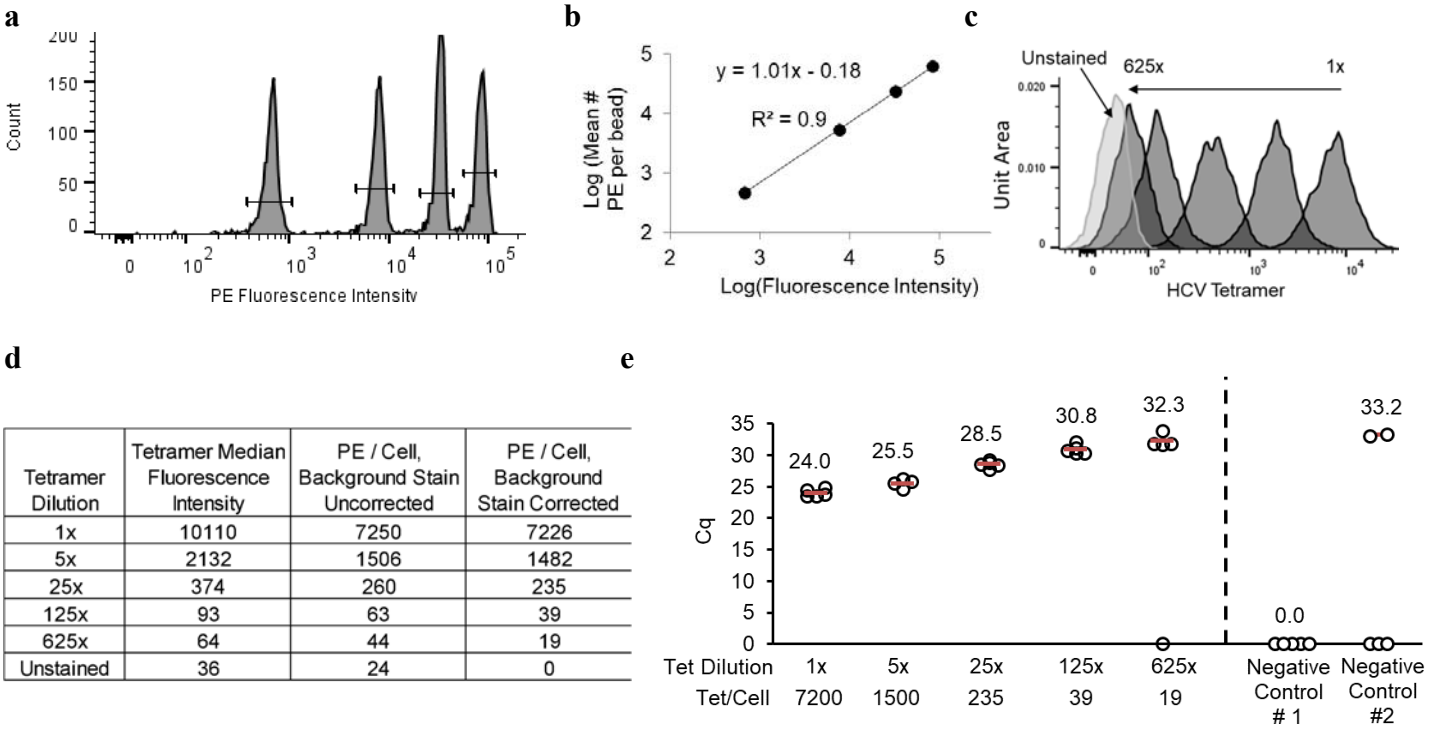
**(a-f)** Peptides generated from either IVTT or the traditional, synthetic peptide method, were diluted at different ratios and were used to form PE labeled pMHC tetramers. Starting concentration of synthetic peptide is 100 μM for all peptides. These pMHC tetramers were used to stain a cognate T cell clone. Anti-CD8a (RPA-T8) was present throughout the staining. MFI, Median Fluorescence Intensity. a.u., arbitrary unit. IVTT can generate 20-100 μM of the desired peptide, which is in the concentration range commonly used for UV-mediated peptide exchange[1]. With other published UV-exchangeable peptides for other HLA alleles[2], we estimated that ~83% of individuals in the world carry at least one HLA allele that is amenable to UV-mediated peptide exchange technology. Panel (b),(d), and (f) was repeated independently once with similar results. Experiments in panel (a), (c), (e) was performed once.




Supplementary Figure 3

Covalent attachment of DNA-BC to PE and APC streptavidin does not affect staining intensity of the resulting tetramers.

**(a-d)** PE and APC labeled streptavidin were covalently attached with DNA linker. An oligonucleotide encoding HCV-KLV(WT) was annealed to streptavidin-conjugated DNA linker and extended to form DNA-BC. DNA-BC pMHC tetramers were formed with either the HCV-KLV(WT) or TYR-YMD peptide and with either PE or APC streptavidin scaffold, as indicated. Resulting tetramers were used to stain a cognate and non-cognate T cell clone. Anti-CD8a (RPA-T8) was present throughout the staining. Fl, fluorescence Intensity. a.u., arbitrary unit. Experiments were repeated once independently with similar results.




**d**

| Tetramer Dilution | Tetramer Median Fluorescence Intensity | PE / Cell, Background Stain Uncorrected | PE / Cell, Background Stain Corrected |
|---|---|---|---|
| 1x | 10110 | 7250 | 7226 |
| 5x | 2132 | 1506 | 1482 |
| 25x | 374 | 260 | 235 |
| 125x | 93 | 63 | 39 |
| 625x | 64 | 44 | 19 |
| Unstained | 36 | 24 | 0 |

Supplementary Figure 4

Quantification of the detection limit of DNA-BC pMHC tetramers

**(a)** Fluorescence of PE-Quantibrite beads that were used for **(b)** calibration of PE fluorescence intensity to protein abundance. Brackets in (a) denote the gates for each the four Quantibrite bead populations (n = 4 bead populations). **(c)** PE labeled, DNA-BC pMHC tetramers containing the HCV-KLV(WT) peptide (with the DNA-BC corresponding to HCV-KLV(WT) sequence) was used to stain a cognate T cell clone at the indicated tetramers dilutions starting at 5 µg/ml for 1x. Anti-CD8a (RPA-T8) was present throughout the staining. **(d)** Calculation of tetramer abundance on each of the staining dilutions from (c) using the calibration curve from (b). Corrected value indicates subtraction of background value from the unstained cell population. **(e)** qPCR of DNA-BC on single cells sorted from various populations.

Detailed information on panel (e):

Tet Dilution 1x – 625x: The 5 tetramer dilutions from (c), amplified with primers specific for DNA-BC encoding the HCV-KLV(WT) sequence

Negative control #1: GP100-IMD binding T cell clone that has been stained with 1x dilution of the DNA-BC HCV-KLV(WT) tetramer as in (c), amplified with primers specific for DNA-BC encoding the HCV-KLV(WT) sequence

Negative control #2: Two PE labeled DNA-BC pMHC tetramer were made containing the HCV-KLV(WT) or GP100-IMD peptide. Each tetramer contains a DNA-BC sequence that corresponds to the peptide. The two tetramers were pooled and used to stain the HCV-KLV(WT) binding clone in (c) at 5 µg/ml each (none diluted). qPCR was performed using primers specific for DNA-BC encoding GP100-IMD only (which corresponds to bound GP100-IMD tetramer). 2 of the 5 cells picked showed a Cq value of 33.2, which is one cycle away from the results obtained using 625x diluted tetramer. This result suggests that in a mixture, DNA-BC hopping from one tetramer to another is very low.

Each circle indicates a qPCR reaction with one sorted cell. 0 Cq value represents no detected amplification after 40 cycles. Red bars indicate the mean Cq value for positively amplified cells.

