



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

Shu-Qi Zhang^{1,7}, Ke-Yue Ma^{2,7}, Alexandra A Schonnesen^{3,7}, Mingliang Zhang^{4,5,7}, Chenfeng He³, Eric Sun³, Chad M Williams³, Weiping Jia^{4,5}  & Ning Jiang^{2,3,6} 

We present tetramer-associated T-cell receptor sequencing (TetTCR-seq) to link T cell receptor (TCR) sequences to their cognate antigens in single cells at high throughput. Binding is determined using a library of DNA-barcoded antigen tetramers that is rapidly generated by *in vitro* transcription and translation. We applied TetTCR-seq to identify patterns in TCR cross-reactivity with cancer neoantigens and to rapidly isolate neoantigen-specific TCRs with no cross-reactivity to the wild-type antigen.

The ability to link T cell antigens, peptides bound by the major histocompatibility complex (pMHC), to TCR sequences is essential for monitoring and treating immune-related diseases. Fluorescently labeled T cell antigen oligomers, such as pMHC tetramers, are widely used to identify antigen-binding T cells¹. However, spectral overlap limits the number of pMHC tetramer species that can be examined in parallel and the extent of cross-reactivity that can be examined¹. Time-of-flight mass cytometry (CyTOF) with isotope-labeled pMHC tetramers can interrogate a larger number of antigen species simultaneously, but its destructive nature prevents the linkage of antigens bound to TCR sequences¹.

DNA-barcoded pMHC dextramer technology has been used for the analysis of antigen-binding T cell frequencies to samples of more than 1,000 pMHCs for T cells sorted in bulk². However, with bulk analysis, information on the binding of single or multiple peptides to individual T cells is lost. In addition, antigen-linked TCR sequences cannot be obtained, which would be valuable for tracking antigen-binding T cell lineages in disease settings, for TCR-based therapeutics development³, and for uncovering patterns in TCR–antigen recognition⁴. Another limitation is the high cost and long duration associated with synthesizing peptides chemically⁵, which prevents the quick generation of a pMHC library that can be tailored to specific pathogens or neoantigens in an individual.

To address these challenges, we developed TetTCR-seq for the high-throughput pairing of TCR sequences with potentially multiple pMHC

species bound on single T cells. First, we constructed a large library of fluorescently labeled, DNA-barcoded (DNA-BC) pMHC tetramers in a rapid and inexpensive manner using *in vitro* transcription and translation (IVTT) (Fig. 1a). Next, tetramer-stained cells were single-cell sorted and the DNA-BC and TCR $\alpha\beta$ genes were amplified by RT-PCR (Fig. 1b). A molecular identifier (MID) was included in the DNA-BC to provide absolute counting of the copy number for each species of tetramers bound to the cell. Finally, nucleotide-based cell barcodes were used to link multiple peptide specificities with their bound TCR $\alpha\beta$ sequences (Fig. 1b). DNA-BC pMHC tetramers are compatible with isolation of rare antigen-binding precursor T cells⁶, making TetTCR-seq a versatile platform for analyzing both clonally expanded and precursor T cells.

As constructing large pMHC libraries via ultraviolet-mediated peptide exchange using traditionally synthesized peptide is costly, with long turnaround times⁵, we use a set of peptide-encoding oligonucleotides that serve as both the DNA-BCs for identifying antigen specificities and as DNA templates for peptide generation via IVTT (Fig. 1a). The IVTT step only adds a few hours once oligonucleotides are synthesized. This substantially reduces the cost (about 20-fold) and time (2–3 d instead of weeks) compared to synthesizing peptides chemically.

pMHC tetramers generated by our IVTT method have similar performance to their synthetic-peptide counterparts (Fig. 1c and Supplementary Figs. 1 and 2). DNA-BC conjugation did not interfere with staining and has comparable sensitivity to fluorescence readouts (Supplementary Figs. 3 and 4). We performed six main TetTCR-seq experiments (Supplementary Fig. 5). We first assessed the ability of TetTCR-seq to accurately link TCR $\alpha\beta$ sequences with pMHC binding from primary CD8⁺ T cells in human peripheral blood. In experiment 1, we constructed a 96-peptide library consisting of well-documented foreign and endogenous peptides bound to HLA-A2 (Supplementary Table 1) and isolated dominant pathogen-specific T cells, as well as rare precursor antigen-binding T cells, from a healthy cytomegalovirus (CMV)-seropositive donor (Fig. 1 and Supplementary Figs. 6 and 7). To test whether TetTCR-seq can detect cross-reactive peptides, we included a documented hepatitis C virus (HCV) wild-type (WT) peptide, HCV-KLV(WT)⁷, and four candidate altered peptide ligands (APL) with one or two amino acid substitutions (Supplementary Table 1). An established HCV-KLV (WT) T cell clone⁷ was spiked into the donor's sample to test its cross-reactive potential.

Bound peptides were classified by an MID threshold derived from tetramer-negative controls and a ratio of MID counts between the peptides above and below this threshold (Fig. 1d and Supplementary Note). Using this classification scheme, we identified the HCV-KLV(WT) epitope from all spike-in cells sorted (Fig. 1e and Supplementary Fig. 8). In addition, we discovered that all four APLs were also classified as binders. A separate experiment confirmed the

¹McKetta Department of Chemical Engineering, University of Texas at Austin, Austin, Texas, USA. ²Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas, USA. ³Department of Biomedical Engineering, University of Texas at Austin, Austin, Texas, USA. ⁴Department of Endocrinology and Metabolism, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, China. ⁵Shanghai Key Laboratory of Diabetes Mellitus, Shanghai Clinical Center of Diabetes, Shanghai, China. ⁶LIVESTRONG Cancer Institutes, Dell Medical School, University of Texas at Austin, Austin, Texas, USA. ⁷These authors contributed equally to this work. Correspondence should be addressed to N.J. (jiang@austin.utexas.edu) or W.J. (wpjia@sjtu.edu.cn).

