High-throughput Biomolecular Technologies and Computational Analysis for Antigen-specific CD8 T cells

Jingyi Xie

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

James R. Heath, Chair

Philip D. Greenberg

Helen Y. Chu

Program Authorized to Offer Degree:

Molecular Engineering and Sciences

University of Washington

Abstract

High-throughput Biomolecular Technologies and Computational Analysis for Antigen-specific CD8 T cells

Jingyi Xie

Chair of the Supervisory Committee:

James Heath

Department of Bioengineering

Cytotoxic CD8 T cell responses represent a major element of the adaptive immune response. Elucidating the relationships between the antigen specificity, TCR clonotype, and T cell functionality, however, has remained an unsolved problem. Here, we present an integrated experimental-computational framework designed for the high-throughput capturing and analysis of antigen-specific CD8 T cells. First, we introduce Antigen-TCR Pairing and Multiomic Analysis of T-cells (APMAT) to profile SARS-CoV-2 specific T cells from a large cohort of COVID-19 participants. We systematically demonstrate how distinct physicochemical features of the antigen-TCR pairs strongly associate with both T cell phenotype and persistence. Next, we present a high-throughput cancer-specific TCR discovery strategy against neoantigens using artificial antigen-

presenting cells based on single-chain-trimer (SCT) peptide - major histocompatibility complex (pMHC), identifying potential TCR targets for adoptive cell therapy. Additionally, we introduce peptide-loaded cleavable-SCT (pCSCT), a highly stable, versatile, and easy-to-use reagent for efficient capturing of antigen-specific CD8 T cells. We then demonstrate the high-throughput application of pCSCT to profile antigen-specific T cells in Human papillomavirus-16 (HPV16)-positive individuals. In summary, we have leveraged high-throughput peptide-MHC technologies, single-cell multi-omics analysis, and TCR engineering to better understand the CD8 T cell response. Our findings offer valuable insights into adaptive immunity across infectious disease, vaccination, and cancer immunotherapy.

Table of contents

ACKNOWLEDGEMENTS

First and foremost, I am eternally grateful to my advisor, Professor James Heath, for his invaluable guidance throughout my doctoral journey. I vividly remember the first time I met Jim, in his signature Hawaiian shirt, as he showed me around the lab. You ignited my passion for science, and provided me with the remarkable opportunity to explore interdisciplinary research, opening up a myriad of exciting future possibilities for my life and career.

I would also like to express my sincere gratitude to the other members of my committee: Professor Philip Greenberg, Professor Helen Chu, and Professor Linda Shapiro, for their generous support and invaluable feedback. I want to extend my thanks to my M.S. advisor, Professor Shaoyi Jiang, without whose support I would not have developed a keen interest in pursuing a Ph.D. degree.

During my doctoral studies at the Institute for Systems Biology, I was surrounded by many brilliant and inspiring colleagues from the Heath lab, Wei lab, and facility teams. I am deeply grateful to Dr. William Chour, Dr. Jongchan Choi, and Dr. Yapeng Su, for their mentorship and guidance at various stages of my Ph.D. journey. I feel fortunate to have collaborated with many talented individuals over the years. I extend my gratitude to all members of the Heath lab for their unwavering support. Special thanks to my lab sisters Rongyu Zhang, Dan Yuan, Rachel Ng, and Jingqi Qi, who made my time spent in and outside the lab memorable. For the ups and downs, laughter and tears, I couldn't have made it this far without these lovely ladies.

To my friends in Seattle, including Shihan Xu, Yue Guo, and many others, being far away from home, you all made Seattle a new home for me.

DEDICATION

To my beloved parents, Hua Xie and Liping Wang, thank you for your endless love. You not only brought me into this world but also empowered me to pursue my dreams. To my younger sister Jingxuan Xie - no matter what challenges we face, we stand united.

To my wonderful husband, Zhehao Li, who came into my life during a period of darkness, bringing endless positive energy. My life has been enriched by your unconditional love and support. This journey would have been impossible without you by my side.

Chapter 1. INTRODUCTION

1.1 ADAPTIVE IMMUNITY AND ANTIGEN-SPECIFIC CD8 T CELL RESPONSE

The adaptive immune response against pathogens is highly antigen-specific. The hallmark of protective adaptive immunity is its ability to recognize and target a wide range of antigens. There are three major components in adaptive immunity: B cells (mainly antibody producing cells), CD4 T cells (helper T cells) and CD8 T cells (cytotoxic T cells)[1]. CD8 T cells are essential to combat intracellular bacteria, viral infection, and cancer. CD8 T cells recognize targeted cells through the interaction between T cell receptor (TCR) and major histocompatibility complex I (MHC-I) molecules that present a disease associated peptide[1]. MHC-I molecules are expressed on the surface of all nucleated cells (with certain exceptions), allowing CD8 T cells to survey intracellular infections or altered MHC-presenting peptides.

The sequence and antigen specificity of the TCR that each cell carries are determined through the rearrangement of the TCR α and β loci during early T cell development. Roughly 95% of the TCRs on circulating T cells are αβ heterodimers[2,3]. The antigen specificity is mainly dependent on the complementarity-determining region 3 (CDR3). The diversity and composition of the TCR repertoire is highly dynamic during acute and chronic infection[4], cancer development and treatment[5]. During primary activation, T cells become activated through antigen recognition, resulting in rapidly proliferation and clonal expansion[6,7]. To eliminate virus-infected cells, the expanded CD8 T cells perform effector functions by secreting perforin/granzyme, inducing apoptosis, and secreting proinflammatory cytokines such as IFN- γ[1,6]. Following antigen elimination, while a few antigen-specific CD8 T cells survive and establish a memory phenotype,

most of the CD8 T cells die via apoptosis during the immune contraction phase. Mechanisms underlying this T cell fate choice remains an active focus.

1.2 Detection and analysis of CD8 T cells: impact, challenges and needs

MHC-I molecules present "non-self" peptides derived from virus, tumor-specific mutations (neoantigens), or aberrantly expressed tumor-associated proteins. The high level of polymorphism in MHC molecules, especially in the peptide binding region, leads to allele-specific repertoires for presented peptides. Cytotoxic CD8 T cell responses against a variety of immunogenic antigens represent a major element of the adaptive immune response. However, it is extremely challenging to fully evaluate the breadth of TCR affinities, antigen specificities, and functionality during infectious disease or cancer development. One major technology limitation is the need for high throughput matching between TCRs and cognate antigens in the context of the low abundancy of each antigen-specific T cell. Such technology could provide insight into interpreting TCR specificity and advantage the development of precise cancer immunotherapies, novel therapeutics in infectious disease, and T cell-directed vaccines.

Over the years, a majority of assays for identifying antigen-specific CD8 T cells have been developed that monitor functional readouts. For example, SARS-CoV-2 specific CD8 T cells were measured using *in vitro* stimulation with peptide pools derived from viral sequences followed by flow cytometry based on activation markers (surface CD69, CD137/4-1BB), or intracellular cytokine staining[8]. However, activated T cells identified via this approach provide limited information of the exact minimal epitope recognized by the cognate TCR, and might include non-specifically activated TCRs. Peptide-MHC molecules have been used extensively for the detection

and isolation of such antigen-specific T cells via TCR-pMHC interaction[9–13]. A library of single chain trimer peptide-MHCs (SCT pMHCs) represents a set of soluble pMHCs in which peptide, β2-microglobulin and an HLA subunit were covalently linked and expressed by mammalian cells. Compared with standard refolded pMHCs, SCT technology has multiple advantages, including large-scale expression of Class I pMHCs with higher stability. Additionally, mammalian expression systems allow possible beneficial post-translational modifications such as N-linked glycosylation. SCT pMHC multimers have been a powerful and versatile tool for the detection of antigen-specific T cells[14]. Most importantly, this strategy has allowed the direct capturing of CD8 T cells specific for vast numbers of potential epitopes from biological material, allowing for the phenotype, TCR, and exact epitope to be simultaneously obtained to provide biological insights.

1.3 CD8 T Cell Receptor (TCR) Engineered Therapies for the Treatment of Cancer

Within recent decades, the field of immunotherapy has revolutionized the treatment of many cancers. The approaches include various immune-checkpoint inhibitors, autologous tumor-infiltrating lymphocyte (TIL) therapy, antigen-specific endogenous T-cell therapy (ETC), chimeric antigen receptor T cell (CAR-T) therapies, T-cell receptor engineered T cell therapy (TCR-T), and personalized neoantigen-based therapeutic cancer vaccines. The ultimate goal of treating cancer is to target and kill cancer cells without adverse effects on normal cells. While CAR-T therapies are highly promising, their use is limited to targeting surface proteins on tumor cells, such as CD19, that usually do not show a high degree of tumor specificity compared to healthy tissue and have demonstrated little clinical efficacy in the setting of solid tumors[15]. In contrast, TCR-T cell approaches can possibly overcome these obstacles. Neoantigens are a major

class of targetable cancer antigens resulting from non-synonymous somatic mutations. Mutational neoantigens have been a main target for TCR-T therapeutics[16]. Interactions between the TCR and their cognate cancer antigens are central to studying the antitumor response and evaluating the therapeutic potential of adoptive T cell transfer therapy. Recent research has revealed the phenotype and reactivity of anti-tumor CD8 T cells in cancer patients. Such tumor-reactive T cells have been enriched from circulating PBMCs or tumor infiltrated lymphocytes (TILs) of cancer patients[5,17–31]. However, the efficacy of naturally isolated and expanded T cells in therapy has been limited by a number of factors, including low TCR avidity towards the neoantigen, off-target toxicities on wild-type self-antigen, and target antigen loss or HLA loss leading to tumor escape. The large-scale generation of neoantigen-HLA T cell capturing reagents and single cell TCR sequencing provides a new approach to isolate and evaluate the therapeutic potential of cancer-specific TCRs. An object of future research includes the identification of clinical-grade tumor-specific TCRs that can undergo potential modifications for enhanced TCR signaling and effector function and expressed in a large number of patient T cells to create effective anti-tumor responses[32–34].

1.4 OVERVIEW OF PROJECT CHAPTERS

In chapter 2, we present a high-throughput cancer-specific TCR discovery strategy against neoantigens using artificial antigen-presenting cells based on single-chain-trimer (SCT) peptide - major histocompatibility complex (pMHC).

In chapter 3, we introduce Antigen-TCR Pairing and Multiomic Analysis of T-cells (APMAT) to profile SARS-CoV-2 specific T cells from a large cohort of COVID-19 participants.

In chapter 4, we introduce peptide-loaded cleavable-SCT (pCSCT), a highly stable, versatile, and easy-to-use reagent for efficient capturing of antigen-specific CD8 T cells.

In chapter 5, we demonstrate the high-throughput application of pCSCT to profile antigen-specific T cells in Human papillomavirus-16 (HPV16)-positive individuals.

Chapter 2. HIGH-THROUGHPUT DISCOVERY OF TCRS AGAINST PUBLIC NEOANTIGENS

2.1 ABSTRACT

The cytotoxicity against tumors by CD8 T cells in an antigen-specific manner have driven the rapid advancement of adoptive cell therapy (ACT) in cancer treatment. TCR-T therapy targeting tumor-specific antigens (TSAs), particularly shared public neoantigens (Pub-NeoAg), could enable broader "off-the-shelf" therapeutic application across patients with common driver mutations. However, the discovery of such TCRs remains challenging due to the rarity of these T cells and lack of effective screening methods. This study describes a high-throughput platform combining pMHC-like libraries and TCR cloning to identify and validate neoantigen-specific CD8 T cells from healthy donor peripheral blood mononuclear cells (PBMCs) using SCT-based artificial antigen-presenting cells. Specificity and therapeutic potential were further characterized for lead TCRs targeting the P53 neoantigens G245S and R248W. Overall, these findings validate a powerful strategy to systematically expand the TCR repertoire against public neoantigens and facilitate the development of broadly applicable TCR-T therapies.

2.2 INTRODUCTION

CD8 T cells recognize and eliminate tumors in an antigen-specific manner. Over the past few decades, adoptive cell therapy (ACT) has emerged as a rapidly growing field that utilizes immune cells to eliminate cancer. Currently, there are mainly three types of ACT that use either tumor-infiltrating lymphocytes (TILs) isolated from patient tumor samples, chimeric antigen receptor (CAR) engineered T cells, or T cell receptor (TCR) engineered T cells[1,2]. For TCR-T and CAR-T

therapies, peripheral T cells are generally isolated from a patient's blood, genetically modified to express either a TCR or a CAR molecule, and expanded for infusion.

There are several limitations for CAR-T therapy. Firstly, CARs typically target highly expressed molecules on the surface of cancer cells, but this is often complicated by off-target effects on healthy cells that also express these markers. As of 2023, FDA-approved CAR-T therapies include targeting CD19 or BCMA, both of which are antigens expressed on normal B cells[3]. Thus, their applicability is currently limited to treating B cell-related blood cancers and is associated with depletion of normal B cells. Additionally, the effectiveness of CAR-T therapy against solid tumors has been very limited due to a lack of unique antigen markers to definitively differentiate cancer tissue from normal tissue.

TCR-T therapy targeting tumor-associated antigens (TAAs) such as MART-1 and MAGE-A3 has resulted in toxicity to normal tissues, in part because these or cross-reactive antigens are detectable in normal tissues. Alternatively, tumor-specific antigens (TSAs) represent an elite class of TCR-T targets. Specifically, TCRs targeting patient-specific, private neoantigens may lead to personalized treatment. However, translating private neoantigen-targeted ACT to a broad range of patients is very challenging due to interpersonal variations in somatic mutations, which are commonly random rather than oncogenic events from replication infidelity, and the cost of customization. Furthermore, private neoantigens are subject to tumor genetic heterogeneity and resulting immunotherapy resistance[4].

In contrast, TCRs targeting shared public neoantigens (Pub-NeoAgs) may benefit a wide range of patients sharing the same mutation. Despite the unique neoantigen repertoire of each patient, Pub-NeoAgs are often generated from gain-of-function mutations in driver genes that directly contribute to cancer cell fitness, such as KRAS[5,6], TP53[7–9], PIK3CA[10] (which encodes the

p110α catalytic subunit of phosphoinositide 3-kinases), BRAF[11], and others[11]. Most importantly, these shared public neoantigens frequently occur in cancers and are shared among patients with certain "hotspot" mutations, such as KRAS-G12V, KRAS-G12D, TP53-G245S, TP53-R248W, and PIK3CA-H1047R. These "hotspot" mutations have been associated with high frequencies of lung cancer, breast cancer, prostate cancer, and others, thus representing a broad patient community. The identification of these shared neoantigens and their cognate T cell receptor (TCR) sequences would facilitate “off-the-shelf” ACTs that could be broadly applied to patients with matching HLA and driver mutations (Figure 2.1).

Interactions between TCRs and their cognate cancer antigens are central to studying the antitumor response and evaluating the therapeutic potential of ACT therapy. However, the identification of TCRs targeting Pub-NeoAgs with exact epitope specificity remains challenging due to the low frequency of these T cells in circulating blood and the lack of effective capture reagents. Only a limited number of public neoantigens, with their HLA restrictions, and corresponding TCRs have been identified to date[4,12–15]. There is an outstanding need for high-throughput discovery of a library of novel TCRs that can target a comprehensive list of public neoantigens.

The use of Single-Chain-Trimer (SCTs) as T cell capture reagents is demonstrated in this thesis as well as by others previously[16,17]. In this chapter, we explore the potential of adapting SCT reagents as artificial antigen-presenting cells for the sensitization and enrichment of neoantigen-specific CD8 T cells in healthy donor (HD) peripheral blood mononuclear cells (PBMCs).

2.3 RESULTS

2.3.1 *A high-throughput TCR discovery platform for public neoantigens.*

We developed a high-throughput platform for the discovery of neoantigen-specific TCRs from healthy donors. The platform proceeds in four steps, as illustrated in Figure 2.2. First, we built a library of SCTs representing high-prevalence public neoantigens (PubNeos), including KRAS, TP53, and PIK3CA, for multiple HLA alleles. Second, CD8 T cells isolated from HLA-matched healthy donors (HDs) undergo *in vitro* sensitization (IVS) and expansion using SCT-based artificial antigen-presenting cells (SCT-aAPCs). Third, enriched, PubNeo-SCT tetramer-positive cells are single-cell sorted into 96-well plates for single-cell TCR sequencing (scTCRseq). Finally, potential TCR clonotypes are reconstructed for T cell specificity and functional validation.

2.3.2 *SCT-based nanoparticle for the effective isolation of antigen-specific T cells*

Prior to the high-throughput platform, we first investigated the feasibility of using individual SCT-based nanoparticles (SCT-NPs) for antigen-specific T cell isolation and expansion. SCT-NPs for unique HLA A*02:01-restricted antigens were assembled as previously described[18]. TCR-transduced Jurkat or primary T cells were incubated with cognate SCT-NPs to allow effective identification of antigen-specific CD8 T cells. The NP-bound cells were then isolated with a magnet for cell culture and expansion (Figure 2.3A). We started with TCR-transduced Jurkat cells that showed varying levels of cognate tetramer signals, possibly due to inconsistent TCR-transduction efficiency or downregulation of TCR expression during cell passaging. After SCT-NP enrichment, we cultured the NP-bound cells for 3-5 days and observed a consistently high purity of antigen-specific cells when using 2.5-5 μl of SCT-NPs. For example, the CMV-tetramer positive cells were enriched from 30.1% to 96.3% without the need for FACS sorting (Figure 2.3B).

We next tested the efficiency of using SCT-NPs to capture low-frequency antigen-specific CD8 T cells. We've previously validated TCR079 identified from a COVID-19 patient , which targets A*02:01 restricted SARS-CoV-2 Spike antigen RLITGRLQSL. Here, TCR079-transduced primary CD8 T cells were spiked into non-cognate healthy donor CD8 T cells. SCT-NPs assembled with the A*02:01 Spike-46-RLITGRLQSL SCT (S46, the cognate antigen for TCR079) were used to isolate TCR079 cells from the mixture. The TCR079 cell frequencies (measured by S46-tetramer) were compared before SCT-NP enrichment and for NP-free cells right after isolation. The SCT-NP isolation showed consistent efficiencies ranging from 45.06% to 60.33% when the expected target cell frequencies ranged from 10% to 1% (Figure 2.S3A).

We then extended the use of SCT-NPs to polyclonal CD8 T cells directly isolated from patient PBMCs. Originally, the A*02:01 CMV-NLVP specific CD8 T cell frequency was 12.4%. After SCT-NP enrichment, the NP-free flow-through contained only 6.05% of such cells, consistent with the enrichment efficiency we observed earlier (around 50%) (Figure 2.S3B). The NP-bound cells (without washing and may contain non-specific cells) were then expanded via rapid expansion protocol (REP) and cultured for 12 days. The enriched CMV-specific T cell population increased to 54.1% with a nearly 400-fold increase in cell number (when estimated start cell number = 831). Overall, we demonstrated that SCT-coated magnetic nanoparticles are an effective tool for the isolation and enrichment of antigen-specific CD8 T cells.

2.3.3 *Enrichment of PubNeo$^+$ cells from multiple healthy donors after in vitro sensitization*

Based on the effective antigen-specific T cell enrichment using SCT-NPs, we next expanded application of SCTs by incorporation into SCT-aAPCs for high-throughput TCR discovery. We built a SCT panel representing a broad range of public neoantigens and common antigens for HLA A*02:01, B*07:02, A*03:01, A*24:02, B*08:01, and C*07:01 (Table 2.1, 2.2). Each SCT-aAPC

was individually prepared and then pooled by HLA for *in vitro* sensitization of CD8 T cells isolated from nine healthy donors (Figure 2.2). For each healthy donor, we only included those SCT-aAPCs with matched HLA for stimulation on Day 0 and Day 11. The frequency of PubNeo$^+$ cells was monitored via SCT-tetramer pools during stimulation. As expected, PubNeo$^+$ cells were below the detection threshold for the SCT-tetramer assay in most donors on Day 0. On Day 15, we observed a stimulation-induced enrichment of PubNeo$^+$ cells from multiple healthy donors (Figure 2.4). For example, A*02:01 KRAS-G12/G/D/V/R cocktail positive cells were identified from four HDs, with cell frequencies ranging from 0.78% to 6.63 (Top row, double-positive populations in Figure 2.4). Other SCT-NP cocktails that induced notable cell enrichment include B*07:02 P53-Y220C, A*02:01 MT1-ELAGIGILTV/NYESO-SLLMWITQV/WT1-RMFPNAPYL/PIK3CA-H1047R, A*02:01 P53-G245S, A*02:01 P53-Y220C, B*08:01 KRAS-G12V/R, etc. To further verify the antigen specificity of these PubNeo$^+$ cells, SCT-cocktail-positive cells were FACS-sorted and expanded. For example, PubNeo$^+$ cells were derived from HD-1 and HD-6 with frequencies ranging from 2.89% - 26.1% against tetramer panels of A*02:01-P53-G245S and A*02:01-P53-R248W SCTs (Figure 2.5). Hence, the HLA-matched SCT-aAPC sensitization strategy can effectively generate and allow enrichment of T cells that confer specificity to public neoantigens.

2.3.4 *TCR repertoire analysis and cloning for PubNeo$^+$ CD8 T cells*

After 20 days of stimulation with SCT-aAPCs, PubNeo$^+$ CD8 T cells were single-cell sorted into 96-well plates for indexed single-cell TCR sequencing. As a proof-of-concept study, we only performed TCR sequencing and cloning for those cells sorted by the A*02:01 P53-G245S and P53-R248W tetramer panels from HD-6, HD-7, and HD-1. Among all tetramer-positive cells, we sorted 300 cells and successfully sequenced productive αβ chains from 200 of them. We resolved a panel of 76 candidate unique TCR clonotypes using distinct variable segments and unique CDR3

sequences (Figures 2.6A-B). In addition to the clones detected in single cells (32.2%), we also observed several TCR sequences shared among multiple cells, indicating clonal expansion of these TCRs.

To preliminarily validate the antigen specificity of PubNeo^{+} TCRs, we selected a panel of 32 TCRs for cloning (Figure 2.6C). Genes for *TCR* α/β chains were transduced in TCR-KO Jurkat cells (CD8-negative) and pre-screened with the SCT-tetramer cocktail, followed by individual SCT-tetramer staining. We identified candidate TCRs with strong tetramer binding against A*02:01 P53-G245S (Figure 2.7A) and R248W tetramers (Figure 2.7B), despite the lack of a contribution by CD8, and without non-specific binding to an irrelevant A2-CMV tetramer. Interestingly, we observed that the tetramer signal for TCRP03 was significantly lower compared to other G245S-specific TCRs, indicating a weaker binding strength.

We also tested similar epitopes with varied lengths or peptide cores. The P53-G245S tetramer cocktail includes four elements that all contain the G245S mutation. The GSM_9-mer, GSM_10-mer, GSM_11-mer peptides all have the G245S mutation at the P2 position, while the YMC-11-mer peptide has the G245S mutation at the P10 position. Among the four shown TCRs, none of them were tetramer-positive for GSM_9-mer or YMC-11-mer (Figure 2.7A). TCR002 and TCR023 showed the highest tetramer binding signal for GSM_10-mer and a decreased signal for GSM_11-mer. TCR005 and TCRP03 only showed binding for GSM_10-mer. For P53-R248W, TCR027 showed tetramer binding to GGM_11-mer but no other R248W-tetramers (Figure 2.7B).

We further confirmed that these tetramer binding signals are TCR-dependent. For example, we observed an elevated tetramer signal with increased surface expression levels of TCRαβ for TCR002 (Figure 2.8).

2.3.5 *CD8-Coreceptor dependency of candidate PubNeo$^+$ TCRs*

The capacity of a TCR to function in a coreceptor-independent manner correlates with the TCR's intrinsic affinity to the peptide-HLA complex. Coreceptor-independent, high-affinity TCRs potentially have improved clinical anti-tumor efficacy. We hypothesized that some of the TCR candidates likely were lower affinity TCRs that require the CD8 coreceptor for peptide-MHC binding and activation. We next repeated the TCR-transduction using a CD8-expressing, TCR-knockout, Nur77-GFP Jurkat cell line that becomes GFP-positive upon TCR activation. TCR^+CD8^{pos} and TCR^+CD8^{neg} populations were compared for their tetramer binding intensity against the cognate tetramer (Figure 2.9A-B). We identified additional TCRs (such as TCR008, TCR013, TCR020) that did not pass our screening when transduced into the non-CD8-expressing Jurkat line. These TCRs showed relatively lower tetramer signals in a highly CD8-dependent manner (Figures 2.9C-D). Consistently, previously verified TCR001 and TCR005 both showed strong binding regardless of CD8 levels, indicating the capacity for CD8 independence. This result indicated that the SCT-aAPC sensitization method can identify both CD8-dependent and CD8-independent TCR targets.

2.3.6 *Activation of candidate PubNeo$^+$ TCRs using peptide-pulsed APCs*

Based on the differences in tetramer binding to multiple epitopes for each TCR, we hypothesized that these peptides also activate T cells differently. To determine if these TCRs are mutation-specific, we performed an activation assay using either mutant P53-G245S peptides (G245S Mut) or wild-type P53-G245 peptides (G245 WT) (Figure 2.10A). First, none of these TCRs were activated by the 9-mer to 11-mer G245 WT peptides, indicating that these TCRs are not likely to induce self-reactive autoimmune responses (Figure 2.10C). None of the TCR-transduced cells showed GFP activation in the absence of peptide. Additionally, we observed the

strongest GFP activation for TCR002, TCR005, and TCRP03 against the G245S 10-mer, which also showed the highest SCT tetramer binding intensity (Figure 2.10B). We observed slight GFP activation by the G245S 9-mer for TCR002 and TCR005, although these TCRs did not bind to the 9-mer SCT-tetramer. TCR008, with low tetramer binding and strong CD8-coreceptor dependency, did not show GFP activation for any G245S peptide, possibly indicating a higher threshold for activation. TCR-transduced Nur77-GFP and NFAT-GFP cells both showed GFP activation in response to titrated concentrations of the G245s 10-mer peptide (Figure 2.10D). As future work, peptide titration and GFP activation assays for all TCRs will be performed to access the activation kinetics and EC50 for each TCR.

2.3.7 *TCR recognition motif analysis using alanine and glycine scanning*

We next investigated the fine specificity of TCR002, TCR005, TCRP03, and TCR023 using amino acid (AA) scanning to define each TCR's binding motif (Figures 2.11A-D). TCR-transduced Nur77-GFP cells were incubated with A*02:01 T2 cells pulsed with distinct peptides differing from the original G245S 10-mer by substitutions with alanine (Ala) or glycine (Gly) at one of the residues, and the change in GFP activation was measured. For TCR002, replacement with either AA at P2-P7 resulted in near-complete loss of activation, while it showed partial tolerance to changes at P8-P10. Similarly, TCR005, TCRP03, and TCRP23 were partially permissive of AA substitutions at P6 and P8-P10, suggesting less contact with this region for the pMHC-TCR interaction (Figure 2.11E). For future work, to establish each TCR's cross-reactivity potential for self-antigens, peptide motif "GSMNRXPXXX" can be used to survey the human proteome for self-antigens that contain residues that permitted GFP activation when altered. A peptide-pulsed T2 assay can be performed to assess the cross-reactivity and rule out potential self-reactive TCRs.

2.4 CONCLUSION AND DISCUSSION

In this chapter, we presented a high-throughput TCR discovery platform for shared public neoantigens. The platform was supported by large-scale production of SCTs to accommodate the scale of epitopes and alleles. We developed a strategy using SCT-based artificial antigen-presenting cells (aAPCs) for *in vitro* sensitization and enrichment of public neoantigen (Pub-NeoAg)-specific TCRs from peripheral blood mononuclear cells (PBMCs) from healthy donors. Using this strategy, we discovered a panel of TCRs that recognize diverse public neoantigens restricted by prevalent HLA alleles. We assessed a panel of candidate TCRs for their antigen specificity, CD8 coreceptor dependency, T cell activation, and TCR recognition motif. Together, these findings demonstrated the potential of this high-throughput platform to facilitate "off-the-shelf" neoantigen-targeting TCR-T therapies.

There are several directions for future work: **First**, expanding the TCR panel and HLA coverage. Although we focused on two of the high-prevalent P53 mutations for TCR functional validation, the modular and high-throughput nature of the SCT-aAPC platform can be readily applied to other PubNeo-specific TCRs, thus covering other oncogenic driver genes and other HLAs. **Second**, exploring the mechanism of immunogenicity for the discovered epitopes. To discuss the physical basis for the immunogenicity of the identified G245S-10mer and R248-11mer epitopes, crystal structure determination and peptide-MHC melting temperature analyses of mutant and wild-type peptides bound to HLA*A:02:01 could be performed. Conformational changes and the features of the TCR-interacting region might suggest whether the immunogenicity derives from structural differences between the mutant and wild-type peptides. Structural analysis of the TCR-pMHC interaction will likely contribute to an optimized TCR product. **Third**, exploring the kinetics of binding by candidate TCRs. TCR-transduced T cells can be stained with

titrated concentrations of cognate tetramers, and MFI values could be fit to equilibrium binding curves to calculate relative KD values. **Fourth**, analyzing the functionality of candidate TCRs. Candidate TCRs can be transduced into primary CD8 T cells to quantify EC50 and functional avidity by measuring the secretion of IFN-γ and TNF-α. **Lastly**, assessing the cytotoxicity of TCR-transduced CD8 T cells in mediating tumor-specific cell killing via live-cell imaging (IncuCyte assay) by incubating them with mutation-matched, patient-derived cancer cell lines.

2.5 METHODS

2.5.1 *Leukopaks processing*

Leukopaks from HLA-matched healthy donors were purchased from STEMCELL and processed immediately to obtain PBMCs. Briefly, leukopaks were gently mixed with ammonium chloride solution (STEMCELL, #07800) at a volume: volume ratio of 1:4 and incubated at 4°C for 15 mins for red blood cell lysis. Cells were then centrifuged at 500 x g for 10 mins at room temperature (15 -25C) with the brake on (deceleration = 9). Platelets were removed by resuspending cells (PBS with 2% FBS) and centrifuging at 150 x g for 10 mins at room temperature with the brake off (deceleration = 0). PBMCs were then cryopreserved using CryoStor CS-10 (SIGMA, C2874) and stored in liquid nitrogen for future use.

2.5.2 *Endogenous TCR knock-out for primary T cells*

PBMCs were isolated from healthy donors leukopaks described previously. Primary CD8 T cells were purified by magnetic enrichment (Miltenyi Biotec) and activated with Dynabeads™ CD3/CD28 (Life Technologies) at a ratio of 1:1 of bead to cell for 48 hours in the presence of IL2 (100 IU/mL). Activated CD8 T cells were subsequently used for endogenous TCR knock-out (KO) suing CRIPSR/Cas9 systems. A mixture of Cas9 and two sgRNAs targeting the first exon of TRAC

(AGAGTCTCTCAGCTGGTACA) and TRBC (GCTGTCAAGTCCAGTTCTAC) is electroporated into CD8 T cells at 3 pulses of 1600C and 10ms using Neo electroporator (Thermo Fisher Scientific).

2.5.3 *SCT-based artificial antigen presenting nanoparticles*

PBMCs were obtained from healthy donors leukopaks described previously. Primary CD8 T cells were purified by magnetic enrichment (Miltenyi Biotec). A panel of SCT based artificial antigen-presenting nanoparticles (SCT-aAPCs) were individually prepared in the 96 well plate. In each well, magnetic nanoparticles (MV4500) were incubated with individual SCT monomer and T cell stimulators (including CD28, CD137, CD2, CD81, CD134, CD11a) for 1 hour to form SCT-aAPCs. Biotin was added into SCT-aAPCs for 30 minutes to block unoccupied streptavidin sites on the nanoparticle. SCT-aAPCs were then washed twice with wash buffer (PBS with 10% BSA). SCT-aAPCs with same HLA restriction were counted and pooled equally to form a SCT-aAPC cocktail for the specific HLA. For each SCT-aAPC cocktail, a set of dual-color labeled SCT tetramers were also prepared for flow analysis. HLA-matched SCT-aAPC cocktails were then spiked into CD8 T cells from each healthy donor at a bead to cell ratio of 1:2. Cells were cultured at 37°C and maintained every 3 days with CTL media (with IL-2, IL-7, and IL-15). On day 10, SCT-aAPCs in the cell culture were removed via magnetic rack. Cells were counted again and then spiked with fresh SCT-aAPC cocktails for another round of stimulation. Tetramer staining for each cell culture were performed on day 0, 10, and 20.

2.5.4 *Tetramer staining and 96-well single cell sorting*

Cocktails of SCT tetramers were prepared for each HLA or driver gene specificity as described previously. individually tetramerized with PE or APC labeled streptavidin at a 4:1 molar

ratio for 30 minutes at 4°C. Biotin was added at an 8:1 molar ratio to streptavidin to block unoccupied biotin binding sites on streptavidin. CD8 T cells co-cultured with SCT-aAPCs were incubated with 50nM protein kinase inhibitor (Dasatinib) for 30 minutes in 37°C prior to tetramer staining. CD8 T cells were then washed and stained with a pool of SCT tetramers for 20 minutes at 4°C, followed by staining with Apotracker Green (Biolegend) and FITC anti-human CD8 Antibody (Biolegend, 344704). Samples were incubated for 15 minutes at 4°C and washed 3 times before sorting. Single, live, CD8-positive, tetramer-positive T cells were indexed sorted by SonyMA900 cell sorter into 96-well plates containing 20 ul of lysis buffer (Takarabio, with 1U/μl RNase inhibitor, Promega).

2.5.5 *Single cell TCR sequencing*

Reverse transcription (RT) and sequence-specific amplification were performed in a series of three nested PCR analyses before sequencing on a MiSeq (Illumina) instrument. Briefly, following 1 hour of lysis at −80°C, the cell lysates were split in half to perform separate TCR α and β amplification. TCR α and β domain genes were cloned from split cell lysates using a OneStep RT PCR kit (QIAGEN) with primers listed in Table 4.2 and Table 4.3 followed by a second round of nested PCR amplification. The PCR product was purified by 0.8X AMPure beads (Beckman Coulter). Individual well was barcoded via third PCR using Nextera XT index Kit (Illumine). Finally, the generated amplified sequences were combined, gel purified and sequenced using 500 cycle v2 MiSeq reagents (Illumina). Demultiplexing data for individual wells were performed following standard Illumina protocols.

2.5.6 *Lentiviral transduction*

Lentiviral vectors encoding TCR α and β chain were described in a recent paper (Chour, et al). Briefly, codon-optimized DNA fragments encoding TCR α and β were PCR amplified, Gibson assembled and ligated into pRRL-SIN Lentiviral vector (gifted by Dr. Phil Greenburg). Primary T cells with endogenous TCR knock-out or Jurkat cells were transduced with lentiviral particles encoding the TCR of interest. Briefly, lentiviral vectors were produced by the transient transfection of HEK 293T cells with the lentiviral constructs above and packaging DNA (mixture of pMD2-G VSVG, pMDL g/pRRE, pRSV-REV at a 1:2:2 ratio) at a 5:2 ratio using Mirus TransIT LT1 (Mirus Bio). At 48- and 72-hours post-transfection, viral supernatant was collected and filtered through 0.45 uM PES filters and concentrated using Lenti-X concentrator (TAKARA). Primary T cells with endogenous TCR knock-out or Jurkat cells were infected with concentrated lentiviral particles in the presence of 2-4 ug/ml polyprene (Sigma-Aldrich) overnight. Cells were maintained every 2-3 days with CTL media with IL-2 supplemented for primary T cells.

2.5.7 *GFP activation assay*

TCR-transduced Nur77-GFP CD8+ cells were co-cultured with HLA-matched T2 cells pulsed with 1uM of the peptide at 1:1 ratio. After 16 hours of incubation, cells were spin down and stain for flow cytometry analysis

2.6 FIGURES

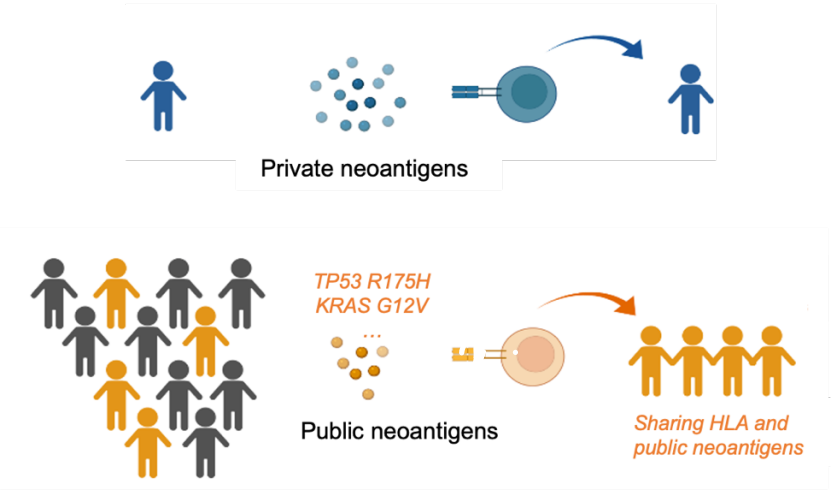


Figure 2.1 Targeting private or public neoantigens for adoptive cell therapy

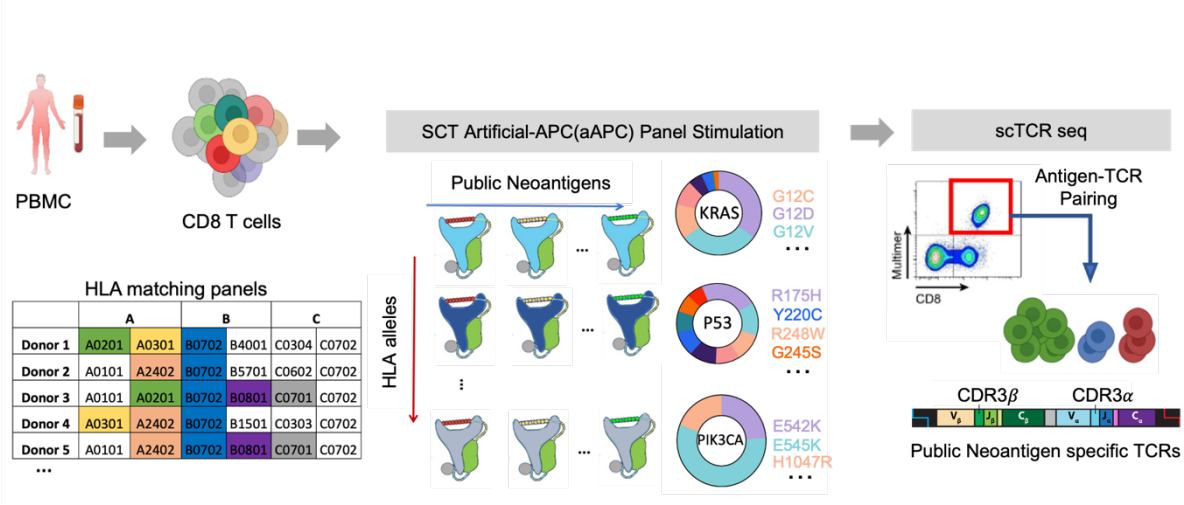


Figure 2.2 SCT-aAPC-based high-throughput TCR discovery platform for public neoantigens.