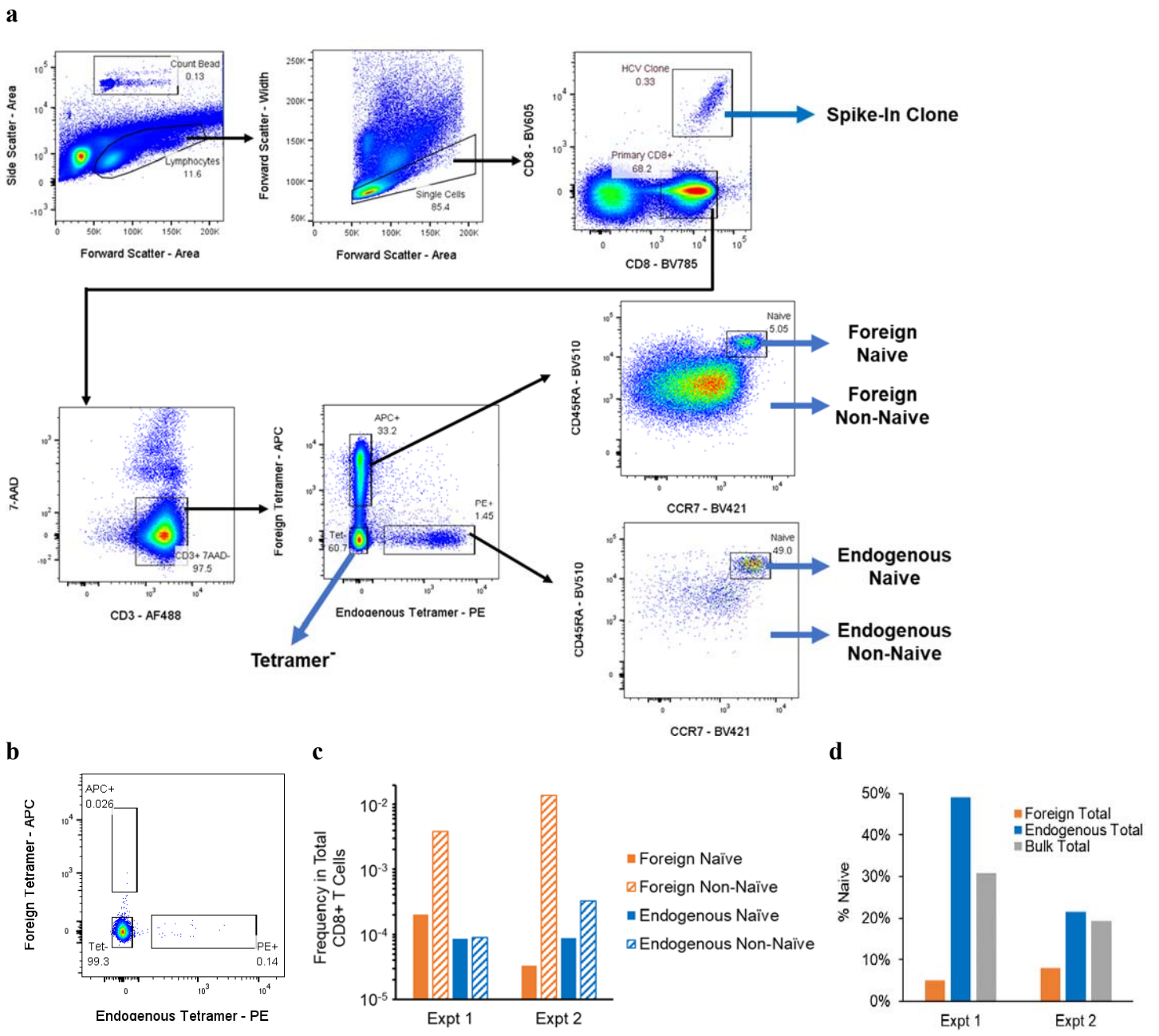
| | | | | Primary CD8+ T Cells (Millions) | | | | Sequence Recovery Efficiency | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Expt | Expt Type | Age, Gender, CMV[a] | pMHC Library[b] | Flow-Through | Tetramer Enriched | Sorted Population | Cells Sorted | TCRα[c] | TCRβ[c] | TCRαβ[c] | DNA-BC[d] | Antigen Detection[e] | Relevant Figures |
| 1 | 96 Foreign Endogenous | 30M(+) | 29 Foreign (APC)<br>61 Endogenous (PE)<br>5 HCV-KLV + Mut. (APC)<br>1 Neg. Ctrl (PE, APC)[f] | 42 | 0.51 | Foreign Naïve | 56 | 28 (50%) | 36 (64%) | 20 (36%) | 56 (100%) | 50 (89%) | Main: 1b,1d,1e,1g, 1h,1i<br>Supp: 6,7,8 |
| | | | | | | Foreign Non-Naïve | 32 | 13 (41%) | 19 (59%) | 10 (31%) | 32 (100%) | 32 (100%) | |
| | | | | | | Endogenous Naïve | 56 | 37 (66%) | 45 (80%) | 34 (61%) | 56 (100%) | 55 (98%) | |
| | | | | | | Endogenous Non-Naïve | 23 | 9 (39%) | 12 (52%) | 4 (17%) | 23 (100%) | 23 (100%) | |
| | | | | | | HCV-KLV Specific Clone | 8 | 8 (100%) | 8 (100%) | 8 (100%) | 8 (100%) | 8 (100%) | |
| | | | | | | Tetramer$^-$ | 8 | n/a | n/a | n/a | 5 (63%) | 0 (0%) | |
| 2 | 96 Foreign Endogenous | 51M(+) | 29 Foreign (APC)<br>61 Endogenous (PE)<br>6 HCV-KLV + Mut. (APC)[g] | 72 | 1.14 | Foreign Naïve | 96 | 74 (77%) | 78 (81%) | 59 (61%) | 96 (100%) | 85 (79%) | Supp: 6,9 |
| | | | | | | Foreign Non-Naïve | 88 | 67 (76%) | 62 (70%) | 54 (61%) | 88 (100%) | 84 (95%) | |
| | | | | | | Endogenous Naïve | 96 | 75 (78%) | 81 (84%) | 64 (67%) | 96 (100%) | 92 (96%) | |
| | | | | | | Endogenous Non-Naïve | 88 | 79 (90%) | 83 (94%) | 77 (88%) | 87 (99%) | 75 (85%) | |
| | | | | | | HCV-KLV Specific Clone | 8 | 7 (88%) | 7 (88%) | 7 (88%) | 8 (100%) | 7 (88%) | |
| | | | | | | Tetramer$^-$ | 8 | n/a | n/a | n/a | 7 (88%) | 0 (0%) | |
| 3[h] | 40 Neoantigen Wildtype | 56M(-)<br>65M(-) | 20 Neoantigen (APC)<br>20 Wildtype (PE)<br>1 HCV-KLV (PE, APC)<br>1 Neg. Ctrl (PE, APC)[f] | 57 | 0.35 | Neo$^+$WT$^-$ | 142 | 112 (79%) | 130 (92%) | 106 (75%) | 142 (100%) | 127 (89%) | Main: 2a-c<br>Supp: 10-12,14 |
| | | | | | | Neo$^-$WT$^+$ | 43 | 36 (84%) | 34 (79%) | 30 (70%) | 43 (100%) | 43 (100%) | |
| | | | | | | Neo$^+$WT$^+$ | 76 | 61 (80%) | 71 (93%) | 59 (78%) | 76 (100%) | 71 (93%) | |
| | | | | | | HCV-KLV Specific Clone | 12 | 12 (100%) | 12 (100%) | 12 (100%) | 12 (100%) | 12 (100%) | |
| | | | | | | Tetramer$^-$ | 12 | n/a | n/a | n/a | 10 (83%) | 0 (0%) | |
| 4[h] | 40 Neoantigen Wildtype | 50M(-)<br>56F(-) | 20 Neoantigen (APC)<br>20 Wildtype (PE)<br>4 MAGE-A (PE, APC)[i] | 107 | 0.23 | Neo$^+$WT$^-$ | 144 | 34 (24%) | 33 (23%) | 12 (8%) | 144 (100%) | 144 (100%) | Supp: 10,13,14 |
| | | | | | | Neo$^-$WT$^+$ | 44 | 16 (36%) | 11 (25%) | 6 (14%) | 44 (100%) | 44 (100%) | |
| | | | | | | Neo$^+$WT$^+$ | 108 | 30 (28%) | 31 (29%) | 11 (10%) | 108 (100%) | 95 (88%) | |
| | | | | | | Tetramer$^-$ | 35 | n/a | n/a | n/a | 13 (37%) | 0 (0%) | |
| 5 | 315 Neoantigen Wildtype | 47F(-) | 158 Neoantigen (PE)[j]<br>157 Wildtype (APC)<br>1 HCV-KLV (PE, APC)<br>1 Neg. Ctrl (PE, APC)[f] | 80 | 0.38 | Neo$^+$WT$^-$ | 221 | 136 (62%) | 137 (62%) | 112 (51%) | 215 (97%) | 197 (89%) | Main: 2d-f<br>Supp: 15-17 |
| | | | | | | Neo$^-$WT$^+$ | 312 | 172 (55%) | 183 (59%) | 134 (43%) | 301 (96%) | 186 (60%) | |
| | | | | | | Neo$^+$WT$^+$ | 255 | 140 (55%) | 150 (59%) | 108 (42%) | 249 (98%) | 189 (74%) | |
| | | | | | | HCV-KLV Specific Clone | 8 | 6 (75%) | 6 (75%) | 6 (75%) | 7 (88%) | 7 (88%) | |
| | | | | | | Tetramer$^-$ | 8 | n/a | n/a | n/a | 7 (88%) | 0 (0%) | |
| 6 | 315 Neoantigen Wildtype | 58M(-) | 158 Neoantigen (PE)[j]<br>157 Wildtype (APC)<br>1 HCV-KLV (PE, APC)<br>1 Neg. Ctrl (PE, APC)[f] | 30 | n/a[k] | Neo$^+$WT$^-$ | 118 | 97 (82%) | 99 (84%) | 86 (73%) | 118 (100%) | 118 (100%) | |
| | | | | | | Neo$^-$WT$^+$ | 68 | 53 (78%) | 58 (85%) | 46 (68%) | 68 (100%) | 66 (97%) | |
| | | | | | | Neo$^+$WT$^+$ | 82 | 62 (76%) | 67 (82%) | 52 (63%) | 82 (100%) | 72 (88%) | |
| | | | | | | Tetramer$^-$ | 6 | n/a | n/a | n/a | 1 (17%) | 0 (0%) | |

[a]Number denote age, M: male, F: female, (-) denote CMV seronegative, (+) denote CMV seropositive
[b]Detailed summary in Supplementary Table. Shown is the number of peptides, peptide category, and fluorescent encoding.
[c]Includes only cells containing productive TCRα and/or TCRβ sequences are included
[d]Includes only cells with at least 100 reads of DNA-BC and this applies to Tetramer$^-$ cells as well.
[e]Includes only cells with at least one detected antigen from the MID threshold criteria
[f]A DNA-BC pMHC tetramer UV-exchanged with a non HLA-A2 binding peptide, RLFAFVRFT
[g]The library is the same as Expt 1, except for the replacement of the negative control peptide with an additional HCV-KLV mutant peptide, HCV-A9N. This peptide did not bind to our HCV-KLV Specific clone in a separate tetramer staining, and serves as a negative control.

[h]Blood samples from two donors were pooled together in Experiment 3 and 4
[i]The library is the same as Expt 3, except for the replacement of the negative control and HCV-KLV peptide with 4 peptides from the MAGE-A antigen family. 3 MAGE-A specific T cells were detected out of 298 cells and were not used for subsequent analysis.
[j]Neo-antigen/WT pairs are used for all antigens except for DHX33-LLA, which have two neo-antigens with substitutions K5T and M4I. One T cell was found to be cross-reactive to all three peptides
[k]Cell count could not be determined because counting beads were not added and the sample was not recorded in its entirety.

## Supplementary Figure 5

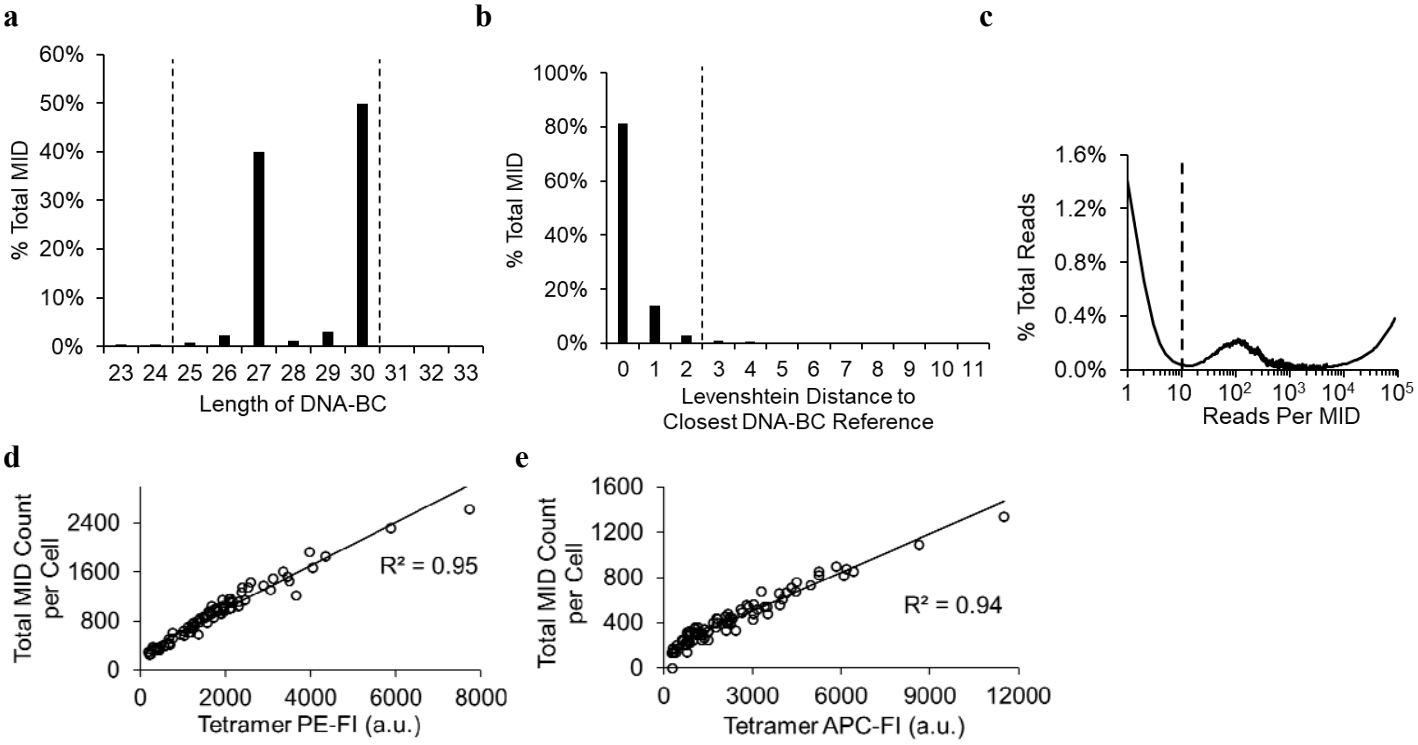
Summary of the 6 main TetTCR-Seq experiments performed and blood donor characteristics. The percentage difference between "DNA-BC" column and "Antigen Detection" column are those T cells without identified binding antigen based on the criteria listed at the beginning of the Supplementary Information. These T cells correspond to grey lines in all the peptide rank curves.




Supplementary Figure 6

Gating scheme and sorting strategy for Experiment 1 and 2

**(a)** Representative gating scheme for Experiment 1 and 2. Shown is gating scheme for Experiment 1. Single-cell lymphocytes were first gated. The HCV-specific T cell clone spike-in, pre-stained with BV605-CD8a, and the primary T cell population, stained with BV785-CD8a, were isolated. $CD8^+$ T cells were gated to be 7-$AAD^-CD3^+$. Naïve and non-naïve antigen-binding cells were sorted from the $PE^+$, endogenous peptides and $APC^+$, foreign peptides. The same antibody panel and gating scheme is used for Experiment 2. **(b)** Tetramer staining of flow-through fraction was used to set the PE and APC tetramer negative and positive gates. An example from Experiment 1 is shown, and a similar result was obtained in the repeat, Experiment 2. **(c)** Frequency of the four antigen-binding T cell populations for Experiment 1 and 2. **(d)** Percent of naïve cells from Foreign and Endogenous $Tetramer^+$ CD8+ T cells for Experiment 1 and 2. Bulk indicates flow-through CD8+ T cells from the same experiment. **(d)** Frequency of the four antigen-binding T cell populations for Experiment 1 and 2.




Supplementary Figure 7

Processing of DNA-BC sequencing reads for Experiment 1

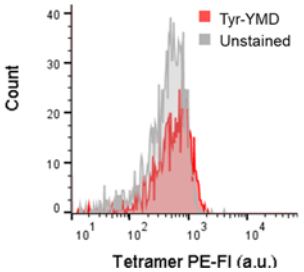
Reads within the same cell barcode that have the same MID sequence were clustered together and were considered as one MID. A consensus peptide-encoding sequence was generated for each cluster. **(a)** MIDs were filtered to only include those having the peptide-encoding sequence be a length of 25-30. All peptides used were 9-10 AA in length, so the DNA length should be 27 and 30. **(b)** MIDs were then filtered such that the closest Levenshtein distance of the peptide-encoding sequence to the reference DNA-BC list is no greater than 2. **(c)** Percent of total reads belonging to each group of MIDs sharing the same read count. MIDs with low read counts (left of the vertical dashed line) were discarded as sequencing error using published strategy[3]. The resulting MIDs can then be assigned to each sorted T cell according to the cell barcode. **(d,e)** Total MID counts associated with each cell from the $PE^+$ (d) and $APC^+$ (e) populations from Experiment 1 were compared to their corresponding tetramer staining intensity from index sorting analysis. Each circle denotes one cell. Line indicates linear regression and the associated R-squared value (n = 79 cells for d, 88 cells for e).