Received 12 November 2017; accepted 26 September 2018; published online 12 November 2018; doi:10.1038/nbt.4282

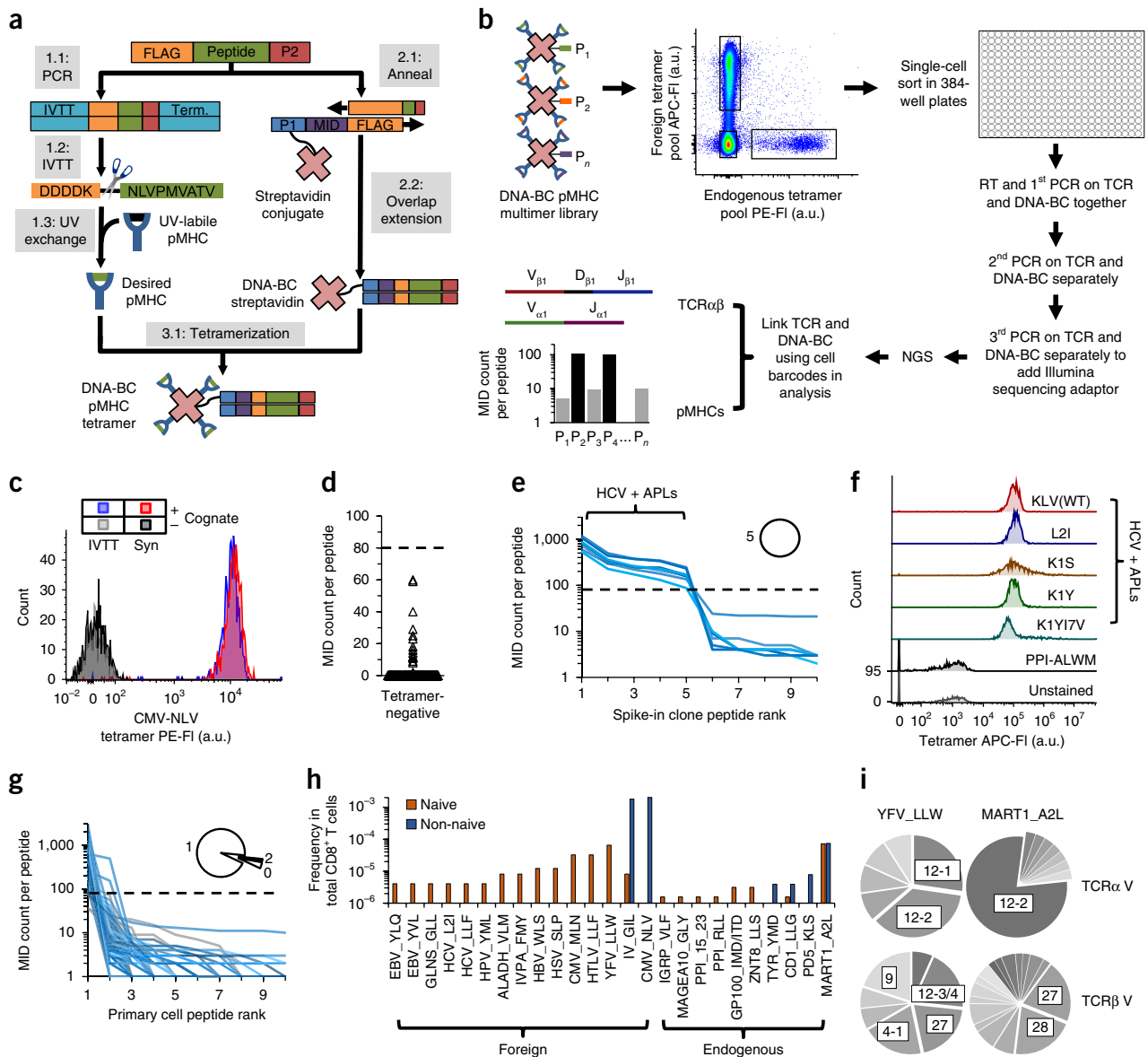


Figure 1 Workflow for generation of DNA-BC pMHC tetramer library and proof of concept for using TetTCR-seq for high-throughput linking of antigen binding to TCR sequences for single T cells. (a) Workflow for generation of DNA-BC pMHC tetramers. (b) Workflow of TetTCR-seq. (c) Comparison of staining performance for IVTT- and synthetic (Syn)-peptide-generated pMHC tetramers on T cell clones. Experiment repeated independently once with similar results. (d) MID counts per peptide detected on single T cells sorted from the tetramer-negative fraction in experiment 1 (768 peptides from 8 cells). Dashed line, MID threshold. (e) Peptide rank curve by MID counts for each of top ten ranked peptides for single sorted cells from the spike-in clone (8 cells) in experiment 1. Inset, proportion of cells with the indicated number of positively binding peptides. (f) Fluorescence intensity of the HCV-KLV(WT) binding T cell clone, used as spike-in in experiment 1, stained individually with the indicated pMHC tetramers in a separate experiment. Experiment performed once. (g) Peptide rank curve by MID counts as in e for the tetramer-positive primary T cell populations (167 cells) in experiment 1. Gray solid lines indicate cells with no detected peptides. (h) Calculated frequencies of antigen-binding T cell populations in total CD8⁺ T cells for peptides with at least one detected T cell, separated by phenotype, in experiment 1. (i) V-gene usage of unique TCR sequences that are specific for YFV_LLW (naive and non-naive combined, $n = 11$ cells for TRAV, $n = 15$ cells for TRBV) or MART1_A2L (naive and non-naive combined, $n = 33$ cells for TRAV, $n = 43$ cells for TRBV). P₁, P₂ and P_n, unique peptide ligands; NGS, next-generation sequencing; PE, phycoerythrin; APC, allophycocyanin; FI, fluorescence intensity; a.u., arbitrary units.

binding of these APLs to the T cell clone (Fig. 1f). These results show that TetTCR-seq is able to resolve positively bound peptides in primary T cells and identify up to five cross-reactive peptides per cell.