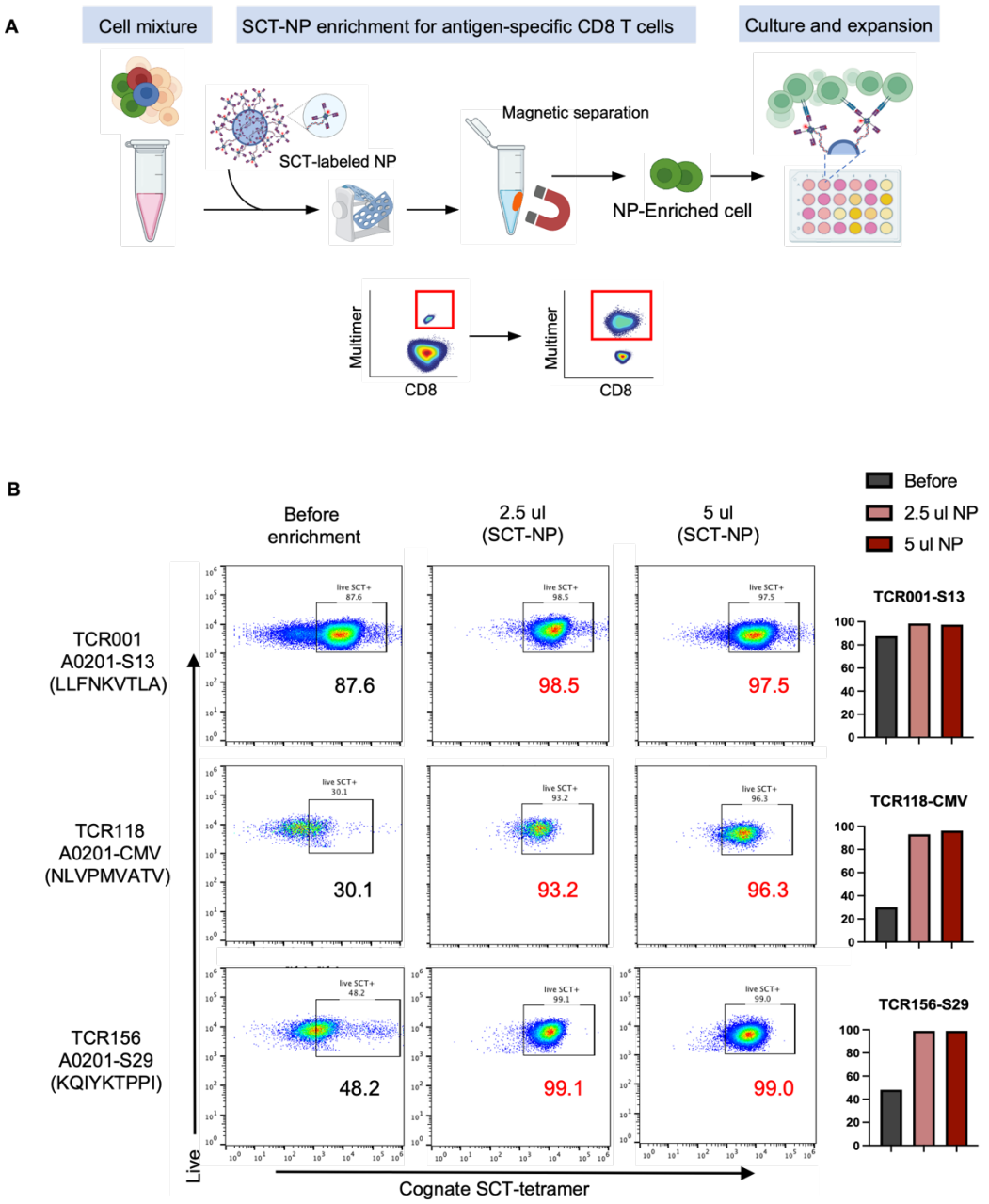


Figure 2.3 SCT-based nanoparticle for the effective isolation of antigen-specific T cells

A. Schematic illustration. TCR-transduced Jurkat or primary T cells with impurities were incubated with cognate SCT-NP to allow effective capture of antigen-specific CD8 T cells. The NP-bound cells are then isolated with a magnet for cell culture and expansion.

B. Tetramer staining for TCR-transduced Jurkat cells before and after 3 days of SCT-NP enrichment. Each row represents a unique TCR-antigen pair. Y-axis: live staining. X-axis: cognate SCT tetramer (S13, CMV, S29, respectively). SCT-tetramer positive frequencies are labeled in each panel, and on the bar plot.

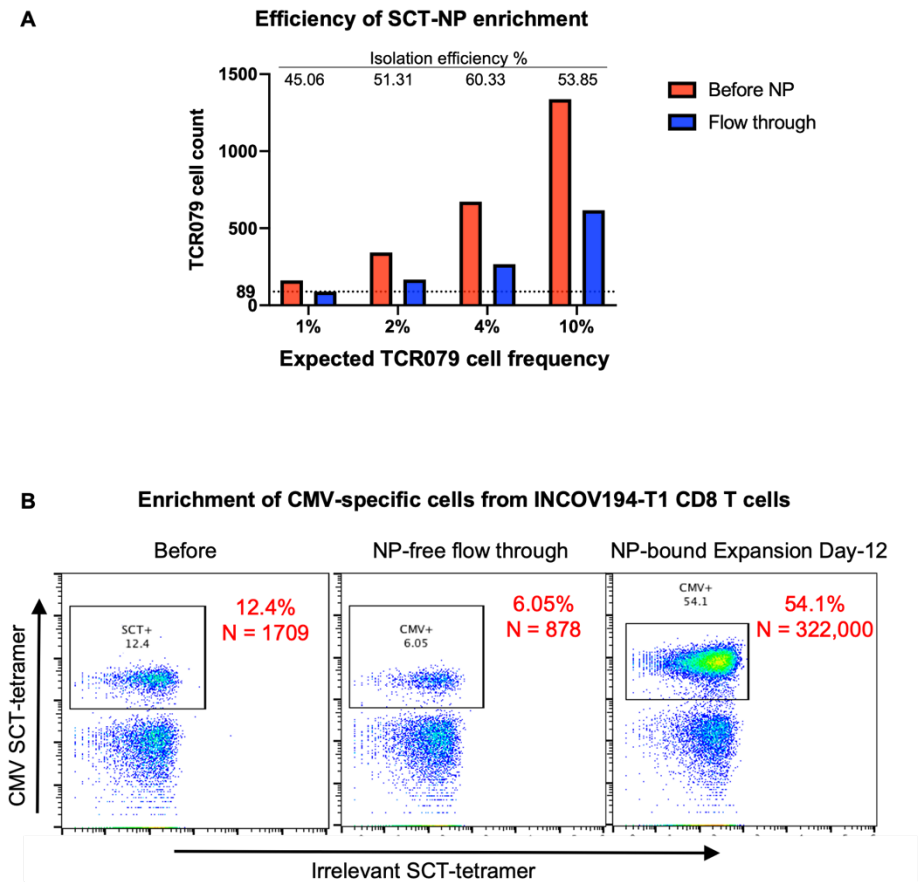


Figure 2.S3 Individual SCT-aAPC for effective antigen-specific T cell enrichment and expansion

A. Tetramer staining for TCR-transduced Jurkat cells before and after SCT-NP enrichment.

B. CMV-specific cell frequencies in CD8 T cells from the donor PBMCs. CMV-SCT-tetramer positive frequencies and cell counts are labeled in each panel.

List of HLA-A*02:01 SCTs for aAPC (N = 131)

Virus (3)	1	NLVPMVATV	CMV-pp65
	2	LLFNKVTLA	SARS-CoV-2 S
	3	RLITGRLQSL	SARS-CoV-2 S
TAAs (3)	4	ELAGIGILTV	MART-1
	5	RMFPNAPYL	WT1
	6	SLLMWITQV	NYESO
Self-antigen (5)	7	VLFGLGFAI	T1D-Islet related
	8	MVWESGCTV	T1D-IGRP
	9	HLVEALYLV	T1D-Islet related
	10	ILHDDEVTV	RPLP1
	11	YLAPENGYL	U1-snRNP
KRAS G12 (4)	12	KLVVVGAGGV	G12
	13	KLVVVGADGV	G12D
	14	KLVVVGAVGV	G12V
	15	KLVVVGARGV	G12R

TP53 (26)			
	16	HMTEVVRHC	R175H
	17	PCEPPEVGS	Y220C
	18	VVPCEPPEV	Y220C
	19	GSMNRRPIL	G245S
	20	NWRPILTII	R248W
	21	MNWRPILTI	R248W
	22	GGMNWRPIL	R248W
	23	NSFEVHVCA	R273H
	24	ACPGRDWRT	R282W
	25	HMTEVVRHCP	R175H
	26	PCEPPEVGSD	Y220C
	27	VVPCEPPEVG	Y220C
	28	VVVPCEPPEV	Y220C
	29	GSMNRRPILT	G245S
	30	CMGGMNWRPI	R248W
	31	HMTEVVRHCPH	R175H
	32	PCEPPEVGSDC	Y220C
	33	VVPCEPPEVGS	Y220C
	34	VVVPCEPPEVG	Y220C
	35	SVVVPCEPPEV	Y220C
	36	GSMNRRPILTI	G245S
	37	YMCNSSCMGSM	G245S
	38	YMCNSSCMGDM	G245D
	39	GGMNWRPILTI	R248W
	40	CMGGMNWRPIL	R248W
	41	LLGRNSFEVHV	R273H

PIK3CA (90)			
	42	KQEKDFLWSHR	E545K
	43	TKQEKDFLWSH	E545K
	44	ITKQEKDFLWS	E545K
	45	EITKQEKDFLW	E545K
	46	SEITKQEKDFL	E545K
	47	LSEITKQEKDF	E545K
	48	PLSEITKQEKD	E545K
	49	DPLSEITKQEK	E545K
	50	RDPLSEITKQE	E545K
	51	TRDPLSEITKQ	E545K
	52	STRDPLSEITK	E545K
	53	KQEKDFLWSH	E545K
	54	TKQEKDFLWS	E545K
	55	ITKQEKDFLW	E545K
	56	EITKQEKDFL	E545K
	57	SEITKQEKDF	E545K
	58	LSEITKQEKD	E545K
	59	PLSEITKQEK	E545K
	60	DPLSEITKQE	E545K
	61	RDPLSEITKQ	E545K
	62	TRDPLSEITK	E545K
	63	KQEKDFLWS	E545K
	64	TKQEKDFLW	E545K
	65	ITKQEKDFL	E545K
	66	EITKQEKDF	E545K
	67	SEITKQEKD	E545K
	68	LSEITKQEK	E545K
	69	PLSEITKQE	E545K
	70	DPLSEITKQ	E545K
	71	RDPLSEITK	E545K

	72	KITEQEKDFLW	E542K
	73	SKITEQEKDFL	E542K
	74	LSKITEQEKDF	E542K
	75	PLSKITEQEKD	E542K
	76	DPLSKITEQEK	E542K
	77	RDPLSKITEQE	E542K
	78	TRDPLSKITEQ	E542K
	79	STRDPLSKITE	E542K
	80	ISTRDPLSKIT	E542K
	81	AISTRDPLSKI	E542K
	82	KAISTRDPLSK	E542K
	83	KITEQEKDFL	E542K
	84	SKITEQEKDF	E542K
	85	LSKITEQEKD	E542K
	86	PLSKITEQEK	E542K
	87	DPLSKITEQE	E542K
	88	RDPLSKITEQ	E542K
	89	TRDPLSKITE	E542K
	90	STRDPLSKIT	E542K
	91	ISTRDPLSKI	E542K
	92	AISTRDPLSK	E542K
	93	KITEQEKDF	E542K
	94	SKITEQEKD	E542K
	95	LSKITEQEK	E542K
	96	PLSKITEQE	E542K
	97	DPLSKITEQ	E542K
	98	RDPLSKITE	E542K
	99	TRDPLSKIT	E542K
	100	STRDPLSKI	E542K
	101	ISTRDPLSK	E542K

	102	RHGGWTTKMDW	H1047R
	103	ARHGGWTTKMD	H1047R
	104	DARHGGWTTKM	H1047R
	105	NDARHGGWTTK	H1047R
	106	MNDARHGGWTT	H1047R
	107	QMNDARHGGWT	H1047R
	108	KQMNDARHGGW	H1047R
	109	MKQMNDARHGG	H1047R
	110	FMKQMNDARHG	H1047R
	111	YFMKQMNDARH	H1047R
	112	EYFMKQMNDAR	H1047R
	113	RHGGWTTKMD	H1047R
	114	ARHGGWTTKM	H1047R
	115	DARHGGWTTK	H1047R
	116	NDARHGGWTT	H1047R
	117	MNDARHGGWT	H1047R
	118	QMNDARHGGW	H1047R
	119	KQMNDARHGG	H1047R
	120	MKQMNDARHG	H1047R
	121	FMKQMNDARH	H1047R
	122	YFMKQMNDAR	H1047R
	123	RHGGWTTKM	H1047R
	124	ARHGGWTTK	H1047R
	125	DARHGGWTT	H1047R
	126	NDARHGGWT	H1047R
	127	MNDARHGGW	H1047R
	128	QMNDARHGG	H1047R
	129	KQMNDARHG	H1047R
	130	MKQMNDARH	H1047R
	131	FMKQMNDAR	H1047R

Table 2.1 List of HLA-A*02:01 SCTs for aAPC (N = 131)

List of Non-A*02:01 SCTs for aAPC (N = 58)

	1	YPDKVFRSSV	SARS-CoV-2 S
	2	FPQSAPHGV	SARS-CoV-2 S
Virus	3	LPPAYTNSF	SARS-CoV-2 S
	4	TPRVTGGGAM	CMV-pp65
Self-antigen	5	GPVPGMPPM	U1-snRNP
KRAS	6	ARGVGKSAL	G12R
	7	PCEPPEVGS	Y220C
	8	VPCEPPEVG	Y220C
	9	VVPCEPPEV	Y220C
	10	VVVPCEPPE	Y220C
	11	SVVVPCEPP	Y220C
	12	HSVVVPCEP	Y220C
	13	RHSVVVPCE	Y220C
	14	PCEPPEVGSD	Y220C
B*07:02 (N = 30)	15	VPCEPPEVGS	Y220C
	16	VVPCEPPEVG	Y220C
	17	VVVPCEPPEV	Y220C
	18	SVVVPCEPPE	Y220C
TP53	19	HSVVVPCEPP	Y220C
	20	RHSVVVPCEP	Y220C
	21	FRHSVVVPCE	Y220C
	22	PCEPPEVGSDC	Y220C
	23	VPCEPPEVGSD	Y220C
	24	VVPCEPPEVGS	Y220C
	25	VVVPCEPPEVG	Y220C
	26	SVVVPCEPPEV	Y220C
	27	HSVVVPCEPPE	Y220C
	28	RHSVVVPCEPP	Y220C
	29	FRHSVVVPCEP	Y220C
	30	TFRHSVVVPCE	Y220C

A*03:01 (N = 4)		KRAS
	1	VVVGAVGVGKS	G12V
	2	VVVGARGVGKS	G12R
	3	VVVGAGGVGK	G12
	4	VVVGARGVGK	G12R

A*24:02 (N = 6)		KRAS
	1	VGVGKSALTI	G12V
	2	EYKLVVVGAV	G12V
	3	RGVGKSALTI	G12R
	4	EYKLVVVGAR	G12R
	5	AVGVGKSAL	G12V
	6	ARGVGKSAL	G12R

B*08:01 (N =16)		KRAS
	1	VGAGGVGKSAL	G12
	2	VGADGVGKSAL	G12D
	3	VGAVGVGKSAL	G12V
	4	VGARGVGKSAL	G12R
	5	GAGGVGKSAL	G12
	6	DGVGKSALTI	G12V
	7	GADGVGKSAL	G12D
	8	GAVGVGKSAL	G12V
	9	GARGVGKSAL	G12R
	10	AGGVGKSAL	G12
	11	DGVGKSALT	G12D
	12	ADGVGKSAL	G12D
	13	AVGVGKSAL	G12V
	14	YKLVVVGAV	G12V
	15	ARGVGKSAL	G12R
	16	GARGVGKSA	G12R

C*07:01 (N=2)		KRAS
	1	ARGVGKSALTI	G12R
	2	VGARGVGKSAL	G12R

Table 2.2 List of Non-A*02:01 SCTs for aAPC (N = 58)

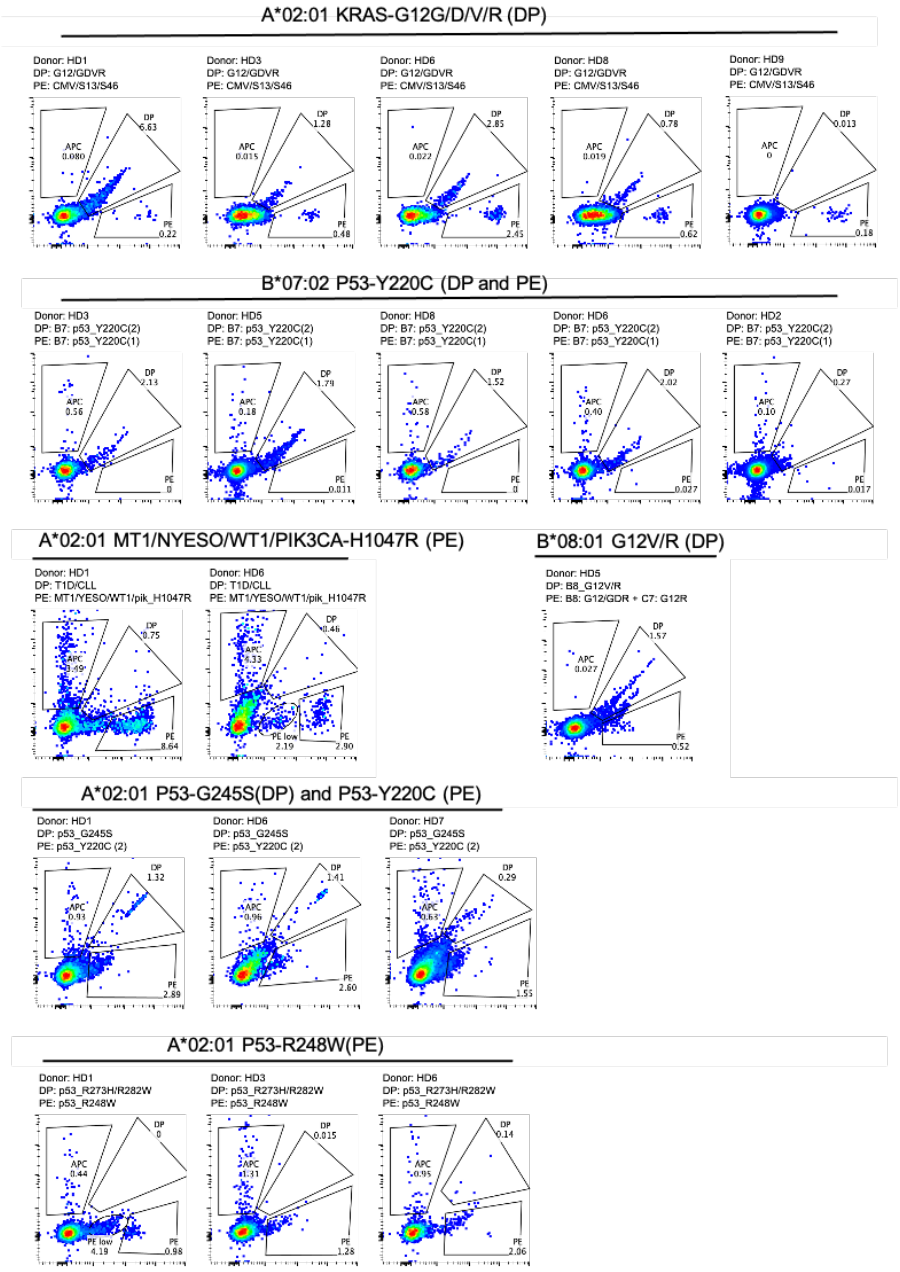


Figure 2.4 Example of SCT-aAPC stimulation-induced PubNeo+ CD8 T populations from multiple HDs.

X and Y axis represent signals from pools of PubNeo-SCT tetramers. Double-positive (DP) and PE-positive populations are PubNeo+ cells with tetramer specificities listed on the top of each panel.

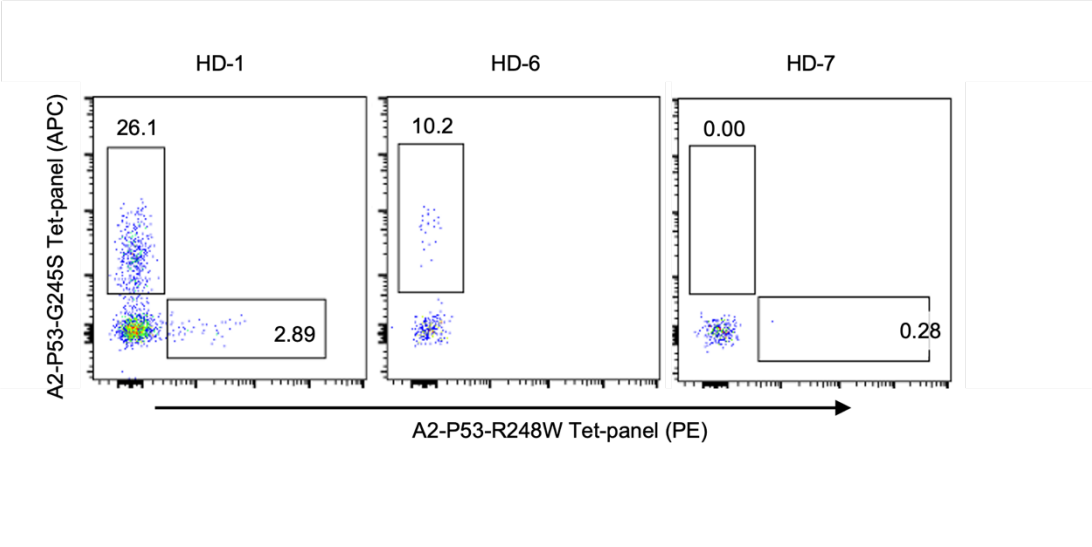
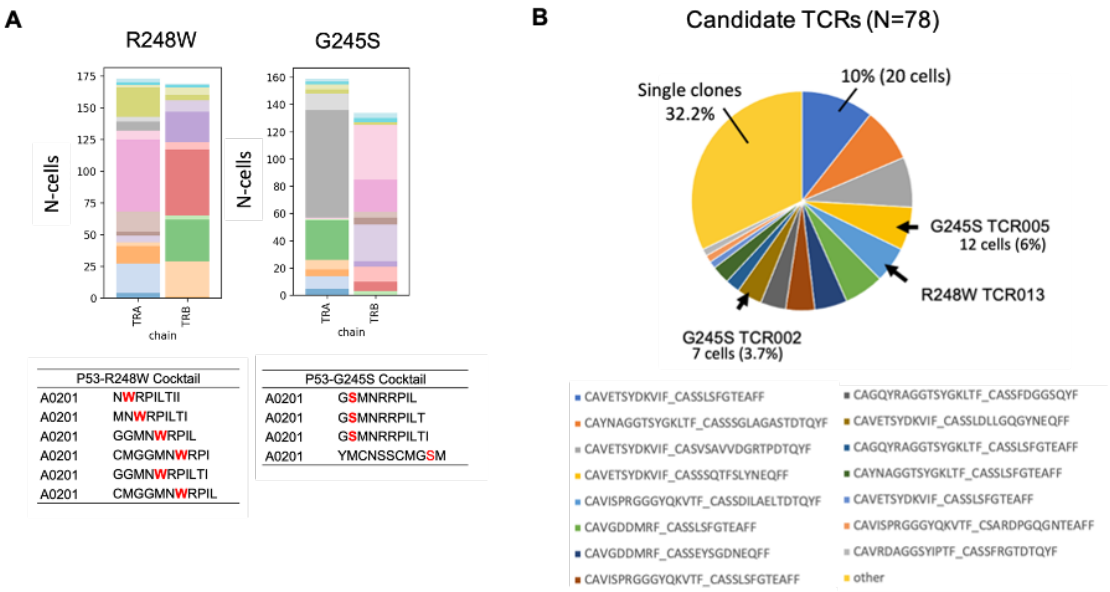


Figure 2.5 Stimulation and enrichment of PubNeo+ cells in HDs.

Examples shown here: CD8 T cells isolated from HD-1, 6, 7 staining with A*02:01 P53-G245S tetramer panel (Y-axis, APC) and A*02:01 P53-R248W tetramer panel (X-axis, PE) after 15 days of SCT-aAPCs stimulation and enrichment.



C

TCR_ID	Potential antigen	TCR
1	p53_G245S	TRAV19_CALSDPPYGANSKLTF_TRAJ56_TRBV20-1_CSARDPGQGNTEAFF_TRBJ1-1_TRBC1
2	p53_G245S	TRAV22_CAVETSYDKVIF_TRAJ50_TRBV11-3_CASSLDLLGQGYNEQFF_TRBJ2-1_TRBC2
3	p53_G245S	TRAV19_CALNPFSSNTGKLIF_TRAJ37_TRBV5-3_CARSLDIGPPRSQSAPAF_TRBJ1-5_TRBC1
4	p53_G245S	TRAV22_CAVETSYDKVIF_TRAJ50_TRBV19_CASSRPGGRGDQPQHF_TRBJ1-5_TRBC1
5	p53_G245S	TRAV22_CAVETSYDKVIF_TRAJ50_TRBV11-3_CASSSQTFSLYNEQFF_TRBJ2-1_TRBC2
6	p53_G245S	TRAV22_CAVESLGGSEKLVF_TRAJ57_TRBV20-1_CSASGGDSYEQYF_TRBJ2-7_TRBC2
7	p53_G245S	TRAV8-1_CAVMNDYKLSF_TRAJ20_TRBV4-3_CASSQEEGPYNEQFF_TRBJ2-1_TRBC2
8	p53_G245S	TRAV12-2_CAVSTNAGKSTF_TRAJ27_TRBV28_CASSFGQSYEQYF_TRBJ2-7_TRBC2
9	p53_G245S	TRAV38-2--DV8_CAYNDEKLTF_TRAJ48_TRBV2_CASSDWASRGDTEAFF_TRBJ1-1_TRBC1
10	p53_G245S	TRAV12-3_CAMGNAGNNRKLIW_TRAJ38_TRBV20-1_CSARDLVDVARFAFF_TRBJ1-1_TRBC1
11	p53_G245S	TRAV38-2--DV8_CAYLGGGADGLTF_TRAJ45_TRBV4-2_CASSQAAESYEQYF_TRBJ2-7_TRBC2
12	p53_R248W	TRAV12-2_CAVISPRGGGYQKVTF_TRAJ13_TRBV6-4_CASSARLGTGELFF_TRBJ2-2_TRBC2
13	p53_R248W	TRAV12-2_CAVISPRGGGYQKVTF_TRAJ13_TRBV9_CASSDILAELTDTQYF_TRBJ2-3_TRBC2
14	p53_G245S	TRAV19_CALSDPPYGANSKLTF_TRAJ56_TRBV20-1-pcr-TCR1_CSARDPGQGNTEAFF_TRBJ1-1_TRBC1
15	p53_G245S	TRAV22_CAVETSYDKVIF_TRAJ50_TRBV11-3-pcr-TCR2_CASSLDLLGQGYNEQFF_TRBJ2-1_TRBC2
16	p53_G245S	TRAV22_CAVETSYDKVIF_TRAJ50_TRBV11-3-pcr-TCR5_CASSSQTFSLYNEQFF_TRBJ2-1_TRBC2
17	p53_G245S	TRAV12-2-pcr-TCR8_CAVSTNAGKSTF_TRAJ27_TRBV28_CASSFGQSYEQYF_TRBJ2-7_TRBC2
18	p53_G245S	TRAV38-2--DV8_CAYNDEKLTF_TRAJ48_TRBV2-pcr-TCR9_CASSDWASRGDTEAFF_TRBJ1-1_TRBC1
19	p53_R248W	TRAV12-2-pcr-TCR12_CAVISPRGGGYQKVTF_TRAJ13_TRBV6-4-pcr-TCR12_CASSARLGTGELFF_TRBJ2-2_TRBC2
20	p53_R248W	TRAV12-2_CAVISPRGGGYQKVTF_TRAJ13_TRBV9-pcr-TCR13_CASSDILAELTDTQYF_TRBJ2-3_TRBC2
21	G245S-R248W	TRAV22_CAVETSYDKVIF_TRAJ50_TRBV6-4_CASSLSFGTEAFF_TRBJ1-1_TRBC1
22	R248W	TRAV38-1_CAYNAGGTSYGKLTF_TRAJ52_TRBV13_CASSSGLAGASTDTQYF_TRBJ2-3_TRBC2
23	p53_G245S	TRAV22_CAVETSYDKVIF_TRAJ50_TRBV28_CASVSAVVDGRTPDTQYF_TRBJ2-3_TRBC2
24	R248W	TRAV12-2_CAVGDDMRF_TRAJ43_TRBV6-4_CASSLSFGTEAFF_TRBJ1-1_TRBC1
25	R248W	TRAV12-2_CAVGDDMRF_TRAJ43_TRBV9_CASSEYSGDNEQFF_TRBJ2-1_TRBC2
26	R248W	TRAV12-2_CAVISPRGGGYQKVTF_TRAJ13_TRBV6-4_CASSLSFGTEAFF_TRBJ1-1_TRBC1
27	p53_R248W	TRAV35_CAGQYRAGGTSYGKLTF_TRAJ52_TRBV28_CASSFDGGSQYF_TRBJ2-3_TRBC2
28	R248W	TRAV35_CAGQYRAGGTSYGKLTF_TRAJ52_TRBV6-4_CASSLSFGTEAFF_TRBJ1-1_TRBC1
29	R248W	TRAV38-1_CAYNAGGTSYGKLTF_TRAJ52_TRBV6-4_CASSLSFGTEAFF_TRBJ1-1_TRBC1
30	G245S-R248W	TRAV22_CAVETSYDKVIF_TRAJ50_TRBV6-4_CASSLSFGTEAFF_TRBJ1-1_TRBC1
31	R248W	TRAV12-2_CAVISPRGGGYQKVTF_TRAJ13_TRBV20-1_CSARDPGQGNTEAFF_TRBJ1-1_TRBC1
32	G245S	TRAV3_CAVRDAGGSYIPTF_TRAJ6_TRBV12-4_CASSFRGTDTQYF_TRBJ2-3_TRBC2

Figure 2.6 TCR repertoire analysis for selected A*02:01 P53 PubNeo+ cells in HDs.

A. Stacked bar plot of unique TCR α and β chains for A*02:01 P53 R248W and G245S tetramer positive cells after indexed single-cell sorting and TCR sequencing.

B. Cell distribution for candidate TCR clones. 300 sorted cells, 200 cells with productive a/b pairs, 76 Unique TCR clones.

C. List of selected TCRs for cloning and evaluation with their potential antigen based on the SCT panel for flow sorting.

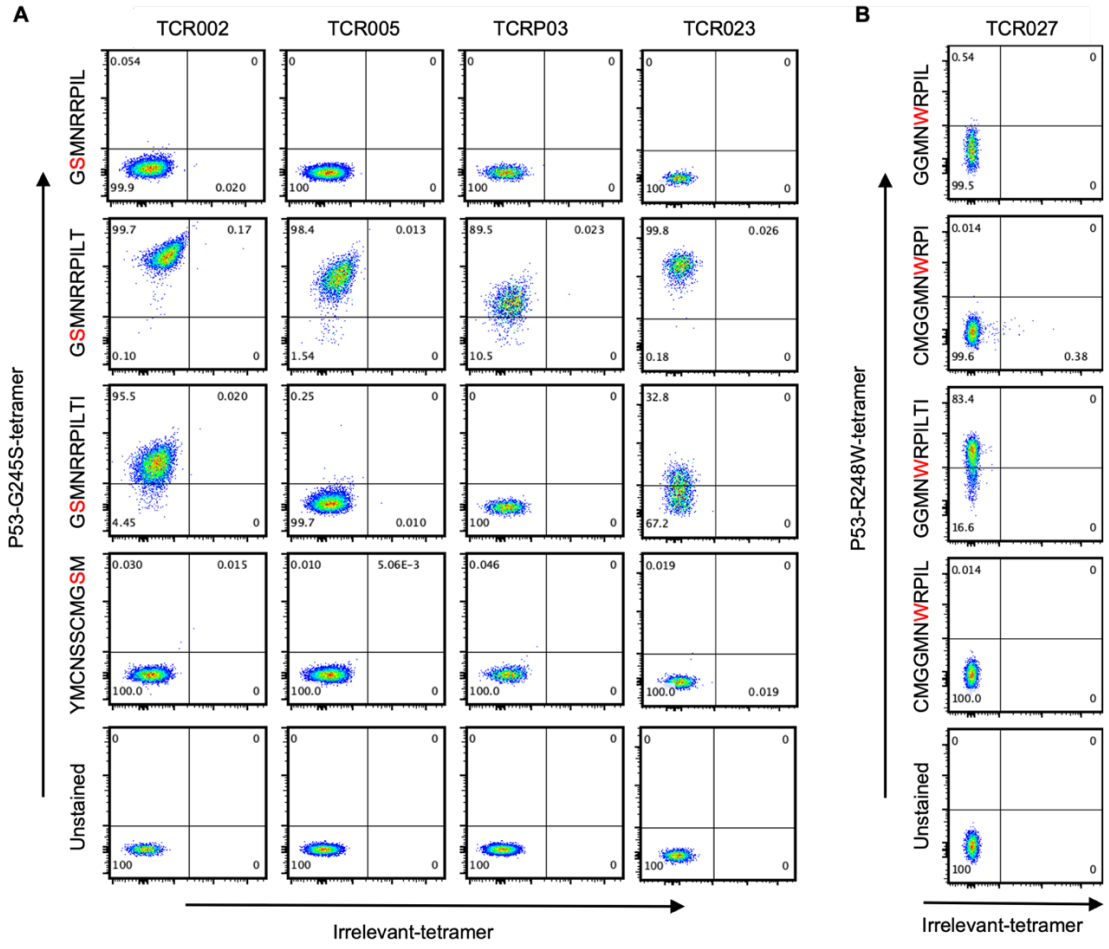


Figure 2.7 Individual tetramer staining of TCR-transduced CD8-negative Jurkat cells. Candidate TCR α/β chains were transduced into TCR-KO Jurkat cells (CD8-negative), and stained with individual tetramer. Only selected TCRs shown here. Cells are gated on live/single/CD3+TCR+ population.

A. Selected TCRs for A*02:01 P53-G245S panel. Each row represents one tetramer specificity from the A2- P53-G245S pool. Mutated amino acid 'S' is highlighted in red. Each column represents one candidate TCR. Y-axis: individual tetramers in P53-G245S panel. X-axis: irrelevant tetramer.

B. Selected TCRs for A*02:01 P53-R248 panel. Y-axis: individual tetramers P53- R248 panel, two additional non-binding tetramers are not shown. X-axis: irrelevant tetramer.