The total MID counts for each single cell correlates with the tetramer fluorescent staining intensity.

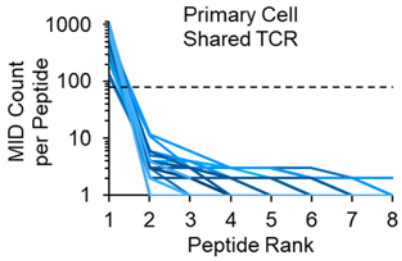
**a**

| Peptide Rank | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| HCV-K1Y | HCV-L2I | HCV-K1Y17V | HCV-K1S | HCV-KLV(WT) | PD5-KLS | SODA-DMW | IA2-SLY | HSV-SLP | GLNS-GLL |
| HCV-K1Y | HCV-L2I | HCV-K1Y17V | HCV-K1S | HCV-KLV(WT) | FBA-ALS | MAGEA10-GLY | EMPTY | HAFP-GLS | IA2-SLY |
| HCV-K1Y | HCV-L2I | HCV-K1Y17V | HCV-K1S | HCV-KLV(WT) | PD5-KLS | HTLV-GLL | ZNT8-VVT | CMV-NLV | IGRP-FLF |
| HCV-K1Y | HCV-L2I | HCV-K1S | HCV-K1Y17V | HCV-KLV(WT) | HAFP-GLS | TYR-YMD | IGRP-RLL | KER-ALL | ALADH-VLM |
| HCV-K1Y | HCV-L2I | HCV-K1Y17V | HCV-K1S | HCV-KLV(WT) | EBV-CLG | HAFP-GLS | HSV-SLP | HTLV-LLF | ZNT8-VAA |
| HCV-K1Y | HCV-L2I | HCV-K1S | HCV-K1Y17V | HCV-KLV(WT) | HCV-YLL | CMV-VLE | INS-HLV | HAFP-GLS | EF2-ILT |
| HCV-K1Y | HCV-L2I | HCV-K1Y17V | HCV-K1S | HCV-KLV(WT) | HCV-YLL | AGL-GLI | WT1-RMF | PPI-15-24 | HAFP-GLS |
| HCV-K1Y | HCV-L2I | HCV-K1Y17V | HCV-K1S | HCV-KLV(WT) | TYR-CLL | HPV-YML | EBV-CLG | SNPG-IML | PPI-ALWM |

**b**




**c**




**d**

| | Peptide Rank | | | | | TCRα | | TCRβ | |
|---|---|---|---|---|---|---|---|---|---|
| Cell Name | 1 | 2 | 3 | 4 | 5 | TRAV | TCRα | TRBV | CDR3β |
| B_E8 | MART1_A2L | IGRP_FLF | GLNS_GLL | PPI_15_23 | HBV_WLS | 12-2*01 | CAVTAGGTSYGKLTF | | |
| B_E9 | MART1_A2L | HCHGA_LLC | GAD65_VMN | MEA_SMY | FBA_ALS | 12-2*01 | CAVTAGGTSYGKLTF | | |
| B_F11 | MART1_A2L | GAD_RMM | IAPP_FLI | IGRP_FLF | | 12-2*01 | CAVTAGGTSYGKLTF | | |
| B_F12 | MART1_A2L | HAFP_GLS | INS_SHL | FBA_ALS | NYESO1_V165 | 12-2*01 | CAVTAGGTSYGKLTF | | |
| B_F5 | MART1_A2L | GAD65_VMN | IA2_SLS | L19_ILM | EMPTY | 12-2*01 | CAVTAGGTSYGKLTF | | |
| A_H1 | IV_GIL | HCV_LLF | SODA_DMW | IGRP_RLL | IAPP_FLI | 27*01 | CAGGGSQGNLIF | | |
| B_D5 | IV_GIL | SODA_DMW | EBV_CLG | HA_VLH | HCV_CIN | 27*01 | CAGGGSQGNLIF | | |
| B_C1 | IV_GIL | EMPTY | SMCY_FID | | | | | 19*01 | CASWDRGNEQFF |
| B_D8 | IV_GIL | EBV_YLQ | BRA_AG | EBV_CLG | DDX5_YLL | 30*01 | CGTLRNNNARLMF | 19*01 | CASWDRGNEQFF |
| B_D3 | IV_GIL | MEA_SMY | IV_AIM | MAGEC2_LLF | HCV_L2I | 30*01 | CGTLRNNNARLMF | | |
| B_C11 | CMV_NLV | IGRP_RLL | PPI_ALWM | HCV_L2I | HCV_K1Y | 3*01 | CADYYGQNFVF | 28*01 | CASSFQGYTEAFF |
| B_C5 | CMV_NLV | BRA_AG | HCV_K1S | HSV_SLP | EBV_YLQ | 3*01 | CADYYGQNFVF | 28*01 | CASSFQGYTEAFF |
| B_C9 | CMV_NLV | IAPP_KLQ | HCV_CIN | HAFP_GLS | IV_GIL | 3*01 | CADYYGQNFVF | 28*01 | CASSFQGYTEAFF |
| B_B11 | CMV_NLV | IAPP_KLQ | IV_GIL | IVPA_FMY | CMV_VLE | | | 28*01 | CASSFQGYTEAFF |
| B_C4 | CMV_NLV | IV_GIL | EBV_YVL | ZNT8_LLS | ALADH_VLM | | | 28*01 | CASSFQGYTEAFF |
| B_E4 | CMV_NLV | PP1_SII | HCV_FLP | MEA_ILP | EMPTY | | | 28*01 | CASSFQGYTEAFF |
| B_C2 | CMV_NLV | EBV_YVL | TYR_CLL | HSV_SLP | EF2_ILT | 35*01 | CAGPMKTSYDKVIF | 12-3*01/12 | CASSSANYGYTF |
| B_D1 | CMV_NLV | GP100_YLE | MEA_ILP | IGRP_VLF | HCV_CIN | 35*01 | CAGPMKTSYDKVIF | 12-3*01/12 | CASSSANYGYTF |
| B_D2 | CMV_NLV | WT1_RMF | HBV_WLS | HCV_KLV | EBV_YLQ | 35*01 | CAGPMKTSYDKVIF | 12-3*01/12 | CASSSANYGYTF |
| B_C3 | CMV_NLV | ZNT8_VAA | INS_SHL | IA2_SLA | MEA_ILP | | | 12-3*01/12 | CASSSANYGYTF |
| B_D10 | CMV_NLV | HCV_FLP | HCV_KLV | HCV_K1Y | HTLV_GLL | | | 12-3*01/12 | CASSSANYGYTF |
| B_D7 | CMV_NLV | AGL_GLI | MEA_SMY | CMV_MLN | EBV_YLQ | | | 12-3*01/12 | CASSSANYGYTF |
| B_E12 | MART1_A2L | GAD65_VMN | HA_VLH | | | | | 27*01 | CASSFAGTTEAFF |
| B_E7 | MART1_A2L | | | | | | | 27*01 | CASSFAGTTEAFF |
| B_F1 | MART1_A2L | MART1_ALM | WT1_RMF | | | | | 27*01 | CASSFAGTTEAFF |
| B_F3 | MART1_A2L | INS_HLV | IGRP_RLL | HTLV_LLF | SODA_DMW | | | 27*01 | CASSFAGTTEAFF |
| B_F4 | MART1_A2L | KER_ALL | GFAP_NLA | HTLV_GLL | MEA_ILP | | | 27*01 | CASSFAGTTEAFF |
| B_F7 | MART1_A2L | GAD65_VMN | HIV_ILK | GP100_KTW | PPI_15_24 | | | 27*01 | CASSFAGTTEAFF |
| B_G2 | MART1_A2L | DDX5_YLL | HIV_SLY | MART1_ALM | HTLV_GLL | 12-2*01 | CAVTAGGTSYGKLTF | 27*01 | CASSFAGTTEAFF |

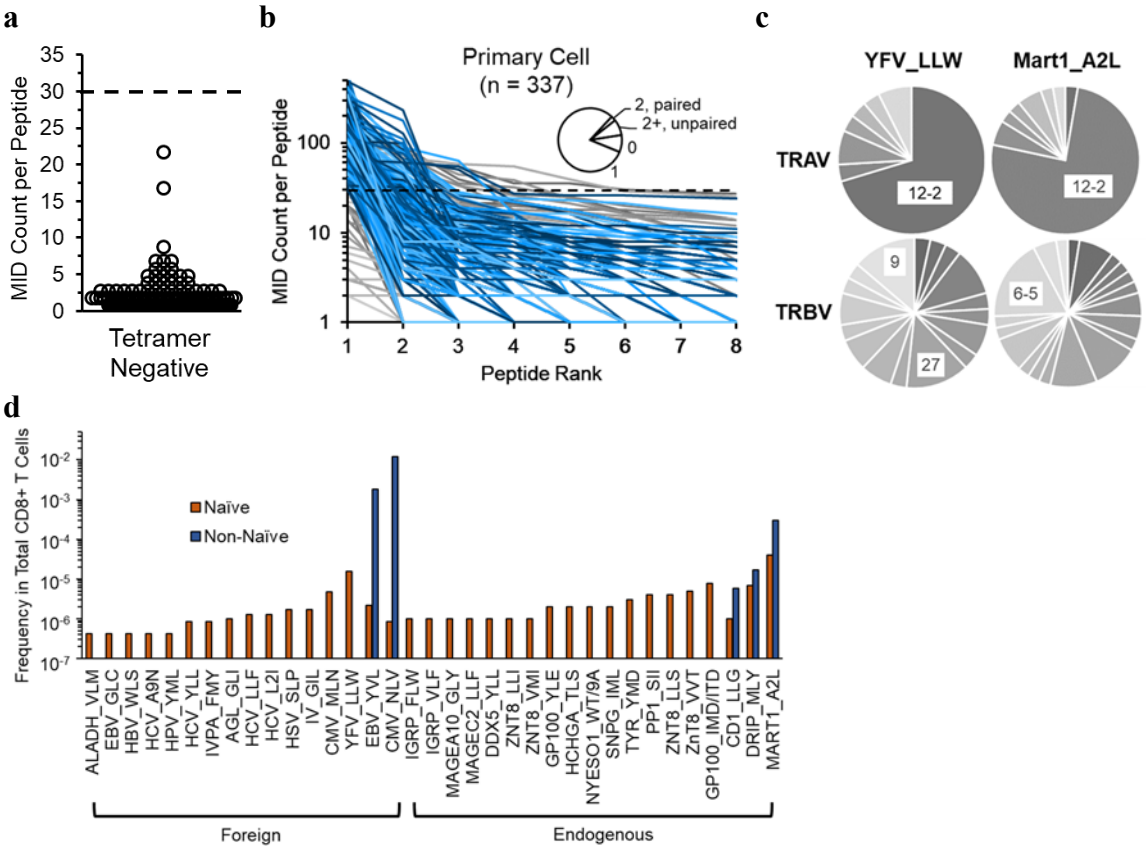
Supplementary Figure 8

Verification of pMHC classification using the spike-in HCV-KLV(WT) binding clone and primary cells with shared TCRs for Experiment 1

**(a)** Raw data of top 10 pMHC specificities of the sorted spike-in HCV-KLV(WT) binding clone, ordered by MID count from high-to-low. Bold border separates detected and non-detected binding peptides by our peptide classification criteria (see the beginning part of the Supplementary Information). The 6$^{th}$ ranked peptide and beyond had significantly lower MID counts than the top 5 (Fig. 1e) and the order of their identities varied from cell-to-cell, which further supports their classification as non-specific binders. Bolded peptides outside the true binding peptide threshold were tested for pMHC tetramer staining in Figure 1f and in panel (b). **(b)** TYR-YMD pMHC tetramer staining of the spike-in clone from Experiment 1 in a separate experiment. Experiment performed once. **(c)** MID count for the top 8 ranked peptides for the Tetramer$^{+}$ primary T cells with shared TCRα and/or TCRβ sequence. Dashed line indicates MID count threshold for identifying positive binding peptides. Experiment

was repeated independently in Experiment 2 with similar results. **(d)** Raw data of top 5 peptides by MID count for T cells sharing at least one TCRα or β chain from (c). Bold border separates positive and non-specific binding peptides.