Most primary T cells were classified as binding one peptide (Fig. 1g). This result is expected because the probability of TCR cross-reactivity between similar peptides is higher than disparate ones^{8,9}. Among the peptides surveyed, we found a high degree of peptide diversity in the foreign-antigen-binding naive T cells (Fig. 1h). This diversity reduced

to two dominant peptides for CMV and influenza in the non-naive repertoire¹⁰ (Fig. 1h). This is expected given the CMV seropositive status and the high probability of influenza exposure or vaccination for this donor. The majority of cells within the endogenous-antigen-binding population bound MART1-A2L, which corroborates its high documented frequency^{6,10} (Fig. 1h). Linked TCR and DNA-BC analysis revealed that TCR α V genes 12-2 and 12-1/12-2 dominated in MART1-A2L- and YFV-LLW-specific TCRs, respectively (Fig. 1i),

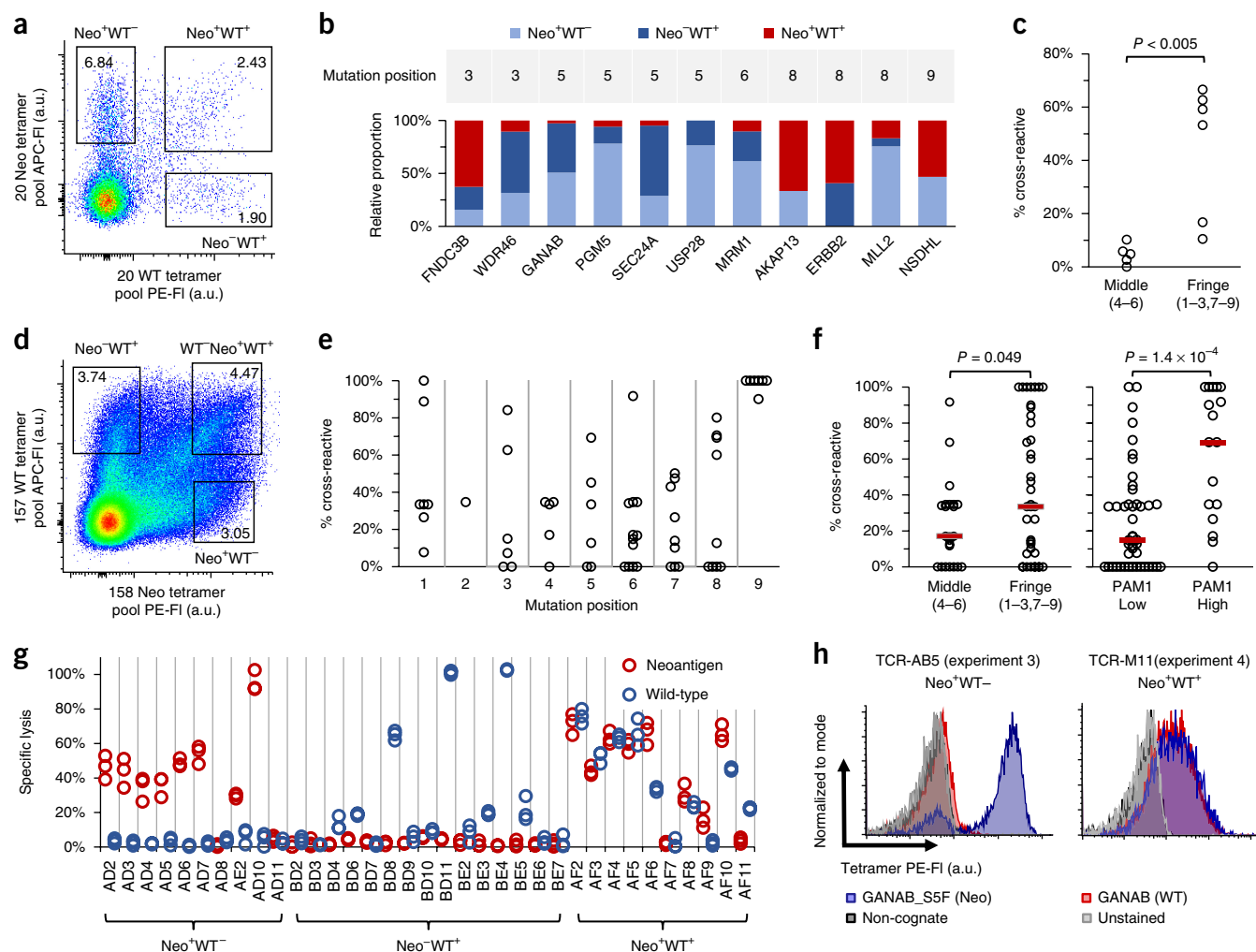


Figure 2 High prevalence of neoantigen-binding T cells that cross-react to WT counterpart peptides, and high-throughput isolation of neoantigen-specific TCRs for multiple specificities in parallel using TetTCR-seq. **(a–c)** Experiment 3, isolation of single Neo- and/or WT-binding T cells from a healthy donor using a library of 20 Neo–WT antigen pairs. **(a)** DNA-BC pMHC tetramer staining profile of naive CD8⁺ T cells from the tetramer pool-enriched fraction. See **Supplementary Figure 10** for gating scheme. **(b)** Relative proportion of T cells among the three possible antigen binding combinations (Neo⁺WT[−], Neo⁺WT⁺, Neo⁺WT⁺) for each Neo–WT antigen pair from experiment 3. Only antigen pairs in which both peptides were detected in at least one cell and have at least three detected cells in total (149 cells; see Online Methods) were included. **(c)** Effect of neoantigen mutation position (indicated in parentheses) on the proportion of cross-reactive T cells from red bars in **b**. (5 Neo–WT pairs for middle and 6 for fringe, one-tailed Mann–Whitney *U*-test). **(d–f)** Experiments 5 and 6, isolation of Neo and/or WT binding T cells using a library of 315 Neo–WT antigen pairs. **(d)** Staining profile as in **a** for experiment 5. See **Supplementary Figure 15** for gating scheme. **(e)** Proportion of cross-reactive T cells for Neo–WT antigen pairs based on mutation position. Same data filter as **b** (62 Neo–WT pairs from 517 cells). **(f)** Effect of neoantigen mutation position as in **c** or PAM1 value on the proportion of cross-reactive T cells in **e**. Red bars denote median (left to right, *n* = 23, 39, 45 and 17 Neo–WT pairs, one-tailed Mann–Whitney *U*-test). Alternative analysis using contingency tables is shown in **Supplementary Figure 16**. **(g)** LDH cytotoxicity assay on *in vitro* expanded primary T cell lines sorted using DNA-BC pMHC tetramers as in **a** interacting with a lymphoblastoid cell line, T2 cells, pulsed with the 20-neoantigen peptide pool or 20-WT counterpart peptide pool. Each condition was performed in triplicates derived from separate wells of cells. **(h)** Staining of Jurkat 76 cell line transduced with TCRs from experiments 3 and 4 with the indicated tetramers. Experiment was performed once. Based on TetTCR-seq of the original T cells, TCR-AB5 recognized the neoantigen GANAB_S5F while TCR-M11 recognized both GANAB_S5F and its WT counterpart, GANAB.