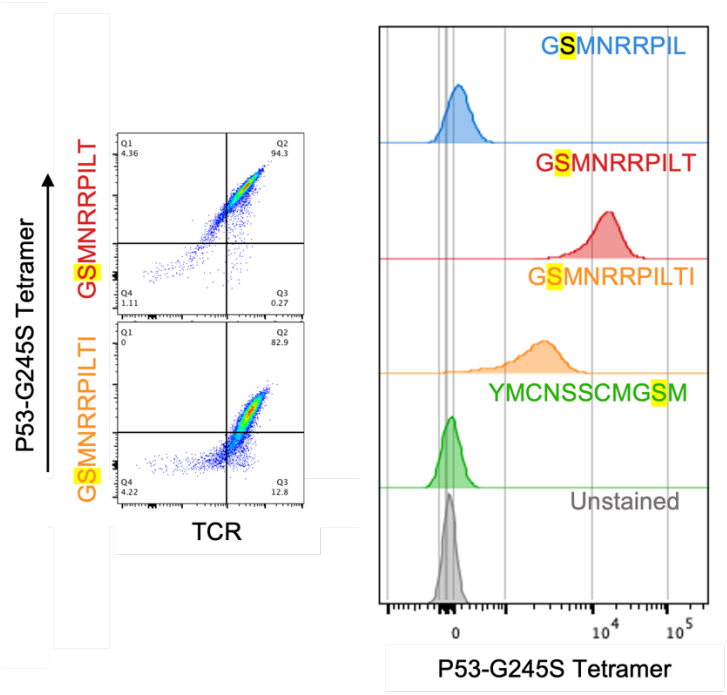


Figure 2.8 TCR-dependent tetramer binding.

Left: Flow cytometry staining of TCR002-transduced CD8-negative Jurkat T cells. Y-axis: A*02:01 P53-G245S tetramer, X-axis: anti-TCR, gated on live/single population.

Right: Overlayed tetramer signal for each SCT within the A*02:01 P53-G245S panel, gated on live/single/CD3+TCR+ population.

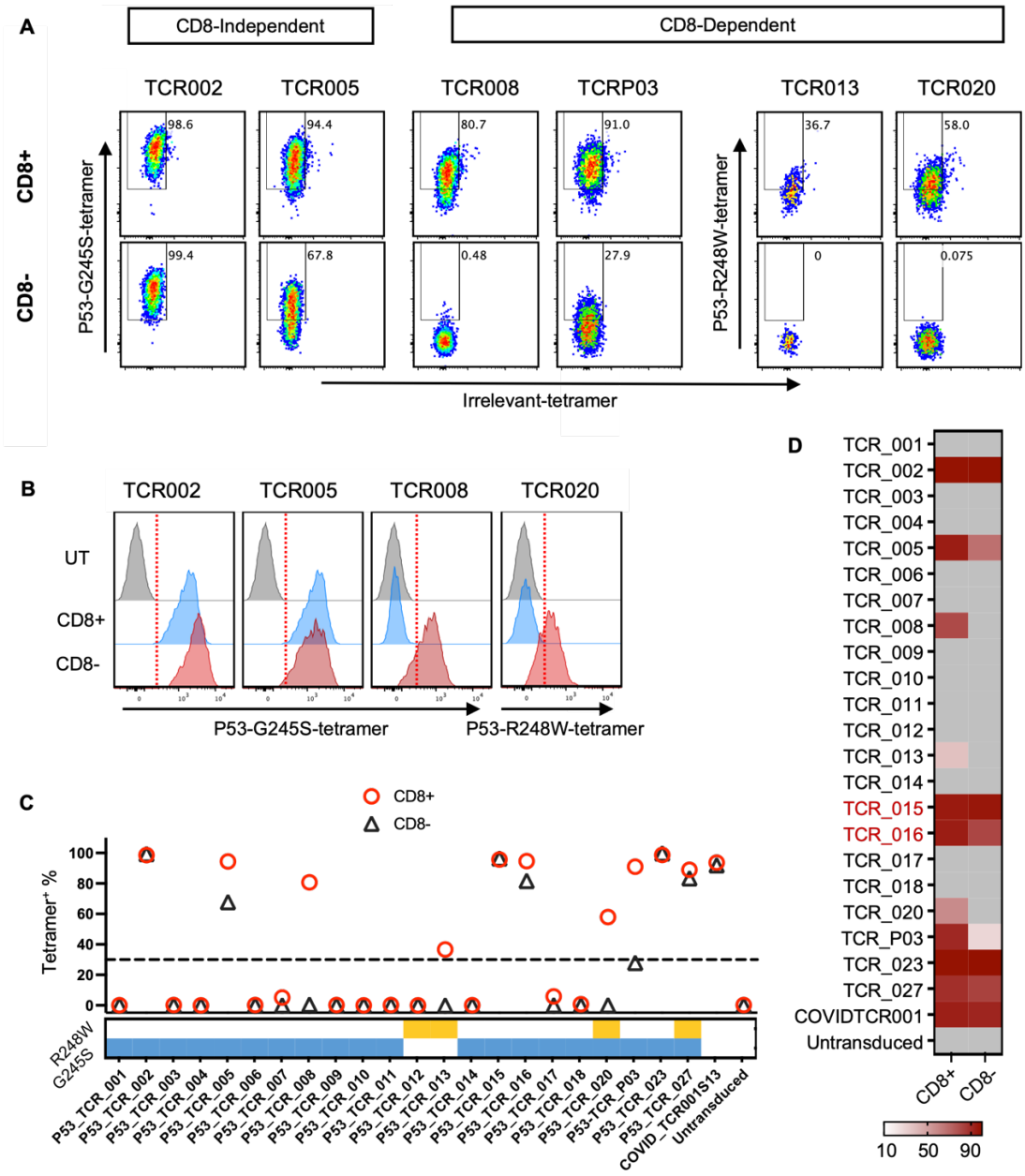


Figure 2.9 CD8-coreceptor dependency of candidate PubNeo+ TCRs.

A. Tetramer staining of TCR-transduced Nur77-GFP Jurkat T cells, gated on live/single cell population. TCR+CD8^{pos} and TCR+CD8^{neg} populations were compared for their tetramer binding intensity. Y-axis: A*02:01 P53-G245S or R248W tetramer. X-axis: Irrelevant tetramer

B. Tetramer staining intensity for selected TCRs, compared with untransduced (UT) cells.

C. Tetramer-positive cell frequencies for each TCR-transduced cell line, gated on CD8 expression level. Antigen specificity tested (G245S or R248W) are labeled at bottom.

D. Heatmap representation of C, showing tetramer-positive cell frequencies for each TCR-transduced cell line. Grey: <10%.

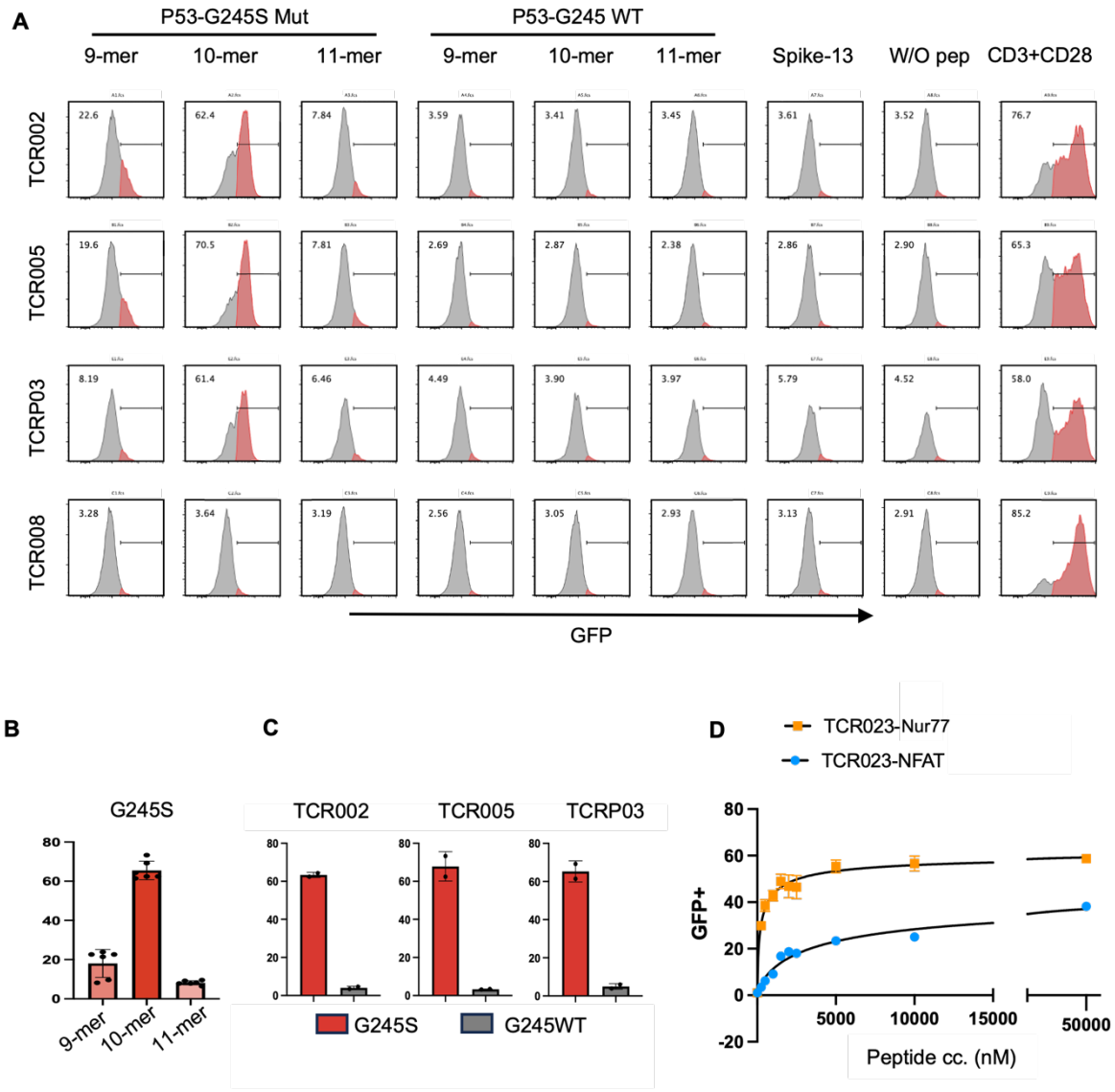


Figure 2.10 TCR activation of TCR-cloned Nur77-GFP-CD8$^+$ cells.

A. Histogram of GFP activation signal. Each TCR was transduced into Nur77-GFP-CD8$^+$ cell line, and incubated with A*02:01 T2 cells pulsed with peptide overnight. Positive control was prepared by incubating with CD3/CD28 microbeads. Gated on live/single/CD8 cell population.

B. GFP-positive percentage for G245S 9-mer, 10-mer, 11-mer peptides. Each dot represents a TCR that showed GFP activation.

C. GFP-positive percentage comparing G245S 10-mer, or G245WT 10-mer.

D. GFP-positive percentage for TCR024-transduced Nur77-GFP and NFAT-GFP cell line, after overnight incubation with T2 cells pulsed with titrated concentrations of G245S 10-mer peptide. Line fitted by Prism 10.

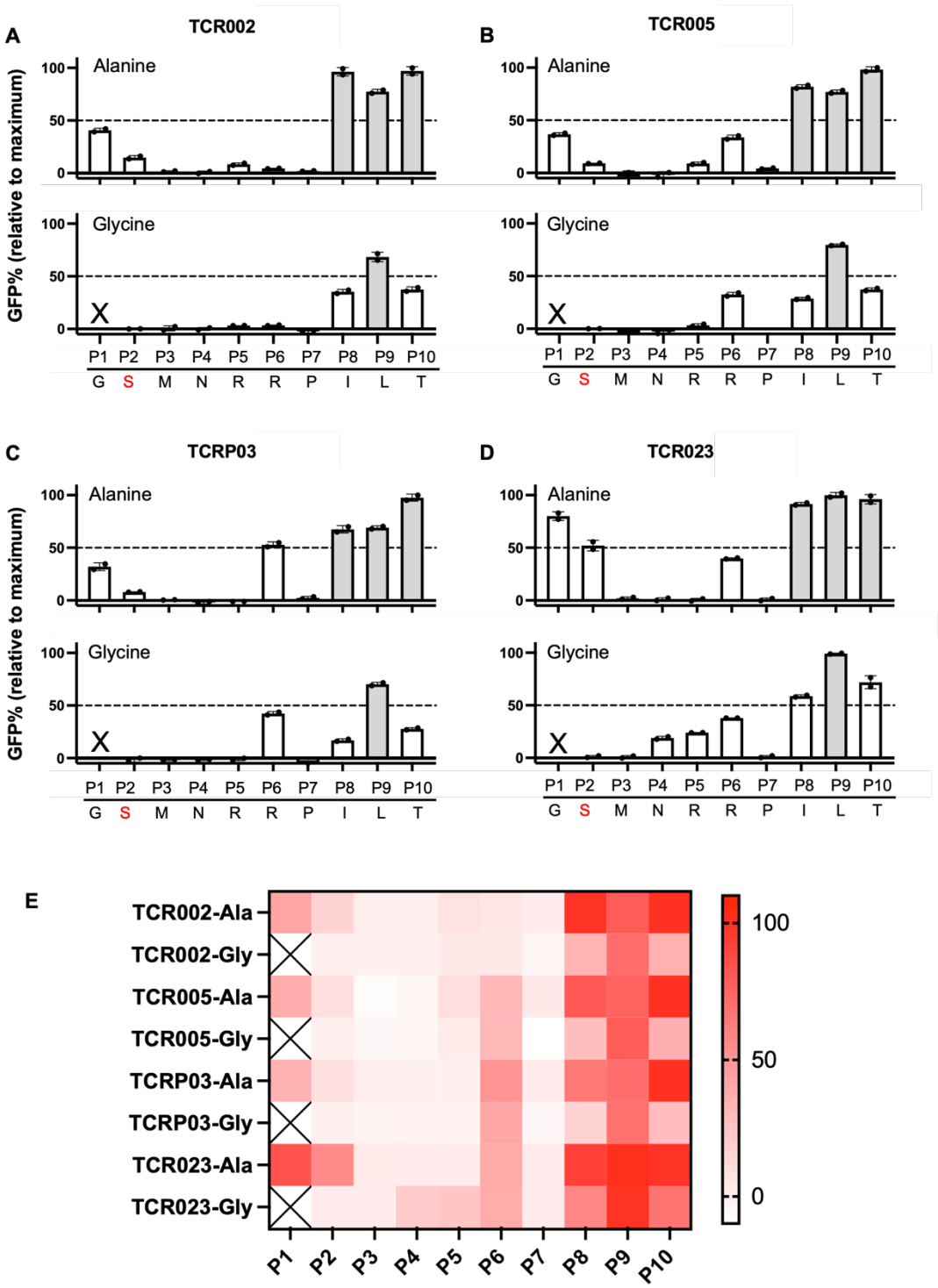


Figure 2.11 AA scan of TCR-transduced Nur-77-GFP-CD8$^+$ Jurkat T cells.

Gated on live/single/CD8 cell population. Cells were incubated with T2 cells pulsed with indicated peptides.

A-D. GFP activation strength to Ala (upper) or Gly (lower) substituted G245S 10-mer peptides. For each TCR, values are shown as the GFP-positive cell frequency relative to the maximum GFP signal for all AA scanning peptides. “X” denotes positions not amenable for substitution.

E. Heatmap representation of data in panel A-D, color-scale of GFP activation for AA scan.

2.7 REFERENCES

1. Kristensen, N. P. *et al.* Neoantigen-reactive CD8 T cells affect clinical outcome of adoptive cell therapy with tumor-infiltrating lymphocytes in melanoma. *J. Clin. Invest.* **132**, e150535.
2. Stem-like CD8 T cells mediate response of adoptive cell immunotherapy against human cancer. https://www.science.org/doi/10.1126/science.abb9847 doi:10.1126/science.abb9847.
3. Sheih, A. *et al.* Clonal kinetics and single-cell transcriptional profiling of CAR-T cells in patients undergoing CD19 CAR-T immunotherapy. *Nat. Commun.* **11**, 219 (2020).
4. Chandran, S. S. *et al.* Immunogenicity and therapeutic targeting of a public neoantigen derived from mutated PIK3CA. *Nat. Med.* **28**, 946–957 (2022).
5. Merz, V. *et al.* Targeting KRAS: The Elephant in the Room of Epithelial Cancers. *Front. Oncol.* **11**, (2021).
6. Waters, A. M. & Der, C. J. KRAS: The Critical Driver and Therapeutic Target for Pancreatic Cancer. *Cold Spring Harb. Perspect. Med.* **8**, a031435 (2018).
7. Albers, A. E. *et al.* Phenotype of p53 wild-type epitope-specific T cells in the circulation of patients with head and neck cancer. *Sci. Rep.* **8**, 10716 (2018).
8. Chasov, V. *et al.* Promising New Tools for Targeting p53 Mutant Cancers: Humoral and Cell-Based Immunotherapies. *Front. Immunol.* **12**, (2021).
9. Huang, J. Current developments of targeting the p53 signaling pathway for cancer treatment. *Pharmacol. Ther.* **220**, 107720 (2021).
10. Martínez-Sáez, O. *et al.* Frequency and spectrum of PIK3CA somatic mutations in breast cancer. *Breast Cancer Res.* **22**, 45 (2020).
11. Løes, I. M. *et al.* Impact of KRAS, BRAF, PIK3CA, TP53 status and intraindividual mutation heterogeneity on outcome after liver resection for colorectal cancer metastases. *Int. J. Cancer* **139**, 647–656 (2016).
12. Pearlman, A. H. *et al.* Targeting public neoantigens for cancer immunotherapy. *Nat. Cancer* **2**, 487–497 (2021).
13. Malekzadeh, P. *et al.* Neoantigen screening identifies broad TP53 mutant immunogenicity in patients with epithelial cancers. *J. Clin. Invest.* **129**, e123791.
14. Wu, D., Gowathaman, R., Pierce, B. G. & Mariuzza, R. A. T cell receptors employ diverse strategies to target a p53 cancer neoantigen. *J. Biol. Chem.* **298**, 101684 (2022).
15. Wu, D., Gallagher, D. T., Gowthaman, R., Pierce, B. G. & Mariuzza, R. A. Structural basis for oligoclonal T cell recognition of a shared p53 cancer neoantigen. *Nat. Commun.* **11**, 2908 (2020).
16. Chour, W. *et al.* Large libraries of single-chain trimer peptide-MHCs enable antigen-specific CD8+ T cell discovery and analysis. *Commun. Biol.* **6**, 1–13 (2023).
17. Puig-Saus, C. *et al.* Neoantigen-targeted CD8+ T cell responses with PD-1 blockade therapy. *Nature* **615**, 697–704 (2023).
18. Peng, S. *et al.* Sensitive Detection and Analysis of Neoantigen-Specific T Cell Populations from Tumors and Blood. *Cell Rep.* **28**, 2728-2738.e7 (2019).

Chapter 3. APMAT ANALYSIS REVEALS ASSOCIATIONS BETWEEN ANTIGEN-TCR SEQUENCES AND CD8 T CELL PHENOTYPE

3.1 ABSTRACT

Elucidating the relationships between a class I peptide antigen, a CD8 T cell receptor (TCR) specific to that antigen, and the T cell phenotype that emerges following antigen stimulation, has remained a mostly unsolved problem, largely due to the lack of large data sets that can be mined to resolve such relationships. Here, we describe Antigen-TCR Pairing and Multiomic Analysis of T-cells (APMAT), which is an integrated experimental-computational framework designed for the high-throughput capture and analysis of antigen specific CD8 T cells, with paired antigen, TCR sequence, and transcriptome information from the same single cells, from many patient samples in parallel. Starting with 951 putative antigens representing a comprehensive survey of the SARS-CoV-2 viral proteome, we utilized APMAT for the capture and single cell analysis of CD8 T cells from 63 HLA A*02:01 COVID-19 participants. We systematically show how distinct physicochemical features of the antigen-TCR pairs strongly associate with both T cell phenotype and T cell persistence. APMAT can be a general approach for analyzing antigen-specific CD8 T cells across a broad spectrum of biological and disease contexts.

3.2 INTRODUCTION

CD8 T cells play a major role in adaptive immunity against pathogens, exhibiting a functional diversity that includes, as major phenotypes, naïve, memory, effector memory, effector, and exhausted, with each of those phenotypes encompassing multiple sub-phenotypes[1,2]. Recent literature has suggested the existence of relationships between a given antigen-specific T cell

clonotype and the phenotypic trajectory that clonotype can take following initial antigen stimulation. For example, tissue-resident, antigen-specific T cell clonotypes in both tumor and chronic viral infection settings have been intimately associated with specific phenotype differentiation trajectories[3–5]. Other recent literature has indicated that, for at least certain immunogenic viral antigens, T cell clonotype is the dominant factor in determining T cell phenotype[6–8]. These similar results, from tissue and peripheral blood, and in very different disease settings, suggest that there may be 'rules' that underlie T cell clonotype-T cell phenotype relationships. Elucidating such relationships requires experimental methods for collecting a large, longitudinal data set in which the transcriptome-level phenotypes of antigen-specific T cells are co-measured across a large number and diversity of antigens, coupled with computational methods for elucidating what, if any, relationships between the T cell receptor (TCR) α/β genes, cognate peptide antigen – major histocompatibility complex (pMHC), and T cell phenotype exist.

Initial hints at what such a rule set might contain can be found in the literature. For example, the importance of hydrophobicity at certain TCR residues is known to associate with cytotoxicity and self-reactivity in CD4 T cells[6,9–11], and the hydrophobicity of certain residues is known to associate with the immunogenicity of antigens[12]. These observations suggest that the detailed biochemical nature of the TCR:pMHC interface can play important roles in determining T cell phenotype and perhaps the cell fate trajectory. This question is not just of fundamental importance to T cell immunology, but it is also highly relevant to the engineering of T cells for use as therapeutic agents for treatments of both cancers[13] and autoimmune diseases[14,15]. However, existing studies predominantly explore only one aspect of the Antigen-TCR-phenotype interplay: either resolving the antigen specificity of T cells from biological samples[16–19], or conducting T cell

receptor analysis on existing datasets lacking antigen specificity (or containing only limited number of antigens)[6,20,21].

We describe an integrated experimental-computational framework, APMAT, which measures and integrates the antigen-specific capture of T cells from many patient biospecimens in parallel with single-cell multi-omics profiling that captures paired antigen and TCR sequences. For antigen-specific T cell capture, we utilize large libraries of single-chain-trimers (SCT)[22,23]. We have recently reported on the feasibility of these libraries to capture and characterize both viral antigen-specific and (tumor-associated) neoantigen-specific T cell populations. Here we use SCTs to construct a 951-element pMHC library that represents a comprehensive survey of putative Class I antigens, presented by HLA A*02:01, from across the full SARS-CoV-2 genome. The DNA-barcoded SCT-pMHC library is used to identify and, through single cell (sc) RNA-seq analysis, concurrently characterize SARS-CoV-2 specific T cells from 63 HLA A*02:01 participants at the stages of acute and post-acute COVID-19. We identify several viral antigen-specific T cell populations observed in previous work by ourselves and others[24–27] demonstrating the feasibility of this framework on a whole viral proteome scale. Moreover, the size of the resultant data set allows for an in-depth exploration and reveals many novel biologic principles of how the physicochemical properties of the TCR-pMHC interface is associated with T cell phenotype and T cell persistence.

Here we captured the CD8 T cells against Sars-Cov-2 and cytomegalovirus (CMV) antigens from a large number (n=61) of HLA-matched (A*02.01) COVID-19 patients. Antigen-specific T cells were characterized by single cell RNA sequencing (scRNAseq) and de-orphaned with a paired TCR α/β chain and cognate peptide epitope. We first annotate how these viral-specific CD8 T cell phenotypes vary across parameters such as the viral genome, patient demographics, and

disease severity. We then characterize the dynamics of SARS-CoV-2 and CMV specific T cell clonotypes over three time points, from early in the infection (T1) to the stage of acute disease around a week later (T2), to the time of post-acute (T3) 2-3 months later. From this time series data, we identified and deeply characterized the TCR clonotypes that expand upon acute infection and recovery, and those that contract. We find that expanding versus contracting viral-specific T cell clonotypes can be differentiated by memory and cytotoxic signatures, which is expected. However, we also find that exhaustion, regulatory and Natural Killer (NK)-like signatures play interesting roles in separating these two populations.

3.3 RESULTS

3.3.1 *APMAT enables integrated multi-modal analysis of antigen-specific CD8 T cells*

We utilized APMAT to identify SARs-CoV-2 antigen-specific T cell responses to COVID-19 infection in a previously described longitudinal cohort of 209 COVID-19 participants[25,26]. From this cohort, we selected 63 HLA-matched participants representing a range of COVID-19 disease severities, with longitudinal reference datasets available at acute (diagnosis and approximately 1-week post-diagnosis), and post-acute (2-3 months post-diagnosis) timepoints (**Supplementary Table 1**)[25,26].

For antigen-specific CD8 T cell capture, we utilized DNA-barcoded, pMHC-like SCT multimers (SCT-dextramers). For library construction, we first utilized NetMHCpan[28] to analyze the full SARS-CoV-2 viral genome (original strain) to resolve a list of 951 putative 9 - 10 mer antigens (**Supplementary Table 2**) for HLA-A*02:01. This peptide list was converted into PCR-optimized DNA primers and used to build the plasmid library. Plasmids were then transfected into Expi293 cells over four days to induce secretion of the SCT protein product. 558 of the 951 putative antigens yielded usable SCT product (**Figure 3.1** top panel). Each SCT was then purified,

biotinylated, and assembled into a fluorophore-labeled and DNA-barcoded dextramer, as previously reported[23]. To minimize batch effects and enable integrated analysis, PBMCs from all 63 participants, collected at the stage of acute disease, were pooled and profiled in parallel within one experiment (**Figure 3.1** middle left). SCT-dextramer-positive CD8 T cells were sorted for scRNAseq to profile gene expression, TCRα/β genes, and antigen specificity (**Figure 3.1** middle right). For each cell captured, antigen-specificity was identified through the dominant pool hashtag (**Supplemental Figure 3.1A**) and dextramer (**Supplemental Figure 3.1B**). Each cell was associated with a specific patient through comparison of *de novo* computed single nucleotide polymorphisms (SNPs) from scRNA-seq with germline whole genome sequence (WGS) profiles of all patients (**Supplemental Figure 3.1C**). Sex gene validation further confirmed patient assignments (**Supplemental Figure 3.1D**). With such a multi-modular dataset, the characteristics of the peptide antigens and the TCR CDR3 sequences can be integrated with T cell phenotype, clinical data derived from electronic health record (EHR), and longitudinal multi-omics datasets (**Figure 3.1** bottom).

3.3.2 *APMAT enables high-throughput representation of whole SARS-CoV-2 genome*

We graphed the distribution of putative antigens against the SARS-CoV-2 genome, and highlighted the 558 (58.7%) constructs that were expressed and used for the T cell capture experiment, and then the 102 (18.3%) that captured at least 1 cell (**Figure 3.2A**, **Supplemental Figure 3.2A**). While APMAT identified CD8 T cells against antigens from across the SARS-CoV-2 genome, antigens from the Spike protein (S) exhibited the highest rate (26.6%) of cell capture (SCTs that captured at least 1 cell) (**Figure 3.2B, Supplementary Table 2**).

We next investigated the antigen sequences associated with SCT expression and cell capturing. In **Figure 3.2C** (left) we provide a sequence logo representation of the 9-mer SCTs that were or

were not expressed, and those expressed as SCTs that did or did not capture cells (**Figure 3.2C** right). We observed an enrichment of hydrophobic residues in non-anchor residues for non-expressed SCT constructs (**Figure 3.2C** left bottom). For SCTs that captured cells, polar and charged residues such as Threonine (T) and Arginine (R) were enriched in non-anchor positions (**Figure 3.2C** right top). We next generated a matrix for each peptide containing NetMHCpan prediction, amino acid identity and numeric values of physicochemical properties (such as hydrophobicity, polarity, etc) for peptide anchor residues ("Anchor") and those residues exposed to the TCR ("Exposed") (**Figure 3.2D, Supplementary Table 3**). As expected, expressed SCTs showed better NetMHCpan prediction (lower prediction rank and lower predicted binding affinity) (**Figure 3.2E** left, **Supplemental Figure 3.2B**). Notably SCT expression yield did not correlate with predicted affinity (**Supplemental Figure 3.2C**). However, the physicochemical properties analysis allowed for a quantitative validation of the hydrophobic trends of non-expressed SCTs (**Figure 3.2E** right), and the polar/charged residue trends of cell-capturing SCTs (**Figure 3.2F** right).

Overall, APMAT enabled a direct and comprehensive analysis of putative epitopes that may be useful for refining prediction algorithms. The data further suggests that an analysis of the physicochemical properties of the antigens (and possibly their cognate TCRs) may prove useful for interpreting the large multimodal data set generated through APMAT.

3.3.3 *Three peptide groups distinguished by sequence physicochemical properties*

We probed for potential relationships between the physicochemical properties of putative epitopes and the antigenicity of those epitopes. We encoded the 951 peptides by residue-level descriptors of their physicochemical properties (**Figure 3.3A**, Methods). Unsupervised clustering based on peptide amino acid identity and properties resulted in a two-dimensional peptide Uniform

Manifold Approximation and Projection (Pep-UMAP) (**Figure 3.3B, Supplementary Table 4**). The upper right wing of this UMAP exhibits higher hydrophobicity and bulkiness, and is dominated by non-expressed SCTs. Conversely, the lower left wing displays greater polarity, and is enriched for expressed SCTs (**Figure 3.3C, Supplemental Figure 3.3A**). However, SCTs that successfully captured T cells are uniformly distributed across the UMAP, indicating that additional factors beyond hydrophobicity modulate antigenicity.

We next utilized unsupervised clustering to resolve whether combinations of physicochemical properties could be used to further classify the peptide antigens. Such analysis clearly distinguished three peptide groups (Pep-Groups), PG1-3 (**Figure 3.3D**). Specifically, PG1 exhibited higher hydrophobicity yet lower charge and polarity (Figure 3.3d left top). PG2 and PG3 both showed higher polarity, but differ in anchor and secondary anchor (Position 6) properties (**Figure 3.3D** left bottom). When analyzing only the cell-capturing SCTs, the hydrophobicity of PG1 becomes even more prominent while PG2-3 distinctions diminish (**Figure 3.3D** right). Accordingly, Pep-Groups occupied different regions on the Pep-UMAP (**Figure 3.3E, Supplemental Figure 3.3B**). Comparisons of PG1-3 for hydrophobicity, SCT expression rate, and cell-capture rate (**Figure 3.3F**) reveal the most hydrophobic group (PG1) is the most challenging to express, but also exhibits the highest rate for cell capture. Thus, we categorized all putative antigen peptides into unique Pep-Groups for downstream analysis.

3.3.4 *Pep-groups associate with different T cell phenotypes*

We next investigated whether the PG1-3 peptide groups associate with distinct CD8 T cell phenotypes during the acute COVID-19 response. The SARS-CoV-2 SCT-dextramer-positive CD8 T cells were filtered to only include those with an assigned patient ID, antigen specificity, and paired TCRα/β sequences. These cells were then projected onto a gene expression UMAP

(GEX-UMAP) based on the scRNAseq data (**Figure 3.4A**). Unbiased clustering and differential gene expression analysis defined canonical CD8 T cell phenotypes including naïve, central memory (CM), effector memory (EM), hybrid, and cytotoxic phenotypes (**Figure 3.4B**). For instance, CCR7, LEF1, TCF7, and SELL were up-regulated in naïve and central memory (CM) cells, while memory markers such as IL7R, GZMK were elevated in memory phenotypes (**Figure 3.4B**, **Supplemental Figure 3.4A-B**). As expected, cytotoxic and EM phenotypes dominate the SARS-Cov-2-specific CD8 T cell response during acute disease[26,29,30], with elevated effector markers such as GZMB, GZMA and PRF1. In addition to gene expression, in **Figure 3.4C** we projected the polarity of exposed residues on the antigen recognized by the individual T cells (left), the distribution of T cells captured by a given antigen across multiple participants (middle), and all captured SARS-CoV-2 specific T cells for a given participant.

The projection of antigen polarity on the GEX-UMAP (**Figure 3.4C)** showed a strong skewing towards cytotoxic CD8 T cell phenotypes, suggesting that antigens with particular physicochemical properties may associate with specific T cell phenotypes. We explored this concept by projecting each cell's antigen specificity (Pep-Groups) onto the GEX-UMAP (**Figure 3.4D** top**).** PG1-specific cells were dominated by naïve-like phenotype and exhibited less clonal expansion relative to the other two groups (**Supplemental Figure 3.4C**). PG2-specific cells were enriched with EM and CM phenotypes, and PG3-specific cells were mainly cytotoxic (**Figure 3.4D** bottom**)**. Furthermore, differential expressed genes (DEGs) in PG3-captured cells have enriched pathway signatures related with immune synapse formation, PD-1 signaling, CD28 co-stimulation, which further highlighted the effector state of PG3-specific cells relative to PG2 (**Supplementary Table 5-6)**. This analysis suggests strong associations exist between the

physicochemical properties of the peptide antigens, and the phenotypic characteristics of the T cells specific to those antigens.

3.3.5 *TCR hydrophobicity is an important factor for effector function*

Intrigued by the link between antigenic peptides and T cell phenotypes, we investigated whether a similar connection exists for the physicochemical features of each antigen-associated TCR-CDR3β by overlaying those features on the GEX-UMAP (**Figure 3.5A, Supplemental Figure 3.5a**). Specifically, we categorized CDR3β residues into V, J, and CDR3βmer (central) regions: The V and J regions comprise the highly conserved n- and c-terminal motifs respectively; while the central CDR3βmer region, which primarily contacts the antigen, is the most diverse in length and amino acid usage. As expected, cytotoxic cells showed maximal clonal expansion followed by effector memory (EM) cells (**Supplemental Figure 3.5B**). We compared the CDR3β features of effector cells (cytotoxic and EM) relative to other cell types (**Supplementary Table 7-8)**. TCR feature differences were not significant for the V and J regions as expected (**Figure 3.5C** middle). However, for the CDR3βmer region, a preference of sequence features was observed. Effector T cells were marked by higher CDR3βmer hydrophobicity and bulkiness, lower polarity and charge, and shorter length (**Figure 3.5C** top and bottom). Note that CDR3β hydrophobicity was defined both by the percentage of hydrophobic residues, and by the average hydrophobicity across central residues (See method). We validated the above trends by plotting how the selected CDR3β physicochemical properties varied across all cell phenotypes. The CDR3β sequences displayed the trend of increased hydrophobicity and decrease in charge and length from naïve to EM and cytotoxic phenotypes (**Figure 3.5D**).

The hydrophobicity of the CDR3β exhibited the strongest significant association with T cell phenotype. This prompted us to define a binary classifier (HPhobic-High and HPhobic-low) based

on the percentage of hydrophobic residues in CDR3βmer (cutoff = 25%) (**Figure 3.5E** top, method). Cells expressing HPhobic-High TCRs were more clonally expanded than HPhobic-Low ones (**Figure 3.5E** bottom). We first validated that TCR-Groups still preserve the physicochemical features observed earlier: Indeed, HPhobic-High TCRs exhibited higher hydrophobicity, shorter CDR3β length, and lower charge (**Supplemental Figure 3.5C**). Density mapping validated that HPhobic-High TCRs were more prevalent in cytotoxic cells than in memory and naïve subsets (**Figure 3.5F**), with elevated exhaustion markers such as LAG3 and TIGIT (**Supplemental Figure 3.5D**). Gene set enrichment analysis linked HPhobic-High clonotypes to TCR activation (e.g. CD3 and TCR Zeta-chain phosphorylation), inflammation, and apoptosis pathways (**Figure 3.5G, Supplementary Table 9-10**). Overall, we demonstrated that, the physicochemical properties of both the peptide antigen and the TCR CDR3β exhibit strong associations with T cell phenotype for SARS-CoV-2 specific CD8 T cells.

3.3.6 *Integrated analysis of both peptide and TCR physicochemical features orchestrate phenotypes of SARS-CoV-2-specific CD8 T cells in acute disease*

To elucidate how interplay between antigen and TCR interactions influence T cell function, we systematically linked peptide and TCR features for the subset of SARS-CoV-2 CD8 T cells with fully paired antigen-TCR information (See Method). This encompassed 87 unique antigenic peptides (SCT-pMHCs) and 440 paired TCR clones (**Figure 3.6A**). Notably, distinct TCR Groups (HPhobic-Low and HPhobic-High) were discerned within each Peptide Group, prompting a refined categorization into PG-TCR Groups based on combined antigen peptide and TCR features (**Figure 3.6B**). For example, PG3-High denotes cells captured by PG3 peptides with HPhobic-High TCRs.

We then further evaluated the combinatorial effect of peptide and TCR features for each PG-TCR group. We depicted key physicochemical properties identified earlier on radar plots, revealing a significant shift in overall characteristics between PG-TCR groups (**Figure 3.6C, Supplemental Figure 3.6A**). For example, PG3:High exhibits high peptide charge and low hydrophobicity on exposed residues, combined with high CDR3βmer hydrophobicity and bulkiness (**Figure 3.6C** left). In contrast, PG2:Low (**Figure 3.6C** middle) emphasizes CDR3βmer length and charge, while PG1:Low (**Figure 3.6C** right) emphasizes distinct characteristics compared to PG3:High.

Building on our findings, we hypothesized combinations of PG-TCR groups associate with T cell phenotypes. PG3:High cells displayed the strongest cytotoxicity – marked by the highest percentage of effector cells and elevated cytotoxic cytokines (e.g. GZMB, PRF1, GZMH) (**Figure 3.6D, Supplemental Figure 3.6B-C**). PG1:Low describe cells with TCR:pMHC interfacial properties that are opposite that of PG3:High, and those cells similarly exhibit phenotypes that contrast with PG3:High – marked by highest frequency of naïve cells and elevated naïve-associated markers (e.g. CCR7, LEF1, TCF7). Notably, PG1 consistently exhibits a Naïve-like phenotype, regardless of TCR-group. With intermediate peptide-TCR properties, PG2:Low and PG3:Low represent transitional phenotypes. We performed further analysis by directly linking the key physicochemical properties to phenotypes independent of the PG-TCR groups. This heatmap highlights the balance between the properties of the antigen and the TCR CDR3b sequence (**Figure 3.6E**), and illustrates the near orthogonal relationships between naïve/CM cells relative to cytotoxic cells. Hence, by integrating peptide and TCR sequence features, APMAT reveals fundamental rules in the peptide-TCR-phenotype synergy.