Results from (c) and (d) shows that, in primary T cells, all cells with shared TCR were classified as binding the same peptide. In addition, similar to (a), peptides that were classified as non-specific varied in their identity for cells with shared TCR sequences and had significantly lower MID counts than the binding peptides.

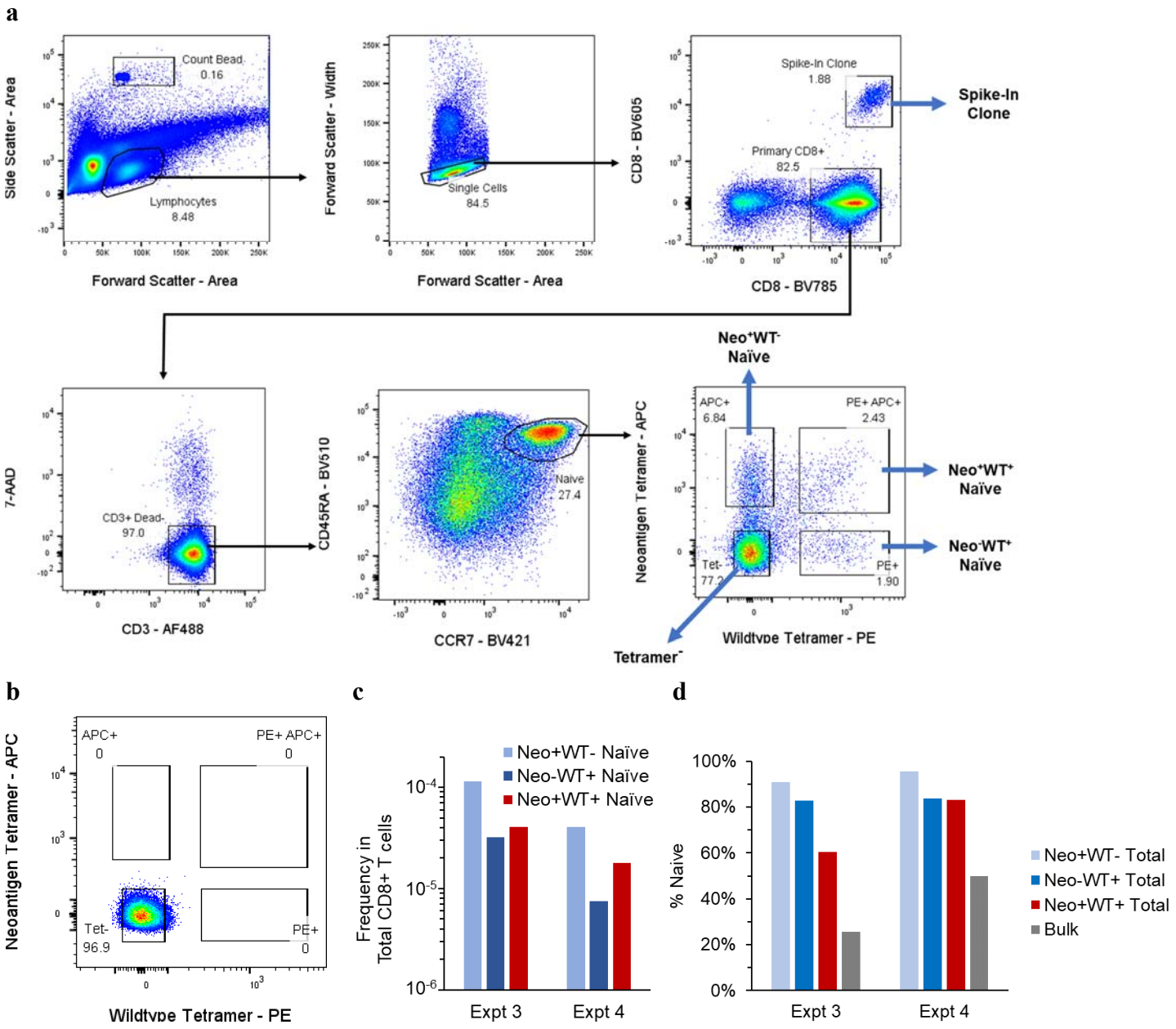



Supplementary Figure 9

Analysis of Experiment 2

**(a)** MID counts greater than 0 from peptides in the Tetramer$^-$ population (n = 8 cells). **(b)** Peptide rank curve by MID counts for all primary T cells. Dashed lines indicate MID threshold for identifying positively bound peptides. Each solid line indicates a cell and only the top 8 peptides were shown ranked by their MID counts. Blue solid lines indicate cells with at least one positively binding peptide; grey solid lines indicate cells that did not positively bind any peptides based on the criteria discussed at the beginning of the supplementary information. Insert pie chart indicate proportion of cells with the indicated number of positively bound peptides. In the insert, paired indicates detection of 2 antigens; one for a wildtype antigen and one for an altered peptide ligand with one amino acid substitution. This was found for GP100 and NY-ESO-1 (Supplementary Table) **(c)** V-gene usage of TCR sequences that are specific for YFV_LLW (n = 27 cells for TRAV, n = 29 cells for TRBV) or MART1_A2L (n = 37 cells for TRAV, n = 39 cells for TRBV). Only distinct TCR sequences were used (one clonal population counts for only one TRAV and/or one TRBV). **(d)** Estimated frequencies of antigen-binding T cell populations in total CD8$^+$ T cells with at least 1 detected cell, separated by phenotype. The same gating strategy in Supplementary Fig. 6a and b is used for Experiment 2.

We found that CMV and EBV-specific T cells accounted for the majority of this donor's non-naïve repertoire, which corroborates the CMV and EBV seropositive status of this individual. In agreement with Experiment 1, we found that, among peptides surveyed, naïve T cells contained greater diversity of antigen specific T cell populations compared to the non-naïve compartment, which is highly skewed towards a select few antigen specific T cell populations. We also found the same dominance in TCRα V gene usage among the MART1-A2L and YFV-LLW specific TCRs in this donor compared to Experiment 1.

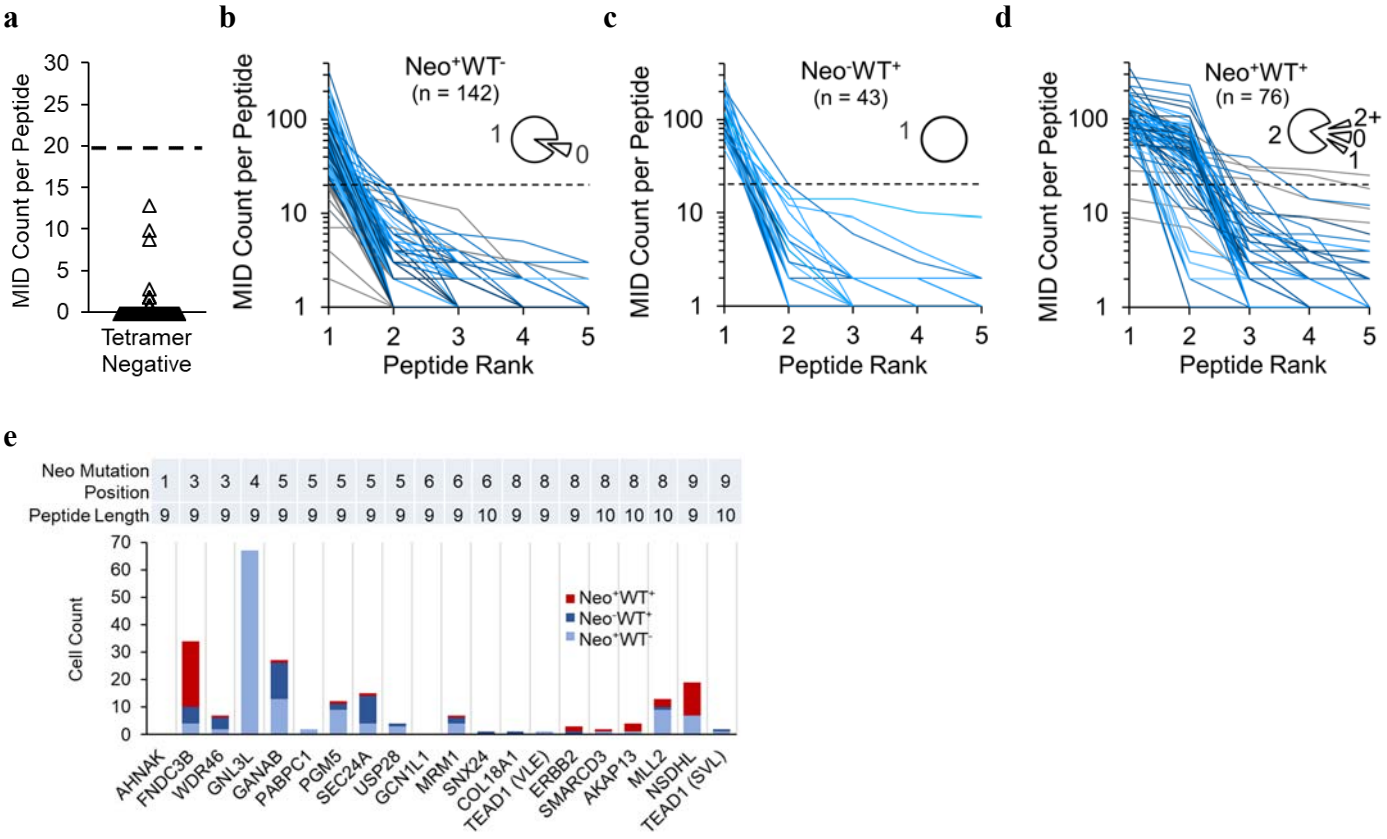



Supplementary Figure 10

Gating scheme and sorting strategy for Experiment 3 and 4

**(a)** Representative gating and sorting scheme for Experiment 3 and 4. Gating scheme for Experiment 3 is shown. **(b)** Tetramer gating on the flow-through fraction of Experiment 3. A similar result was obtained in experimental replicate Experiment 4. **(c)** Estimated frequency of the sorted Tetramer$^+$ populations for Experiment 3 and 4. **(d)** Percentage of naive cells of the indicated Tetramer$^+$ CD8+ T cell population of total Tetramer$^+$ T cells for Experiment 3 and 4. Bulk refers to the flow-through from the same experiment.

A greater proportion of Tetramer binding T cells were found in the naïve phenotype compared to bulk, indicative of no prior exposure to the surveyed antigens.




Supplementary Figure 11

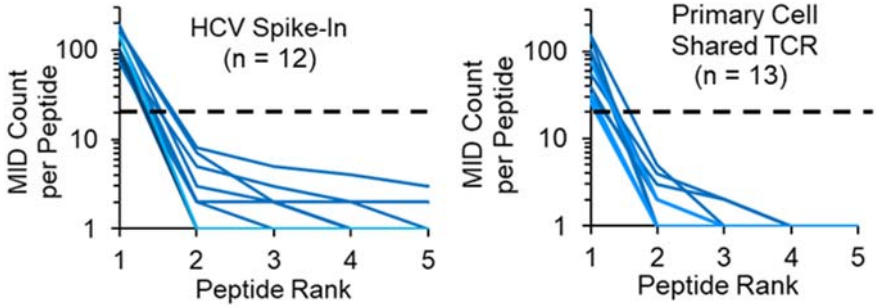
Analysis for Experiment 3

**(a)** MID counts for each peptide from each cell from the Tetramer$^-$ population (12 cells, 42 peptides each). **(b-d)** Peptide rank curve by MID counts for the top 5 peptides for Neo$^+$WT$^-$ (b), Neo$^-$WT$^+$ (c), and Neo$^+$WT$^+$ population (d) for Experiment 3. Dashed lines indicate MID threshold for identifying positively bound peptides. Each solid line indicates a cell and only the top 5 peptides were shown raked by their MID counts. Blue solid lines indicate cells with at least one positively binding peptide; grey solid lines indicate cells that did not positively bind any peptides based on the criteria discussed at the beginning of the supplementary information. Insert pie charts for all three panels indicate proportion of cells with the indicated number of positively bound peptides. **(e)** Cell count for all detected peptides for each Neo-WT antigen pair (n = 223 cells)