which is consistent with other reports^{11,12}. TetTCR-seq on a second CMV-seropositive donor (experiment 2) verified the findings from experiment 1 (**Supplementary Fig. 9**). These results highlight the ability of TetTCR-seq to accurately link pMHC binding with TCR sequences.

We next applied TetTCR-seq to study the extent of cancer antigen cross-reactivity in healthy donor peripheral blood T cells and isolate neoantigen (Neo)-specific TCRs with no cross-reactivity to WT counterpart antigen. Naive T cells from healthy donors are a useful source of neoantigen-binding TCRs³. However, most neoantigens are one amino acid away from the WT sequence, meaning that neoantigen-binding TCRs may cross-react with host cells to cause autoimmunity.

Although clinical adverse effects caused by neoantigen-recognizing T cells cross-reacting with endogenous tissue have not been reported, possibly owing to the lack of technology development, other forms of cross-reactivity have been reported to cause death in clinical trials¹³. In experiment 3, we surveyed 20 pairs of Neo–WT peptides that bind HLA-A2. Since pMHC tetramers can select T cells reactive to peptides that are not naturally processed, peptides were chosen on the basis of previous evidence of tumor expression and T cell targeting^{3,14–16} (**Supplementary Table 4**). Neo and WT pMHC pools were labeled using two separate fluorophores, allowing sorting of three cell populations: Neo⁺WT[−], Neo⁺WT⁺ and Neo⁺WT⁺ (**Fig. 2a** and **Supplementary Fig. 10**).

T cells with two detected peptide binders accounted for 84% of the Neo⁺WT⁺ population, 98% of which belonged to a Neo–WT antigen pair (**Supplementary Fig. 11**). All cells with shared TCR $\alpha\beta$ amino acid sequence had the same peptide specificity (**Supplementary Fig. 12**). Cells in the Neo⁺WT⁺ population bound 11 of the 20 Neo–WT antigen-pairs, indicating that Neo–WT cross-reactivity is widespread in the precursor T cell repertoire (**Fig. 2b**). We noted that neoantigens with mutations at fringe positions 3, 8 and 9 elicited significantly more cross-reactive T cells than the ones at center positions 4, 5 and 6 (**Fig. 2c**, $P < 0.005$). This is consistent with previous reports using alanine scanning in a mouse model¹⁷. TetTCR-seq on a separate donor (experiment 4) showed the same trend (**Supplementary Fig. 13g**, $P < 0.05$). Undetected peptides from both experiments were due to their low T cell frequencies (**Supplementary Fig. 14**).

To test the feasibility of TetTCR-seq to screen larger libraries, we assembled a library of 157 Neo–WT antigen pairs and profiled T cell cross-reactivity in more than 1,000 tetramer⁺CD8⁺ sorted single T cells from two donors (experiments 5 and 6, **Fig. 2d** and **Supplementary Figs. 15** and **16**). ELISA on all 315 pMHC species showed no difference in pMHC ultraviolet exchange efficiency between detected and undetected peptides (**Supplementary Fig. 17**). As in experiment 3 and 4, neoantigen mutations in the fringes had elevated percentages of cross-reactive T cells relative to mutations in the middle (**Fig. 2e,f** and **Supplementary Fig. 16j**). Using this larger dataset, we also found that neoantigen mutations with high PAM1 values, a surrogate for chemical similarity related to evolutionary mutational probability¹⁸, had a significantly higher percentage of cross-reactive T cells than those with low PAM1 values (**Fig. 2f**, $P < 0.005$). Thus, in addition to mutation position, WT-binding T cells are more likely to recognize the neoantigen if the mutated amino acid is chemically similar to the original. Additional, so far unaccounted-for variations still exist between peptides, highlighting the necessity for experimental screening against WT cross-reactivity when using neoantigen-based therapy in cancer.

We also assessed the utility of TetTCR-seq for isolating neoantigen-specific TCRs with no cross-reactivity to WT. We generated primary cell lines, each derived from five sorted cells in the Neo⁺WT[–], Neo⁺WT⁺ and Neo⁺WT⁺ populations from experiments 3 and 4. These cells lysed antigen-pulsed target cells in a manner that matched their gating scheme during sorting, independent of the choice of pMHC tetramer fluorophore (**Fig. 2g** and **Supplementary Fig. 18**). Further TetTCR-seq analysis of Neo⁺WT[–] and Neo⁺WT⁺ T cell lines showed unique TCRs in each cell line targeting a limited range of antigens (**Supplementary Fig. 19a,b**). Cytotoxicity testing confirmed the cross-reactivity of Neo⁺WT⁺ cell lines as identified by TetTCR-seq (**Supplementary Fig. 19c**). Lastly, the antigen recognition of Jurkat76 cell lines transduced with TCRs identified from experiments 3 and 4 matched their original specificities (**Fig. 2h** and **Supplementary Fig. 20**). Together, our T cell line and TCR-transduced Jurkat experiments show that TetTCR-seq is not only capable of identifying cross-reactive TCRs on a large scale but can also identify functionally reactive neoantigen-specific TCRs that are not cross-reactive to WT peptide in a high-throughput manner, which could be valuable in TCR redirected adoptive cell transfer therapy^{3,19}.