3.3.7 *Distinct peptide-TCR groups associate with distinct longitudinal fates of SARS-CoV-2-Specific CD8 T cells*

T cell activation strength through pMHC-TCR interaction can influence cell fate over time[31–33]. Certain T cell phenotypes exhibit long-term *in vivo* persistence following acute illness, while highly cytotoxic phenotypes can undergo activation-induced cell death (AICD) and contract after clearance of pathogen[34–36]. We hypothesized that distinct Pep-TCR groups may associate with T cell fate decisions across longitudinal disease trajectories as well. Leveraging the longitudinal CD8 T cell atlas (scRNA-seq and scCITE-seq) generated from the same COVID-19 cohort[25], we tracked SARS-CoV-2-specific CD8 T cells matched on patient ID and TCR sequences from acute to post-acute timepoints **(Figure 3.7A, Supplementary Table 11**). This analysis further revealed longitudinal differences between PG-TCR groups. As expected, the overall percentage of SARS-CoV-2 specific CD8 T cells identified from the reference dataset decreased from 3% to 1.3% at the post-acute timepoint. Specifically, PG3:High cells were short-lived, and showed the greatest decrease in abundance at the post-acute timepoint. By contrast, PG1 cells (including PG1:High and PG1:Low) were the most persistent, showing an increased abundance at the later timepoint (**Figure 3.7B**, **Supplemental Figure 3.6d)**.

In addition to abundance changes over time, we examined gene expression changes for SARS-CoV-2 specific CD8 T cells. Combining all PG-TCR groups, we observed that persisting cells at the post-acute timepoint showed higher expression of genes that were associated with long-lived memory signatures (such as CCR7, IL7R, etc) (**Supplemental Figure 3.6E**), in agreement with previous studies[16,19,21]. In addition, our analysis further suggested a few other trends. Specifically, PG1 cells identified at the late timepoint showed a relatively lower naïve signature (CCR7, TCF7, etc) and slightly higher effector functions (GZMB, GZMH, etc) than those cells at the earlier timepoint (**Figure 3.7C,** right). By contrast, the few remaining post-acute PG3:High cells evolve

to express lower cytotoxicity (GZMB, PRF1, etc) and higher CCR7, indicating a shift towards less effector or more central memory phenotypes[16,37,38] (**Figure 3.7C**, left). In summary, APMAT revealed that distinct combinations of peptide and TCR physicochemical properties exhibit clear associations with not only cellular phenotypes at acute disease, but also divergent cell fates over time (**Figure 3.7D**).

3.4 CONCLUSION AND DISCUSSION

In this chapter, we present APMAT as a framework for the integrated experimental and computational analysis of antigen-specific CD8 T cells paired with phenotypic data on those T cells. We applied APMAT to investigate circulating CD8 T cells collected from 63 participants at the acute and convalescent stages of COVID-19. These cells exhibited specificity to SARS-CoV-2 antigens presented by HLA A*02:01 from across the full viral proteome. Our analysis uncovered relationships between the physicochemical characteristics of the pMHC:TCR interface and the corresponding T cell phenotypes and T cell phenotypic evolution. Our analysis is, in several aspects, in alignment with existing literature. We identified T cell populations targeting previously reported immunogenic antigens[18,24,39,40], and we find that the prediction rankings from the commonly-used NetMHCpan algorithm effectively assist in designing pMHC multimer constructs that will likely capture T cells. In addition, the longitudinal behavior of the antigen-specific clonotypes is consistent with current literature. Highly-cytotoxic effector clonotypes, which are expanded at acute disease, contract significantly at the time of convalescence. This is in line with the hypothesis that these cells undergo antigen-induced cell death. In contrast, those persisting clonotypes with mild effector potential traverse a phenotypic evolution towards central or effector memory phenotypes[41]. Further, clonotypes that exhibit memory and progenitor-like phenotypes at acute disease persist or expand at convalescence[16,42,43]. These results further emphasize the

importance of identifying early determinants of divergent clonal differentiation trajectories in antigen-specific T cell response.

We leveraged physicochemical metrics to characterize the pMHC:TCR interface to uncover a novel view of T cell specificity-clonotype-phenotype relationships. Take, for example, the above-described case of effector T cells that expand during an acute infection and contract at convalescence[41]. Our analysis further suggests that T cells possessing TCRs characterized by a hydrophobic CDR3β, which recognize antigens featuring hydrophobic anchor residues alongside charged or polar exposed residues, are more inclined towards effector phenotypes at the stage of acute COVID-19, and contract during convalescence. An analogous description of the pMHC:TCR interface (but with near opposite characteristics) can be used to identify those clonotypes that exhibit naïve or central memory phenotypes at acute disease and remain at convalescence. Notably, our analysis reveals that the anchor and exposed residues of the antigen exert different influences on T cell phenotype. Unlike many prior studies that relied on a rough annotation of the antigen identity, our approach takes into account the position of each individual residue, as well as the classification of antigens that may be comprised of distinct sequences, and yet exhibit biochemical similarities at the TCR:pMHC interface.

Whether or not these physicochemical determinants are general to other disease context is not resolved here, although the consistency of the picture painted here with what has been observed in tissue settings in murine models of chronic viral infection,[4] as well as in tumors,[5,6] does suggest some level of generality. Further, an analysis of a separate study on SARS-CoV-2 reactive T cells (**Supplemental Figure 3.7B**)[44]**,** as well as our recently reported phenotypic analysis of T cell clonotypes specific to two highly immunogenic viral antigens from influenza and CMV[8], (**Supplemental Figure 3.7A**) also suggest a degree of generality. Datasets of similar breadth and

depth for a range of diseases and for antigens presented by different HLA alleles are needed to more fully resolve this picture, and such work represents an exciting future direction.

We hypothesize that an analogous study as the one reported here, but directed at antigens presented by different HLA alleles, might yield a different set of 'rules' that are dependent upon HLA-specific docking geometries. While Class I HLA A, B, and C alleles are highly polymorphic, over 90% of the world's population carries at least one of the dozen most common. An APMAT analysis centered around those most common alleles would offer valuable insights for assessing the therapeutic potential of tumor-targeting TCRs and T cell vaccines.

3.5 METHODS

3.5.1 *COVID-19 individuals*

The study cohort is a subset of the INCOV cohort published previously[25,26]. Procedures for the INCOV study were approved by the Institutional Review Board (IRB) at Providence St. Joseph Health with IRB study number STUDY2020000175 and the Western Institutional Review Board with IRB study number 20170658. Potential participants were identified at five hospitals of Swedish Medical Center and affiliated clinics located in the Puget Sound region near Seattle, WA. All enrolled participants provided written in-person informed consent and samples were de-identified prior to analysis. 63 HLA A*02:01 individuals from the INCOV cohort were selected for this study. PBMCs collected at enrollment close to diagnosis (Acute-1), 1 week (Acute-2) were used for antigen-TCR paring assay in this study.

3.5.2 *Large-scale preparation of peptide-HLA complex libraries*

Single chain trimer (SCT) peptide-MHC (pMHC) libraries of the virus antigens were generated as described previously[23]6/7/24 2:12:00 PM. Briefly, potential HLA A*02:01 binding

epitopes (9-11 mer peptides) were generated from the complete SARS-CoV-2 genome (Wuhan-Hu-1 strain, GenBank ID: MN908947.3) and filtered by NetMHCpan-4.1 prediction[28]. A plasmid library of pcDNA3.1 vectors encoding covalently linked peptide antigen, β2M, HLA was built and verified by SANGER sequencing. Plasmids were transfected into Exp293 cells (ExpiFectamine™, Thermo Fisher) following manufacturer protocol. SCT expression yield was measured and normalized. Expressed pMHC-like SCTs were biotinylated (BirA ligase Kit, Avidity) and Histag purified (IMAC PhyTip columns, PhyNexus) in 96-well format. Individual SCT concentration was measured by protein absorbance at 290nm.

3.5.3 *SCT-dextramer generation and cell staining*

SCT dextramers were individually DNA barcoded using dCODE Klickmers (dCODE Klickmer, Immudex). Briefly, SCT pMHC monomer was mixed with barcoded dCODE-PE-dextramer at a ratio of 20 ligands per dextran and incubated for at least 1h on ice before adding biotin (100 μM) to block free binding sites. Dextramer cocktails were prepared by mixing 31-65 unique SARS-CoV-2 SCT dextramers and CMV (NLVPMVATV) SCT dextramers freshly before cell staining. PBMCs from 63 participants were thawed for CD8 T cell enrichment (Human CD8 T cell Isolation Kit, Miltenyi Biotec) according to the manufacturer's protocol then incubated with Human TruStain FcX blocking reagent (422302, BioLegend) for 10 min at 4 °C before wash. Cells were then divided into tubes and processed simultaneously on ice. Each tube of CD8 T cells were stained with a cocktail of dextramers for 25 min on ice in the presence of herring sperm DNA according to the manufacturer's instructions, with individual dextramer concentration at 1.1 nM. Cells were washed three times before surface antibody staining. FITC anti-human CD8 Antibody (BioLegend, 344704, clone SK1) and Apotracker™ Green viability dye (Biolegend, 427403) was added into each tube, in addition to one unique TotalSeq-C anti-human hashtag antibody

(BioLegend) to identify each tube. Samples were incubated for 30 min on ice and washed 3 times before sorting.

3.5.4 *10X genomics single cell sequencing*

Single, live, CD8, dextramer-positive T cells were sorted into FACS buffer (PBS, 2%FBS, 2mM EDTA and 10mM HEPES) using a BD FACSAriaII cell sorter. Sorted cells were immediately were pelleted, resuspended and loaded into a 10X Chromium reaction for single cell RNA sequencing (scRNA-seq). GEX, VDJ and Surface Protein libraries were generated using Chromium Next GEM Single-Cell 5′ kits v2 (10X Genomics) according to the manufacturer's protocol. Libraries were sequenced on an Illumina NovaSeq at a read length of 26x90 bp.

3.5.5 *Whole genome sequencing*

DNA extraction from whole blood was performed via bead-based enrichment on an automated extraction platform (Qiagen Qiasymphony and/or Promega Maxwell). The resultant extracts were quantified by Nanodrop. WGS library preparation was performed using Illumina DNA Prep kits and the final barcoded libraries were quantified by fluorometer. Libraries were multiplexed and loaded onto an Illumina flow cell for sequencing at 30x or higher coverage on a NovaSeq 6000 instrument. Raw sequencing data was analyzed for sequence variants using the Illumina DRAGEN field-programmable gate array (FPGA) platform. Briefly this platform performs the following automated steps: conversion of raw sequencing image data into demultiplexed fastq files, alignment to the reference human genome (hg19), analysis of single nucleotide variants, indels and copy number/structural variants using variant calling algorithms as well as assessment of sequencing data quality. Analyses with hg38 were computed after a liftover was done using the UCSC browser[46]. WGS information was de-identified.

3.5.6 *Single-cell sequencing data processing and phenotype assignment*

Transcriptome, TCR, surface protein levels and antigen specificity were simultaneously analyzed for each cell. Raw data were processed via Cell Ranger Single-Cell Software Suite (v3.1.0, 10X Genomics) using GRCh38 as a reference. Cells that fit any of the following filters were excluded due to low quality: n-counts <1000 or >10,000, n-genes <250 or >2500, mitochondrial percentage >10%. Gene counts for each cell were normalized by total expression, multiplied by a scale factor of 10,000 and transformed to log scale. Single cells were assigned phenotypes by clusters determined through the leiden algorithm. Phenotype associated transcripts were acquired from literature as follows[16,26,38,47]. Naïve/memory: *LEF1*, *TCF7*, *CCR7*. Memory: *IL7R*. Effector Memory: *GZMK*. Cytotoxic: *PRF1*, *GZMB*. Exhaustion: *TIGIT*, *PDCD1*. Proliferation: *MKI67*.

3.5.7 *Demultiplexing using genetic variants*

We wrote a Snakemake workflow (https://github.com/racng/snakemake-merge-wgs) for processing GVCF files from WGS to generate a multi-sample VCF file of exon variants. GVCFs are combined for specified genomics region (autosomes and sex chromosomes) using GATK (v4.1.9.0) GenomicsDBImport to generate a GenomicsDB datastore, which is then used by GenotypeGVCFs for joint calling of variants. After removing indels using vcftools (v0.1.16)[48] and excluding intron variants via bcftools (v1.8)[49] the remaining exon variants were lifted to GRCh38 (hg38) using CrossMap (v0.5.2)[50] and filtered again to remove KI27 contigs and duplicated variants. For each 10x library, BAM alignment files from cellranger were filtered for reads from autosomes and sex chromosomes. Using the processed VCF file (929,678 SNPs), single cell variants were extracted from the filtered BAM files via 10x Genomics VarTrix (https://github.com/10XGenomics/vartrix) with the coverage scoring method. To reduce memory

usage, single cell variants were kept if both ALT and REF alleles are detected in the dataset. We then used vireo (v0.5.0)[51] to assign donor identity and doublets based on the processed single-cell and WGS variants. Doublets were removed. Cells that unassigned with donor identity were removed.

3.5.8 *Antigen assignment based on dextramers and hashtags*

Raw reads for each Hashtags and Dextramers were normalized. Cells were assigned to their maximally expressed Hashtag. Cells that expressed multiple Hashtags at a high level were removed as potential doublets. Hashtag identities were then used to identify cells' SCT-dextramer cocktail. For each cell, we calculated the number of unique molecular identifiers (UMIs) for each dextramer, and the percentage of each dextramer. We assigned each cell an antigen only if their UMI count was >25 and the UMIs specific for that dextramer occupied >25% of that cells' dextramer reads. Antigens were then assigned by the maximally mapped dextramer for each cell. Ambiguous cells that didn't assigned with any dextramer were removed.

3.5.9 *Peptide physicochemical property assignment and Pep-UMAP*

We first transform peptide sequences into a numerical peptide matrix, where each row represents a residue position, and each column represents a feature characterizing amino acids, including amino acid identity, charge, hydrophobicity, weight, bulkiness, polarity, sulfur presence, aromaticity (Supplementary Table 3: Amino acid property scales used for peptide and TCR residues)[12]. The numeric values were scaled to ensure consistency range of values. For quantitative comparison of peptides in Figure 2, we calculated the average value for each property for anchor residues and TCR-exposing residues. Specifically, position 2 and 9 (and occasionally 6) tend to serve as anchors for HLA-A*0201 binding, while other "exposed" residues potentially contact the

TCR (CITE). Logo plot in Figure 2 was generated by Seq2Logo - 2.0[52] using default settings. In figure 3, to visualize peptide features and similarities, we applied UMAP followed by an autoencoder for dimensional reduction. In detail, since peptides have varying lengths, the residue positions are mapped to a normalized scale of 0 to 100. Each amino acid's features are replicated across the corresponding positions in the matrix. We then implemented an autoencoder using an MLPRegressor from scikit-learn to reduce the dimensionality of the peptide matrix. Finally, we computed a two dimensional peptide UMAP (Pep-UMAP) using scanpy.tl.umap to visualize peptide features and similarities.

3.5.10 *TCR physicochemical property assignment and analysis*

TCR-related physicochemical properties were computed for each cell based on its TCR CDR3 sequence of the beta chain. These properties, including charge, hydrophobicity, weight, bulkiness, polarity, sulfur presence, aromaticity, were calculated based on Supplementary Table 3 (Amino acid property scales used for peptide and TCR residues). Charge is absolute value unless specifically indicated. Specifically, we categorized CDR3β residues into V, J, and CDR3bmer regions. The V/J region comprise the first four, and last five amino acids, respectively, while the central CDR3bmer contains the amino acids in between. We evaluated TCR CDR3 hydrophobicity numerically by the average hydrophobicity of the CDR3bmer region. Additionally, we introduced a categorical score called "HPhobic%". HPhobic% represents the percentage of strongly hydrophobic residues ('A', 'V', 'L', 'I', 'F', 'M') in the CDR3β, excluding the first and last four amino acids. We calculated 45 properties for each cell's TCR sequences, including charge, hydrophobicity, weight, bulkiness, polarity, for three regions of CDR3β (V, J, CDR3βmer), as well as full CDR3α chain. The Mann-Whitney U test is applied to compare the distribution of each property between effector (Cytotoxic, Effector Memory, Hybrid) and non-effector (Naïve and

Central Memory) cell phenotypes. Log2 fold change (log2fc) for each property was calculated between the mean values of effector cell types and non-effector cell types. Top selected properties are based on the criteria of log2fc absolute value > 0.05, and p value <0.05. (Supplementary Table 7 and 8)

3.5.11 *Peptide and TCR groups density analysis*

Densities for peptide groups were projected onto GEX-UMAP by matching the antigen specificity of each single CD8 T cell to the peptide group that antigen peptide belongs to. Embedding density was first calculated via sc.tl.embedding_density then a 5 n-neighbor kNN graph was used to diffuse the values via five iterations to create a whole UMAP score for the density scores, as reported previously[8,53]. TCR group density calculations were implemented this same methodology.

3.5.12 *Differential gene expression and signature analysis*

Differentially expressed genes were called via scanpy.tl.rank_genes_groups through the Scanpy package using the ‘‘Wilcoxon’’ method which implements the Mann-Whitney U test[54]. Differentially expressed genes (DEGs) between peptide groups (PG2 vs PG3) were filtered for p values < 0.05 (Supplementary Table 5). Enriched pathways through filtered DEGs were computed via Enrichr[55] and provided in Supplementary Table 6. Top enriched pathways in PG3 than PG2 were reported in Figure 3.4e. DEGs between TCR groups (HPhobic-High vs HPhobic-Low) were called similarly to generate list of DEGs (Supplementary Table 9) and enriched pathways (Supplementary Table 10). Top enriched pathways in HPhobic-High than HPhobic-Low were reported in Figure 3.5g.

3.5.13 *Longitudinal T cell inquiry by GLIPH2 analysis for SARS-CoV-2 specific TCRs*

We utilized a reference dataset - our previously reported longitudinal dataset on 209 COVID-19 participants contained both scRNA-seq and scTCR-seq data from the same single cell along with assigned donors[25]. We perform GLIPH2 analysis on SARS-CoV-2 specific TCRs identified in this study that assigned with antigen specificity, and TCRs from the longitudinal dataset with unknown antigen specificity. GLIPH2 was run on http://50.255.35.37:8080/ using the GLIPH2 algorithm[56], version 1.0 reference for CD8, with all_aa_interchangeable set to YES. Antigen specific CD8 T cells were identified as those that belonged to the same GLIPH group of the SARS-CoV-2 specific TCRs identified in this study. Longitudinal SARS-CoV-2 specific CD8 T cells were then subset from the reference dataset for analysis in Figure 7. Cell count for each PG-TCR groups at Acute and Post-Acute timepoints were provided in Supplementary Table 11.

3.5.14 *Statistical analysis*

Microarray like datasets were analyzed using SCANpy and statistical comparisons were generated using "scanpy.tl.rank_genes_groups" using the "wilcoxon" method. All correlations were calculated using Pearson correlation. All p values were calculated using Mann-Whitney U test unless otherwise specified. Bar charts were provided with error bars when multiple values were present, and these bars represented standard errors. Bar level represented the mean variable value.

3.6 FIGURES

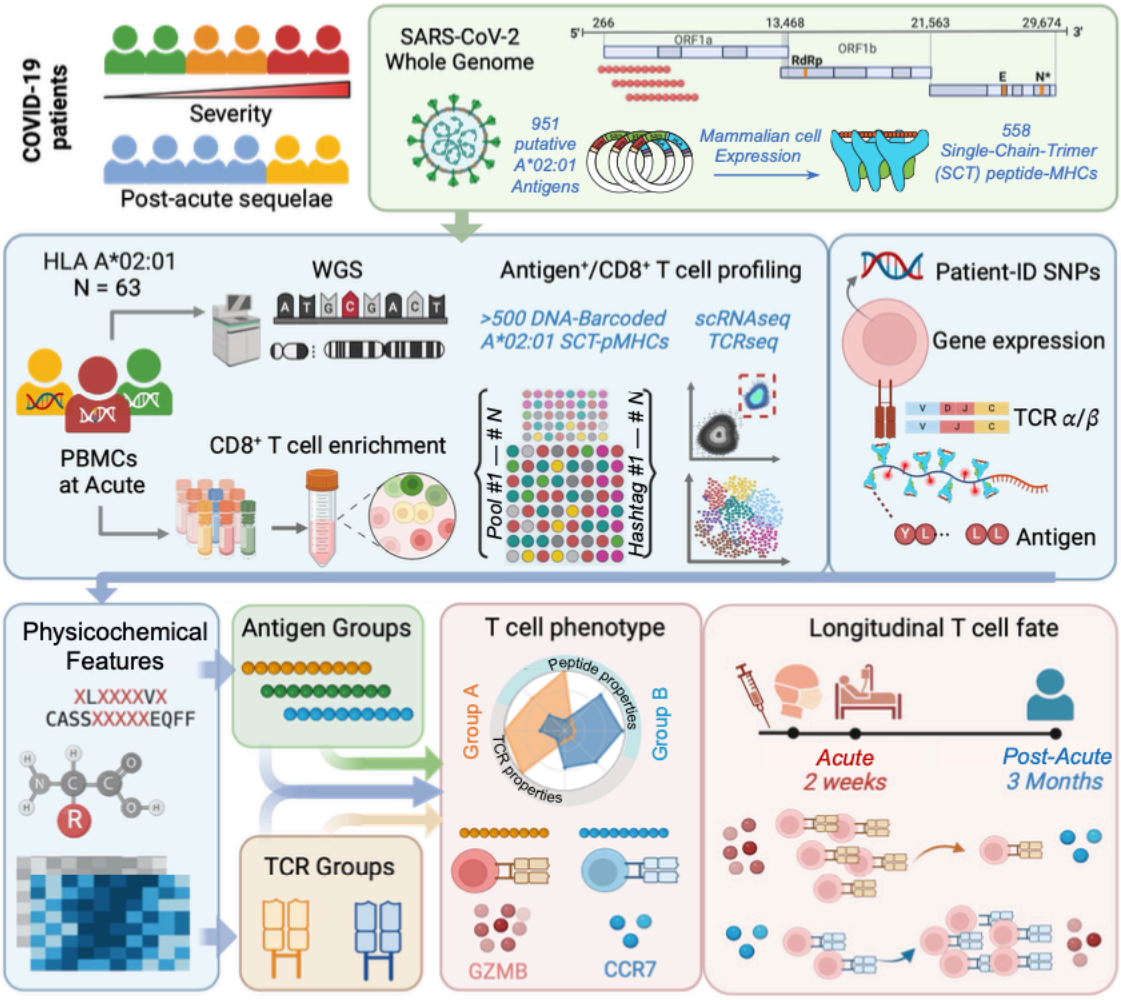


Figure 3.1: Schematic overview of antigen-TCR pairing and multiomic analysis of T-cells (APMAT) in COVID-19 participants

Top: A library of 558 single chain trimers (SCTs) representing HLA-A*02:01 peptide-MHC complexes was constructed to cover antigens across the full SARS-CoV-2 genome. The library started from a list of 951 putative antigens that were then expressed in the form of SCT by mammalian cells.

Middle: CD8 T cells from 63 HLA-A*02:01 donors, derived from PBMCs collected during acute disease stage of COVID-19, were stained with pools of DNA-barcoded SCT-dextramers and hashtags to identify the SCT specificity. SCT-pMHC multimers binding CD8 T cells were sorted for 10X genomics single cell RNA and TCR sequencing. Whole genome sequencing (WGS) data from the same participants was used to generate patient-specific single nucleotide polymorphism (SNP) profiles. The multi-modular barcoding strategy resulted in single cells simultaneously have gene expression, TCR, cognate antigen, and were linked to participants via SNP profiles.

Bottom: Antigen-TCR pairs were sub-grouped based on the physicochemical features of their peptide/TCR sequences. These subgroups were then analyzed to identify associations with gene expression and T cell phenotypes. Longitudinal T cell analysis was performed by TCR-based query from the longitudinal single-cell multi-omics dataset of the same cohort, tracking their abundance and phenotypic transition overtime.

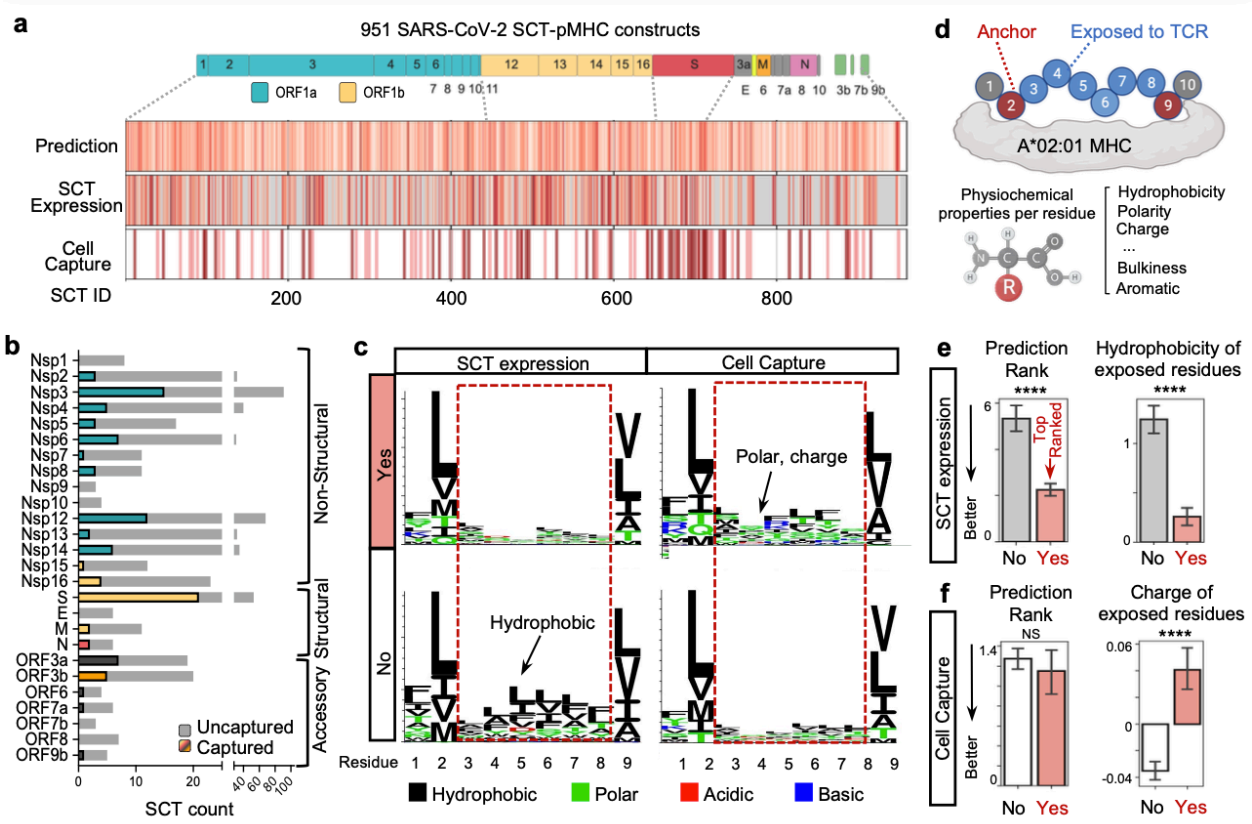


Figure 3.2: APMAT enables high-throughput representation of whole SARS-CoV-2 genome

A. Graphic relating the full SARS-CoV-2 genome to the predicted HLA A*02:01 presenting antigens (top row), to those SCT constructs that were expressed in usable yield (middle row), to those SCTs that captured antigen-specific CD8 T cells from COVID-19 participants (bottom row). For the SCT Expression row, darker red lines correspond to higher SCT expression, while grey means low/no expression. For the Cell Capture row, darker red lines mean more cells captured.

B. Bar plot showing the distribution of expressed SCTs for each SARS-COV-2 protein. SCTs that were used to successfully captured CD8 T cells are shown in colored bars, uncaptured SCTs in light gray. The color code of the captured cells is that used in the top row of panel a.

C. Antigen sequence motif of expressed (top left) vs non-expressed (bottom left) SCT constructs; and those expressed ones that did (top right) or did not capture CD8 T cells (bottom right).

D. For each peptide, conventional anchor positions (R2, R9) and non-anchor residues are assigned. The physicochemical properties are tabulated for each residue.

E. Bar plots comparing expressed vs non-expressed SCT constructs. X-axis: SCT expression status. Y-axis: NetMHC prediction rank (left) and average hydrophobicity of exposed residues (right).

F. Bar plots comparing expressed SCTs that captured CD8 T cells vs those that did not capture cells. X-axis: Cell capture status. Y-axis: NetMHC prediction rank (left) and average charge of exposed residues (right). Charge values represent the average of positive and negative charges rather than the absolute value.

**** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, standard error and mean are utilized for bar plots.

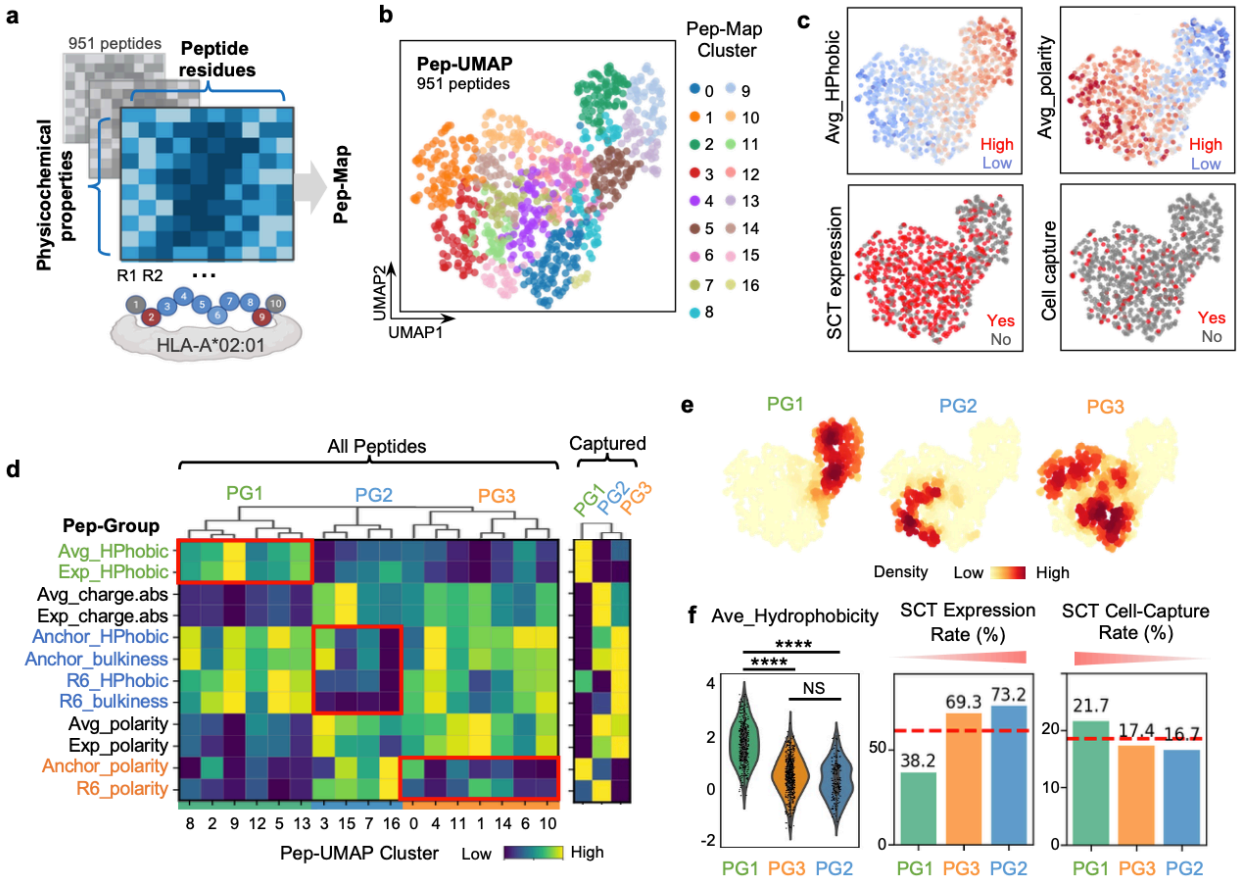


Figure 3.3: Three peptide groups distinguished by sequence physicochemical properties

A. All 951 putative SARS-CoV-2 peptides for HLA-A*02:01 are encoded by the physicochemical properties of amino acids at each position for unsupervised clustering.

B. UMAP embedding of all peptides (Pep-UMAP) based on their physicochemical properties with peptide clusters colored (legend on the right), each dot represents a unique SARS-CoV-2 peptide.

C. Pep-UMAP colored with selected physicochemical properties, including average (Avg) hydrophobicity (HPhobic), average polarity, whether the SCT expressed, and whether the SCT captured CD8 T cells. Average values were calculated for all residues, including “anchor” an “exposed” residues.

D. Left: Clustermap of the 951 peptide antigens by their normalized physicochemical properties, revealing 3 major peptide groups (Pep-Groups). The key signatures that distinguish the individual Pep-Groups are highlighted in red boxes. Right: Clustermap of the Pep-Groups only including those peptides that captured CD8 T cells.

E. Pep-UMAP with densities of PG1-3 depicted, legend on the bottom.

F. Left: Violin plot of peptide hydrophobicity for Pep-Groups, sorted by mean value. Middle: SCT protein expression efficiency for the Pep-Groups. Right: SCT cell capture efficiency for Pep-Groups.

**** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, standard error and mean are utilized for violin and bar plots.

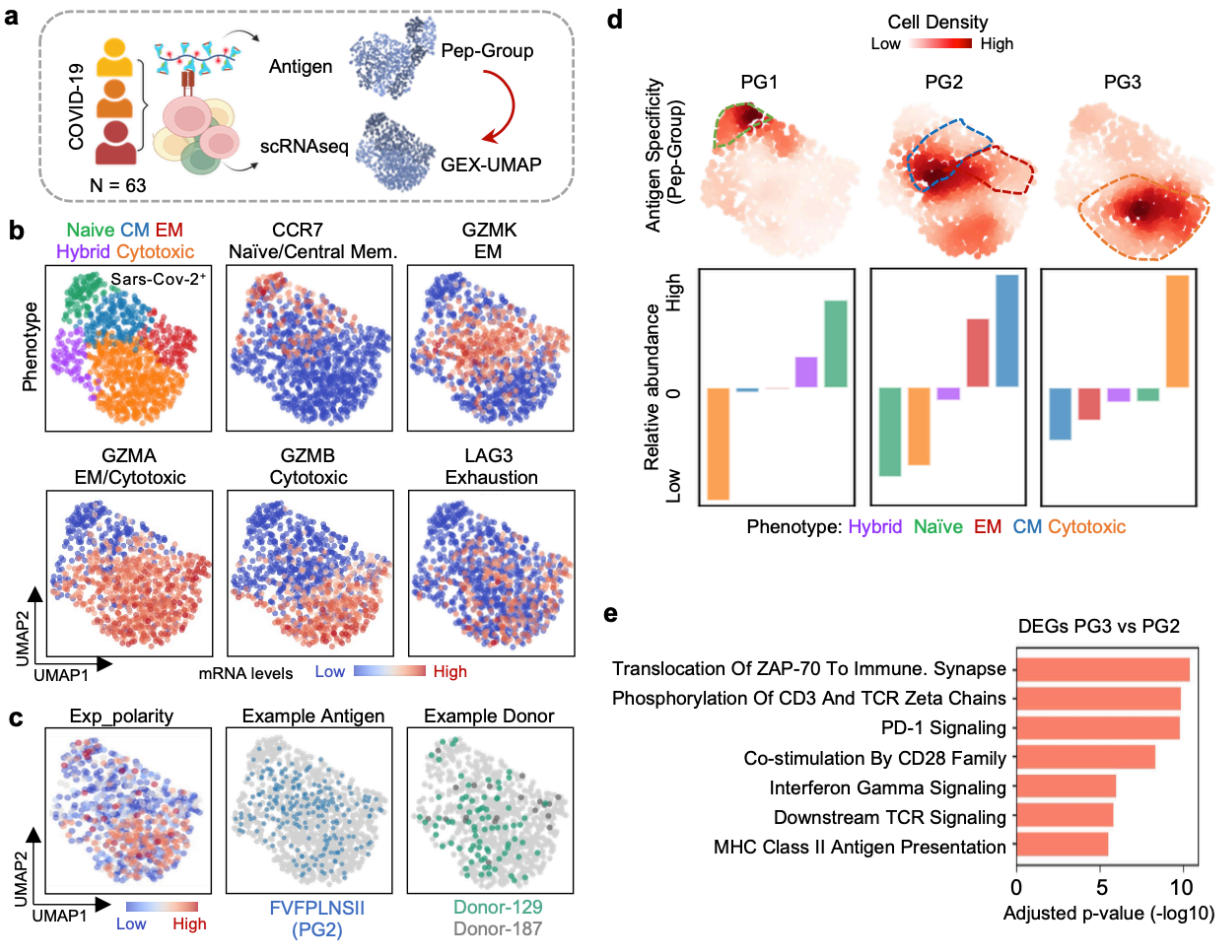


Figure 3.4: Pep-groups associate with different T cell phenotypes.