**a**

| Cluster | Peptide Rank 1 | 2 | 3 | 4 | 5 | TCRα,1 TRAV | TCRα,1 CDR3α | TCRα,2 TRAV | TCRα,2 CDR3α | TCRβ TRBV | TCRβ CDR3β | Cell Name |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1a | GANAB | GCN1L1 | USP28 | | | 12-2*01 | CAVNNARLMF | | | 4-3*01 | CASSQGGGGTDTQYF | BA4 |
| 1a | GANAB | EMPTY | USP28_C5F | | | 12-2*01 | CAVNNARLMF | | | 4-3*01 | CASSQGGGGTDTQYF | BB2 |
| 1a | GANAB | EMPTY | AKAP13_Q8K | USP28_C5F | NSDHL | 12-2*01 | CAVNNARLMF | | | 4-3*01 | CASSQGGGGTDTQYF | CD6 |
| 1a | GANAB | SEC24A_P5L | | | | 12-2*01 | CAVNNARLMF | | | 4-3*01 | CASSQGGGGTDTQYF | CH5 |
| 1a | GANAB | PABPC1_R5Q | AHNAK_S1F | SEC24A | TEAD1_L8F | 12-2*01 | CAVNNARLMF | | | 4-3*01 | CASSQGGGGTDTQYF | CH11 |
| 1b | GANAB | PABPC1 | EMPTY | HCV_KLV | TEAD1_L8F | 12-2*01 | CAVNNARLMF | | | | | BB5 |
| 1b | GANAB | SEC24A_P5L | GANAB_S5F | | | 12-2*01 | CAVNNARLMF | | | | | BE4 |
| 1b | GANAB | SEC24A | FNDC3B_L3M | PABPC1 | EMPTY | 12-2*01 | CAVNNARLMF | | | | | BH12 |
| 1b | GANAB | HCV_KLV | SNX24 | GANAB_S5F | EMPTY | 12-2*01 | CAVNNARLMF | | | | | BH4 |
| 1c | MRM1 | SNX24_P6L | MRM1_T6P | ERBB2 | SNX24 | 12-2*01 | CAVNNARLMF | | | 4-1*01 | CASSPSPGSEQYF | BA5 |
| 2 | MLL2_L8H | GCN1L1_L6P | | | | 12-2*01 | CAVGGYNFNKFYF | 12-1*01 | CVALRGGSQGNLIF | 5-6*01 | CASSFRDSSYEQYF | AD7 |
| 2 | MLL2_L8H | COL18A1 | COL18A1_S8F | TEAD1_L9F | | 12-2*01 | CAVGGYNFNKFYF | 12-1*01 | CVALRGGSQGNLIF | 5-6*01 | CASSFRDSSYEQYF | AE4 |
| 3 | GNL3L_R4C | | | | | 22*01 | CAVKTSYDKVIF | | | 28*01 | CASSRGGHEQYF | AD11 |
| 3 | GNL3L_R4C | PABPC1_R5Q | SEC24A_P5L | EMPTY | | 22*01 | CAVKTSYDKVIF | | | 20-1*01 | CSAGVYEQYF | AF7 |
| 3 | GNL3L_R4C | EMPTY | GCN1L1_L6P | AHNAK | WDR46_T3I | 22*01 | CAVKTSYDKVIF | | | 20-1*01 | CSARVTSGSYEQYF | CE5 |
| 3 | GNL3L_R4C | SMARCD3_H8Y | EMPTY | SNX24_P6L | TEAD1_L9F | 22*01 | CAVKTSYDKVIF | | | 20-1*01 | CSSVTEAFF | CF12 |
| 3 | GNL3L_R4C | AKAP13_Q8K | | | | 22*01 | CAVKTSYDKVIF | | | | | AD6 |
| 4 | GNL3L_R4C | SEC24A_P5L | TEAD1_L8F | | | 22*01 | CAVRTSYDKVIF | | | 19*01 | CASSWDNGGYTF | AF9 |
| 4 | GNL3L_R4C | WDR46_T3I | SMARCD3_H8Y | COL18A1_S8F | NSDHL_A9V | 22*01 | CAVRTSYDKVIF | | | 20-1*01 | CSAPGGSGANVLTF | AE8 |
| 4 | GNL3L_R4C | WDR46 | PGM5_H5Y | TEAD1_L9F | NSDHL_A9V | 22*01 | CAVRTSYDKVIF | | | 29-1*01 | CSVEGPGGRIANTEAFF | CD11 |
| 4 | GNL3L_R4C | AHNAK | GNL3L | AKAP13 | PGM5_H5Y | 22*01 | CAVRTSYDKVIF | | | 20-1*01 | CSVTQGDGTDTQYF | AB6 |
| 5 | USP28_C5F | FNDC3B_L3M | EMPTY | FNDC3B | ERBB2_H8Y | 30*01 | CGTRGGSGNTPLVF | 35*01 | CAGQMYSGGGADGLTF | 12-3*01/12 | CASTATFGVTEAFF | CC2 |
| 5 | USP28_C5F | SEC24A | TEAD1_L8F | MRM1 | USP28 | 30*01 | CGTRGGSGNTPLVF | 35*01 | CAGQMYSGGGADGLTF | 12-3*01/12 | CASTATFGVTEAFF | CD9 |
| 6 | GNL3L_R4C | USP28_C5F | | | | 6*01 | CALQTGANNLFF | | | 27*01 | CASSLWAGETQYF | AA10 |
| 6 | GNL3L_R4C | | | | | 6*01 | CALQTGANNLFF | | | 27*01 | CASSLWAGETQYF | AE2 |
| 6 | GNL3L_R4C | SMARCD3 | PGM5_H5Y | USP28_C5F | | 6*01 | CALQTGANNLFF | | | 27*01 | CASSLWAGETQYF | AE3 |
| 6 | GNL3L_R4C | TEAD1_L8F | AKAP13_Q8K | COL18A1 | NSDHL | 6*01 | CALQTGANNLFF | 14/DV4*01 | CAMIFNDYKLSF | 27*01 | CASSLWAGETQYF | CG2 |
| Spike-In | HCV_KLV | CLP_RLF | SEC24A_P5L | GCN1L1 | TEAD1_L8F | 14/DV4*01 | CAMIFNDYKLSF | | | 19*01 | CASSTGNYGYTF | AG1 |
| Spike-In | HCV_KLV | COL18A1 | TEAD1_L8F | TEAD1_(VLE) | FNDC3B | 14/DV4*01 | CAMIFNDYKLSF | | | 19*01 | CASSTGNYGYTF | AG10 |
| Spike-In | HCV_KLV | COL18A1_S8F | GCN1L1_L6P | | | 14/DV4*01 | CAMIFNDYKLSF | | | 19*01 | CASSTGNYGYTF | AG11 |
| Spike-In | HCV_KLV | PABPC1 | TEAD1_L9F | WDR46_T3I | PGM5 | 14/DV4*01 | CAMIFNDYKLSF | | | 19*01 | CASSTGNYGYTF | AG12 |
| Spike-In | HCV_KLV | FNDC3B | COL18A1_S8F | TEAD1_L9F | NSDHL_A9V | 14/DV4*01 | CAMIFNDYKLSF | | | 19*01 | CASSTGNYGYTF | AG2 |
| Spike-In | HCV_KLV | GNL3L | AKAP13_Q8K | NSDHL | CLP_RLF | 14/DV4*01 | CAMIFNDYKLSF | | | 19*01 | CASSTGNYGYTF | AG3 |
| Spike-In | HCV_KLV | SEC24A | PABPC1 | TEAD1_(SVL) | ERBB2_H8Y | 14/DV4*01 | CAMIFNDYKLSF | | | 19*01 | CASSTGNYGYTF | AG4 |
| Spike-In | HCV_KLV | COL18A1_S8F | TEAD1_L8F | MRM1 | PABPC1_R5Q | 14/DV4*01 | CAMIFNDYKLSF | | | 19*01 | CASSTGNYGYTF | AG5 |
| Spike-In | HCV_KLV | CLP_RLF | GCN1L1_L6P | | | 14/DV4*01 | CAMIFNDYKLSF | | | 19*01 | CASSTGNYGYTF | AG6 |
| Spike-In | HCV_KLV | AHNAK_S1F | SNX24_P6L | TEAD1_L9F | SMARCD3_H8Y | 14/DV4*01 | CAMIFNDYKLSF | | | 19*01 | CASSTGNYGYTF | AG7 |
| Spike-In | HCV_KLV | COL18A1_S8F | NSDHL | MLL2 | | 14/DV4*01 | CAMIFNDYKLSF | | | 19*01 | CASSTGNYGYTF | AG8 |
| Spike-In | HCV_KLV | SNX24 | COL18A1 | SMARCD3 | GANAB_S5F | 14/DV4*01 | CAMIFNDYKLSF | | | 19*01 | CASSTGNYGYTF | AG9 |

**b** **c**

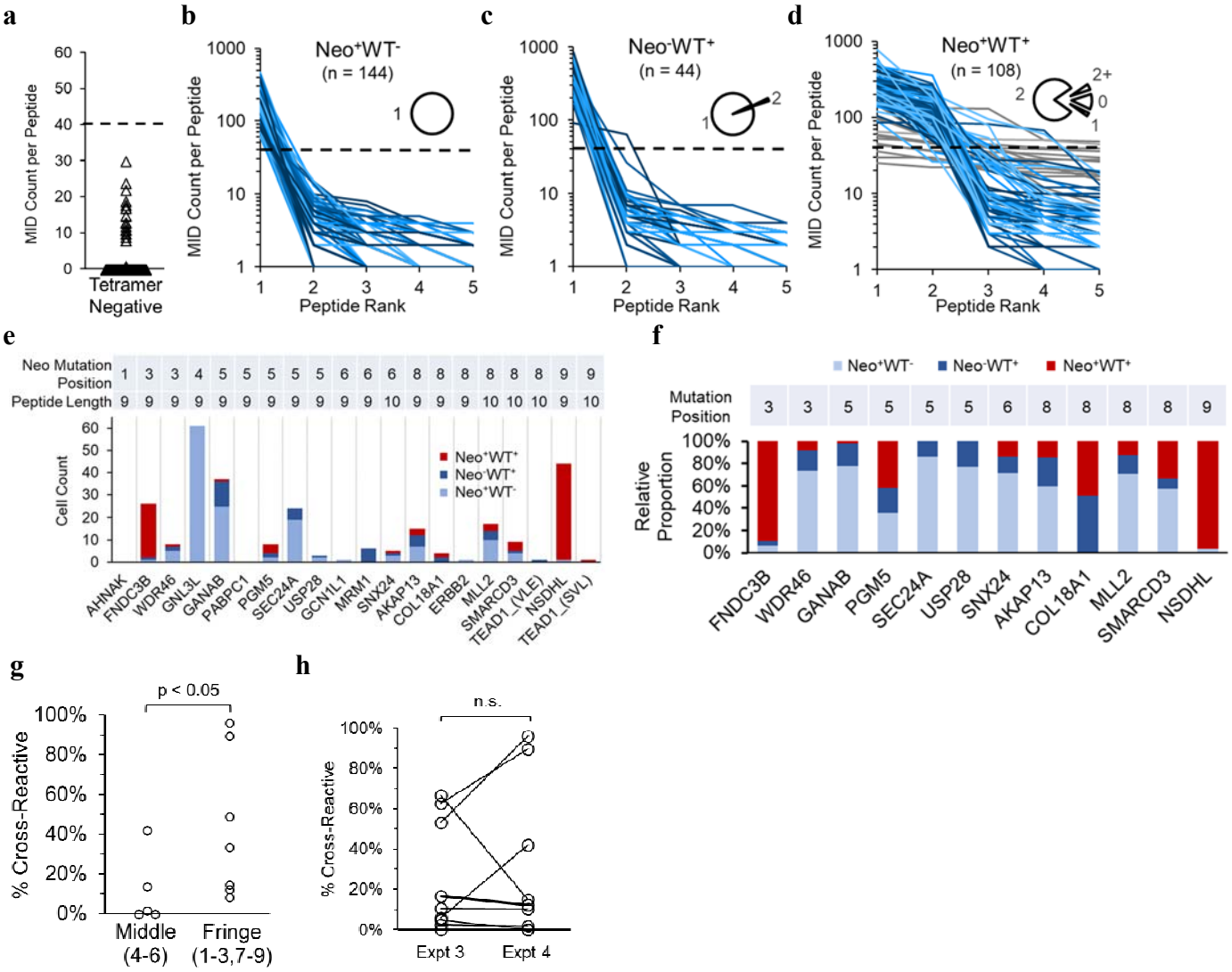



Supplementary Figure 12

Verification of pMHC classification using the spike-in HCV-KLV(WT) binding clone and primary cells with shared TCRs in Experiment 3

**(a)** Top 5 epitopes by MID count for T cells sharing at least one TCRα or β chain. Bold border indicates the positively-classified binding peptides. TCRα or β chains with the same color in the same cluster have the same nucleotide sequence for the respective chain. **(b,c)** Peptide rank curve by MID counts for the HCV-KLV(WT) binding spike-in clone (12 cells) (b) and primary cells with shared TCR (13 cells) (c). Dashed lines indicate MID threshold for identifying positively bound peptides. Each solid blue line indicates a cell and only the top 5 peptides were shown raked by their MID counts. For (c) we only consider cells with identical TCRα and TCRβ sequence on an AA level, corresponding to cluster 1a, 2, 5, and 6 in (a). For WT-antigen, the peptide was named after the protein; for Neo-antigen, the peptide was named as protein name_AA#AA.