In conclusion, we show that TetTCR-seq can accurately link TCR sequences with multiple antigenic pMHC binders in a high-throughput manner, which can be broadly applied to interrogate antigen-binding T cells in T cell populations, from infection to autoimmune disease and cancer immunotherapy, potentially even for individual patients. With methods emerging for predicting antigenic pMHCs for groups of TCR sequences⁴, TetTCR-seq can not only facilitate development in this area but also help to validate informatically predicted antigens. Lastly, TetTCR-seq can be integrated with

single-cell transcriptomics and proteomics to gain further insights into the connections between single T cell phenotype on the one hand and TCR sequence and pMHC-binding landscape on the other²⁰.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank B. Wendel for discussions and for producing recombinant HLA-A2; M.M. Davis and H. Huang at Stanford University for discussion of the lentiviral transduction protocol and providing a template TCR construct and HLA-A2 construct; W. Uckert at the Max Delbrück Center for Molecular Medicine for sharing the Jurkat 76 cell line; A. Brock at University of Texas Austin for sharing the HEK 293T cell line; J. Lou at the Institute of Biophysics, Chinese Academy of Sciences, for helping with HCV APL prediction; P. Parker, K. Patel and H. Pan for assistance with initial prototyping; and the NIH tetramer center for additional pMHC tetramer reagents. We also thank anonymous blood donors and staff members at We Are Blood for sample collection. This work was supported by NIH grants R00AG040149 (N.J.), S10OD020072 (N.J.) and R33CA225539 (N.J.), by NSF CAREER Award 1653866 (N.J.), by Welch Foundation grant F1785 (N.J.), by the Robert J. Kleberg, Jr. and Helen C. Kleberg Foundation (N.J.) and by National Natural Science Foundation of China major international (regional) joint research project 81220108006 (W.J.) and NSFC-NHMRC joint research grant 81561128016 (W.J.). N.J. is a Cancer Prevention and Research Institute of Texas (CPRIT) Scholar and a Damon Runyon-Rachleff Innovator. S.-Q.Z. is a recipient of Thrust 2000 – Archie W. Straiton Endowed Graduate Fellowship in Engineering No. 1. A.A.S. is a recipient of the Cockrell School of Engineering fellowship and the Thrust 2000 – Mario E. Ramirez Endowed Graduate Fellowship in Engineering.

AUTHOR CONTRIBUTIONS

S.-Q.Z. conceived and developed the technology platform. S.-Q.Z. and N.J. conceived and designed the study. S.-Q.Z. and K.-Y.M. designed, performed and analyzed data for the majority of experiments; A.A.S. and M.Z. performed TCR cloning, transduction and pMHC tetramer staining studies; C.H. wrote the script for converting sequencing data into TCR sequences, DNA-BC and MID, and predicted HCV APLs; C.M.W. and E.S. performed *in vitro* cell culture and functionality experiments; W.J. cosupervised study and codesigned some experiments; N.J. supervised the study; S.-Q.Z. and N.J. wrote the manuscript with feedback from all authors.

COMPETING INTERESTS

N.J. is a scientific advisor for ImmuDX LLC and Immune Arch Inc. A provisional patent application has been filed by the University of Texas at Austin for the method described here.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

- Newell, E.W. & Davis, M.M. *Nat. Biotechnol.* **32**, 149–157 (2014).
- Bentzen, A.K. *et al. Nat. Biotechnol.* **34**, 1037–1045 (2016).
- Strønen, E. *et al. Science* **352**, 1337–1341 (2016).
- Glanville, J. *et al. Nature* **547**, 94–98 (2017).
- Rodenko, B. *et al. Nat. Protoc.* **1**, 1120–1132 (2006).
- Yu, W. *et al. Immunity* **42**, 929–941 (2015).
- Zhang, S.Q. *et al. Sci. Transl. Med.* **8**, 341ra377 (2016).
- Birnbaum, M.E. *et al. Cell* **157**, 1073–1087 (2014).
- Bullock, T.N.J., Mullins, D.W., Colella, T.A. & Engelhard, V.H. *J. Immunol.* **167**, 5824–5831 (2001).
- Newell, E.W. *et al. Nat. Biotechnol.* **31**, 623–629 (2013).
- Bovay, A. *et al. Eur. J. Immunol.* **48**, 258–272 (2018).
- Dietrich, P.-Y. *J. Immunol.* **170**, 5103–5109 (2003).
- Cameron, B.J. *et al. Sci. Transl. Med.* **5**, 197ra103 (2013).
- Cohen, C.J. *et al. J. Clin. Invest.* **125**, 3981–3991 (2015).
- Rajasagi, M. *et al. Blood* **124**, 453–462 (2014).
- Carreno, B.M. *et al. Science* **348**, 803–808 (2015).
- Nelson, R.W. *et al. Immunity* **42**, 95–107 (2015).
- Wilbur, W.J. *Mol. Biol. Evol.* **2**, 434–447 (1985).
- Dudley, M.E. *et al. Science* **298**, 850–854 (2002).
- Peterson, V.M. *et al. Nat. Biotechnol.* **35**, 936–939 (2017).

ONLINE METHODS

PE- or APC-labeled streptavidin conjugation to DNA linker. Conjugation of the DNA linker to PE- or APC-labeled streptavidin was performed following manufacturer's protocols (Solulink). Excess unconjugated DNA linker was removed by six wash steps in a Vivaspin 6 100-kDa protein concentrator (GE Healthcare). Conjugates were concentrated, and then passed through a 0.2- μ m centrifugal filter. The molar DNA:protein conjugation ratio was kept between 1:3 and 1:7. DNA:protein conjugation ratio was determined by absorbance using a 1 mg/ml PE- or APC-labeled streptavidin reference solution. The absorbance of the DNA-streptavidin conjugate was then compared with this standard curve to determine the effective protein concentration of the conjugate. The DNA concentration was determined from the difference in the A_{260} absorbance between the DNA-streptavidin conjugate and a protein concentration-matched version of the PE or APC streptavidin.