A. Schematic illustration of mapping peptide physicochemical properties with T cell gene expression profiles. CD8 T cells were captured by SCT-dextramers from 63 HLA-matched COVID-19 participants for scRNAseq and TCR sequencing. Gene expression UMAP (GEX-UMAP) was generated based on scRNAseq. We used the SCT identity to connect Pep-Groups with gene expression.

B. GEX-UMAP of Sars-CoV-2-specific CD8 T cells with different phenotypes color-encoded (legend on the top), and then color-coded by expression levels of selected mRNA transcripts.

C. GEX-UMAP, color-coded by an example physicochemical property of the peptide antigen: the polarity of exposed residues (left); color encoded to reveal all T cells specific to a given antigen (middle); and color encoded to reveal the SARS-CoV-2 specific T cells from example participants.

D. Top: GEX-UMAP color encoded with densities of each cell's antigen specificity by Pep-Groups. Bottom: Bar plot of the relative abundance of phenotypes for cells identified by each Pep-Group, with legend on the bottom.

E. Top enriched biological pathways of genes significantly elevated in cells captured by PG3 relative to PG2.

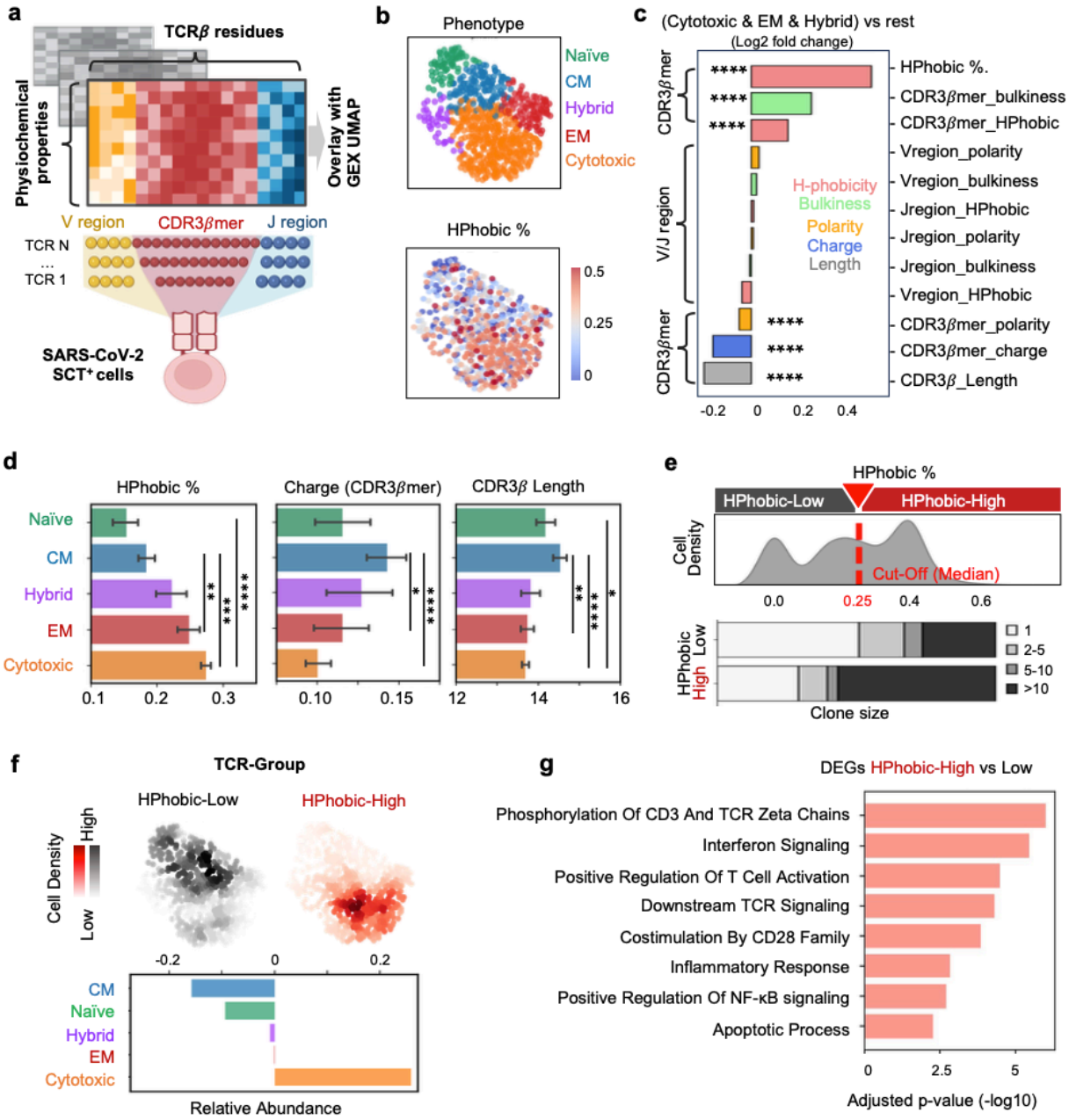


Figure 3.5: TCR hydrophobicity is an important factor for effector function.

A. Each TCR beta chain was split into V, CDR3βmer, and J regions, and then encoded by the physicochemical properties of amino acids at each residue position for overlay with the GEX-UMAP.

B. GEX-UMAP, color-coded by phenotypes (top) and overlayed with an example physicochemical property of the TCR: the percentage of hydrophobic residues within CDR3βmer (HPhobic %) (bottom).

C. The top differential TCR physicochemical properties for effector (Cytotoxic, EM and Hybrid) phenotypes, relative to non-effector (CM and naïve) phenotypes. Each bar colored by types of properties.

D. Bar plots showing the variation of selected TCR physicochemical properties across T cell phenotypes.

E. Top: We define binary TCR-Groups (HPhobic-Low and HPhobic-High) based on the median value of hydrophobic residue percentages for all SARS-CoV-2 cells. Bottom: Stacked bar plot of clonal size distribution for each TCR-Groups with legend on the right.

F. GEX-UMAP with densities of the TCR-Groups projected. The relative abundance of phenotypes in HPhobic-High cells in is plot on the bottom.

G. Top enriched biological pathways of genes significantly elevated in cells with HPhobic-High vs HPhobic-Low TCRs.

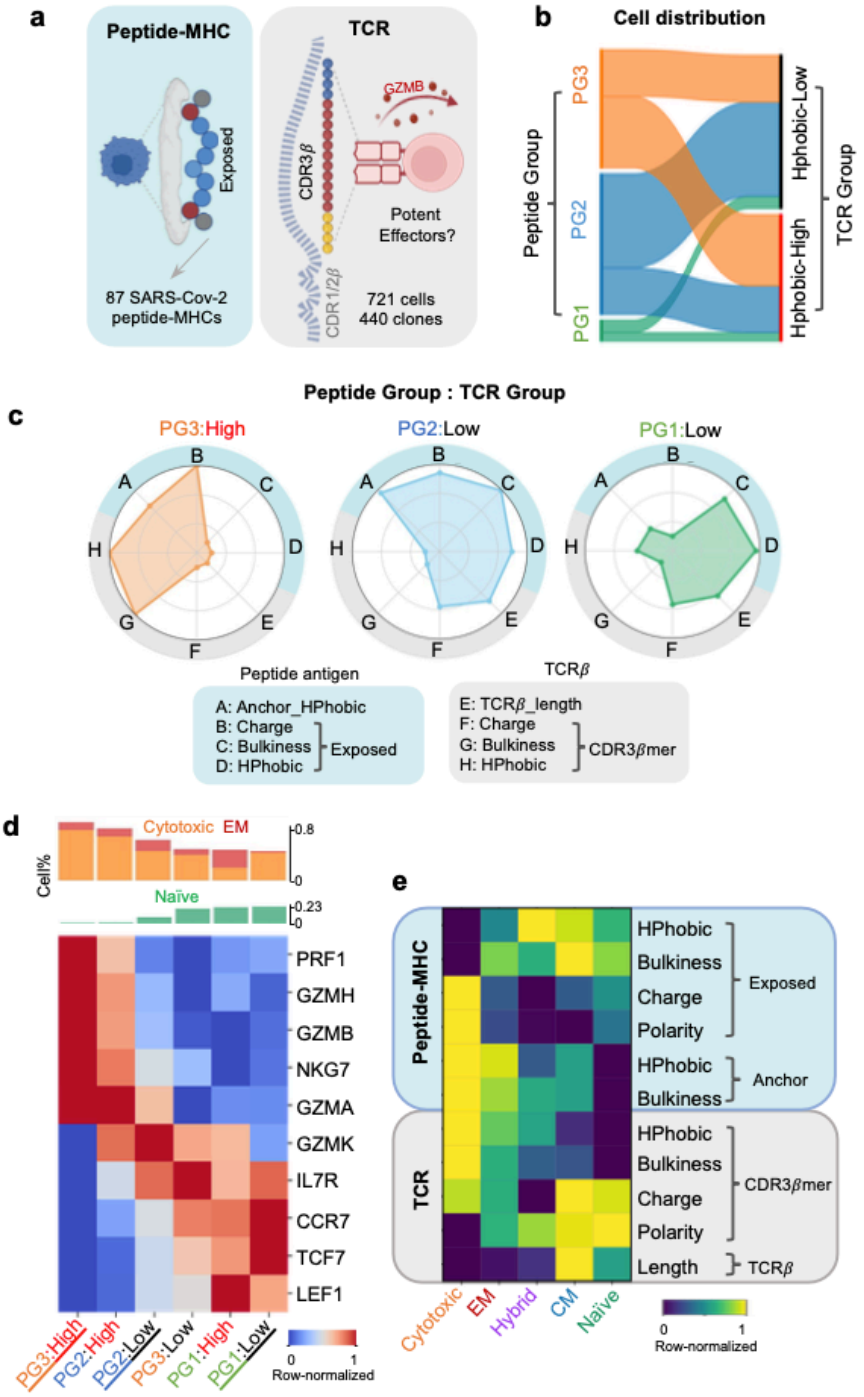


Figure 3.6: Combination of peptide-TCR features associated with cell phenotypes

A. We investigated the combination of peptide and TCR properties for antigen-TCR paired SARS-CoV-2 CD8 T cells.

B. Cell distribution between peptide groups (Left) and TCR groups.

C. Radar graphs of physicochemical properties for example PG-TCR groups. Blue and grey shaded areas of the outer rings indicate Pep-Group, and TCRβ properties, respectively. Each axis displays the normalized average value for each property, with lowest value in the center. The shaded polygons reflect the property space occupied by the peptide-TCRβ groupings. Legend on the bottom.

D. Top: Cell percentage of Cytotoxic, EM and Naïve phenotypes for each PG-TCR group. Bottom: Heatmap showing the selected mRNA level for each PG-TCR group. PG-TCR groups with underline are example ones from Fig.6 c.

E. Heatmap showing average value of peptide and TCR physicochemical properties for each phenotype.

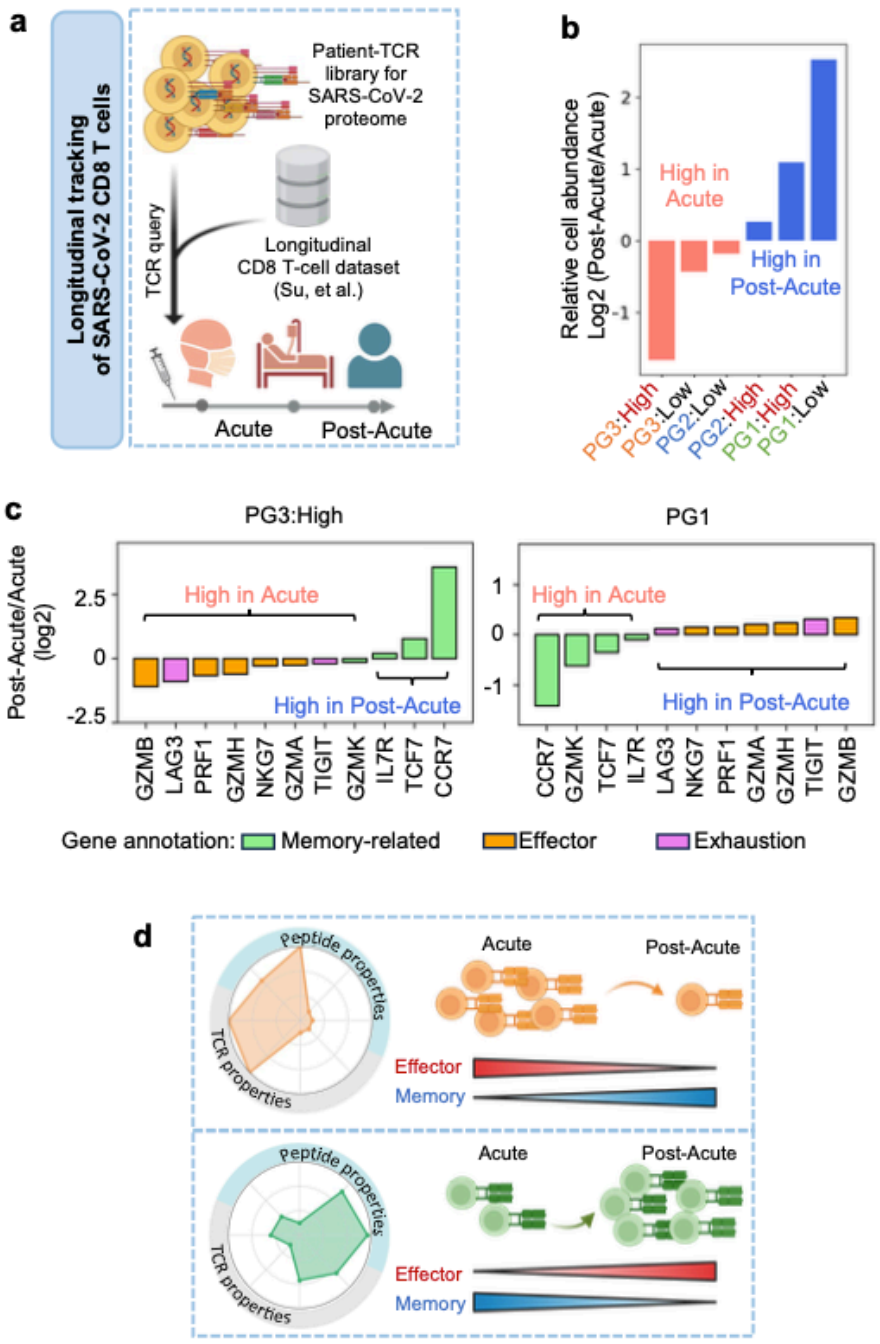


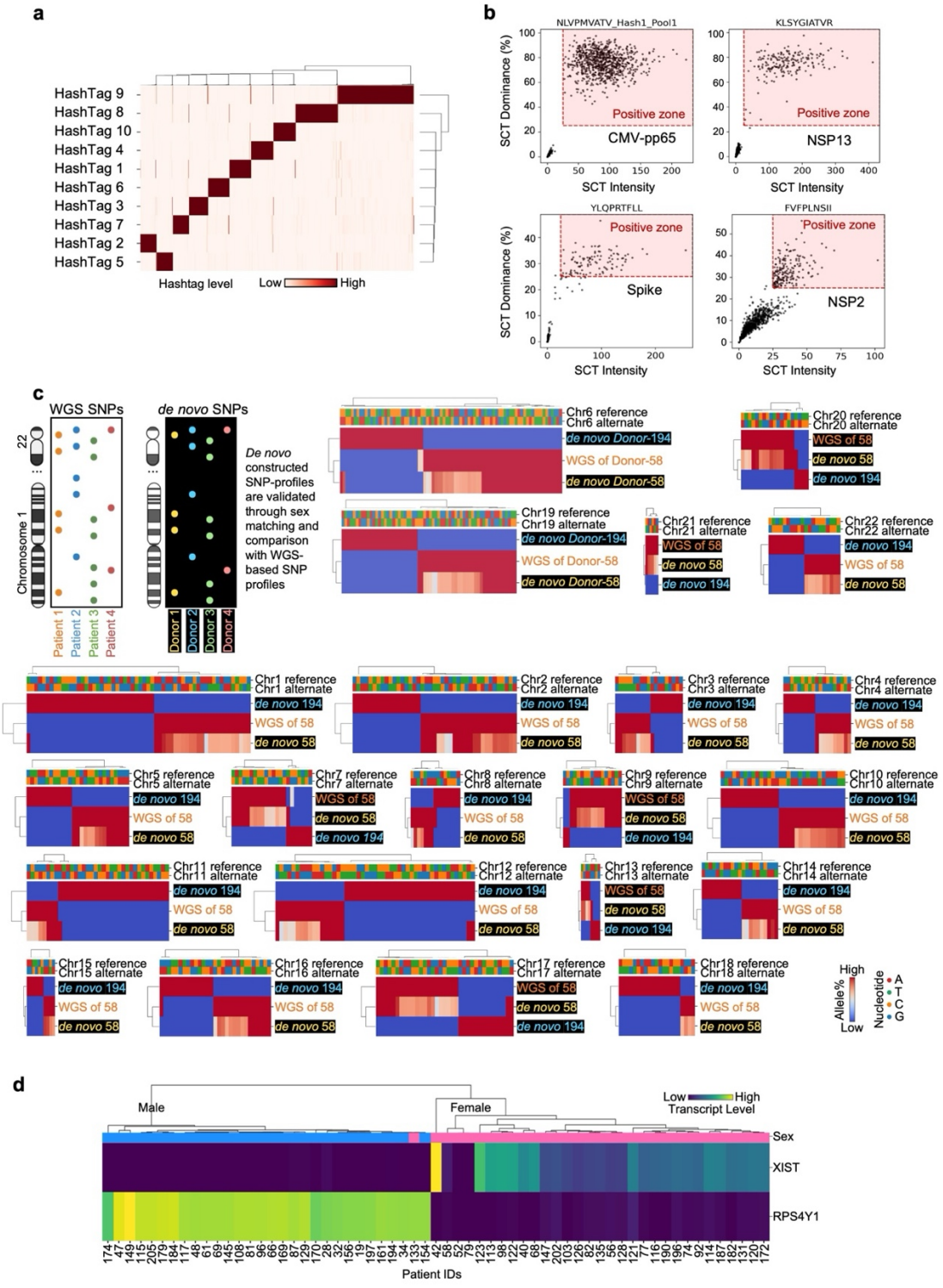
Figure 3.7: Longitudinal analysis for PG-TCR groups

A. Tracking SARS-CoV-2 CD8 T cells from acute to post-acute timepoint based on TCR-GLIPH query from a previously reported longitudinal dataset.

B. Cell abundancy change from post-acute to acute timepoint for each PG-TCR groups. Red or blue means higher abundancy at acute or post-acute timepoint, respectively.

C. Selected gene expression change from post-acute to acute timepoint for each PG-TCR groups. Gene annotations on the bottom.

D. Cartoon summary: The association between distinct peptide-TCR properties and cell fates for antigen-specific CD8 T cells.



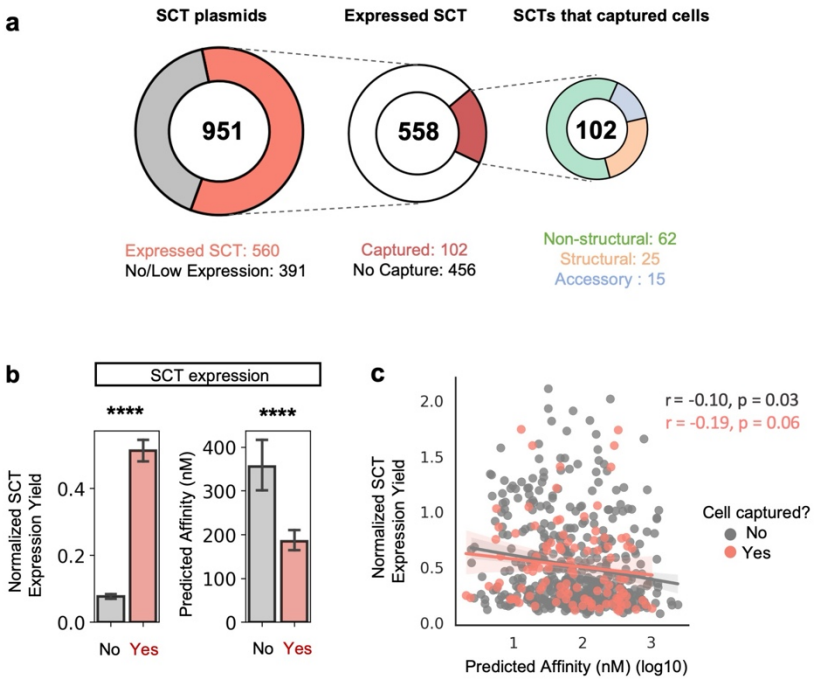
Supplemental Figure 3.1: Technical validation and methods for antigen and patient assignment

A. Heatmap where each row is a Hashtag used for each SCT-dextramer pool and each column is a single cell. Legend at bottom.

B. Scatter plots of antigen assignment for representative CMV and SARS-CoV-2 antigens. X-axis is the SCT intensity, defined by number of UMIs mapped to the SCT-dextramer. Y-axis is the SCT dominance, defined by percent of the cell's SCT-dextramer associated UMIs that mapped to this antigen's SCT-dextramer. Cells assigned with the antigen are shown in red positive zone, and have SCT intensity >25, and SCT dominance >25%.

C. Patient assignment by comparison of the WGS SNPs and derived *de novo* SNPs derived from scRNA-seq data, exampled by Donor-58 and Donor-194. Each panel represents the comparison for each chromosome. Allele frequency and nucleotide identity of reference and alternate are shown in lower right legend.

D. Heatmap with rows as sex specific genes (RPS4Y1 for male, XIST for female), columns are assigned participants. Upper row indicates patient sex from clinical records with blue as male and pink as female. Legend on upper right.



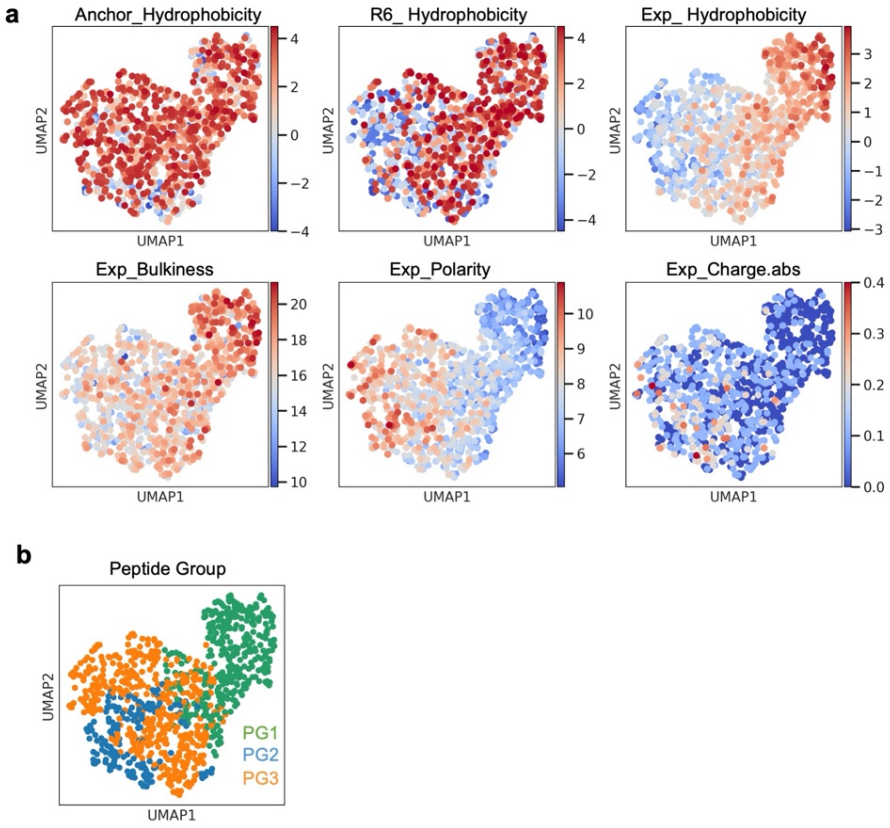
Supplemental Figure 3.2: SARS-CoV-2 SCT expression

A. Number of constructed SCT plasmids for putative peptides (left), expressed SCT constructs with useable protein expression yield (middle), and SCTs that captured SARS-CoV-2 CD8 T cells from COVID-19 participants.

B. Bar plots comparing expressed vs non-expressed SCT constructs. X-axis: SCT expression status. Y-axis: Normalized SCT expression yield (left) and NetMHCpan predicted binding affinity between peptide and HLA-A*02:01 (right).

C. Scatter plot for normalized SCT expression yield (Y-axis) and predicted binding affinity (X-axis) for all expressed SCTs, each dot represents a SCT construct.

**** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, standard error and mean are utilized for bar plots.

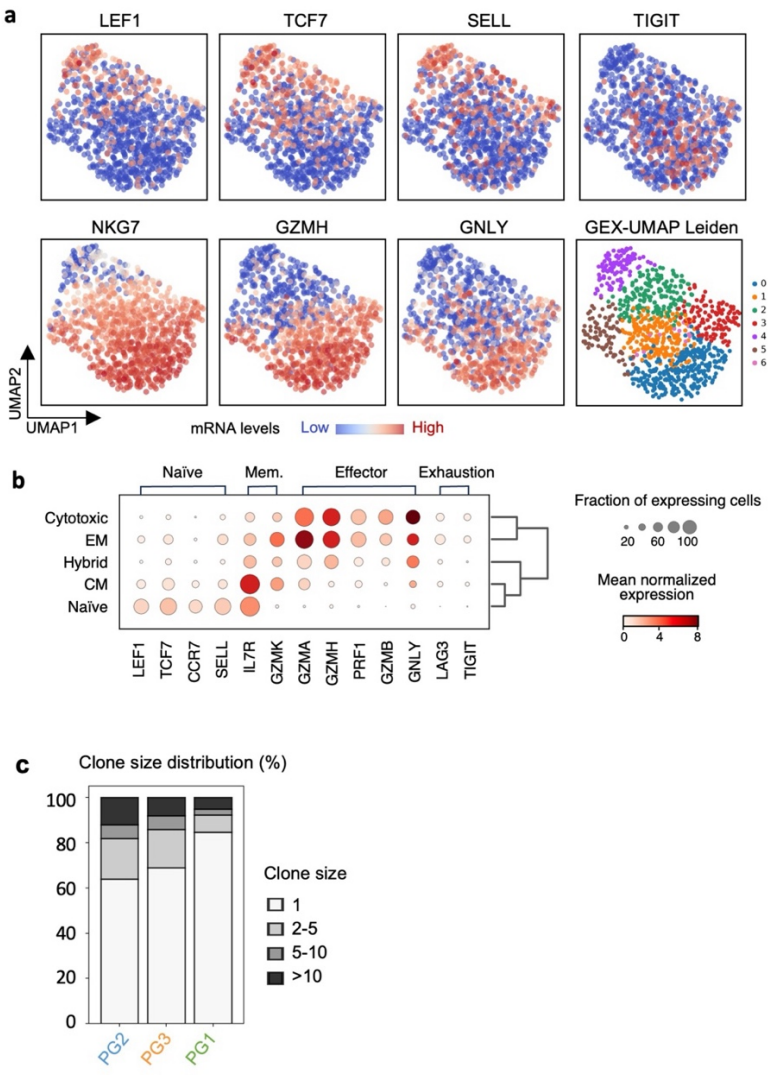


Supplemental Figure 3.3: Characteristics of Pep-UMAP and Pep-Groups

A. Pep-UMAP colored with selected physicochemical properties, including anchor hydrophobicity, hydrophobicity of the 6th residue, average hydrophobicity, bulkiness, polarity, and absolute charge of exposed (Exp) residues.

B. UMAP embedding of Pep-Groups.

**** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, standard error and mean are utilized for bar plots.

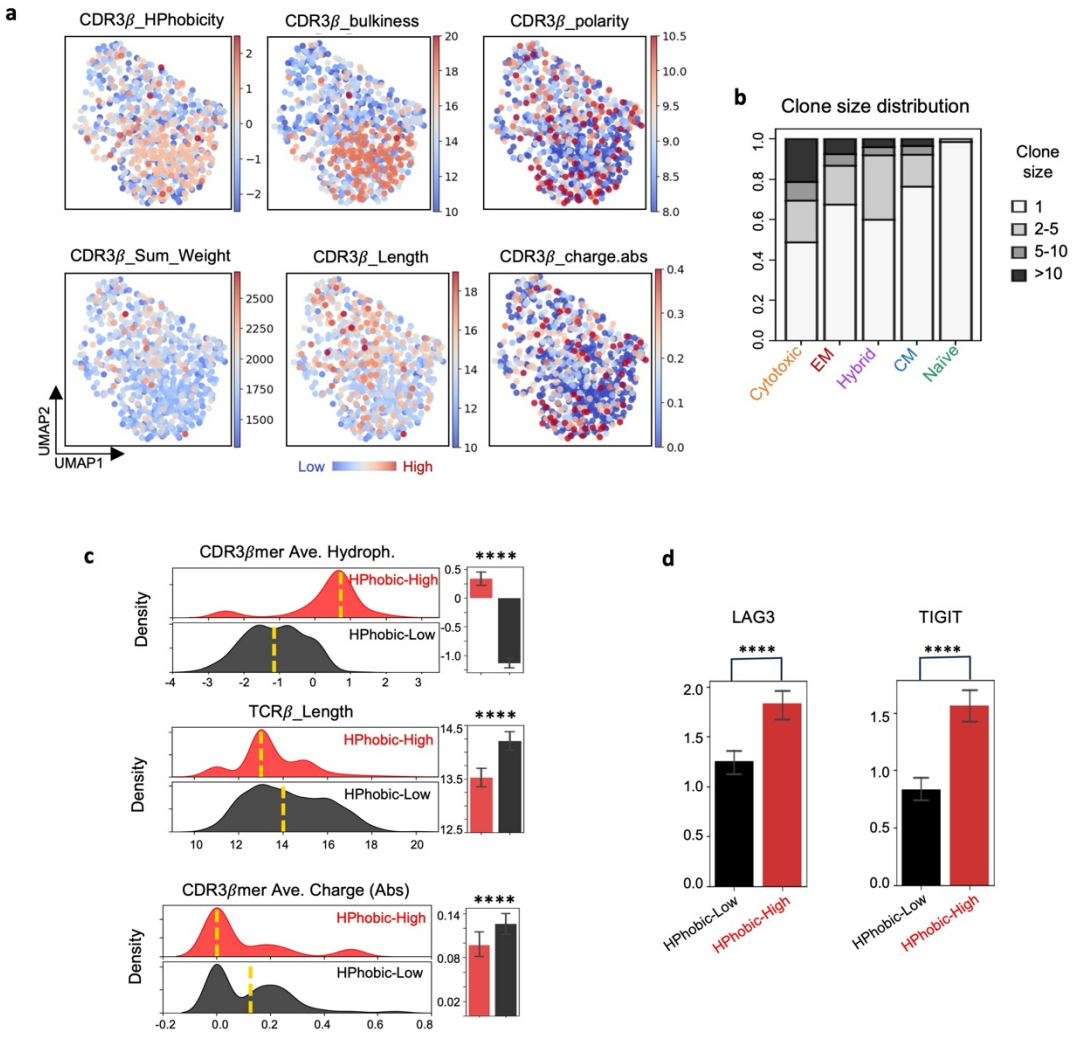


Supplemental Figure 3.4: Characteristics of GEX-UMAP

A. UMAP embedding of gene expression (GEX-UMAP) for SARS-CoV-2 specific CD8 T cells color-coded by expression levels of selected mRNA transcripts, and UMAP leiden groups.

B. Dot plot showing normalized expression levels of selected marker genes in each T cell phenotype. The size and color of each dot represent the fraction of expressing cells and the mean of normalized expression levels in each phenotype.

C. Stacked bar plot of clone size distribution for T cells captured by threes PG-group with legend on the right.



Supplemental Figure 3.5: Characteristics of TCR Groups

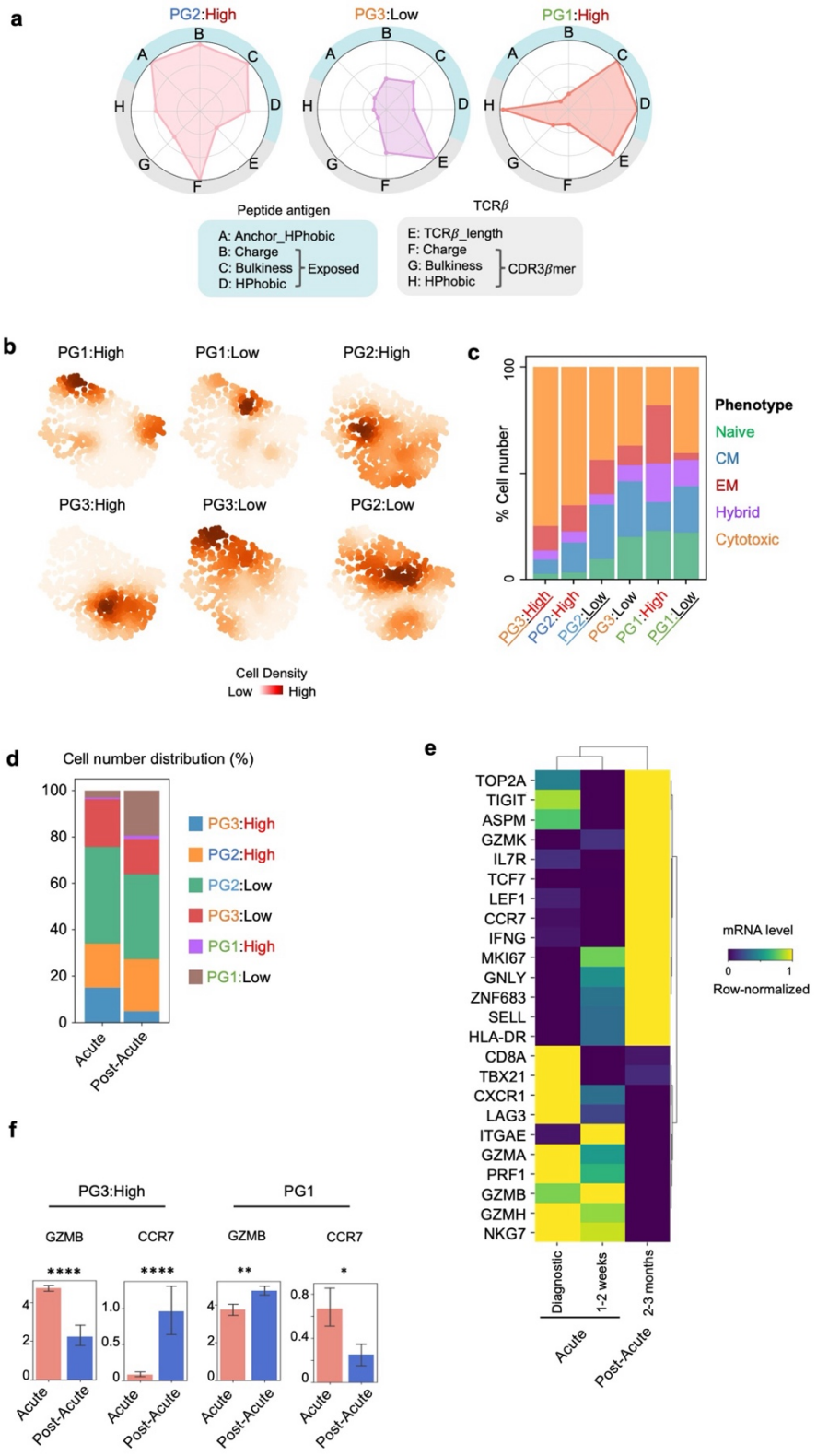
A. GEX-UMAP colored with selected physicochemical properties of the TCRβ sequences.

B. Stacked bar plot of clone size distribution for each of the T cell phenotypes with legend on the right.

C. Distribution of CDR3βmer hydrophobicity (top), CDR3β length (middle), CDR3βmer absolute charge (bottom) for each of the TCR-Groups. Bar plots for the respective property are on the right.

D. Bar plots for the mRNA levels of LAG3 and TIGIT.

**** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, standard error and mean are utilized for bar plots.



Supplemental Figure 3.6: Characteristics of PG-TCR groups

A. Radar graphs of physicochemical properties for example PG-TCR groups. Blue and grey shaded areas of the outer rings indicate Pep-Group, and TCRβ properties, respectively. Each axis displays the normalized average value for each property, with lowest value in the center. The shaded polygons reflect the property space occupied by the peptide-TCRβ groupings. Legend on the bottom.