Just as in Experiment 1, our criteria correctly classified all peptides for the spike-in HCV-binding clone. Interestingly, despite only sorting on the $CCR7^{+}CD45RA^{+}$ naïve phenotype, we detected 6 clusters of primary T cells with shared TCR sequences on the AA level (Clusters 1-6 in Supplementary Fig. 12a). Cells with shared TCR α and β sequences bound the same peptide (Clusters 1a, 2, 5, 6). Many of these TCRs were found to be encoded by different TCRα and TCRβ nucleotide sequences, indicating convergent VDJ recombination. We also found that in some TCRs, the same TCR α chain is sufficient for them to engage the same pMHC, while TCRβ chains are all different (Clusters 3 and 4). However, in other TCRs, the same TCR α paired with a different TCR β chain can lead to different peptide specificity (Compare Cluster 1c to 1a). These results highlight the advantage of high-throughput linking of TCR sequence with its antigenic peptide as a first step in deciphering the TCR repertoire, which could be complementary to bioinformatics analysis[4].



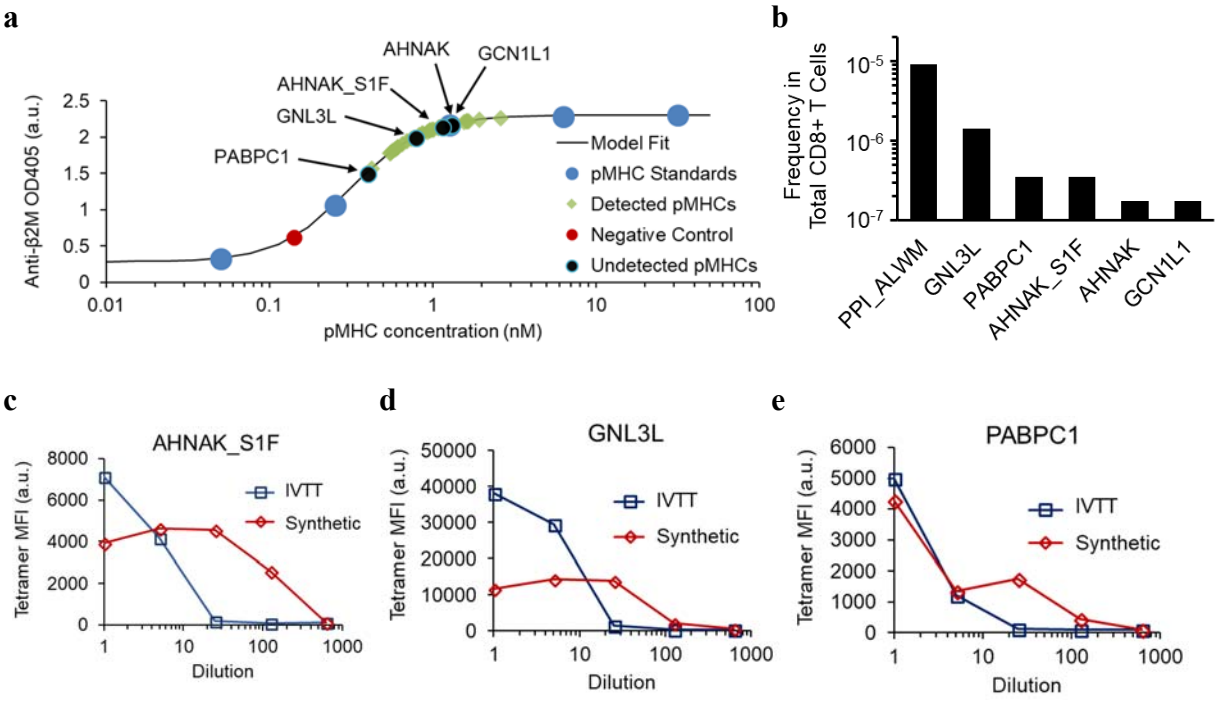

Supplementary Figure 13

DNA-BC analysis for Experiment 4

**(a)** MID counts associated with peptides from the sorted Tetramer$^-$ CD8$^+$ T cells (36 cells). MID threshold for positively binding peptide is designated by the dashed line. **(b-d)** Peptide rank curve by MID counts for the (b) Neo$^+$WT$^-$, (c) Neo$^-$WT$^+$ and (d) Neo$^+$WT$^+$ primary cells. Dashed line indicates MID threshold for identifying positively bound peptides. Each solid line indicates a cell and only the top 5 peptides were shown ranked by their MID counts. Blue solid lines indicate cells with at least one positively binding peptide; grey solid lines indicate cells that did not positively bind any peptides based on the criteria discussed at the beginning of the supplementary information. Insert pie charts for all three panels indicate proportion of cells with the indicated number of positively bound peptides. **(e)** Cell count for all detected peptides for each Neo-WT gene pair (n = 274 cells). **(f)** Relative proportion of the three cell populations for each Neo-WT gene pair from (e), similar to Fig. 2b. Each antigen was normalized by the relative frequency and number of cells sorted from the corresponding Tetramer$^+$ population (see Methods). Only pairs where both the Neo-antigen and Wildtype were detected in at least one cell, and have at least 3 detected cells total were considered (n = 200 cells). **(g)** Comparison of cross-reactivity for Neo-WT antigen-binding T cell populations from (f) that have mutations near the middle or fringes (n = 5 Neo-WT antigen-pairs for middle and 6 pairs for fringe, One-tailed Mann-Whitney U Test). **(h)** Comparison of the percent cross-reactive T cells that exist within each Neo-WT antigen-binding T cell population between Experiment 3 and 4. Only Neo-WT pairs that meet the criteria in (f) and are shared between the two experiments are considered. Dot represents one Neo-WT pair and lines connect the same pair from the two experiments (n =

9 Neo-WT antigen-pairs for Experiment 3 and 9 pairs for Experiment 4, One-tailed Wilcoxon Signed-Rank Test, $W = 18$, not significant at $p < 0.05$). The same gating strategy in Supplementary Fig. 10a and b is used for Experiment 4.

The percentage of cross-reactive T cells for the same Neo-WT antigen pair was not significantly different between Experiment 3 (Fig. 2b, c) and 4, indicating that this property is conserved between donors for the peptides tested.

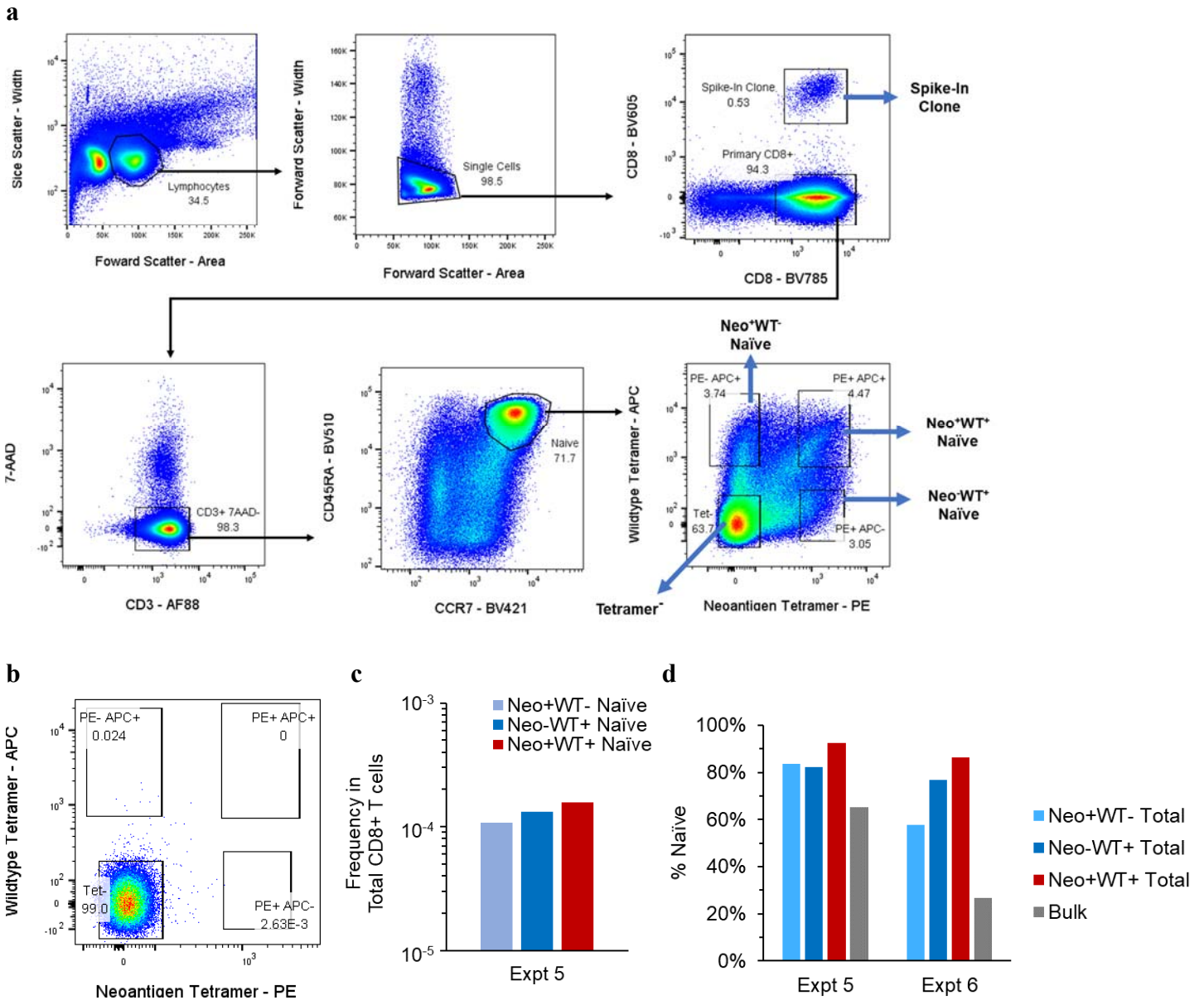



Supplementary Figure 14

Validation for "undetected" peptides in Experiment 3 and 4

**(a)** ELISA for all 40 pMHC monomers UV-exchanged with IVTT-generated Neo or WT peptides. UV-exchanged pMHC monomers are plated at a concentration of 1.6 nM estimated based on the un-exchanged MHC monomer concentration, followed by anti-β2M staining. Blue dots represent un-exchanged MHC monomer diluted at various concentration from lowest to highest (0.05, 0.25, 1.25, 6.25, 31.25 nM). Red dot represents UV-exchanged pMHC in IVTT solution that did not contain a peptide-encoding DNA template. Black dots indicate the 5 "undetected" peptides in Experiment 3 and 4. Solid line is a sigmoidal model fit to the standards. Arrows indicate "undetected" peptides from Experiment 3 and 4. **(b)** TetTCR-Seq experiment on an additional donor's PBMC sample using an IVTT-generated pMHC tetramer library for PPI_ALWM and the five "undetected" peptides. Shown is the estimated frequency of each antigen-binding CD8+ T cell population. **(c-e)** Peptide titration experiments were performed for three of the "undetected" peptides where T cell clones could be generated using Tetramer$^{+}$ T cells from (b). Experiments were performed once. Peptides generated from either IVTT or the traditional, synthetic peptide method, were diluted at different ratios and were used to form PE labeled pMHC tetramers. Starting concentration of synthetic peptide is 100 μM for all peptides. These pMHC tetramers were used to stain a cognate T cell clone. Anti-CD8a (RPA-T8) was present throughout the staining. MFI, Median Fluorescence Intensity. a.u., arbitrary unit. For WT-antigen, the peptide was named after the protein; for neo-antigen, the peptide was named as protein name_AA#AA.