Generation of DNA-barcoded fluorescent streptavidin. Annealing of the peptide-encoding oligonucleotide to the complementary DNA linker on the DNA linker PE or APC streptavidin conjugate was done at 55 °C for 5 min, then cooled to 25 °C at -0.1 °C/s in the presence of 250 μ M dNTP in Cutsmart buffer (NEB). Then 1 μ l of extension mixture consisting of 0.1 μ l Cutsmart 10 \times , and 0.125 μ l Klenow fragment Exo- (5 U/ μ l, NEB) was added before starting the extension at 37 °C for 1 h. The reaction was stopped by adding EDTA. The final DNA-barcoded fluorescent streptavidin conjugate was stored at 4 °C. Corresponds to steps 2.1 and 2.2 in **Figure 1a**.

In vitro transcription and translation. Peptide-encoding DNA oligonucleotides were purchased from IDT and Sigma. DNA templates were generated by PCR with 400 μ M dNTP, 1.05 μ M IVTT_r primer, 1 μ M IVTT_f primer, 25 pM DNA oligonucleotide and 0.0375 U/ μ l Ex Taq HS DNA polymerase. The reaction proceeded at 95 °C 3 min; then 30 cycles of 95 °C 20 s, 52 °C 40 s, 72 °C 45 s; then 72 °C 5 min. The PCR product was diluted with 73.3 μ l of water. Corresponds to step 1.1 in **Figure 1a**.

Twenty microliters of 1.5 \times concentrated PURExpress IVTT master mix (NEB) was made, consisting of 10 μ l solution A, 7.5 μ l solution B, 0.8 μ l of Release Factor 1+2+3 (5 reaction/ μ l, NEB special order), 0.25 μ l enterokinase (16 U/ μ l, NEB), 0.25 μ l murine RNase inhibitor (40 U/ μ l, NEB), and 1.2 μ l H₂O. One microliter of the diluted PCR product was added to 2 μ l of the IVTT master mix on ice and then incubated at 30 °C for 4 h. Corresponds to step 1.2 in **Figure 1a**.

pMHC UV exchange and tetramerization. Biotinylated pMHC containing a UV-labile peptide was directly added to the completed IVTT reaction. pMHC UV exchange and tetramerization followed a previously described protocol^{5,6} (see **Supplementary Note**). The UV exchange was performed for 60 min on ice, and then the sample was incubated at 4 °C for at least 12 h. Confirmation of the quality and concentration of UV-exchanged pMHC monomer was assessed by an ELISA assay as described previously⁵. DNA-barcoded fluorescent streptavidin conjugate was then added to its corresponding UV-exchanged pMHC monomer mix at molar ratio of 1:6.7 and incubated at 4 °C for 1 h to produce the final DNA-BC pMHC tetramer. Corresponds to steps 1.3 and 3.1 in **Figure 1a**.

DNA-BC pMHC tetramer pooling. 500 μ l of staining buffer (PBS, 5 mM EDTA, 2% FBS, 100 μ g/ml salmon sperm DNA, 100 μ M D-biotin, 0.05% sodium azide) was added to a 100 kDa Vivaspin protein concentrator (GE) and incubated for at least 30 min. The concentrator was spun at 10,000g and further staining buffer was added until 1 ml of solution ran through the membrane. Immediately before cell staining, 0.65 μ l of each DNA pMHC tetramer is added to 400 μ l of staining buffer, transferred to the concentrator, and then spun at 7,000g for 10 min or longer until the volume reached \sim 50 μ l. Corresponds to **Figure 1b**, left.

pMHC tetramer staining and sorting. Human Leukocyte Reduction System (LRS) chambers were obtained from deidentified donors by staff members at We Are Blood with informed consent. The use of LRS chamber from deidentified donors for this study was approved by the Institutional Review Board of the University of Texas at Austin and is compliant with all relevant ethical

regulations. Antigen-specific T cell isolation was performed following a previously established protocol⁶. In brief, CD8⁺ T cells were isolated from LRS chambers using the RosetteSep CD8⁺ T cell enrichment cocktail (Stemcell) together with Ficoll-paque (GE Healthcare).

Cells were either kept either on ice or at 4 °C in refrigerator for the remainder of the experiment. In experiments 1, 2, 3 and 5, an HCV-KLV(WT) binding clone, prestained with BV605-CD8a, was spiked into the main sample. Cells were resuspended into staining buffer containing \sim 60 nM of each DNA-BC pMHC tetramer and 0.025 mg/ml of BV785-CD8a (RPA-T8) antibody and incubated for 1 h at 4 °C. Cells were washed and then incubated with anti-PE and anti-APC microbeads (Miltenyi). After washing, tetramer-staining cells were enriched using an LS column (Miltenyi). The enriched fraction was eluted off the column and washed into FACS buffer containing 0.05% sodium azide, and stained with AF488-CD3, 7-AAD, BV421-CCR7, BV510-CD45RA and BV785-CD8a (BioLegend). The tetramer-depleted flow-through fraction was stained with the same antibody panel. Single cells were sorted using BD FACSaria II into 4 μ l lysis buffer following a previously published protocol⁷.

TCR library preparation. Single-cell TCR amplification and sequencing was done following a published protocol⁷ with a minor modification. RT was performed on TCR in the same lysis well when DNA-BCs were present. During the first PCR amplification, P1 and P2 primers were included in the primer mix at 100 nM final concentration for concurrent amplification of TCR and the DNA-BC (**Supplementary Table 10**). During subsequent PCRs, TCR and DNA-BC were amplified separately in parallel wells in 384-well plates. After PCR, multiple cells were pooled, purified by gel electrophoresis and extraction, and then sequenced using Illumina Mi-seq V2 kit. Sequence reads were analyzed following a previous protocol⁷. Corresponds to **Figure 1b**.

DNA-BC library preparation. One microliter of the first PCR product from the TCR and DNA-BC amplification was combined with 100 nM V1f_rxn2 and V1r_rxn2 primers and 0.025 U/ μ l Ex Taq HS (Takara) to 5 μ l volume for a second PCR. PCR proceeded at 95 °C 3 min; then ten cycles of 95 °C 20 s, 55 °C 40 s, and 72 °C 45 s; then 72 °C 5 min. PCR primers include cell barcode sequences to encode wells and a partial Illumina adaptor as previously described⁷.