B. GEX-UMAP color encoded with densities of PG-TCR groups based on each cell's antigen specificity and TCR sequence.

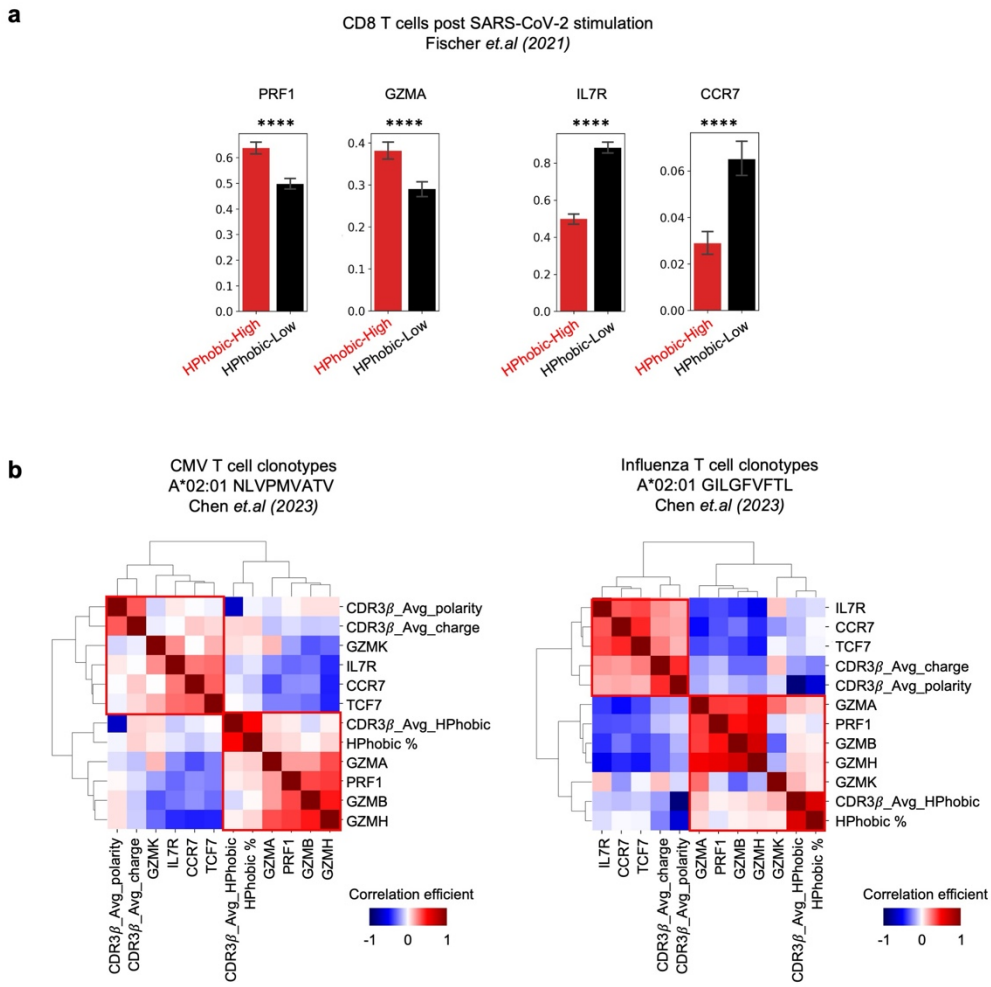
C. Stacked bar plot of phenotype distribution for each of the PG-TCR groups with legend on the right.

D. Stacked bar plot of PG-TCR group distribution for acute and post-acute timepoint with legend on the right.

E. Clustermap of cells by expression levels of selected mRNA transcripts for each timepoint. Legend on the right.

F. Bar plots comparing expression levels of selected mRNA transcripts for cells at acute and post-acute timepoint.

**** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, standard error and mean are utilized for bar plots.



Supplemental Figure 3.7: External dataset validation

A. Bar plots comparing expression levels of selected mRNA for cells within each TCR-Groups. Original dataset from Fischer *et. al* (2021)[45].

B. Pearson correlation coefficients between selected mRNA levels and TCR CDR3β properties for CMV-NLVP (left) and Influenza-GILG (right) specific CD8 T cells. Original dataset from Chen *et. al* (2023)[8].

3.7 REFERENCES

1. Chang, J. T., Wherry, E. J. & Goldrath, A. W. Molecular regulation of effector and memory T cell differentiation. *Nat Immunol* **15**, 1104–1115 (2014).
2. Zhang, N. & Bevan, M. J. CD8+ T Cells: Foot Soldiers of the Immune System. *Immunity* **35**, 161–168 (2011).
3. Connolly, K. A. *et al.* A reservoir of stem-like CD8+ T cells in the tumor-draining lymph node preserves the ongoing antitumor immune response. *Science Immunology* **6**, eabg7836 (2021).
4. Daniel, B. *et al.* Divergent clonal differentiation trajectories of T cell exhaustion. *Nature Immunology* **23**, 1614–1627 (2022).
5. Pai, J. A. *et al.* Lineage tracing reveals clonal progenitors and long-term persistence of tumor-specific T cells during immune checkpoint blockade. *Cancer Cell* **41**, 776-790.e7 (2023).
6. Lagattuta, K. A. *et al.* Repertoire analyses reveal T cell antigen receptor sequence features that influence T cell fate. *Nature Immunology* **23**, 446–457 (2022).
7. Pettmann, J. *et al.* The discriminatory power of the T cell receptor. *eLife* **10**, e67092 (2021).
8. Chen, D. G., Xie, J., Su, Y. & Heath, J. R. T cell receptor sequences are the dominant factor contributing to the phenotype of CD8+ T cells with specificities against immunogenic viral antigens. *Cell Reports* **42**, 113279 (2023).
9. Daley, S. R. *et al.* Cysteine and hydrophobic residues in CDR3 serve as distinct T-cell self-reactivity indices. *Journal of Allergy and Clinical Immunology* **144**, 333–336 (2019).
10. Schmidt, J. *et al.* Prediction of neo-epitope immunogenicity reveals TCR recognition determinants and provides insight into immunoediting. *Cell Reports Medicine* **2**, 100194 (2021).
11. Textor, J. *et al.* Machine learning analysis of the T cell receptor repertoire identifies sequence features of self-reactivity. *Cell Systems* **14**, 1059-1073.e5 (2023).
12. Chowell, D. *et al.* TCR contact residue hydrophobicity is a hallmark of immunogenic CD8 T cell epitopes. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E1754–E1762 (2015).
13. Leidner, R. *et al.* Neoantigen T-Cell Receptor Gene Therapy in Pancreatic Cancer. *New England Journal of Medicine* **386**, 2112–2119 (2022).
14. Pauken, K. E. *et al.* TCR-sequencing in cancer and autoimmunity: barcodes and beyond. *Trends in Immunology* **43**, 180–194 (2022).
15. Shah, K., Al-Haidari, A., Sun, J. & Kazi, J. U. T cell receptor (TCR) signaling in health and disease. *Signal Transduction and Targeted Therapy* **6**, (2021).
16. Adamo, S. *et al.* Signature of long-lived memory CD8 T cells in acute SARS-CoV-2 infection. *Nature* **602**, 148–155 (2022).
17. Ma, K.-Y. *et al.* High-throughput and high-dimensional single-cell analysis of antigen-specific CD8 T cells. *Nat Immunol* **22**, 1590–1598 (2021).
18. Saini, S. K. *et al.* SARS-CoV-2 genome-wide T cell epitope mapping reveals immunodominance and substantial CD8 $^{+}$ T cell activation in COVID-19 patients. *Sci. Immunol.* **6**, eabf7550 (2021).
19. Mold, J. E. *et al.* Divergent clonal differentiation trajectories establish CD8+ memory T cell heterogeneity during acute viral infections in humans. *Cell Reports* **35**, 109174 (2021).
20. Stadinski, B. D. *et al.* Hydrophobic CDR3 residues promote the development of self-reactive T cells. *Nat Immunol* **17**, 946–955 (2016).
21. Minervina, A. A. *et al.* SARS-CoV-2 antigen exposure history shapes phenotypes and specificity of memory CD8 T cells. *Nat Immunol* (2022) doi:10.1038/s41590-022-01184-4.

22. Puig-Saus, C. *et al.* Neoantigen-targeted CD8 T cell responses with PD-1 blockade therapy. *Nature* **615**, (2023).
23. Chour, W. *et al.* Large libraries of single-chain trimer peptide-MHCs enable antigen-specific CD8 T cell discovery and analysis. *Communications Biology* **6**, 1–13 (2023).
24. Grifoni, A. *et al.* SARS-CoV-2 human T cell epitopes: Adaptive immune response against COVID-19. *Cell Host & Microbe* **29**, 1076–1092 (2021).
25. Su, Y. *et al.* Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell* **185**, 881-895.e20 (2022).
26. Su, Y. *et al.* Multi-Omics Resolves a Sharp Disease-State Shift between Mild and Moderate COVID-19. *Cell* **183**, 1479-1495.e20 (2020).
27. Lee, J. W. *et al.* Integrated analysis of plasma and single immune cells uncovers metabolic changes in individuals with COVID-19. *Nature Biotechnology* (2021) doi:10.1038/s41587-021-01020-4.
28. Reynisson, B., Alvarez, B., Paul, S., Peters, B. & Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Research* **48**, W449–W454 (2021).
29. Zheng, H. *et al.* Multi-cohort analysis of host immune response identifies conserved protective and detrimental modules associated with severity across viruses. *Immunity* **54**, 753-768.e5 (2021).
30. Zhang, J.-Y. *et al.* Single-cell landscape of immunological responses in patients with COVID-19. *Nature Immunology* **21**, 1107–1118 (2020).
31. Mckeithan, T. W. Kinetic proofreading in T-cell receptor signal transduction. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 5042–5046 (1995).
32. Irving, M. *et al.* Interplay between T cell receptor binding kinetics and the level of cognate peptide presented by major histocompatibility complexes governs CD8 T cell responsiveness. *Journal of Biological Chemistry* **287**, 23068–23078 (2012).
33. Chin, S. S. *et al.* T cell receptor and IL-2 signaling strength control memory CD8 T cell functional fitness via chromatin remodeling. *Nature Communications* **13**, (2022).
34. Kaech, S. M. & Cui, W. Transcriptional control of effector and memory CD8 T cell differentiation. *Nature Reviews Immunology* **12**, 749–761 (2012).
35. Joshi, N. S. *et al.* Inflammation Directs Memory Precursor and Short-Lived Effector CD8 T Cell Fates via the Graded Expression of T-bet Transcription Factor. *Immunity* **27**, 281–295 (2007).
36. Kaech, S. M. & Wherry, E. J. Heterogeneity and Cell-Fate Decisions in Effector and Memory CD8 T Cell Differentiation during Viral Infection. *Immunity* **27**, 393–405 (2007).
37. Zhou, X. *et al.* Differentiation and Persistence of Memory CD8 T Cells Depend on T Cell Factor 1. *Immunity* **33**, 229–240 (2010).
38. Campbell, J. J. *et al.* CCR7 Expression and Memory T Cell Diversity in Humans. *The Journal of Immunology* **166**, 877 LP – 884 (2001).
39. Dolton, G. *et al.* Emergence of immune escape at dominant SARS-CoV-2 killer T cell epitope. *Cell* **185**, 2936-2951.e19 (2022).
40. Wu, D. *et al.* Structural assessment of HLA-A2-restricted SARS-CoV-2 spike epitopes recognized by public and private T-cell receptors. *Nature Communications* **13**, 1–14 (2022).
41. Wherry, E. J. & Ahmed, R. Memory CD8 T-Cell Differentiation during Viral Infection. *Journal of Virology* **78**, 5535–5545 (2004).

42. Giles, J. R. *et al. Shared and Distinct Biological Circuits in Effector, Memory and Exhausted CD8 T Cells Revealed by Temporal Single-Cell Transcriptomics and Epigenetics*. *Nature Immunology* vol. 23 (Springer US, 2022).
43. Harty, J. T. & Badovinac, V. P. Shaping and reshaping CD8 T-cell memory. *Nature Reviews Immunology* **8**, 107–119 (2008).
44. Fischer, D. S. *et al.* Single-cell RNA sequencing reveals ex vivo signatures of SARS-CoV-2-reactive T cells through 'reverse phenotyping'. *Nature Communications* **12**, (2021).
45. Fischer, D. S. *et al.* Single-cell RNA sequencing reveals ex vivo signatures of SARS-CoV-2-reactive T cells through 'reverse phenotyping'. *Nat Commun* **12**, 4515 (2021).
46. Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Research* **12**, 996–1006 (2002).
47. Szabo, P. A. *et al.* Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nature Communications* **10**, 4706 (2019).
48. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
49. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, 1–4 (2021).
50. Zhao, H. *et al.* CrossMap: A versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006–1007 (2014).
51. Huang, Y., McCarthy, D. J. & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *Genome Biology* **20**, 1–12 (2019).
52. Thomsen, M. C. F. & Nielsen, M. Seq2Logo: A method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Research* **40**, 281–287 (2012).
53. Chen, D. G. *et al.* Integrative systems biology reveals NKG2A-biased immune responses correlate with protection in infectious disease, autoimmune disease, and cancer. *Cell Reports* **43**, 113872 (2024).
54. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* **19**, 15 (2018).
55. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Current Protocols* **1**, 1–51 (2021).
56. Huang, H., Wang, C., Rubelt, F., Scriba, T. J. & Davis, M. M. Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat Biotechnol* **38**, 1194–1202 (2020).

Chapter 4. PEPTIDE-LOADED CLEAVABLE SINGLE-CHAIN TRIMER FOR ANTIGEN-SPECIFIC T CELL DETECTION

4.1 ABSTRACT

The high-throughput detection of antigen-specific T cells is crucial for understanding the adaptive immune response, but has been limited by challenges in rapidly and cost-effectively generating large libraries of peptide-MHC (pMHC) complexes. Here, we introduce peptide-loaded cleavable single-chain trimer (pCSCT) that taking the advantages of conventional disulfide-stabilized SCT, but enables rapid pMHC generation from pre-synthesized peptides using just a thermal cycler. We demonstrate pCSCT reagent's high efficiency in detecting antigen-specific T cells against viral and cancer neoantigens from engineered cell lines and patient PBMCs. It avoids the non-specific binding for some of conventional SCTs and commercial pMHCs. With high stability, versatility and ease-of-use, pCSCT represents a powerful tool for high-throughput analysis of antigen-specific CD8 T cell responses across infectious diseases, vaccines and cancer immunotherapy.

4.2 INTRODUCTION

One major challenge in high-throughput detection of antigen-specific T cells using peptide-MHC (pMHC)-like reagents is the rapid and cost-effective generation of large libraries presenting hundreds to thousands of different peptides with varying binding affinities. The highly modular single-chain trimer (SCT) technology enables the rapid expression of individual pMHC molecules with high purity without synthesizing individual peptides[1–6]. Traditionally, peptide-exchange

technologies haven been developed to utilize a single stock of a conditional peptide MHC molecule and exchange the peptide in a later step[7–9,10,11]. Most peptide-exchange technologies can overcome the need of individual small-scale paralleled refolding reactions for three compartments including peptide, β2-microglobulin (β2M) and HLA. Peptides of interest are loaded in a separate step by UV-mediated[10,12] or temperature-induced[7] peptide exchange or via disulfide-stabilized[8,13] empty MHC-I molecules. However, the β2M and HLA components are more commonly bacterially expressed, or require a CHO-S expression system, and usually with limited reagent yield and stability. Taking advantage of the human-derived Expi-293 SCT expression system, we introduce "peptide-loaded cleavable SCT" (pCSCT) that utilizes disulfide-stabilized MHC molecules, as well as the covalent linkage between β2M and MHC, to improve the yield, stability, and performance of peptide-receptive pMHC reagents.

4.3 RESULTS

4.3.1 *Cleavable SCT template enable the high yield production of peptide-receptive CSCT reagents*

Cleavable SCT (CSCT) template for multiple HLAs were designed by inserting a TEV protease cleavage site between the peptide and β2M of the traditional SCT template (Figure 4.1 top). The CSCT template is treated with TEV protease to remove the linker between the temporary peptide and β2M. Peptide-loaded CSCT (pCSCT) is then produced by peptide exchange in the presence of the target peptide. For pCSCT-based pMHC multimers, we also prepared negative control tetramers using the corresponding CSCT template to exclude any false positive binding signal from the uncleaved CSCT template (see flow gating strategies in Figure 4.S1).

To investigate the potential influence of the TEV protease cleavage site position, we prepared four variations (sites A, B, C, D) of cleavable SCT templates for HLA-A*02:01 encoding a temporary peptide (MART1 antigen: ELAGIGILTV) by modifying the distance between the cleavage site and the peptide (Figure 4.1 bottom). The reduced SCD-PAGE gel showed a similar yield of the MART1-encoding TEV-cleavable SCT compared to the unmodified MART1-SCT (Figure 4.S2A). The slightly elevated size for CSCT templates indicates the inserted TEV cleavage site. All four designs showed similar expression yields around 1.4 mg/ml.

The A*02:01 MART1-CSCT was incubated with TEV protease for one hour at 30 ºC and left at 4ºC overnight to remove the covalent linkage between the temporary peptide (MART1, ELAGIGILTV) and β2M. The peptide-receptive SCD reagents were concentrated using size-exclusive ultra-centrifugation (MWCO = 35kDa) to remove TEV protease and any free peptide. A slightly size change occurred after TEV treatment (Figure 4.S2A). Tetramer staining using variations of MART1-CSCT and unmodified MART1-SCT showed consistent binding on the MART1-specific Jurkat cell line, indicating an intact peptide-MHC structure and proper pMHC-TCR interaction with or without the presence of TEV cleavage site between the peptide-β2M linker (Figure 4.S2B).

4.3.2 *Optimization of the CSCT template design by adjusting the TEV cleavage site*

To determine the optimal design of CSCT templates, we compared four CSCT designs (cleavage sites A, B, C, D) by evaluating the performance of peptide-exchanged CSCT (pCSCT) tetramers on TCR-expressing Jurkat cell lines. The list of tested antigen-specific TCRs is provided in Table 4.1 and Figure 4.1A.

Antigen description	**Antigen ID**	S13	S33	S46
	Source	SARS-CoV-2	SARS-CoV-2	SARS-CoV-2
	Epitope	Spike_821-829	Spike_269-278	Spike_995-1004
	Peptide	LLFNKVTLA	YLQPRTFLLK	RLITGRLQSL
	Predicted affinity (nM)	12.82	108.8	163.63
	% ELRank	0.105	2.371	0.517
TCR features	**TCR ID**	TCR001	TCR008	TCR079
	HLA restriction	A*02:01	A*02:01	A*02:01
	TRBV	TRBV28	TRBV7-9	TRBV20-1
	CDR3b	CASSLTGGGYEQYF	CASSPDIEQYF	CSGEGGDTEAFF
	TRAV	TRAV9-2	TRAV12-2	TRAV19
	CDR3a	CALNDYKLSF	CAVNQDDKIIF	CALSEGYNFNKFYF

Table 4.1 List of antigen-TCR pairs for pCSCT validation in CTR-transduced Jurkat cell lines.

We first tested the effect of TEV cleavage and peptide exchange on the template MART1-CSCT. Tetramers were prepared for MART1-CSCT before and after TEV treatment and tested on the MART1-specific TCR Jurkat cell line. As shown in Figure 4.S2 left, all uncleaved MART1-CSCT templates (including sites A, B, C, D) showed consistent tetramer binding signals similar to the conventional SCT. For the "peptide-receptive" state of the CSCT after TEV cleavage, the tetramer binding signal slightly decreased for sites B, C, D, while site A showed the highest tetramer binding. Despite slight signal intensity differences, all four templates captured the same percentage of MART1-positive cells. This could be explained by the fact that, since MART1 peptide is an optimal binder for HLA-A*02:01, the cleaved free peptide can remain properly seated within the HLA binding groove and stabilize the pMHC structure, despite the TEV treatment disabling the covalent bond between the MART1 peptide and β2M.

After linker cleavage and removal of TEV protease, peptide-receptive CSCTs were incubated with high concentration of target peptide (S13-LLFN) to form the “peptide-exchanged” state of the CSCT (pCSCT), where the target S13 peptide substituted the original MART1 peptide. As shown in Figure 4.S2 right panel, S13-pCSCT showed significantly decreased tetramer binding of

MART1-specific Jurkat cells, where Site-A and Site-C showed completely no tetramer positive signal, indicating high peptide exchange efficiency. Based on the stepwise testing of MART1-CSCT on the MART1 Jurkat cells, we conclude that Site A and Site C showed highest decrease of template peptide (MART1) signal after TEV cleavage and peptide exchange.

Next, we compared the binding performance across four CSCT templates against the cognate TCR expressing Jurkat cell lines (Figure 4.2A). Cleaved HLA A*02:01 MART1-CSCT monomers were exchanged with three SARS-CoV-2 spike peptides (S13, S33, S46), and then prepared into fluorophore-labeled tetramers (PE and APC) to detect antigen-specific cells. All pCSCT tetramers showed no non-specific binding when staining irrelevant TCR cell lines (data not shown). Antigen-specific cells were successfully detected using pCSCT tetramers for all three peptide-TCR pairs (Figure 4.2B). The ratio of tetramer-positive frequencies between each pCSCT template and conventional SCT was calculated for each peptide-TCR pairs (Figure 4.2C). For S33 and S46, pCSCT tetramer-positive frequencies were similar to SCT, regardless of the TEV cleavage site. However, S13-pCSCT tetramer only showed optimal performance using Site-A and Site-C, detecting comparable percentages of S13-specific cells to the conventional SCT tetramer. The site-A pCSCT tetramer for all three peptide-TCR pairs showed the highest tetramer signal intensity compared to sites B, C, and D (Figure 4.2D). Based on these results, we chose site-A as the optimal CSCT template design for further experiments.

4.3.3 *Optimization of the peptide-exchange condition for pCSCT multimer*

To determine the optimal peptide-exchange condition for pCSCT, two conditions were tested to load the peptide-receptive CSCT with the target peptide: 4°C overnight incubation or 32°C for 1-hour incubation, and compared to the conventional SCT. The efficiency of these conditions was

evaluated by tetramer binding of S13-pCSCT and S46-pCSCT tetramers to their cognate TCR-expressing Jurkat cell lines. Both peptide-exchange conditions resulted in pCSCT tetramers efficiently detecting consistent frequencies of antigen-specific cells, comparable to the corresponding SCT tetramer (Figures 4.3A and 4.3B). Therefore, both conditions (4°C overnight or 32°C for 1 hour) allowed effective generation of functional pCSCT multimers for the HLA-A*02:01 allele.

Similar to UV-induced peptide exchange, we used an excessive amount of the target peptide to accelerate the peptide exchange. We then examined the optimal ratio between the target peptide and CSCT monomer using S33 and S46 as examples. Antigen-specific cells were consistently identified by pCSCT-tetramer when the peptide's molar ratio was 100-fold higher than the CSCT (Figure 4.4A). Since we used the A*02:01 MART1-CSCT as the template, allowing us to calculate the peptide exchange efficiency by testing the tetramer binding signal before and after peptide exchange on MART1-specific Jurkat cells. Consistently, increasing the ratio higher than 100-fold did not significantly improve the tetramer binding toward cognate TCR-expressing cells (Figures 4.4A and 4.4C), nor enhance the peptide exchange efficiency, since for both peptides tested, the exchange efficiency was 100%, meaning complete abrogation of binding toward MART1-specific Jurkat cells. Titrating the amount of pCSCT in cell staining showed that 1.2 μL from the pCSCT tetramer reaction resulted in the optimal tetramer signal (Figure 4.4D).

4.3.4 *Sensitivity and stability of pCSCT tetramer for the detection of antigen-specific T cells*

To further evaluate pCSCT's performance detecting low frequency antigen-specific T cells, we investigated the detection limit and sensitivity of pCSCT tetramer. The sensitivity of pCSCT tetramer staining was determined by calculating the recovery of TCR001, TCR008, and TCR079 expressing cells spiked into Jurkat cells expressing an irrelevant, P53-G245S specific TCR. The

limit of detection is 0.156% or lower, based on the lowest frequency tested. Moreover, across all peptide-TCR pairs and antigen-specific T cell frequencies ranging from 0.1% to 20%, we demonstrate a tight correlation in the detected T cell frequency using pCSCT or conventional SCT tetramers.

One limitation of peptide-MHC-like reagents is their stability and limited shelf-life. Here, we tested the long-term stability of the pCSCT reagent in the form of a peptide-loaded monomer. To reduce the effect of varying TCR expression levels in Jurkat cells during cell passaging, which might affect the actual antigen-specific cell frequency, we compared batches of pCSCT reagents with varying storage lengths and tested them against cells in a single experiment to maintain consistency for TCR expression. S33-pCSCT and S46-pCSCT monomers were prepared at 7–10-day intervals by incubating a 100-fold molar excess of S33 or S46 peptide with freshly TEV-treated A*02:01 MART1-CSCT template. Each pCSCT monomer was stored in PBS at 4°C for up to 75 days. All batches of pCSCT monomers were multimerized onto fluorochrome-labeled streptavidin on the same day and then used to stain cognate TCR-expressing Jurkat cells. We observed consistent cell frequencies (Figure 4.6A) and staining intensities (Figure 4.6B) for pCSCT-tetramer-positive cells for up to 75 days at 4°C for S46-pCSCT, whereas S33-pCSCT retained full efficiency up to 62 days with a small loss on day 75 (Figure 4.6C). All tested pCSCT tetramers with varying storage durations showed no non-specific binding for irrelevant TCR-expressing Jurkat cells (Figure 4.6A).

Alternatively, TEV-treated, "peptide-receptive" CSCT monomers can be stored in PBS at 4°C or -20°C containing 20% glycerol for even longer-term storage (see Figure 4.9). This validates the long-term stability of the pCSCT reagent for T cell detection, in the form of either ready-to-use,

peptide-loaded pCSCT tetramers or the "peptide-receptive" CSCT monomer allowing flexible future loading of the desired peptide.

4.3.5 *Two strategies of pCSCT preparation for flexible application*

So far, we have demonstrated the efficiency of pCSCT in a two-step procedure, where peptide-exchange happens after TEV-cleavage. This “Two-step” strategy enables the bulk preparation of “peptide-receptive” CSCT monomer and increases the flexibility of peptide loading as a separate step, requiring only a single ultracentrifuge tube for the removal of TEV protease. However, this strategy has limitations: first, it is a two-step process and often requires two days for a complete reaction; second, it requires a stable "peptide-receptive" structure, and when working with a CSCT template with a low-affinity peptide, the unfolding of the "peptide-receptive" structure might lead to failed peptide exchange. We showed that for at least two cases, we failed to produce pCSCT using the following CSCT templates, possibly due to this reason: A*24:02 with the template peptide KITEQEKDFLW (PIK3CA E542K) and B*07:02 with the template peptide RHSVVVPCEP (P53-Y220C), where both peptides are neoantigens that are not strong binders for their respective HLAs.

We hypothesized that a "one-step" method, where TEV cleavage happens in the presence of the desired peptide, can overcome the stability requirement for the intermediate "peptide-receptive" state (Figure 4.7A). We prepared S33-pCSCT and S46-pCSCT using the A02:01 MART1-CSCT template via both one-step and two-step strategies and tested them on TCR008 and TCR009-expressing Jurkat cells, respectively. When tested against cognate TCR-expressing Jurkat cells, both one-step and two-step pCSCT tetramers showed comparable cell frequencies and staining intensities to SCT, indicating that both strategies are effective for antigen-specific cell staining (Figures 4.7B and 4.7C). For HLA-B*07:02 with the template peptide RHSVVVPCEP (P53-

Y220C), one-step pCSCT showed effective cell capture, while the two-step method did not (data not shown). Thus, we proved that both strategies are effective when the template peptide can stabilize the "peptide-receptive" CSCT, while the one-step strategy showed the benefit of easier and faster preparation, in addition to eliminating the requirement for an optimal template peptide.

4.3.6 *Sensitive virus-specific CD8 T cell detection from donor PBMCs*

Based on the excellent performance of pCSCTs demonstrated in engineered TCR-expressing Jurkat cell lines with known peptide-TCR pairs, we next investigated the feasibility of using pCSCT for identifying antigen-specific CD8 T cells from donor peripheral blood mononuclear cells (PBMCs). We prepared S13, S33, S46, and CMV (NLVPMVATV)-pCSCT tetramers and tested them against freshly thawed donor PBMCs. Antigen-specific CD8 T cell frequencies (gated from live, CD3+CD8+ cells) were determined by dual-colored pCSCT tetramers. We detected antigen-specific T cell responses against S33 from 3 out of 9 COVID-19 donors, with frequencies ranging from 0.097% to 0.81% of total CD8 T cells (Figure 4.8A, top row). CMV-specific CD8 T cells were detected from 4 out of 9 COVID-19 donors, with frequencies ranging from 0.38% to 5.75% of total CD8 T cells (Figure 4.8A, bottom row). Other antigen specificities were not detected for the tested donors. Our results indicate the potential of using pCSCT tetramers for the direct screening of antigen-specific responses from clinical samples, such as PBMCs.

4.3.7 *Highly specific neoantigen-specific T cell detection using pCSCT compared to refolded-pMHC and SCT*

So far, we have demonstrated the use of the peptide-loaded CSCT reagent for identifying antigen-specific T cells against virus epitopes, which generally have higher binding affinities to the HLA molecule than neoantigens. To explore the application of these reagents in cancer-specific TCR

discovery, we investigated the binding performance of neoantigen-loaded CSCT tetramers on cognate TCR cell lines. Specifically, KRAS-G12V peptides were loaded onto the A*11:01 CSCT template and tested against KRAS-G12V TCR-expressing Jurkat cells (Greenberg lab). To evaluate the specificity of pCSCT tetramers, the following peptides were tested: G12V with varying length: G12V (8-16) - VVGA**V**GVGK, G12V (7-16) - **V**VVGA**V**GVGK, and G12D (8-16) - VVGA**D**GVGK. Conventional SCT and commercial pMHC for G12V (8-16) - VVGAVGVGK were compared in parallel. Both the A*11:01 KRAS-G12V SCT and commercial pMHC tetramers showed significant non-specific binding on untransduced cells (Figure 4.9, top), consistent with our previous results[14] . In contrast, the pCSCT tetramer loaded with the same G12V (8-16) - VVGAVGVGK peptide showed highly specific binding against cognate TCRs (G12V TCR#1 and TCR#2) without non-specific binding against untransduced cells (Figure 4.9, middle). Moreover, the pCSCT tetramer loaded with G12V (7-16) - VVVGAVGVGK, containing one extra amino acid compared to G12V (8-16), did not bind to TCR#1 or TCR#2. When loaded with G12D (8-16) - VVGADGVGK, containing one altered amino acid compared to the cognate peptide, it also did not bind to both G12V-specific TCRs (Figure 4.9, bottom). Collectively, these results demonstrated the application of the pCSCT reagent in neoantigen-specific TCR discovery, with superior specificity compared to SCT and commercial pMHC in some cases.

4.3.8 *Peptide-loading efficiency and stability of G12V-pCSCT reagent*

To investigate the effective peptide exchange condition for the G12V-pCSCT reagent, we incubated the TEV-cleaved A*11:01 EBV-CSCT template with 100-fold, 50-fold, and 10-fold molar ratios of the G12V peptide. Each G12V-pCSCT formulation identified comparable frequencies of G12V-TCR expressing T cells (Figure 4.10A), indicating effective peptide-exchange efficiency for the G12V peptide. We next investigated the stability of G12V-pCSCT

reagents in the format of ready-to-use multimers or as peptide-receptive monomers. In condition A, the A*11:01 EBV-CSCT template was treated with TEV protease, stored in PBS at 4°C in the form of a "peptide-receptive" monomer for two weeks before peptide exchange, tetramerization, and cell staining. In condition B, the peptide-exchanged G12V-pCSCT tetramer was stored at 4°C for two weeks before cell staining. Both conditions showed uncompromised tetramer binding signals against the cognate TCR (Figure 4.10B). Furthermore, the TEV protease-treated, "peptide-receptive" A*11:01 EBV-CSCT template showed effective peptide exchange and T cell binding after three months of storage (Figure 4.10C). In conclusion, the A*11:01 G12V-pCSCT reagent demonstrated long-term stability for T cell detection.

4.3.9 *Comparison between CSCT reagents using different template peptide*

So far, we have shown the successful production of the G12V-pCSCT reagent using the A**11:01 EBV-CSCT template (Template #1, peptide: ATIGTAMYK). However, this EBV peptide is a strong binder of A*11*:01 with a NetMHC predicted affinity of 6.03 nM. We next investigated whether using a weaker template peptide will enhance the performance of pCSCT by facilitating the peptide-exchange efficiency. For this purpose, two more A*11*:01 CSCT templates were designed and expressed in Expi293 cells in addition to the original Template #1. The expression yields for all three templates were comparable, ranging from 1.2-1.5 mg/L (Figure 4.11A inset). Template #2 and #3 encoded two HIV peptides that showed relatively lower predicted binding affinities compared to template #1 (Figures 4.11A, B). G12V-pCSCT tetramers prepared using each CSCT template showed comparable tetramer-positive cell frequencies. However, templates #2 and #3 showed higher fluorescent intensities compared to template #1, indicating a better outcome of peptide exchange (Figure 4.11C). Based on these results, we decided to use #3 as the optimal A*11*:01 CSCT template.

4.3.10 *Extension of CSCT reagents for multiple HLA alleles*

The CSCT template design is highly modular, as the optimal template peptide, cleavage site, and HLA subunits can be easily substituted to cover a wide range of HLA alleles. Here, we demonstrated the feasibility of building CSCT reagents for a total of nine common Class I A, B, C alleles, including A*02:01, A*11:01, A*24:02, A*01:01, A*03:01, B*07:02, B*08:01, B*35:01, and C*07:02. The choice of template peptide was determined based on the following criteria: (1) The expression of corresponding SCT reagents. We have observed a similar expression yield for SCT and CSCT when the same peptide was encoded. A higher expression yield is preferred. (2) The NetMHC prediction of the template peptide, which was used to estimate the stability of the "peptide-receptive", TEV-cleaved CSCT in the absence of the rescue (target) peptide. A low-affinity peptide might be helpful for peptide exchange but could also compromise the stability of the "peptide-receptive" CSCT product when using the previously mentioned "two-step" method. (3) The template peptide's antigen specificity, which is associated with the background signal from uncleaved or unexchanged template CSCT. When performing flow cytometry analysis using the peptide-loaded CSCT (pCSCT) reagent, we simultaneously stained with a tetramer using the CSCT template as a background control to gate out the "false-positive" tetramer signal from template CSCT tetramers in case these T cells exist in the sample. However, we still preferred to use a less likely detected peptide antigen in the CSCT template to address "false-positive" concerns. Generally, a cancer neoantigen epitope or HIV epitope is more ideal than "MART-1", which is normally less likely to be detected in healthy donors, and more ideal than common virus epitopes (such as CMV or flu peptides). The protein expression yields for the constructed templates were listed in Table 4.2. In summary, we successfully expressed CSCT templates for eight HLA alleles,

enabling high-throughput screening of antigen-specific CD8 T cell responses across a wide range of populations.

HLA_template	Template peptide	antigen	Yield (μM)
A0101_TEV_Site-A	TDLGQNLLY	Human adenovirus C	6.03
	ATDALMTGF	Hepatitis C virus	3.06
	GDTDSRFISY	TESLA_c4_neoantigen	4.18
A0201_TEV_Site-A	HMTEVVRHCPH	p53_R175H	3.10
	GSMNRRPILT	p53_G245S	2.59
	NSFEVHVCA	p53_R273H	1.79
	HMTEVVRHC	p53_R175H	3.48
	YMCNSSCMGSM	p53_G245S	6.14
	ALQIPFAMQM	Sras-CoV-2_S	2.91
	FVFPLNSII	Sras-CoV-2_ORF1ab	9.83
	TLIVNSVLL	Sras-CoV-2_E	5.71
	MVWESGCTV	T1D-related (IGRP)	11.94
	SLLMWITQV	NYESO	7.46
	ELAGIGILTV	MART1	25.82
A0301_TEV_Site-A	VVVGAVGVGKS	KRAS_G12V	1.63
A1101_TEV_Site-A	QVPLRPMTYK	HIV-nef protein	6.47
	AIFQSSMTKIL	HIV-pol protein	9.86
	ATIGTAMYK	Epstein–Barr virus	16.43
	IVTDFSVIK	Epstein–Barr virus	12.30
	AVFDRKSDAK	Epstein–Barr virus	18.36
A2402_TEV_Site-A	KITEQEKDFLW	PIK3CA_E245K	14.10
B0702_TEV_Site-A	RHSVVVPCEP	P53_Y220C	1.13
B3501_TEV_Site-A	GAVGVGKSAL	KRAS12_G12V	2.45
	LPYPDPSRI	Sras-CoV-2_ORF1ab	7.99
C0702_TEV_Site-A	LTDEMIAQY	Sras-CoV-2_S	1.73
	SLRPDTRYVL	Sras-CoV-2_NSP4	2.45
	SYRSTYMIL	Cytomegalovirus	2.63

Table 4.2 Template CSCT expression yield in Expi-293 cells.

All templates listed in this table are using A_Y84C_A139C template of corresponding HLA, and include TEV protease recognition sequence at site-A. Yield represent the calculated CSCT concentration after His-Tag purification based on the protein A280 concentration (as measured by NanoDrop 8000 Spectrophotometer) when eluted in 60ul PBS. When multiple transfection experiments were performed for the same plasmid, the average yield from all experiments was shown in the table.

HLA	SCT Plasmid ID	Antigen	Peptide
A1101	210930_A1101_A_Y84C_A139C_EBV1-TEV_ATIGTAMYK	EBV	ATIGTAMYK
A1101	210930_A1101_A_Y84C_A139C_EBV3-TEV_AVFDRKSDAK	EBV	AVFDRKSDAK
A1101	210930_A1101_A_Y84C_A139C_EBV2-TEV_IVTDFSVIK	EBV	IVTDFSVIK
A0201	200526_A0201_A_Y84C_A139C_CoVprot_417_MLFTMLRKL	SARS-CoV-2	MLFTMLRKL
A0201	200526_A0201_A_Y84C_A139C_CoVprot_842_SIWNLDYII	SARS-CoV-2	SIWNLDYII
A0201	200526_A0201_A_Y84C_A139C_CoVprot_576_MLSDTLKNL	SARS-CoV-2	MLSDTLKNL
A0301	comPACT66527	E6_HPV16	KFYSKISEY
A0301	comPACT66671	E6_HPV16	SLYGTTLEQ
A1101	comPACT67189	E6_HPV16	KFYSKISEYR

Table 4.3 List of example SCTs that have shown non-specific binding ("sticky" SCTs) towards untransduced T cell line or primary CD8 T cells.