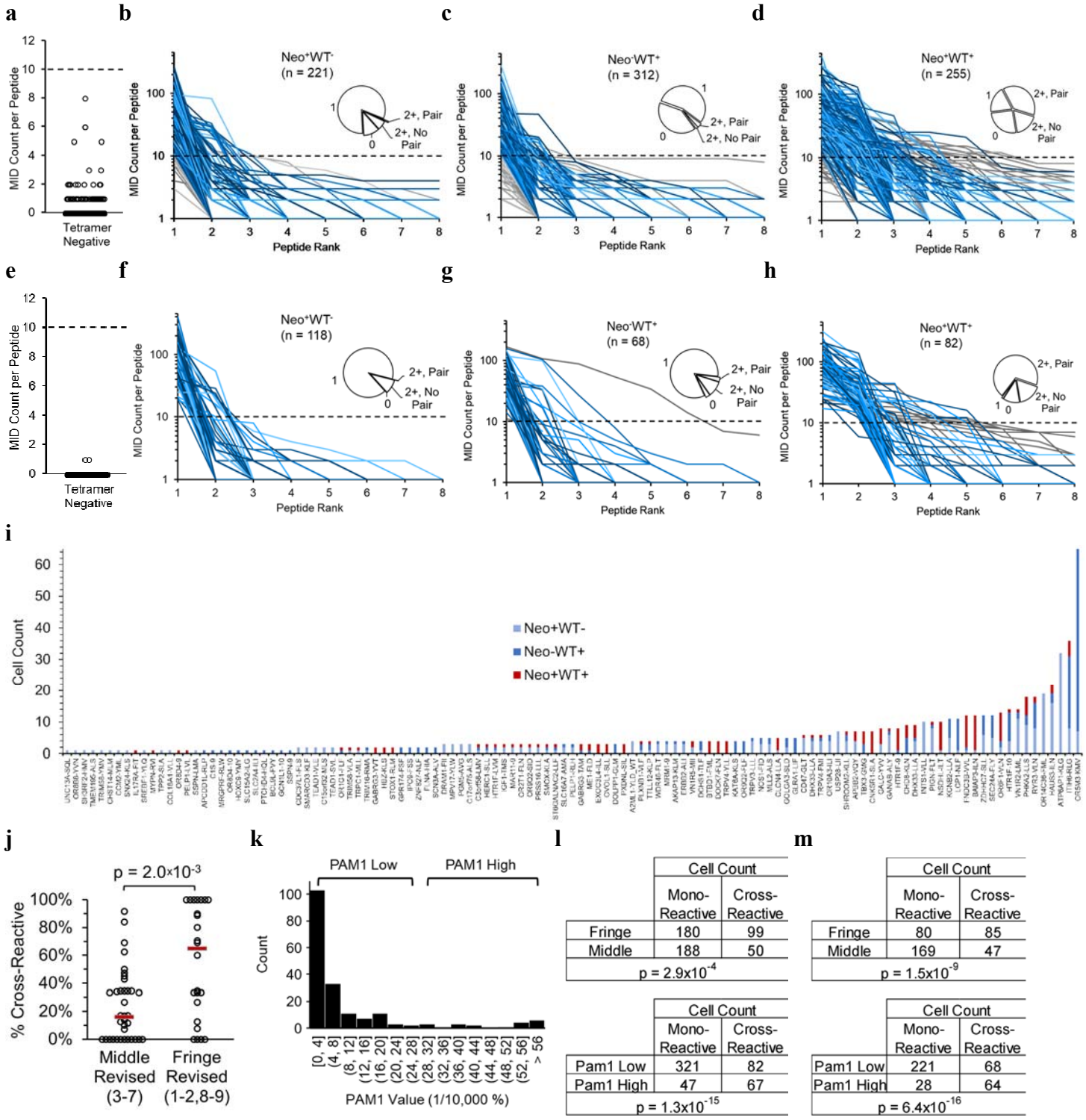
Five peptides in Experiment 3 and 4 had no detected T cell binding. Further analysis showed no difference in the pMHC UV-exchange efficiency associated with detected and undetected peptides (Supplementary Fig. 14a). TetTCR-Seq on a subsequent donor using these 5 peptides showed that these antigen-binding T cells are present at low frequencies in blood (Supplementary Fig. 14b). Furthermore, we successfully generated monoclonal T cell lines specific for 3 of the peptides and found that IVTT-generated pMHC tetramers stained similarly as their synthetic peptide counterparts (Supplementary Fig. 14c-e). These results confirm that "undetected" peptide-binding T cells in Experiment 3 and 4 were more likely caused by low cell frequency rather than inefficient pMHC generation by IVTT.




Supplementary Figure 15

Gating scheme and sorting strategy for Experiment 5 and 6

**(a)** Representative gating scheme for Experiment 5 and 6. Shown is the gating scheme for Experiment 5. **(b)** Tetramer gating on the flow-through fraction from Experiment 5. A similar result was obtained in experimental replicate Experiment 6. **(c)** Estimated frequencies of the three Tetramer$^+$ populations for Experiment 5. Frequencies could not be obtained for Experiment 6. **(d)** Naïve T cell percentages for each of the three Tetramer$^+$ populations and bulk flow-through CD8+ T cells for Experiment 5 and 6.

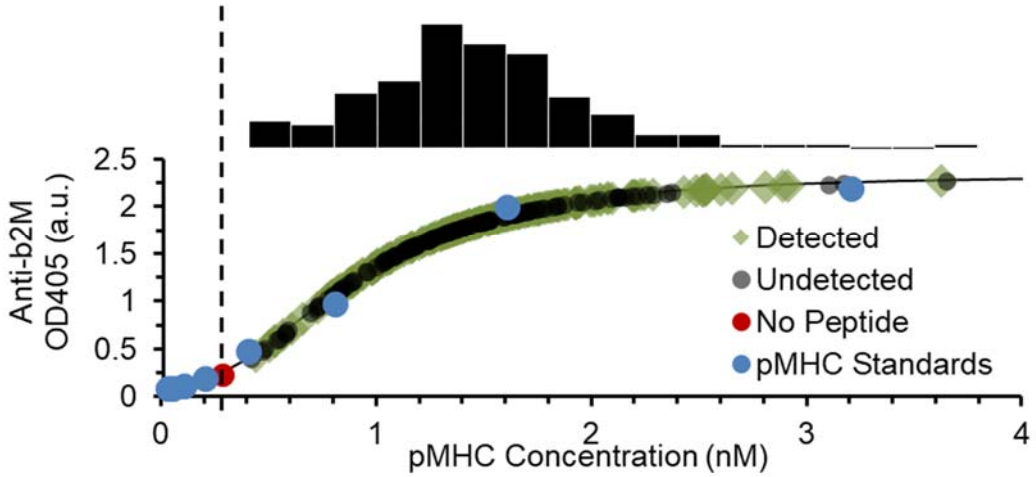



Supplementary Figure 16

Analysis of Experiment 5 and 6

**(a-h)** MID counts associated with peptides from the sorted Tetramer$^{-}$ CD8$^{+}$ T cells for Experiment 5 (a) and 6 (e). Peptide rank curve by MID counts for the indicated Tetramer$^{+}$ cell populations for Experiment 5 (b-d) and 6 (f-h). Dashed line indicates MID threshold for identifying positively bound peptides. Each solid line indicates a cell and only the top 8 peptides were shown ranked by their MID counts. Blue solid lines indicate cells with at least one positively binding peptide; grey solid lines indicate cells that did not positively bind any peptides based on

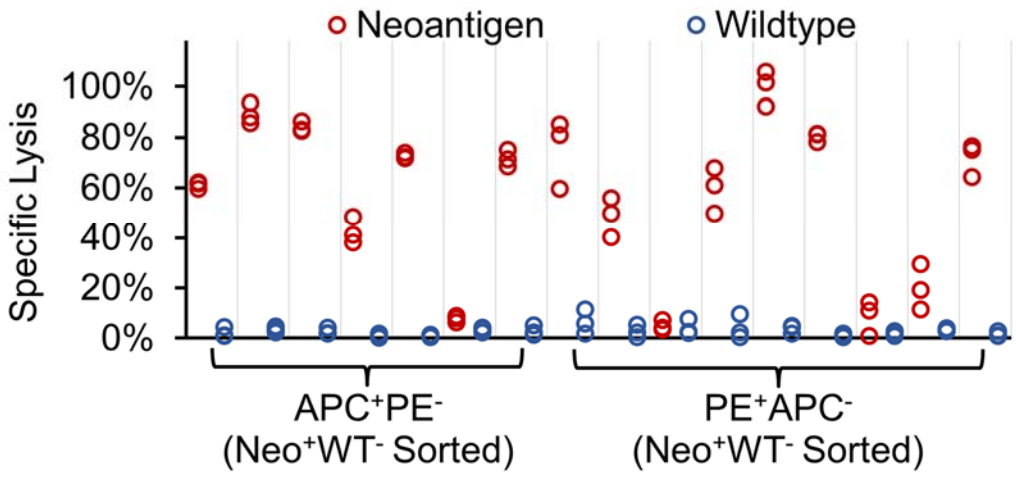
the criteria discussed at the beginning of the Supplementary Information. Insert pie charts for all these panels indicate proportion of cells with the indicated number of positively bound peptides. For insert pie charts, 2+ Paired indicates that all detected peptides from a given cell belong to a particular Neo/WT antigen pair; this has the same meaning as "2" in pie chart inserts of Experiment 3 and 4, but since we included one WT that had two neo-antigens in this library (DHX33-LLA) we found one cell that was cross reactive to all three peptides, which is counted in this category as well. 2+ unpaired indicates at least 2 detected peptides but at least one peptide did not belong to a particular Neo/WT antigen pair. **(i)** Total cell counts for Neo-WT antigen pairs with at least one detected cell (n = 678 cells). **(j)** As in Fig. 2f, a greater difference in the percent of cross-reactive antigen-binding populations is observed when revising the peptide middle position to position 3-7. Each circle represents the percent of cross-reactive T cells observed for one Neo-WT antigen pair. Only antigen pairs where both the Neo and WT peptides were detected in at least one cell, with at least 3 cells total are included. Red bars denote median. (n = 24 Neo-WT antigen-pairs for middle and 38 for fringe, One-tailed Mann-Whitney U Test). **(k)** Definition of PAM1 high/low threshold. PAM1 values for amino acid pairs i and j are calculated by adding the one directional PAM1 values, $PAM1_{ij}$ + $PAM1_{ji}$, as defined by Wilbur et al.[5]. Shown is a histogram of all the possible PAM1 values between non-identical amino acids (n = 190 AA transitions). The top 10% is designated as PAM1 High. **(l)** Contingency table analysis on the dependence of T cell cross-reactivity on mutation position and PAM1substitution value, using the same antigen pairs as in Figure 2f (One-tailed Fisher's Exact Test). **(m)** Same analysis as (l), except the top 5% of antigen pairs by cell count is removed from analysis, which accounted for 24% of the total cell count from (l).




Supplementary Figure 17

ELISA on the 315 pMHC monomer library that's UV-exchanged with IVTT-generated peptides for Experiment 5 and 6

We selected neo-antigens with high predicted affinity for HLA-A2 from recent literature, and preference was given to those with positive binding and/or T cell assays (Supplementary Table).
UV-exchanged pMHC monomer using IVTT-generated peptides are plated on ELISA plates at a concentration of 1.6 nM estimated from unexchanged MHC monomer concentration and then stained with anti-β2m antibody. Blue circles represent pMHC concentration standards. Solid line represents sigmoidal model fit to the standards. Red dot represents UV-exchanged pMHC in IVTT solution that did not contain a peptide-encoding DNA template, thus serves as a negative control. Black dots represent peptides that were not detected in Experiments 5 or 6. Green diamonds represents peptides that were detected in at least one cell in Experiment 5 or 6. Top histogram combines both the detected and undetected peptides in respect to pMHC monomer concentration plotted below. Dashed line represents the minimum threshold for pMHC UV-exchange. The blue dot standard to the right side of the dashed line is 0.4 nM of un-exchanged MHC monomer.




Supplementary Figure 18

Both PE and APC fluorescent DNA-BC pMHC tetramers can be used to sort neo-antigen-specific T cells with no functional reactivity to WT counterpart peptide

A DNA-BC pMHC library was constructed as in Experiment 3 and 4 to sort $APC^{+}PE^{-}$ ($Neo^{+}WT^{-}$) primary T cells. A fluorescence swapped pMHC library compared to Experiment 3 and 4, where neo-antigen pMHCs were on the PE channel and WT pMHCs were on the APC channel, was used to sort $PE^{+}APC^{-}$ ($Neo^{+}WT^{-}$) primary T cells. 5 cells were sorted per well for *in vitro* culture. LDH cytotoxicity assay on *in vitro* expanded primary T cells sorted interacting with T2 cells pulsed with the 20 neo-antigen peptide pool or 20 WT counterpart peptide pool. Each set of red/blue circles represent one T cell line. Each condition was performed in triplicates derived from separate wells of cells.