A third PCR was used to add an Illumina adaptor using ILLU_f and ILLU_r primers. PCR proceeded by the same configuration as the second PCR, but using five cycles. Multiple wells were then pooled and purified by gel electrophoresis and gel extraction. The library was sequenced to a depth of at least 6,000 reads/cell. Refer to **Supplementary Table 10** for primer sequences. Corresponds to **Figure 1b**.

DNA-BC sequence processing. Raw reads were filtered based on the constant region of the DNA-BC. Reads were further separated according to cell barcodes. Within each cell barcode, reads with an identical MID sequence were clustered together and a consensus peptide-encoding sequence was built for each cluster. Each cluster represents one MID count.

Clusters were filtered based on the peptide-encoding region to be 25–30 nt in length, and with a Levenshtein distance no greater than 2 from the nearest known DNA-BC sequence. A histogram was then created expressing the percentage of total reads belonging to each group of clusters sharing the same read count. Clusters with low read counts, which occur as a result of sequencing errors, were removed (**Supplementary Fig. 7a–c**)²¹. The clusters were then collected into their corresponding cell and peptide based on the cell barcode and peptide-encoding DNA sequence, respectively.

Calculation of antigen-binding T cell frequency. The absolute frequency of antigen-binding T cells for peptide a_i is calculated as follows:

$$\text{Freq}(a_i) = \frac{\frac{\text{No. } a_i\text{-specific T cells}}{\text{No. total tetramer}^+ \text{ T cells}} \times \frac{\text{No. beads counted}}{\text{Theoretical No. beads added}}}{\text{Total CD8}^+ \text{ T cell count}}$$

The total CD8⁺ T cell count was determined by measuring the fraction of CD8⁺ T cells in the flow-through using 7-AAD with anti-CD3 and anti-CD8 antibodies, multiplied by the total live cell count of flow-through by a cell

counter. In experiments 2 and 5, counting beads were not added to the sample but the entire enriched fraction was recorded; therefore, we assumed 100% bead recovery in these two experiments. Absolute frequencies could not be calculated for experiment 6 because counting beads were not added and not all of the enriched fraction was recorded.

Calculation of percentage cross-reactive T cells for experiments 3–6. We calculate the relative proportion of T cells belonging to the Neo⁺WT⁺, Neo[−]WT⁺ and Neo⁺WT[−] antigen-binding cell populations for each Neo–WT antigen-pair using cells with positive antigen detection. We restricted our analysis to cells with one identified antigen in the Neo[−]WT⁺ and Neo⁺WT[−] sorted populations and two identified antigens in the Neo⁺WT⁺ sorted population (Supplementary Figs. 11e, 13e and 16i). From this dataset, we performed normalization to account for differences in the frequency and number of cells sorted for the three cell populations. Taking these two normalizations into account, the equation for calculating the relative proportion p of cells binding to peptide a in population b for experiments 3 and 4 is

$$p(a_i, b_j) = \frac{\text{relfreq}(b_j) \times \frac{\text{count}(a_i, b_j)}{\text{totalsort}(b_j)}}{\sum_b \frac{\text{relfreq}(b) \text{count}(a_i, b)}{\text{totalsort}(b)}}$$

a_i refers to a Neo–WT antigen-pair in the Neo⁺WT⁺ population, the corresponding WT peptide only in the Neo[−]WT⁺ population, and the corresponding Neo peptide only in the Neo⁺WT[−] population. b_j refers to one of the three cell populations, Neo⁺WT[−], Neo[−]WT⁺ or Neo⁺WT⁺. $\text{count}(a_i, b_j)$ refers to the antigen-binding T cell count in cell population b_j binding to peptide a_i . $\text{relfreq}(b_j)$ refers to the percentage of cell population b_j taken from the tetramer gating in the tetramer-enriched fraction, which is a measure of the relative cell frequency (Supplementary Fig. 10a). $\text{totalsort}(b_j)$ is the total number of cells sorted for cell population b_j .

The percentage of cross-reactive T cells for any Neo–WT antigen-pair a_i is $p(a_i, b_{\text{Neo}^+ \text{WT}^+})$ (same values as red bars in Fig. 2b). While this calculation can be performed for all Neo–WT antigen pairs, we restricted our analysis to Neo–WT antigen pairs containing at least three cells in which both the Neo and WT antigen were detected in at least one cell. PAM1 values for amino acid pairs i and j are calculated by adding the unidirectional PAM1 values, $\text{PAM1}_{ij} + \text{PAM1}_{ji}$, as defined by Wilbur *et al.*¹⁸. This was calculated for all Neo–WT antigen pairs (Supplementary Fig. 16k).

An aggregate analysis was performed for experiments 5 and 6. Since cells are aggregated from these two experiments, we normalized for the cell counts in the three tetramer⁺ populations, but not the cell frequency, because the relative frequencies of the three cell populations in both experiments were comparable. The altered equation used for experiments 5 and 6 is the following:

$$p(a_i, b_j) = \frac{\text{count}(a_i, b_j) / \text{totalsort}(b_j)}{\sum_{b_1} \frac{\text{count}(a_i, b_1)}{\text{totalsort}(b_1)}}$$

T cell lines and functional assay. T cell lines were generated according to previously published protocol⁶, but using the DNA-BC pMHC tetramer pool. Cells were gated in the same manner as in Supplementary Figure 10 except for the AF488 channel, where CD3-AF488 was replaced by the dump channel CD4,14,16,19,32,56-AF488. Five cells from the same population (Neo⁺WT[−], Neo[−]WT⁺, Neo⁺WT⁺) were sorted into each well. Functional status was analyzed 10–21 d after restimulation.

Functionality was measured and analyzed using the LDH cytotoxicity assay kit (ThermoFisher) following the manufacturer's instructions as described previously. For Figure 2g and Supplementary Figure 18, T2 cells (ATCC) were pulsed with a peptide pool consisting of either the 20 neoantigen peptides (250 μM total, 12.5 μM each peptide) or 20 WT peptides (250 μM total, 12.5 μM each peptide). Background cytotoxicity was subtracted by using T2 cells pulsed with HCV-KLV(WT) peptide (250 μM). For Supplementary Figure 19c, T2 cells were pulsed with 12.5 μM of a single peptide or a peptide pool consisting of the 19 indicated neoantigen or WT peptides at 12.5 μM per peptide. Background cytotoxicity was subtracted by using T2 cells not pulsed with peptide. For each well, 60,000 T cells were incubated with 6,000 peptide-pulsed T2 cells for 4 h at 37 °C. Each condition for each cell line (derived from five single sorted cells) was performed in triplicates.