HLA	antigen	peptide	NetMHC (nM)	SCT expression
A0201	Flu-M1	GILGFVFTL	7.27	Minimum/No
A0201	EBV-LMP2A	CLGGLLTMV	75.6	Minimum/No
A0201	EBV-BLMF1	GLCTLVAML	62.23	Minimum/No

Table 4.4 List of example SCTs encoding well-established immunogenic epitopes, but are not expressed in yield

4.4 Conclusion and Discussion

In this chapter, we introduced peptide-loaded cleavable SCT (pCSCT) reagents as an easy-to-use tool for efficient T cell detection. We demonstrated their potential in the identification of antigen-specific T cells against high-affinity virus antigen and low-affinity cancer-specific neoantigens, from both engineered T cell lines and patient-derived peripheral blood mononuclear cells (PBMCs). We optimized the peptide-loaded CSCT experimental conditions, which allow efficient peptide exchange to generate ready-to-use peptide-pMHC monomers in a one-step reaction. These pCSCT tetramers showed comparable cell staining efficiency and sensitivity to conventional SCT

tetramers in detecting low-frequency antigen-specific T cells. One limitation of traditional peptide-MHC-like reagents is their stability and limited shelf-life. Here, the pCSCT reagent remained effective after months of storage, either in the form of ready-to-use, peptide-loaded pCSCT tetramers or as a "peptide-receptive" CSCT monomer, allowing flexible future loading of the desired peptide. Overall, taking advantage of the conventional disulfide-stabilized SCT's high yield and stability, as well as the flexibility of peptide exchange, pCSCT reagents can be easily converted into large libraries of pCSCT multimers representing a wide range of antigen specificities.

The choice of template peptide is an important factor for the success of the pCSCT reagent, highly related to the CSCT template expression yield and peptide-exchange efficiency. Factors to consider when choosing the CSCT template include: the expression yield of the CSCT template, the possibility of detecting a T cell population against the template CSCT, and the peptide-HLA binding affinity. For HLA-A*02:01, we have built and compared multiple template peptides with a wide range of predicted affinities, including MART1, P53 neoantigens, and viral antigens (Table 4.2). When chosen properly, the uncleaved CSCT can serve as a negative control. For example, the MART1 antigen is normally not detected in PBMCs, although it can exist in a small population at low frequencies. The mutant P53 antigen (such as R175H) would be an even more ideal template since its T cell population is rarely detected. Regarding peptide-HLA binding affinity, we have shown that using a strong binder as the template will increase the stability of the CSCT in the peptide-receptive state since the template peptide can serve to stabilize the pMHC structure. However, for A*11:01 and A*02:01, we have data showing that the peptide exchange efficiency is higher when the target peptide is competing with a weaker placeholder. To address the possible trade-off between stability and exchange efficiency, especially when using a weak peptide

template, we introduced the one-step pCSCT method, where TEV cleavage and peptide exchange occur in the same reaction, skipping the intermediate "peptide-receptive" product. This allows the high concentration of the desired peptide to directly stabilize the structure and form the pCSCT directly.

Temperature-induced peptide exchange has been reported to generate stable peptide-MHC monomers for both human HLA: A*02:01 and mouse H-2K^{b7}. The optimal temperature varies depending on the choice of template peptide and desired peptide, possibly related to the peptide-MHC's thermal stability. In this chapter, we applied a universal temperature (32°C) for peptide exchange, regardless of the template peptide and HLA, to help accelerate the library-approach. Using a relatively mild temperature of 32°C for 1 hour might accelerate the peptide-exchange process without compromising the performance of the pCSCT. For future work, the exchange temperature can be further optimized to accommodate individual CSCT HLA templates. In addition, extending the peptide exchange reaction at 4°C overnight allows a more complete peptide exchange while lowering the rate of unfolding or degradation under thermal changes.

SCT reagents have been a powerful and versatile tool for the detection of antigen-specific T cells. However, there are some limitations to the application of SCT reagents: (a) The SCT template needs to be optimized for individual HLAs for optimal protein expression. In this case, the pCSCT has the benefit of overcoming uneven expression yields of individual antigens by using the universal, highly-expressed CSCT template for each HLA library. (b) In certain cases, the SCT version of the pMHC multimer was found to exhibit non-specific binding to non-cognate TCRs (such as the A*11:01 G12V SCT and the ones listed in Table 4.3). Such non-specific binding is not well-understood yet, but it might be related to HLA modification or structural changes due to

the covalent linkage between pMHC compartments. In this case, SCT reagents provide an excellent alternative for sensitive and specific detection. (c) We have shown that the SCT expression yield does not necessarily reflect the peptide-MHC binding affinity, which means that some low-affinity epitopes will generate high yields of SCT expression, even when the epitope is unlikely to be naturally presented by the HLA. On the other hand, some well-established immunogenic epitopes are not expressible in the form of SCT (Table 4.4). Instead, the peptide loading process for the pCSCT is highly dependent on the binding affinity between the peptide and HLA, which might be beneficial for the discovery of immunogenic epitopes more likely to occur in natural antigen presentation. These reagents might also be an effective tool for understanding antigen-HLA binding. (d) In some cases, a list of target peptides (neoantigens generated from patients or peptides from target viruses) is already being synthesized for peptide-based stimulation assays. The flexible peptide-receptive CSCT reagent would be beneficial since the peptides can be immediately used to prepare the pMHC tetramer in one day with only a thermal cycler (or incubator) needed, reducing the cost and time of preparing SCT libraries, which usually takes 3 weeks depending on the size of the library.

In summary, we herein presented the peptide-loaded cleavable SCT (pCSCT) reagent as an alternative to the conventional SCT. It is highly efficient and more easily accessible. We anticipate that pCSCT reagents can be broadly applied for the high-throughput detection of antigen-specific T cell responses across many patients in parallel and also facilitate the discovery and validation process of cancer-specific TCRs for adoptive cell therapy.

4.5 METHODS

4.5.1 *Cloning of cleavable-SCT (CSCT) template plasmids*

Class I SCT-encoded plasmids were constructed in pcDNA3.1 Zeo(+) plasmid (Thermo Fisher Scientific) as previously reported[1]. Briefly, for each HLA allele, disulfide-stabilized SCT plasmids were cloned by assembling: the IFNα2 signal sequence, the template peptide, L1 (constraining three glycine-serine linker sequence [GGGGS]), β2M, L2, HLA, His-tag, BirA biotin ligase recognition site. Additionally, we included Y84C and A139C mutations for each HLA to form a disulfide bond between two subunits. For A*02:01 MART1-CSCT template plasmids, a TEV protease recognition sequence was included into L1 with four designs:

SCT: [GGGGS] [GGGGS] [GGGGS]

CSCT Site-A: ENLYFQ[GGGGS] [GGGGS] [GGGGS]

CSCT Site-B: [GGGGS] ENLYFQ [GGGGS] [GGGGS]

CSCT Site-C: [GGGGS] [GGGGS] ENLYFQ [GGGGS]

CSCT Site-D: [GGGGS] [GGGGS] [GGGGS] ENLYFQ

Extension PCR was conducted with KOD Hot Start polymerase (MilliporeSigma) using primers listed in Table 4.3. For other CSCT templates with different template peptide or HLA alleles, only Site-A design was built based on its better performance. The PCR product was phosphorylated and ligated using T4 Polynucleotide Kinase and T4 DNA Ligase (New England Biolabs) before digest with DpnI (New England Biolabs) to remove excess PCR template. The TEV-sequence inserted plasmids were then transformed into One-Shot TOP10 Chemically Competent E. coli (Thermo Fisher Scientific). Plasmids were verified by Sanger sequencing for Exp393 transfection following manufacture protocol. The supernatant of the transfection solution was filtered through 0.22 μm filter 96-well plates (AcroPrep 96 filter places, 1 mL - 0.2 μm, PTFE membrane) and then

concentrated for overnight biotinylation with BirA enzyme kit (Avidity). The biotinylated CSCTs were then purified with HisTag resin tips (Phynexus) and stored in PBS containing 20% glycerol and stored at −20 °C.

primer	sequence
TevA_for	ggtGGAggaGGTtctggag
TevB_ for	ggaggtggtggaagcggtg
TevC_ for	ggtggtggcgggagcatccagagaa
TevD_ for	agcATCcagAGAaccCCCA
TevA_rev	ttgaaagtacaggttctcCACGGTCAGGATGCCGATG
TevB_ rev	ttggaaatacaaattctcagaACCtccTCCaccCACG
TevC_ rev	ttgaaaatacaggttctcgcttccaccacctccagaA
TevD_ rev	ttgaaaatacaggttctccccgccaccaccgcttcca

Table 4.3 PCR primers for CSCT template design.

4.5.2 *Preparation of peptide-loaded CSCT tetramers*

Peptides were purchased from GenScript with >85% purification and dissolved in DMSO before use. For two-step pCSCT preparation: biotinylated, purified CSCT was treated with TEV protease (New England Biolabs) for one hour at 30°C in the presence of TEV buffer following manufacture protocol (followed by an optional 4°C overnight). The reaction was ultra-centrifuged using a MWCO of 30 kDa to remove TEV protease and resuspended into PBS to form “peptide-receptive” CSCT monomer, and stored in 4°C before use. Desired peptide and cleaved CSCT monomer were then mixed at a molar ratio of 100:1 (unless other ratios used for experimental optimization) in PBS and incubated at 32ºC for 30 minutes (followed by an optional 4°C overnight) to form peptide-loaded CSCT (pCSCT) monomer. For one-step pCSCT preparation: biotinylated, purified CSCT was mixed with TEV protease, TEV buffer, desired peptide (peptide to CSCT molar ratio of 100:1)

for one hour at 32°C (followed by an optional 4°C overnight) to directly form pCSCT monomer. The reaction was then ultra-centrifuged using a MWCO of 30 kDa to remove TEV protease and resuspended into PBS to form pCSCT. For both two-step and one-step generated pCSCT monomer, tetramers were prepared individually. Specifically, a 5:1 ratio between monomer and fluorophore-labeled streptavidin was used to produce dual-color labeled pCSCT-tetramer-PE and pCSCT-tetramer-APC. Untreated CSCT monomer was tetramerized to generate CSCT-tetramer-PEcy7 as a background control. All tetramers were blocked with biotin for free binding sites before cell staining, and stored in 4C before use.

4.5.3 *TCR-transduced cell lines*

TCR001 (S13-specific) and TCR079 (S46-specific) were previously reported[1]. TCR008 was identified from the work described in Chapter 2 via DNA-barcoded dextramer. Lentiviral vectors encoding TCR α and β chain were designed as previously described[1]. Briefly, codon-optimized DNA fragments encoding TCR α and β chains were PCR amplified, Gibson assembled and ligated into pRRL-SIN Lentiviral vector (obtained from Dr. Philip D. Greenberg). TCR knock-out Jurkat cells were transduced with lentiviral particles encoding the TCR of interest. Briefly, lentiviral vectors were produced by the transient transfection of HEK 293T cells with the lentiviral constructs above and packaging DNA (mixture of pMD2-G VSVG, pMDL g/pRRE, pRSV-REV at a 1:2:2 ratio) at a 5:2 ratio using Mirus TransIT LT1 (Mirus Bio). At 48- and 72-hours post-transfection, viral supernatant was collected and filtered through 0.45 uM PES filters and concentrated using Lenti-X concentrator (TAKARA). The TCRαβ negative Jurkat76 cell line were infected with concentrated lentiviral particles with 2-4 ug/ml polyprene (Sigma-Aldrich) overnight. On Day 5 of transduction, TCR-transduced Jurkat cells were assessed for transduction efficiency

and antigen-specificity via flow cytometry assay described below, and enriched via FACS sorting if needed. Cells were maintained every 2-3 days with R10 media.

4.5.4 *Cell staining and flow cytometry for TCR-transduced Jurkat cell lines*

Enriched, high purity TCR-expressing Jurkat cells were stained with freshly mixed template CSCT-tetramer (PE-cy7) and peptide-loaded pCSCT-tetramers (PE and APC) for 20 minutes at 4°C. Cells were washed three times and stained with a cocktail of antibodies including Apotracker Green (Biolegend), BV421 anti-human CD3 (Biolegend, 344704). Samples were incubated for 15 minutes at 4°C and washed 3 times before flow cytometry (Attune). Dual-color labeled pCSCT-tetramer-positive Jurkat cells were gated from single, live, CD3-positive, template CSCT-tetramer (PE-cy7) negative population. For detection sensitivity assay, cells were prepared by spiking in desired ratios of cognate TCR-expressing cells into P53-specific Jurkat cells that expressing an irrelevant TCR.

4.5.5 *Cell staining and flow cytometry for donor PBMCs*

PBMCs from HLA-A*02:01 COVID-19 patients were freshly thawed and incubated with Human TruStain FcX blocking reagent (422302, BioLegend) for 10 min at 4 °C. Cells were then aliquoted into 96-well plate to simultaneously screen each antigen-specificity per well. Each well was stained with 2ul of each template CSCT-tetramer (PE-cy7) and peptide-loaded pCSCT-tetramers (PE and APC) in 50 ul of SE buffer for 20 minutes at 4°C. Cells were washed three times before surface antibody staining including BV421 anti-human CD3 Antibody (BioLegend, 300434, clone UCHT1), BV605 anti-human CD8 Antibody (BioLegend, 344704, clone SK1) and Apotracker™ Green viability dye (Biolegend, 427403). Samples were incubated for 15 minutes at 4°C and washed 3 times before flow cytometry (Attune). Dual-color labeled pCSCT-tetramer-positive

Jurkat cells were gated from lymphocytes, single, live, CD3/CD8-positive, template CSCT-tetramer (PE-cy7) negative population.

4.5.6 *Statistics and data analysis*

Data and statistical analyses were done in GraphPad Prism. Data were presented as the mean ± standard error of mean (SD) of two or three experimental replicates unless otherwise indicated within figure legends. A p value of <0.05 denoted statistically significant differences.

4.6 FIGURES

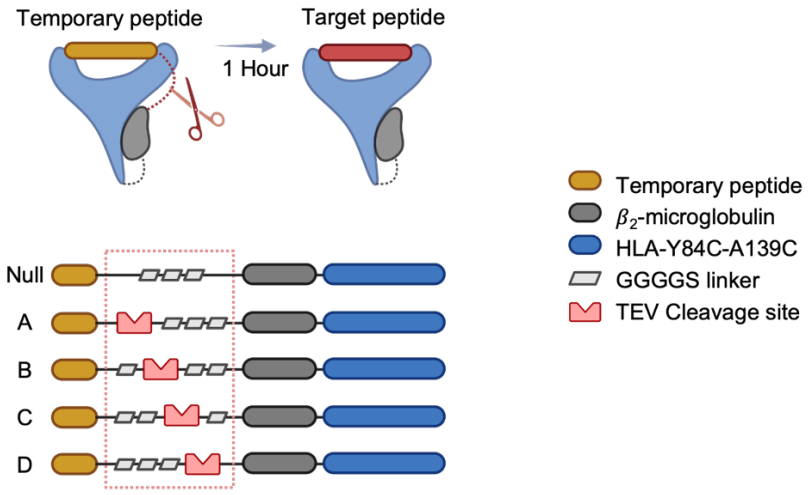


Figure 4.1 Peptide-receptive cleavable SCT

Top: Schematic illustration of pMHC monomer preparation using of peptide-receptive cleavable SCT. CSCT template is treated with TEV protease to remove the linker between temporary peptide and β2M. peptide-loaded CSCT (pCSCT) is produced by peptide exchange in the presence of target peptide.

Bottom: Designs of cleavable SCT (CSCT). CSCT template was modified with a TEV protease cleavage site within the linker between the temporary peptide and β2M. Four insertion sites (namely site-A, B, C, D) were constructed for comparison.

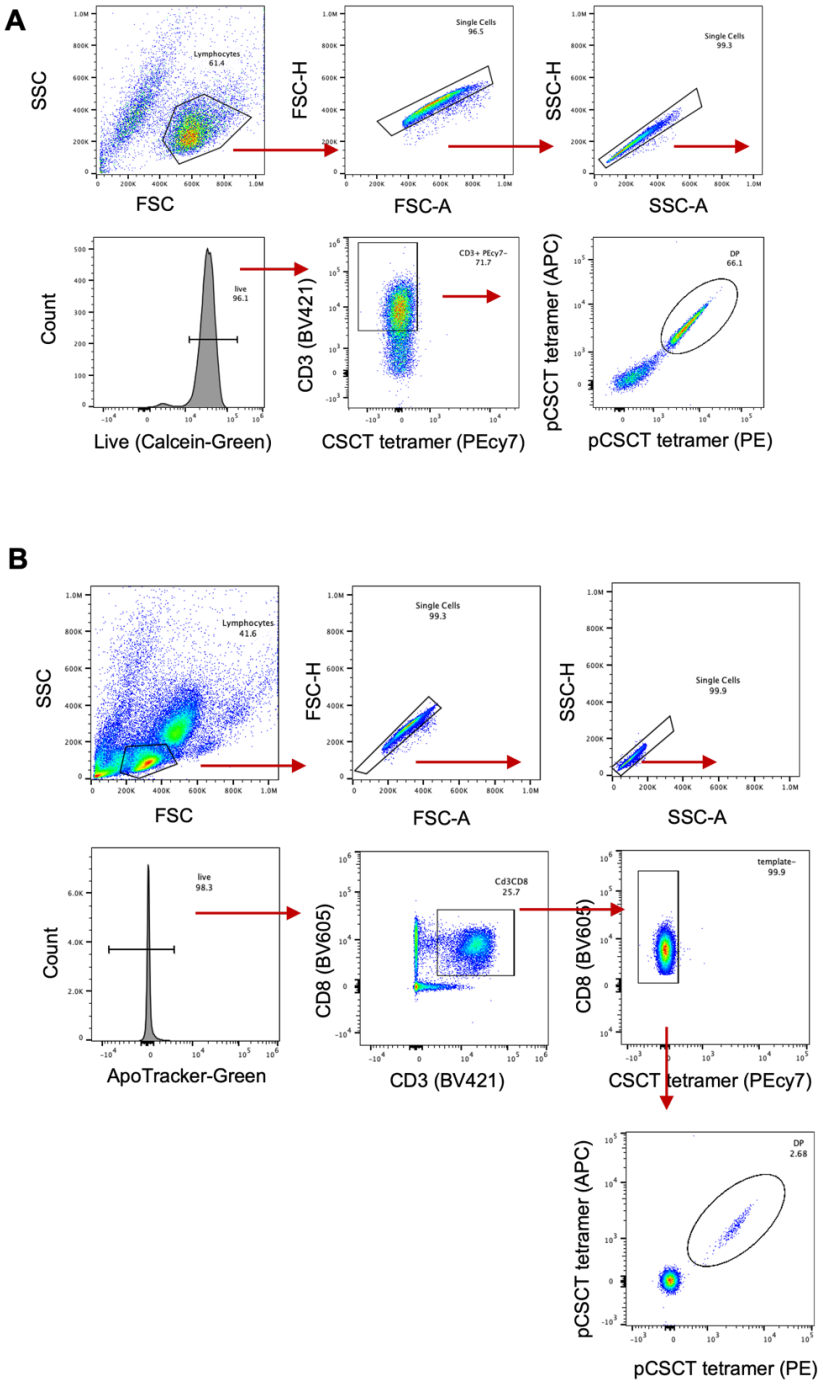


Figure 4.S1 Flow cytometry gating strategies using peptide-loaded cleavable SCT (pCSCT)

A. Gating example for TCR-cloned Jurkat cells

B. Gating example for patient PBMCs

A

Antigen description				TCR features				
Antigen ID	Source	Peptide	Predicted affinity (nM)	TCR ID	TRBV	CDR3b	TRAV	CDR3a
S13	SARS-CoV-2	LLFNKVTLA	12.82	TCR001	TRBV28	CASSLTGGGYEQYF	TRAV9-2	CALNDYKLSF
S33	SARS-CoV-2	YLQPRTFLLK	108.8	TCR008	TRBV7-9	CASSPDIEQYF	TRAV12-2	CAVNQDDKIIF
S46	SARS-CoV-2	RLITGRLQSL	163.63	TCR079	TRBV20-1	CSGEGGDTEAFF	TRAV19	CALSEGYNFNKFYF

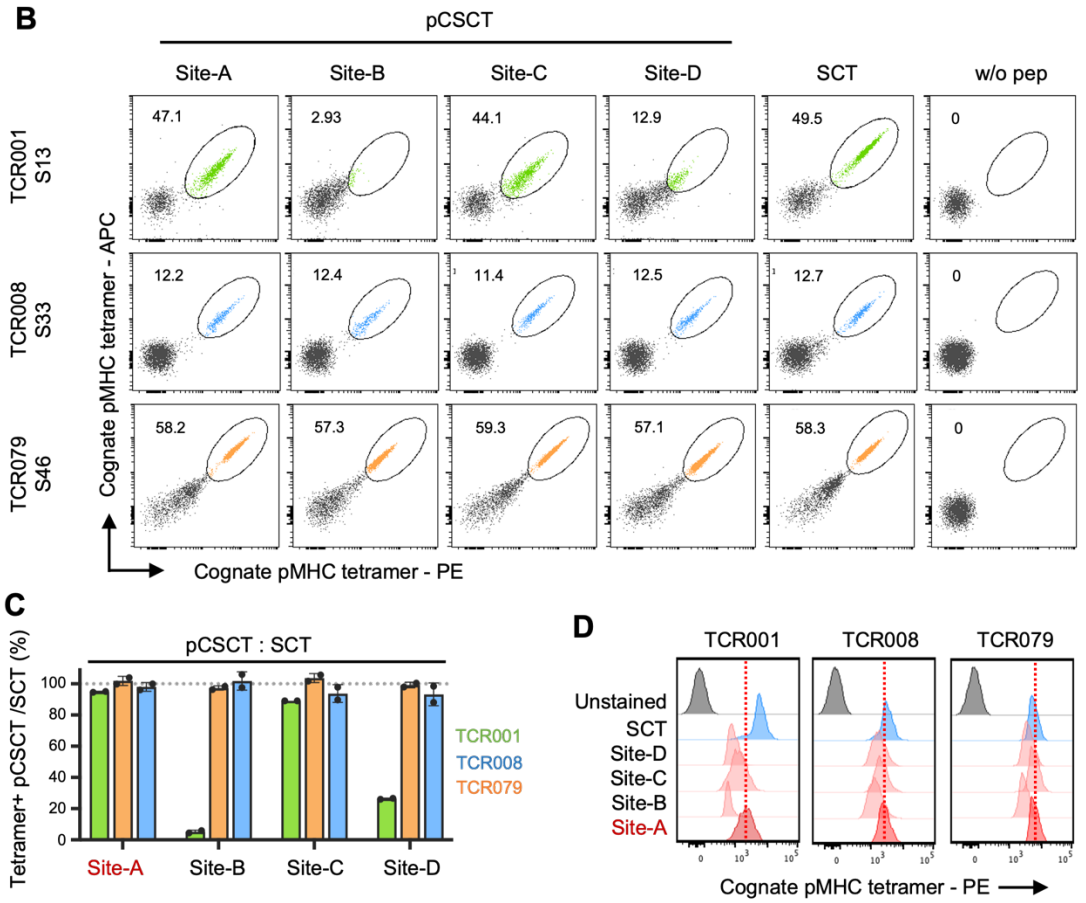


Figure 4.2 Optimization of the pCSCT design

A. List of antigen-TCR pairs for pCSCT validation in Jurkat cell lines.

B. Comparison between four CSCT templates using different TEV cleavage site. HLA A*02:01 MART1-CSCT monomers were exchanged with three SARS-CoV-2 Spike peptides (S13, S33, S46), and then prepared into fluorophore-labeled tetramers (PE and APC) to detect antigen-specific Jurkat cells. For each row, S13, S33, S46 -specific TCR Jurkat cells were mixed with

irrelevant TCR-expressing cells, and stained with cognate pCSCT and SCT tetramers. Double-positive cells were gated and antigen-specific percentage was shown for each panel.

C. Bar plot showing the ratio of tetramer-positive frequencies between each pCSCT template and SCT for each peptide-TCR pairs. Legend on the right. Dash line indicates when pCSCT and SCT captured equal percentage of antigen-specific cells.

D. Histograms of tetramer-PE signal for each antigen-TCR pairs (columns) and each CSCT templates (rows).

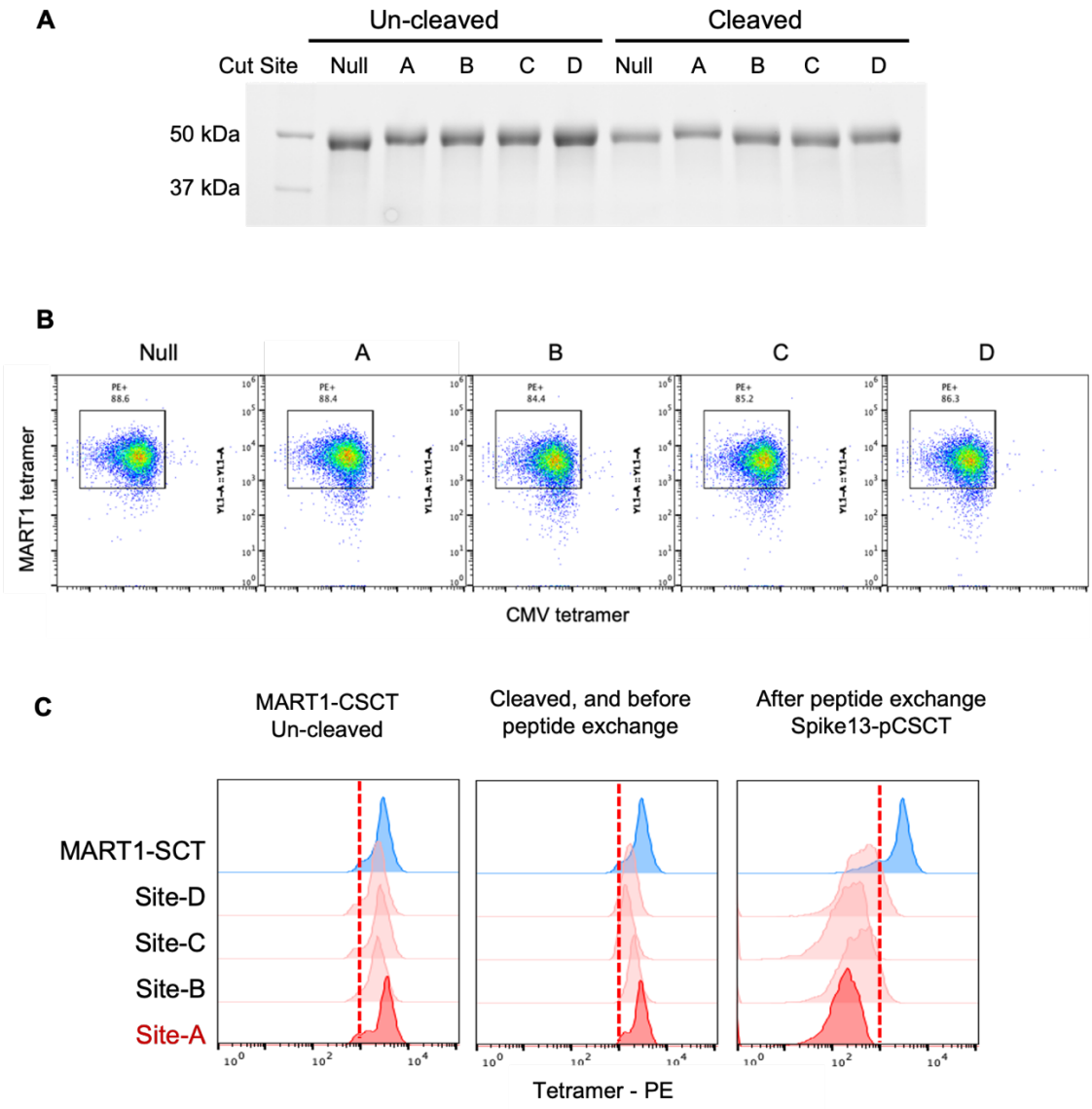


Figure 4.S2 Characterization of A*02:01 MART1-CSCT templates

A. Reduced SDS-PAGE gel for A*02:01 MART1 CSCT templates before and after TEV protease cleavage. Null, A, B, C, D represent SCT design, and four CSCT designs, respectively.

B. Flow cytometry test of pMHC tetramer binding assay. A*02:01 MART1-specific Jurkat cells were stained with conventional MART1-SCT or MART1-CSCT tetramers. Y-axis: MART1 tetramer of version Null, A, B, C, or D, respectively for each column. X-axis: Irrelevant CMV tetramer.

C. Overlaid Tetramer signal for SCT and each CSCT designs when staining MART1-specific Jurkat cells. From left to right: step-wise illustration of MART1-CSCT before TEV cleavage (Left), “peptide-receptive” state of the CSCT after TEV cleavage, and before loading with target peptide (Middle), “peptide-exchanged” state of the CSCT where target peptide (Spike13) substituted the original MART1 peptide, and formed peptide-exchanged CSCT (pCSCT).

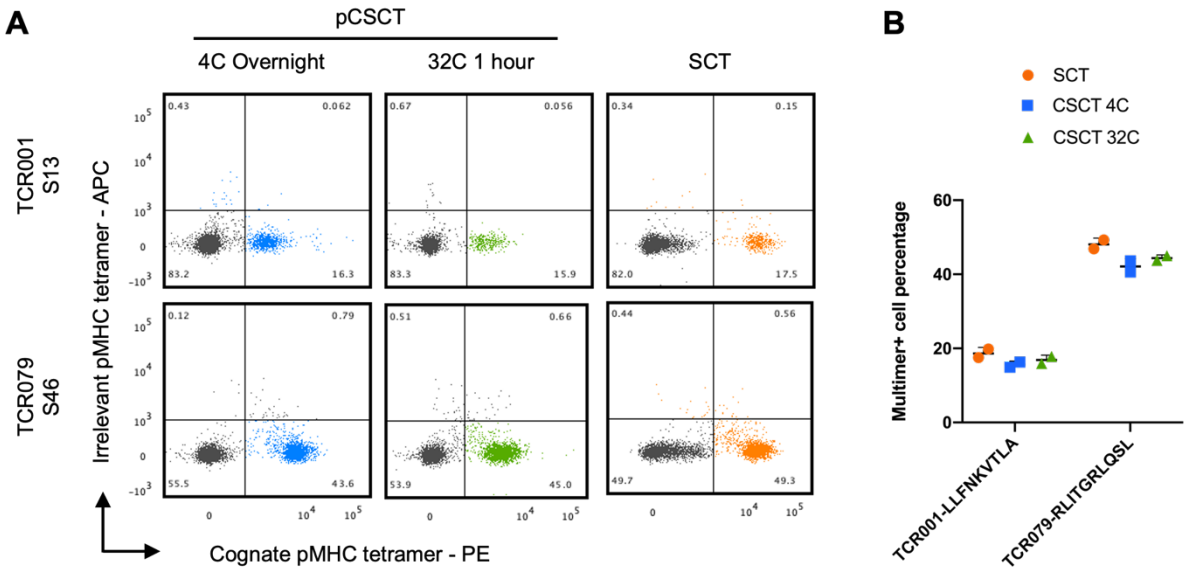


Figure 4.3 Temperature-accelerated peptide exchange for the generation of peptide-loaded CSCTs

A. Representative flow cytometry plots showing the binding of the S13-pCSCT and S46-pCSCT tetramers to their cognate TCR-expressing Jurkat cell line. Each pCSCT was prepared under two reaction conditions for peptide exchange (4°C Overnight or 32°C for one hour) as indicated, and compared with the conventional SCT tetramer. X-axis: cognate pMHC tetramer (PE) using pCSCT or SCT monomer as indicated for each column. Y-axis: Irrelevant pMHC tetramer (APC). B. Tetramer-positive cell frequencies for SCT tetramer, and pCSCT tetramers prepared under two conditions. Mean and SD of two independent measurements were plotted.

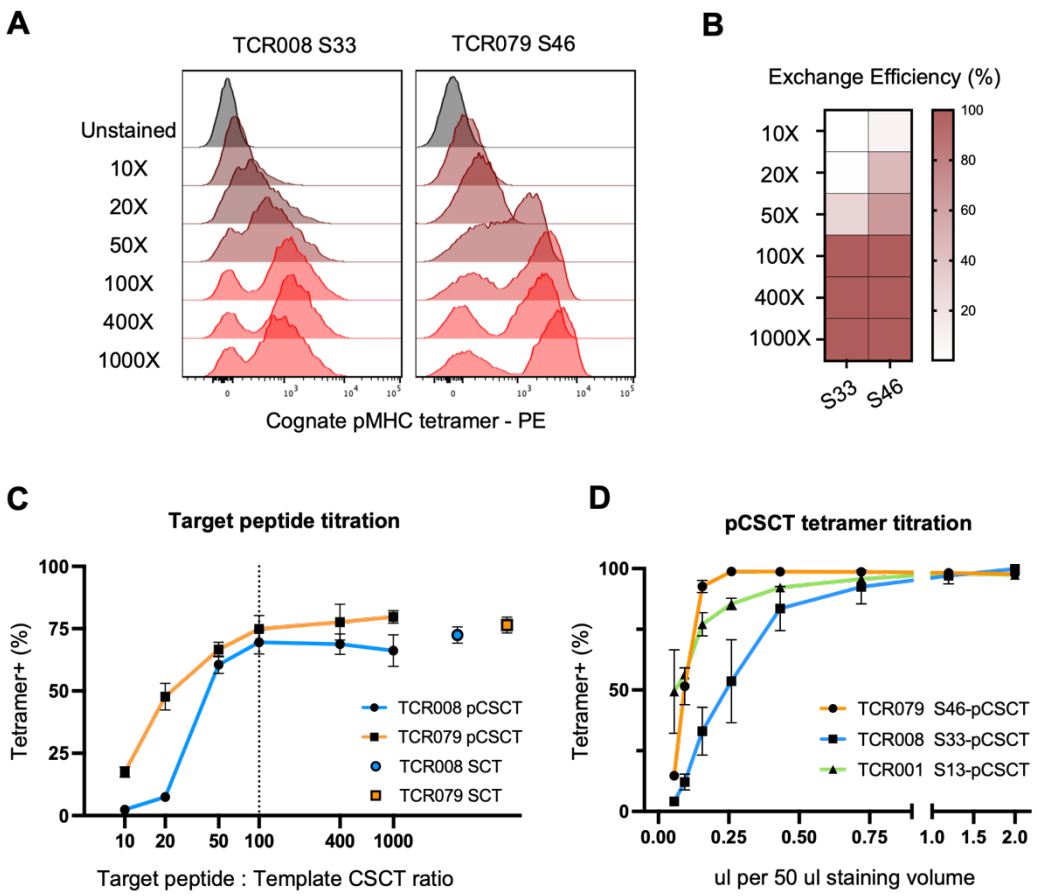


Figure 4.4 Optimization of peptide exchange and pCSCT-tetramer staining protocol

A. Overlaid tetramer signal using pCSCT monomers prepared with titrated amount of S33 or S46 peptide during peptide exchange. A*02:01 MART1-CSCT template was used for all reactions. X-axis: pCSCT-tetramer (PE) signal. Y-Axis: peptide to CSCT molar ratio from 10 to 1000.

B. Peptide-exchange efficiency determined by fold change of MART1-specific cell binding.

C. Antigen-specific cell frequencies measured by pCSCT tetramers when titrating peptide molar ratio during peptide-exchange. Mean and SD of three independent measurements were plotted.

D. Antigen-specific cell frequencies measured by pCSCT tetramers when titrating tetramer amount during cell staining. Mean and SD of three independent measurements were plotted.