The choice of fluorophore did not affect T cell functional profile, as tested by swapping the fluorophore encoding of the DNA-BC pMHC library.

**a**

5-cell-sort T cell lines from Neo$^{+}$WT$^{-}$ population

| Cell Name | TRAV | CDR3α | TRBV | CDR3β | Specificity | Mutation Position | Cell Count |
|---|---|---|---|---|---|---|---|
| AD2 | 8-1*01 | CAARPDSNYQLIW | 4-2*01 | CASSQGGDEQFF | GANAB_S5F | 5 | 6 |
| AD3 | | | 4-1*01 | CASSQDPLLEGYSYNEQFF | GNL3L_R4C | 4 | 1 |
| AD5 | 17*01 | CATDDSGGYQKVTF | 12-3/4 | CASSGELRGAEKLFF | PGM5_H5Y | 5 | 7 |
| | 13-1*01 | CAAPGNYGQNFVF | 6-5*01 | CASSYGGEQPQHF | GNL3L_R4C | 4 | 1 |
| | 29/DV5*01 | CAAYRSGAGSYQLTF | 9*01 | CASSNGYGNQPQHF | GANAB_S5F | 5 | 2 |
| AD7 | 24*01 | CAFYNQGGKLIF | 9*01 | CASSVEGVADEQYF | PGM5_H5Y | 5 | 10 |
| AD8 | 12-2*01 | CAALTDSWGKLQF | 25-1*01 | CASSDLQGGGPGEQYF | GANAB_S5F | 5 | 4 |
| | | | 9*01 | CASSVEGVADTQYF | PGM5_H5Y | 5 | 1 |
| AD10 | | | 5-4*01 | CASSSTGALYEQYF | SNX24_P6L | 6 | 1 |
| | 12-2*01 | CAVISGGGADGLTF | 28*01 | CASTIALGYEQYF | GNL3L_R4C | 4 | 5 |
| AD11 | 6*01 | CALEAGNKLTF | 19*01 | CASSSGGGNEQFF | GNL3L_R4C | 4 | 4 |
| | 8-6*01 | CAVSESRNRDDKIIF | | | | | |

**b**

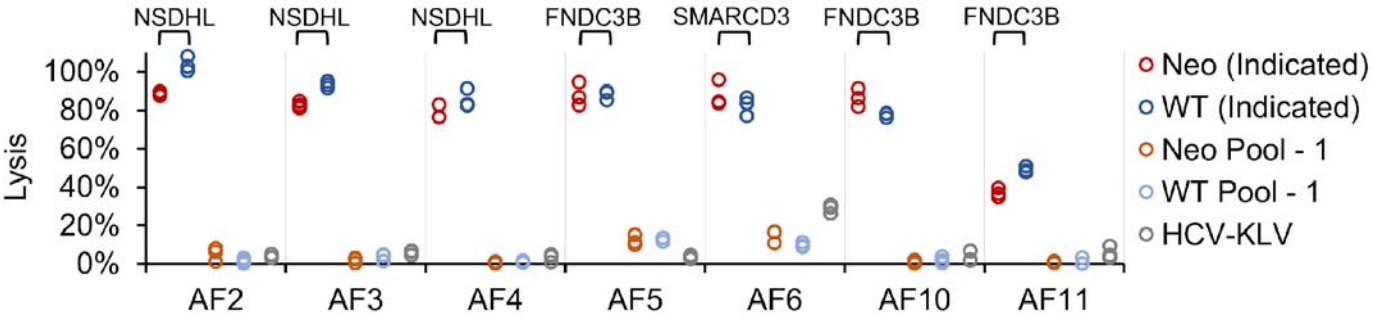
5-cell-sort T cell lines from Neo$^{+}$WT$^{+}$ population

| Cell Name | TRAV | CDR3α | TRBV | CDR3β | Specificity 1 | Specificity 2 | Mutation Position | Cell Count |
|---|---|---|---|---|---|---|---|---|
| AF2 | 5*01 | CAESGGNNNDMRF | 4-1*01 | CASSPSLQVAQHF | NSDHL | NSDHL_A9V | 9 | 16 |
| AF3 | 8-4*01 | CAVSDARNYQLIW | 24-1*01 | CATSDPGTTNTGELFF | NSDHL | NSDHL_A9V | 9 | 21 |
| AF4 | 21*01 | CAVEVLQGAQKLVF | 5-4*01 | CASSLGGGAYEQYF | NSDHL | NSDHL_A9V | 9 | 14 |
| AF5 | 6*01 | CAVTGNQFYF | 20-1*01 | CSASPGYNEQFF | FNDC3B | FNDC3B_L3M | 3 | 1 |
| | 13-1*01 | CAASRDSGYALNF | 11-2*01 | CASSLAGESSYNEQFF | FNDC3B | FNDC3B_L3M | 3 | 23 |
| AF6 | 14/DV4*01 | CAMREANNYGQNFVF | 7-6*01 | CASSLGDPNTEAFF | SMARCD3 | SMARCD3_H8Y | 8 | 23 |
| AF7 | 12-2*01 | CAVNLDQTGANNLFF | 7-7*01 | CASSLGGYSPLHF | FNDC3B | FNDC3B_L3M | 3 | 3 |
| | 35*01 | CAGQVPGNQFYF | 7-8*01 | CASSLVGVSSYEQYF | Multiple (COL18A1, PGM5, PGM5_H5Y) | | | 20 |
| AF8 | 5*01 | CAERKLIYNQGGKLIF | 9*01 | CASSGQRGAYNEQFF | Multiple (COL18A1, TEAD1_VLE, FNDC3B, NSDHL, COL18A1_S8F) | | | 22 |
| AF9[a] | n/a | n/a | n/a | n/a | Non-Specific | | | 16 |
| AF10 | 14/DV4*01 | CAMREGGNYGQNFVF | 19*01 | CASSTQGQPQHF | FNDC3B | FNDC3B_L3M | 3 | 14 |
| | 8-3*01 | CAVGAGEAAGNKLTF | 5-6*01 | CASSLGGSYTF | FNDC3B | FNDC3B_L3M | 3 | 3 |
| | 25*01 | CAGVLDSNYQLIW | 5-1*01 | CASSPWTPTTDTQYF | FNDC3B | FNDC3B_L3M | 3 | 10 |
| AF11[b] | 20*01 | CAVRGYSGGGADGLTF | 3-1*01 | CASNFPNISEGTCSNQPQHF | FNDC3B | FNDC3B_L3M | 3 | 7 |
| | 20*01 | CAVRGYSGGGADGLTF | 20-1*01 | CQAGGRLGELFF | FNDC3B | FNDC3B_L3M | 3 | 1 |
| | 20*01 | CAVRGYSGGGADGLTF | 3-1*01 | CASNFPNISEGTCSNQPQHF | FNDC3B | FNDC3B_L3M | 3 | 5 |
| | | | 20-1*01 | CQAGGRLGELFF | | | | |

[a]TCR amplification failed for all tetramer positive cells from AF9, the resulting DNA-BCs reveal no pattern in antigen specificity

[b]For the same TCRα, we observed two TCRβ sequences that were detected both indvidually and together in single cells. All of these cells bound the same Neo-WT antigen pair listed, but we cannot deduce whether it is a single clone with two functional TCRβ chains or two separate clones that contain the same TCRα. Therefore, we list the cell counts associated with each TCRβ combination individually

**c**

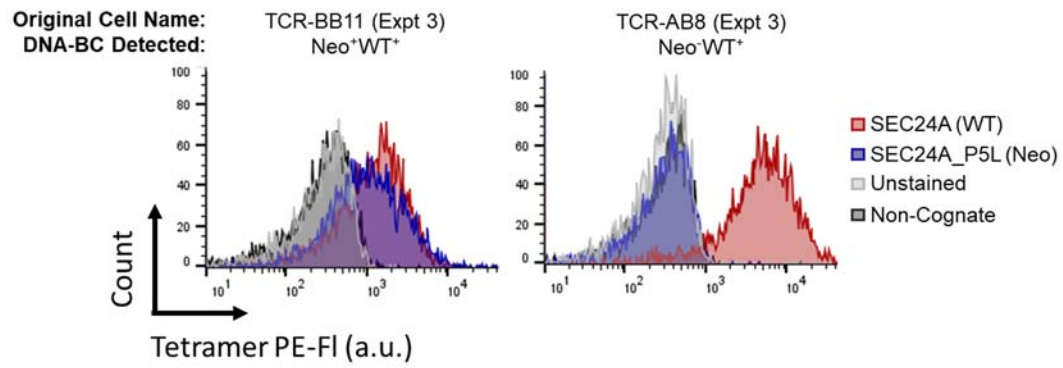



Supplementary Figure 19

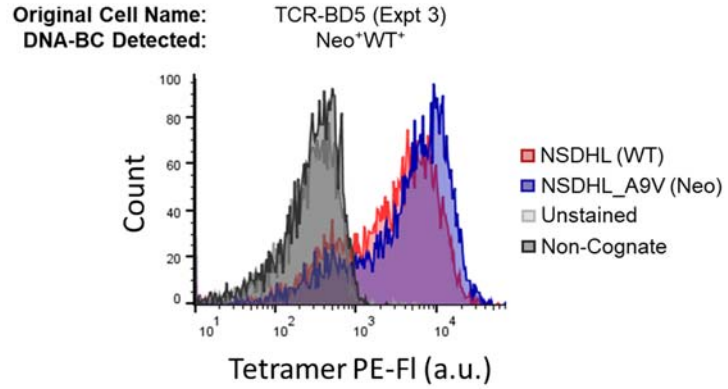
Characterization of the Neo$^{+}$WT$^{-}$ and Neo$^{+}$WT$^{+}$ cell lines in Fig. 2g

**(a,b)** T cell clonal composition as assessed by single cell TCR sequencing and matched pMHC specificity for the T cell lines in the $Neo^{+}WT^{-}$ (a) and $Neo^{+}WT^{+}$ (b) of Fig. 2g. For (a), TetTCR-Seq was performed for pooled cell lines and the resulting single sorted cells were matched to the correct T cell line from bulk TCR sequencing results of each T cell line. For (b), TetTCR-Seq was performed on each T cell line using the 40 Neo-WT DNA-BC pMHC tetramer library. Single cell DNA-BC and TCR sequences were used to tally the T cell clonality and the antigen binding of each T clone within a T cell line. For WT-antigen, the peptide was named after the protein; for neo-antigen, the peptide was named as protein name_AA#AA. **(c)** LDH cytotoxicity assay on the monoclonal T cell $Neo^{+}WT^{+}$ lines, discovered from (b), using the pMHC identified by TetTCR-Seq. Each condition was performed in triplicates derived from separate wells of cells. . "Neo pool – 1" and "WT Pool – 1" refers to the other 19 Neo-antigens and Wildtype peptides, respectively, that were not identified by TetTCR-Seq for the given cell line. HCV-KLV peptide was used as a known-antigen negative control.

**a**




**b**




Supplementary Figure 20

Tetramer staining of additional Jurkat 76 cell lines transduced with TCRs identified from Experiment 3

Jurkat 76 cells were transduced with the indicated TCRs, derived from primary T cell with positively identified antigens from Experiment 3, and then stained with the indicated pMHC tetramers. **(a)** A pair of TCRs that were identified to be cross reactive for both the Neo-antigen and Wildtype versions of SEC24A or just the Wildtype from TetTCR-Seq. **(b)** a TCR identified to be cross reactive for the Neo-antigen and Wildtype versions of NSDHL from TetTCR-Seq. Fl, fluorescence Intensity. a.u., arbitrary unit. For WT-antigen, the peptide was named after the protein; for Neo-antigen, the peptide was named as protein name_AA#AA. Experiments were performed once.

References:


1. Rodenko, B. et al. Generation of peptide-MHC class I complexes through UV-mediated ligand exchange. *Nat. Protocols* **1**, 1120-1132 (2006).
2. Chang, C.X. et al. Conditional ligands for Asian HLA variants facilitate the definition of CD8+ T-cell responses in acute and chronic viral diseases. *Eur J Immunol* **43**, 1109-1120 (2013).
3. Fu, G.K., Wilhelmy, J., Stern, D., Fan, H.C. & Fodor, S.P.A. Digital Encoding of Cellular mRNAs Enabling Precise and Absolute Gene Expression Measurement by Single-Molecule Counting. *Analytical Chemistry* **86**, 2867-2870 (2014).
4. Glanville, J. et al. Identifying specificity groups in the T cell receptor repertoire. *Nature* **547**, 94-98 (2017).
5. Wilbur, W.J. On the PAM matrix model of protein evolution. *Molecular biology and evolution* **2**, 434-447 (1985).