Lentiviral TCR transduction. Lentivirus production and TCR transduction was performed as previously described⁴ with the following modifications. TCRs were synthesized as GenParts (GenScript) and cloned into pLEX_307 (a gift from David Root via Addgene) under an EF-1a promoter. The vector also confers puromycin resistance. All vector sequences were confirmed via Sanger sequencing before viral production. Seventy-two hours after transduction, expression of the TCR was analyzed by flow cytometry. Antigen binding of the transduced cells was confirmed by pMHC tetramer and anti-CD3 antibody (BioLegend) staining.

Statistics and reproducibility. The relevant statistical test, sample size, replicate type, and P values for each figure are found in the figure and/or corresponding figure legend. For functional assays, an LDH value was measured for each of three separate wells of cells, all derived from the same cell lines, that were subjected to the same condition (Fig. 2g and Supplementary Figs. 18 and 19c). Representative experiments 1, 3 and 5 were each repeated once with similar results, as described for experiments 2, 4 and 6, respectively.

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

Code availability. Custom analysis code can be downloaded from GitHub at <https://github.com/utjianglab/TetTCR>.

Data availability. All TCR and peptide information is given in Supplementary Tables 1–10. Sequencing data can be accessed with accession code phs001678.v1.p1 from dbGaP.

21. Fu, G.K., Wilhelmy, J., Stern, D., Fan, H.C. & Fodor, S.P.A. *Anal. Chem.* **86**, 2867–2870 (2014).

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

► Experimental design

1. Sample size

Describe how sample size was determined.

no sample size calculation was performed

2. Data exclusions

Describe any data exclusions.

For Experiment 1-6, we excluded cells that were classified as not binding any peptide from further downstream analyses using a pre-established exclusion criteria that can be found in Supplementary Information.

For tallying of antigen-binding cell counts (Supplementary Fig. 11e,13e,16i), We included cells with 1 detected antigen specificity in the Neo+WT- and Neo-WT+ cell populations and cells with 2 detected antigen specificity in the Neo+WT+ cell population.

For analysis of percent cross-reactivity in Experiment 3-6 (Fig 2b,c,e,f and Supplemental Fig. 13f-h, 16j), we take the tallied cell counts and further analyze only Neo-WT antigen pairs that contain at least 3 cells total and have positive detection of both the Neo and WT peptide.

In Experiment 4, 4 peptides were included from the MAGE-A antigen family that was not included in Experiment 3. 3 MAGE-A specific T cells were detected out of 298 cells and were not used for subsequent analysis.

Experiment 5 and 6, some cells did not meet the gating thresholds and were removed prior to downstream analysis. This removed 6.9% (60/864) and 4.8% (14/289) of sorted cells from experiment 5 and 6, respectively.

In Experiment 2, 5, and 6, we removed the consensus TCR sequences CIEHTNSGGSNYKLTF for CDR3alpha and CASSLEETQYF for CDR3beta prior to TCR analysis. These TCR sequences are from a feeder cell contamination resulting from culturing the HCV-specific T cell clones used as spike-in.

3. Replication

Describe whether the experimental findings were reliably reproduced.

all attempts at replication were successful

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Not relevant to our study because subjects were not allocated into experimental groups

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

blinding was not relevant to this study because we were not running a controlled trial

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- ☐ ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
- ☐ ☒ A statement indicating how many times each experiment was replicated
- ☐ ☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- ☒ ☐ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- ☒ ☐ The test results (e.g. p values) given as exact values whenever possible and with confidence intervals noted
- ☐ ☒ A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- ☐ ☒ Clearly defined error bars

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

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

7. Software

Describe the software used to analyze the data in this study.

FACS diva and flowjo V10 were used for flow cytometry analysis. Microsoft Excel 2016 used for data analysis. Perl, Java, Groovy, and Python were used for DNA-BC analysis.

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* [guidance for providing algorithms and software for publication](#) may be useful for any submission.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

Release Factor 1+2+3 are a special order from NEB; there are no special restrictions on this item. All other reagents are commercially available as indicated in online methods

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

all antibodies are supplied by Biolegend and are validated for use in Flow Cytometry according to their website. Antibody titration was performed for each antibody on human peripheral blood mononuclear cells to determine the correct concentration to use.
 CD8a-BV605 (Cat #: 301040, Clone RPA-T8)
 CD8a-BV785 (Cat #: 301046, Clone RPA-T8)
 CD8a (Cat #: 301002, Clone RPA-T8)
 CCR7-BV421 (Cat #: 353208, Clone G043H7)
 CD45RA-BV510 (Cat #: 304142, Clone HI100)
 HLA-A2-PE (Cat # 343306, clone BB7.2)
 CD3-AF488 (Cat # 317310, Clone OKT3)

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

T2 were from ATCC.
 Jurkat 76 cell line from Dr. Wolfgang Uckert.
 All T cell clones and lines are in vitro expanded primary T cells from healthy de-identified human blood donors.

b. Describe the method of cell line authentication used.

All T cell clones and lines were authenticated by antibody staining with anti-CD8. T2 cell lines were verified by anti-HLA-A2 antibody staining. Jurkat cells validated by anti-CD3 and anti-TCR antibody staining.

c. Report whether the cell lines were tested for mycoplasma contamination.

cells lines were not tested for mycoplasma contamination

d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

No commonly misidentified cell lines were used

► **Animals and human research participants**

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

no animals were used

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

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Human Leukocyte Reduction System (LRS) chambers were obtained from de-identified donors at the We are Blood with strict adherence to guidelines from the Institutional Review Board of the University of Texas at Austin. Refer to Supplementary Figure 5 for age, gender and CMV-serostatus information.