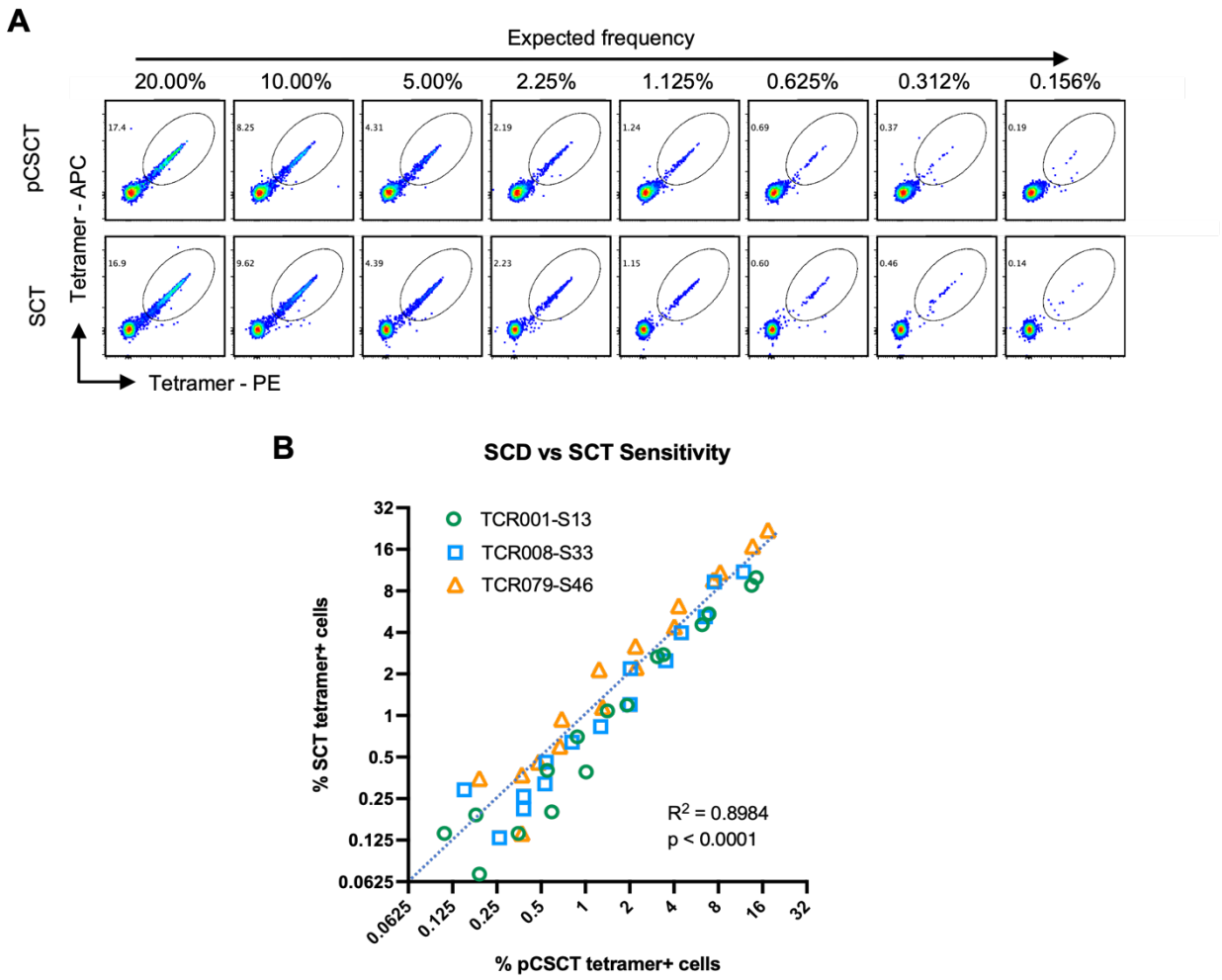


Figure 4.5 Sensitivity of pCSCT tetramer

A. Representative flow cytometry result for irrelevant TCR-expressing Jurkat cells spiked in with titrated frequencies of TCR079-expressing Jurkat cells, and stained with cognate S46-pCSCT or S46-SCT tetramers. Calculated (Expected) cell frequency for each column was shown on the top. S46-specific cells were gated by dual-fluorophores (PE and APC) and detected frequency was shown within each panel.

B. Correlation of antigen-specific T cell frequencies identified using pCSCT tetramers of conventional SCTs for TCR-expressing Jurkat cells with titrated cell frequencies.

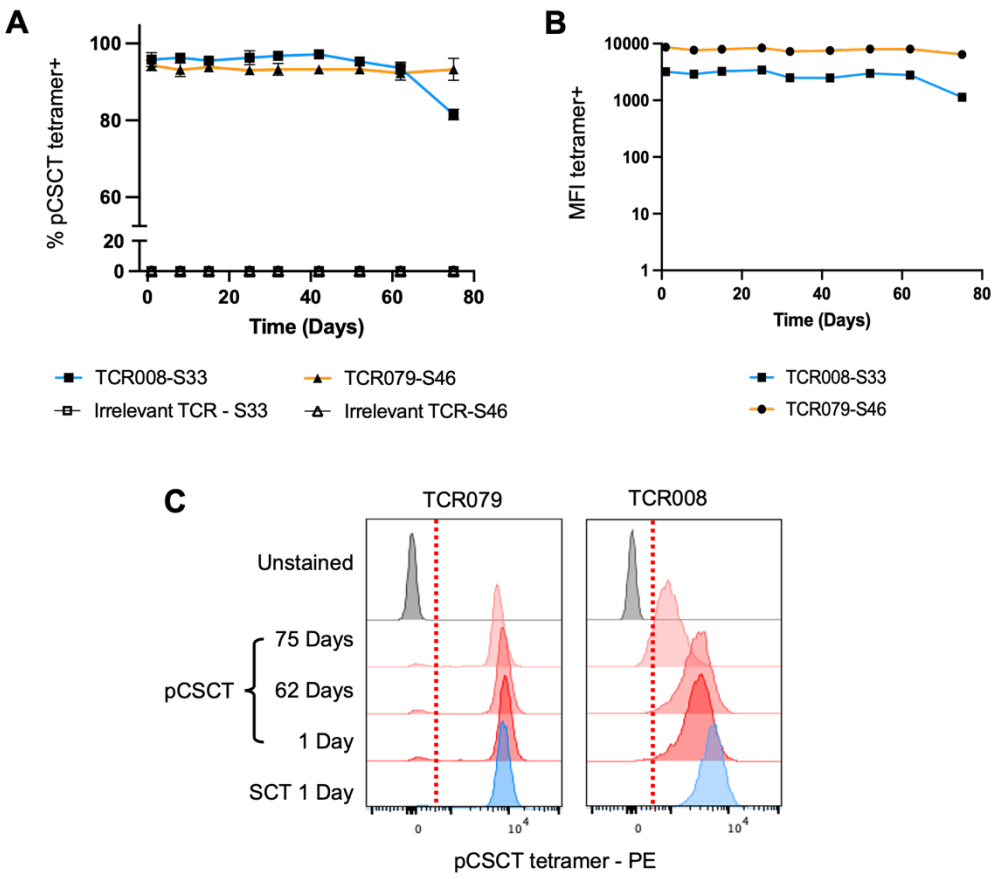


Figure 4.6 Stability of pCSCT tetramer for long-term storage

A. The frequency of detected TCR-expressing Jurkat cells using cognate S33-pCSCT or S46-pCSCT tetramers. S33-pCSCT and S46-pCSCT monomers were prepared on a 7-10-day interval, and then stored in PBS and 4°C for up to 75 days. All batches of pCSCT monomers were multimerized onto fluorochrome-labeled streptavidin on the same day and then used to stain cognate TCR-expressing Jurkat cells. Legend on the bottom.

B. Mean fluorescent intensity (MFI) of detected TCR-expressing Jurkat cells. Legend on the bottom.

C. Selected histograms of pCSCT-tetramer signal with varying storage length, and compared with freshly-prepared conventional SCT-tetramer. Red dash line indicates the threshold for positive gating.

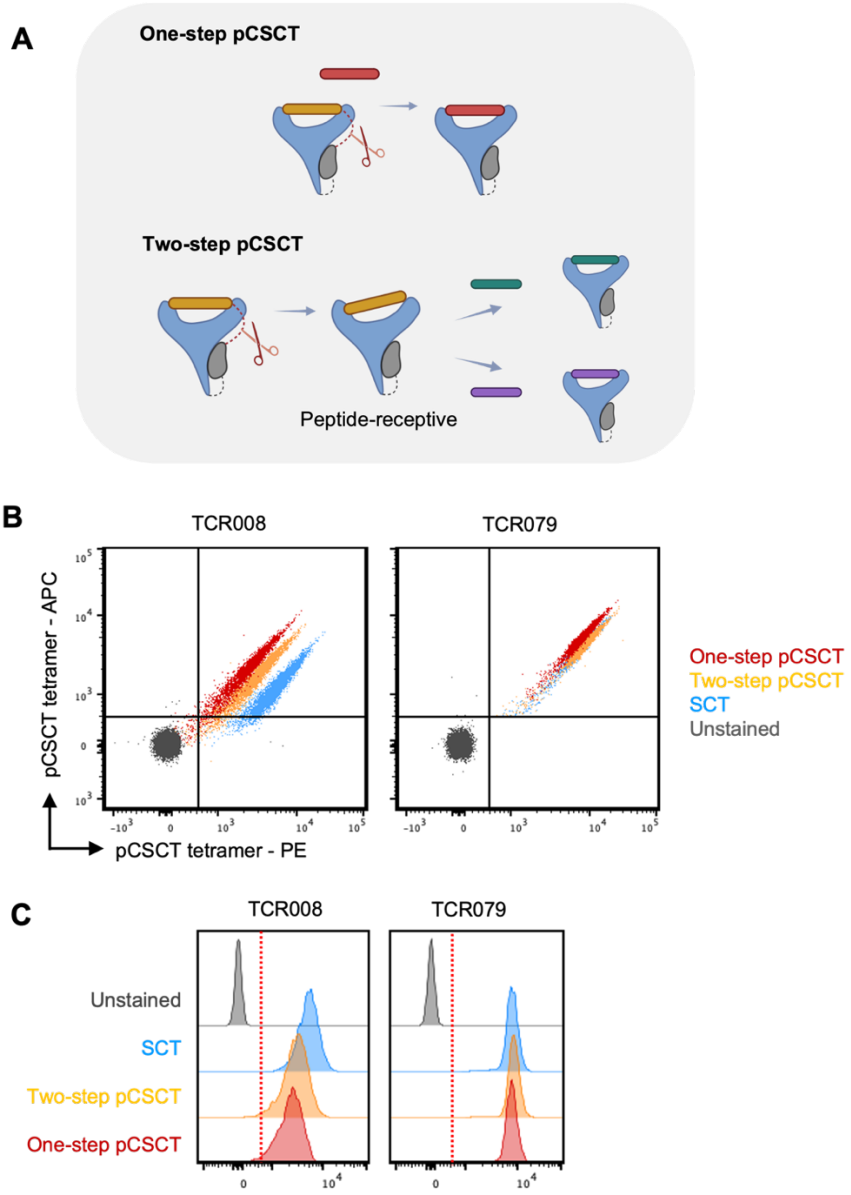


Figure 4.7 Two strategies of pCSCT preparation enable flexible application

A. Schematic illustration of one-step and two-step pCSCT strategies. For one-step pCSCT, TEV-cleavage of template CSCT happens in the presence of the desired peptide, while two-step pCSCT generated "peptide-receptive" state before peptide-exchange.

B. Representative flow cytometry plots showing the overlay of one-step and two-step pCSCT-tetramers, comparing with conventional SCT-tetramer. Legend on the left.

C. Histograms of tetramer signal. Red dash line indicates the threshold for positive gating.

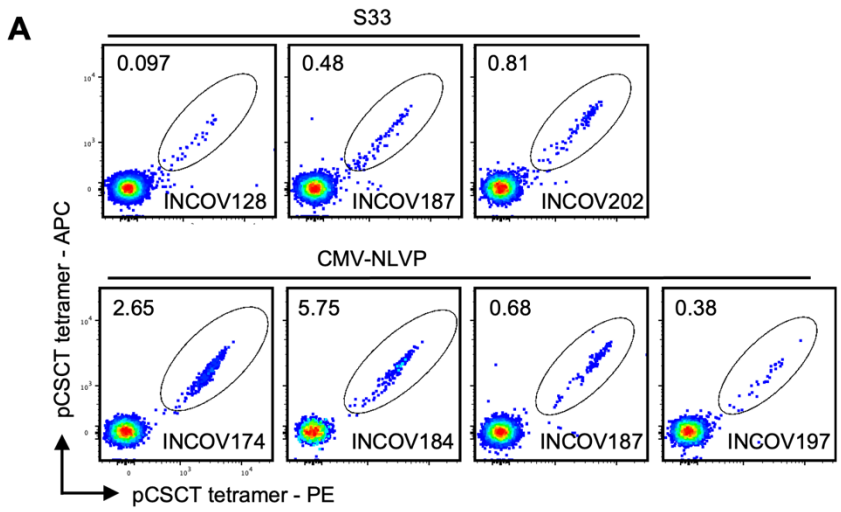


Figure 4.8 Sensitive detection of virus-specific CD8 T cells from donor PBMCs

A. Representative flow cytometry plots of A*02:01 S33-pCSCT and CMV-pCSCT tetramer staining from freshly thawed PBMCs from COVID-19 donors. Cells were gated from lymphocytes/live/single cell/$CD3^{+}CD8^{+}$ population. Antigen-specific CD8 T cells were gated on double-positive pCSCT-tetramer staining with frequencies shown in each panel.

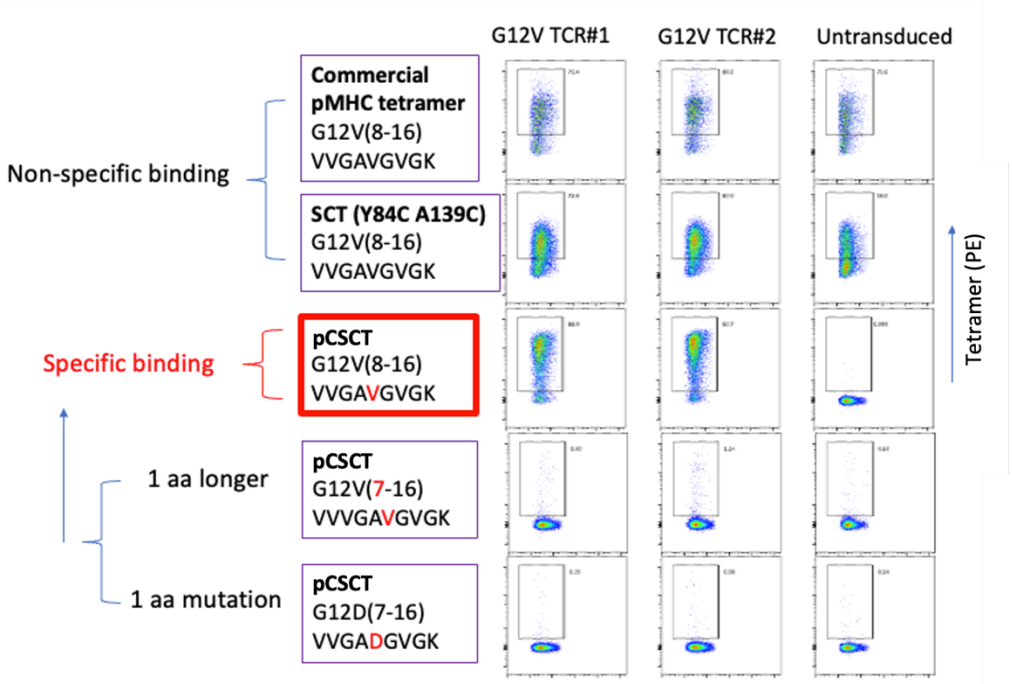


Figure 4.9 Specific binding of pCSCT tetramer with cognate KRAS-G12V-specific T cells. Flow cytometry result for tetramer staining of HLA-A*11:01 KRAS-G12V-specific T cells using pCSCT, conventional SCT, or commercial pMHC tetramers. Gated on live/single cells. Y-axis: Tetramer-PE. X-axis: FSC.

Note: All G12V-pCSCT reagents in this figure were prepared using HLA-A*11:01 EBV-CSCT template (template peptide: ATIGTAMYK).

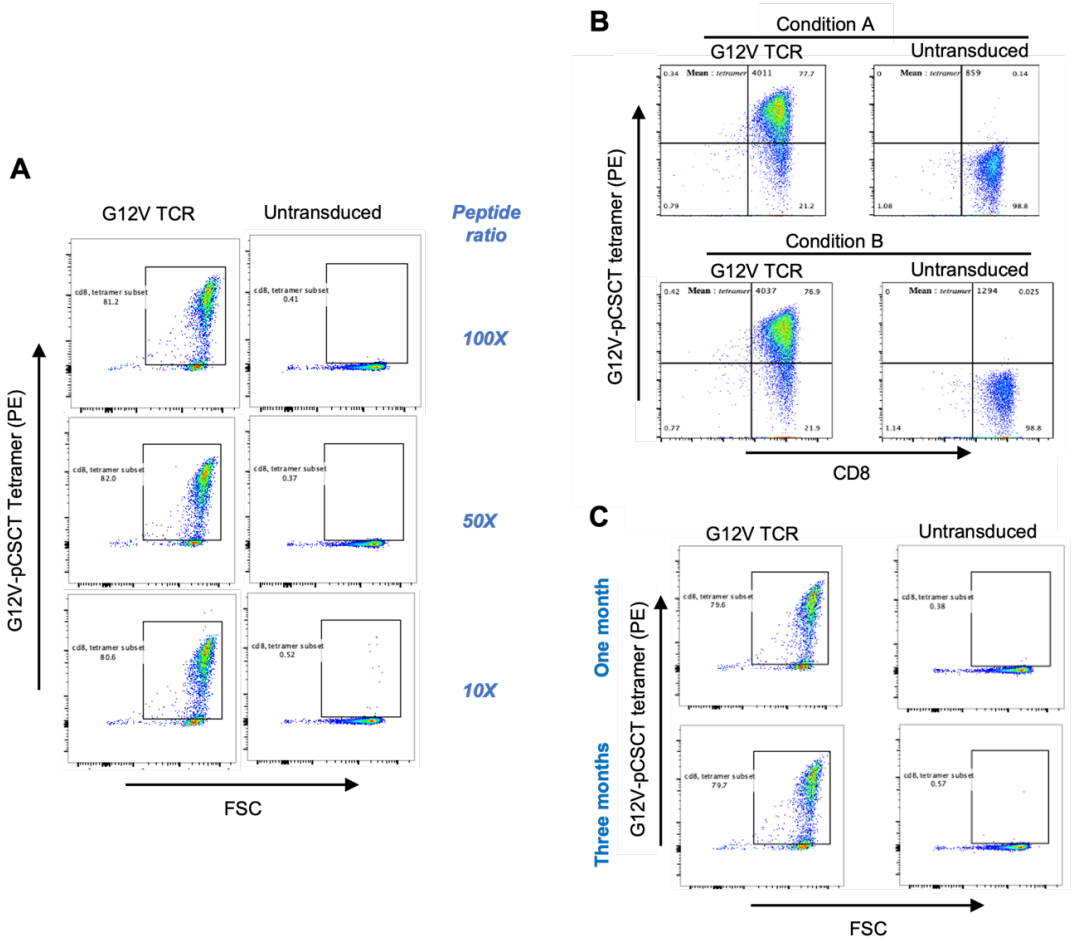


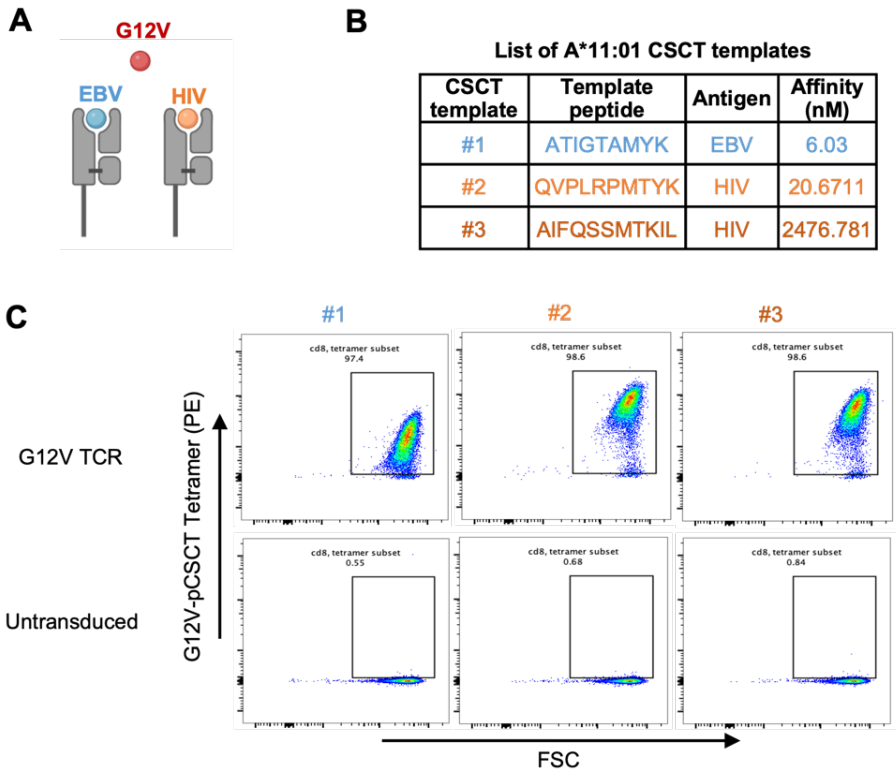
Figure 4.10 Peptide-loading efficiency and stability of A*11:01 G12V pCSCT reagent.

A. Flow cytometry result for staining HLA-A*11:01 KRAS G12V-specific T cells using pCSCT tetramers prepared with a titrated amount of G12V peptide. From top to bottom: peptide to CSCT template ratio of 100:1, 50:1, 10:1, respectively. Gated on live/single cells. Y-axis: G12V-pCSCT tetramer signal, X-axis: FSC.

B. Stability test of "peptide-receptive" HLA-A*11:01 CSCT monomer (Condition A) or peptide-loaded G12V-pCSCT tetramer (Condition B) after two weeks. Y-axis: G12V-pCSCT tetramer signal, X-axis: CD8.

C. G12V-pCSCT tetramer prepared by "Peptide-receptive" HLA-A*11:01 CSCT monomer after 1 month or 3 months of TEV cleavage. Y-axis: G12V-pCSCT tetramer signal, X-axis: FSC.

Note: All G12V-pCSCT reagents in this figure were prepared using HLA-A*11:01 EBV-CSCT template (template peptide: ATIGTAMYK).



CSCT template	Template peptide	Antigen	Affinity (nM)
#1	ATIGTAMYK	EBV	6.03
#2	QVPLRPMTYK	HIV	20.6711
#3	AIFQSSMTKIL	HIV	2476.781

Figure 4.11 ***Comparison between different A*11:01 CSCT templates***

A. Schematic illustration of peptide-exchange using CSCT templates with different place-holder peptides.

B. List of three A*11:01 CSCT templates.

C. Flow cytometry result of G12V-pCSCT tetramer staining, each tetramer was prepared using indicated CSCT template. Gated on live/single cells. Y-axis: G12V-pCSCT tetramer signal, X-axis: FSC.

4.7 REFERENCES

1. Chour, W. *et al.* Large libraries of single-chain trimer peptide-MHCs enable antigen-specific CD8 T cell discovery and analysis. *Commun. Biol.* **6**, 1–13 (2023).
2. Puig-Saus, C. *et al.* Neoantigen-targeted CD8 T cell responses with PD-1 blockade therapy. *Nature* **615**, 697–704 (2023).
3. Foy, S. P. *et al.* Non-viral precision T cell receptor replacement for personalized cell therapy. *Nature* 1–10 (2022) doi:10.1038/s41586-022-05531-1.
4. Yu, Y. Y. L., Netuschil, N., Lybarger, L., Connolly, J. M. & Hansen, T. H. Cutting Edge: Single-Chain Trimers of MHC Class I Molecules Form Stable Structures That Potently Stimulate Antigen-Specific T Cells and B Cells1. *J. Immunol.* **168**, 3145–3149 (2002).
5. Truscott, S. M. *et al.* Disulfide Bond Engineering to Trap Peptides in the MHC Class I Binding Groove1. *J. Immunol.* **178**, 6280–6289 (2007).
6. Mitaksov, V. *et al.* Structural Engineering of pMHC Reagents for T Cell Vaccines and Diagnostics. *Chem. Biol.* **14**, 909–922 (2007).
7. Luimstra, J. J. *et al.* A flexible MHC class I multimer loading system for large-scale detection of antigen-specific T cells. *J. Exp. Med.* **215**, 1493–1504 (2018).
8. Saini, S. K. *et al.* Empty peptide-receptive MHC class I molecules for efficient detection of antigen-specific T cells. *Sci. Immunol.* **4**, eaau9039 (2019).
9. Overall, S. A. *et al.* High throughput pMHC-I tetramer library production using chaperone-mediated peptide exchange. *Nat. Commun.* **11**, 1909 (2020).
10. Toebes, M. *et al.* Design and use of conditional MHC class I ligands. *Nat. Med.* **12**, 246–251 (2006).
11. Saini, S. K. *et al.* Dipeptides promote folding and peptide binding of MHC class I molecules. *Proc. Natl. Acad. Sci.* **110**, 15383–15388 (2013).
12. Rodenko, B. *et al.* Generation of peptide–MHC class I complexes through UV-mediated ligand exchange. *Nat. Protoc.* **1**, 1120–1132 (2006).
13. Meyer, M. *et al.* MediMer: a versatile do-it-yourself peptide-receptive MHC class I multimer platform for tumor neoantigen-specific T cell detection. *Front. Immunol.* **14**, (2024).
14. Chour, W. Molecular Technologies for Antigen-Based Immunity. (California Institute of Technology, 2021). doi:10.7907/z20t-nq62.

Chapter 5. VIRUS-SPECIFIC T CELL LANDSCAPE IN HPV16-POSITIVE INDIVIDUALS

5.1 ABSTRACT

Human papillomaviruses (HPVs) infection is a major contributor of cervical cancers, with HPV16 and HPV18 being the most oncogenic and prevalent types. The oncoproteins E6 and E7 play crucial roles in tumor development and are promising vaccine targets for therapeutic purposes. Additionally, T cell receptors (TCRs) specific to epitopes within E6 and E7 offer great potential for adoptive cell therapy in HPV-associated cancers. Here, we employed a high-throughput peptide-loaded cleavable-SCT (pCSCT) platform integrated with single-cell multi-omics to comprehensively profile antigen-specific CD8 T cells against immunogenic epitopes from CMV, EBV, influenza, and HPV16. Furthermore, we identified low-frequency circulating HPV16-specific TCR clones from precancerous HPV16-positive patients during the interventions of a therapeutic HPV DNA vaccine. This approach enables paired antigen-TCR discovery and phenotypic characterization to facilitate development of TCR-based adoptive cell therapies against HPV-associated cancers.

5.2 INTRODUCTION

Human papillomaviruses (HPV) are a group of small (8 kbp), non-enveloped DNA viruses. High-risk HPV types (such as HPV16 and HPV18) account for >70% of all cervical cancer cases reported worldwide[1–3]. There are three major genome regions for HPV viruses: (1) the early region (E), which has six common open reading frames (E1, E2, E4, E5, E6, and E7) encoding non-structural proteins; (2) the late region (L1 and L2), encoding structural capsid proteins; and

(3) the non-coding region. E1, E2, and E3 are involved in DNA and virus replication, transcription, and genome division during the viral life cycle. E4 and E5 are involved in cell cycle entry, immune evasion and virus release[1,2,4,5]. The E6 (150 amino acids) and E7 (98 amino acids) proteins are the major oncoproteins in high-risk HPV types (such as HPV16). Specifically, E6 promotes degradation of the tumor suppressor p53, while E7 interacts with retinoblastoma-associated (pRB) proteins, and both promote HPV replication[6]. Consequently, while L1, which is expressed on the surface of the encapsidated virus, has been the target for prophylactic HPV vaccines (e.g., Gardasil, Gardasil9, Cervarix), E6 and E7 have been a particular interest in cancer immunology for treating HPV-related cancers[7]. T cell receptors (TCRs) targeting epitopes in E6 and E7 hold high potential for treating HPV-positive cancers[4,5,8]. In a recent trial, an HLA-A*02-restricted TCR selective for HPV16 E7 (epitope 11-19) was administered to 12 patients with metastatic HPV-associated epithelial cancers, with 6 experiencing partial responses[4]. These encouraging results highlight the need to further improve cell therapy platforms for HPV-related tumors and facilitate identification of potent TCRs against HPV.

Here, we demonstrate the integration of a user-friendly, high-throughput peptide-loaded cleavable cleavable-SCT (pCSCT) platform (described in Chapter 4) with single-cell multi-omics assays for rapid epitope-based antigen-specific T cell profiling. In addition to HPV, common viruses such as EBV, CMV and influenza are associated with persistent infection, and can contribute to cancer development either via chronic inflammation or immune suppression, or directly by dysregulating oncogenes in the host. Hence, we aim to assess the antigen specific CD8 T cell response against these common viruses simultaneously in this work. We show that DNA-barcoded pCSCT multimers identified antigen-specific CD8 T cells against CMV, EBV, and influenza epitopes in a multiplexed assay from multiple individuals in parallel. Antigen-

specific T cells against different viruses exhibited divergent phenotypic distributions. We deconvoluted low-frequency circulating HPV16-specific CD8 T cell clones from low sample volumes of multiple precancerous patients undergoing interventions with an HPV DNA vaccine. This high-throughput approach enables the identification of antigen-TCR pairs with phenotypic readouts, directly facilitating the development of TCR-based adoptive cell therapy for HPV-related cancers.

5.3 RESULTS

5.3.1 *High-throughput preparation of peptide-loaded cleavable SCT library representing common virus and HPV epitopes*

To comprehensively characterize T cell immunity in HPV16-positive patients, we constructed a peptide-MHC (pMHC) library spanning epitopes from a diverse set of common viruses, including cytomegalovirus (CMV), Epstein-Barr virus (EBV), influenza viruses (Flu), as well as HPV16 epitopes (Table 5.1). We generated peptide-loaded cleavable SCTs (pCSCTs) (N = 52) for predicted epitopes presented by prevalent MHC-I molecules, including HLA-A*02:01, A*03:01, A*11:01, A*24:02, and B*07:02. Conventional SCTs (N = 44) were also prepared in house or by a collaborator using a previously described method[9–11]. Each pCSCT and SCT underwent pre-screening with peripheral blood mononuclear cells (PBMCs) from healthy donors to exclude those exhibiting non-specific binding. As previously mentioned, pCSCTs can overcome the issue of non-specific binding sometimes seen with conventional SCTs, such as in the case of the KRAS G12V tetramer. Here, 6 out of the 44 conventional SCTs were removed from the 10X experiment due to non-specific binding, while all of the pCSCTs passed the pre-screening. Each SCT and pCSCT was then assembled into PE-labeled and DNA-barcoded dextramer for antigen-specific CD8 T cell staining.

Conventional single-chain_trimer (SCT) (N=44)

ID	HLA	Antigen	Peptide	In_10X
1	A0201	HPV_E7	LLMGTLGIV	Yes
2	A0201	HPV_E7	YMLDLQPET	Yes
3	A0201	HPV_E6	LQTTIHDII	Yes
4	A0201	HPV_E6	TTIHDIILE	Yes
5	A0201	HPV_E7	MLDLQPETT	Yes
6	A0201	HPV_E6	ELQTTIHDI	Yes
7	A0201	HPV_E7	TLEDLLMGTL	Yes
8	A0201	HPV_E7	TLHEYMLDL	Yes
9	A0201	HPV_E6	SLYGTTLEQ	Yes
10	A0201	HPV_E6	AMFQDPQER	Yes
11	A0201	HPV_E6	IVYRDGNPYA	Yes
12	A0201	HPV_E6	TLEQQYNKP	Yes
13	A0201	HPV_E7	HVDIRTLED	Yes
14	A0201	HPV_E7	VQSTHVDIR	Yes
15	A0201	HPV_E6	EVYDFAFRD	Yes
16	A0201	HPV_E6	TTLEQQYNK	Yes
17	A0201	HPV_E6	YRDGNPYAV	Yes
18	A0201	HPV_E7	QLNDSSEEE	Yes
19	A0201	HPV_E6	LLRREVYDF	Yes
20	A0201	HPV_E7	DLQPETTDLY	Yes
21	A0201	HPV_E6	NIRGRWTGR	Yes
22	A0201	HPV_E7	RAHYNIVTF	Yes
23	A0201	HPV_E6	QLLRREVYD	Yes
24	A0201	HPV_E6	HLDKKQRFH	Yes
25	A0301	HPV_E6	HLDKKQRFH	Yes
26	A0301	HPV_E6	KFYSKISEY	No
27	A0301	HPV_E6	AMFQDPQER	Yes
28	A0301	HPV_E6	IVYRDGNPY	Yes
29	A0301	HPV_E6	SLYGTTLEQ	No
30	A1101	HPV_E6	TTLEQQYNK	Yes
31	A1101	HPV_E6	RTAMFQDPQ	Yes
32	A1101	HPV_E6	SLYGTTLEQ	Yes
33	A1101	HPV_E6	AMFQDPQER	Yes
34	A1101	HPV_E6	FQDPQERPRK	Yes
35	A1101	HPV_E6	IVYRDGNPY	Yes
36	A1101	HPV_E6	GTTLEQQYNK	Yes
37	A1101	HPV_E7	VQSTHVDIR	Yes
38	A1101	HPV_E6	KFYSKISEYR	No
39	A0201	cef_3-CMV	NLVPMVATV	Yes
40	A0201	Influenza A virus	FMYSDFHFI	Yes
41	A1101	cef_7-EBV	ATIGTAMYK	No
42	A1101	cef_8-EBV	AVFDRKSDAK	No
43	A1101	cef_9-EBV	IVTDFSVIK	No
44	B0702	cef_20-CMV	TPRVTGGGAM	Yes

Peptide-loaded Cleavable SCT (pCSCT) (N=52)

ID	HLA	Antigen	Peptide	In_10X
45	A0201	HPV	FAFRDLCIV	Yes
46	A0201	HPV	KLPQLCTEL	Yes
47	A0201	HPV	LLMGTLGIV	Yes
48	A0201	HPV	TIHDIILECV	Yes
49	A0201	HPV	TLEDLLMGTL	Yes
50	A0201	HPV	TLHEYMLDL	Yes
51	A0201	HPV	TLQDVSLEV	Yes
52	A0201	HPV	VLLLWITAA	Yes
53	A0201	HPV	YICEEASVTV	Yes
54	A0201	HPV	YMLDLQPET	Yes
55	A0201	CMV	NLVPMVATV	Yes
56	A0201	Influenza A virus	GILGFVFTL	Yes
57	A0201	Influenza A virus	FMYSDFHFI	Yes
58	A0201	EBV	CLGGLLTMV	Yes
59	A0201	EBV	GLCTLVAML	Yes
60	A1101	Protein E7	GIVCPICSQK	Yes
61	A1101	E6_HPV16_sp_P03	GTTLEQQYNK	Yes
62	A1101	Protein E6	IILECVYCK	Yes
63	A1101	Protein E7	IVCPICSQK	Yes
64	A1101	E6_HPV16_sp_P03	IVYRDGNPY	Yes
65	A1101	Protein E6	LLIRCINCQK	Yes
66	A1101	E6_HPV16_sp_P03	TTLEQQYNK	Yes
67	A1101	EBV	ATIGTAMYK	Yes
68	A1101	EBV	AVFDRKSDAK	Yes
69	A1101	EBV	IVTDFSVIK	Yes
70	A1101	Influenza A virus	SIIPSGPLK	Yes
71	A2402	hpv_30_CDST	CDSTLRLCV	Yes
72	A2402	hpv_31_LCVQ	LCVQSTHVDI	Yes
73	A2402	hpv_32_HYNI	HYNIVTFCC	Yes
74	A2402	cef_10_AYAQ	AYAQKIFKIL	Yes
75	A2402	cef_11_DYNF	DYNFVKQLF	Yes
76	A2402	cef_13_QYDP	QYDPVAALF	Yes
77	A2402	cef_14_TYGP	TYGPVFMCL	Yes
78	A2402	cef_15_TYPV	TYPVLEEMF	Yes
79	A2402	cef_16_VYAL	VYALPLKML	Yes
80	A2402	cef_17_VYYT	VYYTSAFVF	Yes
81	A2402	cef_43_DYCN	DYCNVLNKEF	Yes
82	B0702	hpv_22_TPTL	TPTLHEYML	Yes
83	B0702	hpv_26_RPRK	RPRKLPDLC	Yes
84	B0702	hpv_27_CPEE	CPEEKQRHL	Yes
85	B0702	cef_18_RPHE	RPHERNGFTV	Yes
86	B0702	cef_19_RPPI	RPPIFIRLL	Yes
87	B0702	cef_20_TPRV	TPRVTGGGAM	Yes
88	A0301	hpv_19_KFYS	KFYSKISEY	Yes
89	A0301	hpv_20_IVYR	IVYRDGNPY	Yes
90	A0301	hpv_21_SLYG	SLYGTTLEQ	Yes
91	A0301	cef_4_ILRG	ILRGSVAHK	Yes
92	A0301	cef_5_KLGG	KLGGALQAK	Yes
93	A0301	cef_6_RLRA	RLRAEAQVK	Yes
94	A0301	cef_40_RVLS	RVLSFIKGTK	Yes
95	A0301	cef_41_RVRA	RVRAYTYSK	Yes
96	A0301	cef_42_SIIP	SIIPSGPLK	Yes

Table 5.1 List of peptide-MHC reagents (SCTs and pCSCTs) used in the 10X experiment

5.3.2 *The magnitude of antigen-specific T cell response in HPV-vaccinated patients*

HPV16-positive patient samples were kindly provided by Dr. Connie L. Trimble, MD. Patients were enrolled in a clinical trial for the assessment of VGX-3100 in CIN2/3 associated with HPV-16 and HPV-18 as previously described[12]. VGX-3100 is a synthetic therapeutic DNA

vaccine targeting HPV-16 and HPV-18 E6 and E7 proteins[12,13], given by electroporation at study weeks 0, 4, and 12. For selected patients, longitudinal PBMCs collected at baseline, weeks 6, 14, 24, and 36 were included in this analysis (Table 5.2). We selected 11 patients based on their HLA type and inventory availability, prioritizing those who express A*02:01, A*11:01. Five donor PBMCs (purchased from CTL ePBMC) were also included as positive control given their known positive ELISPOT result for CEF peptide pools. To facilitate the high-throughput cell labeling strategy, each vial of PBMCs was thawed and labeled with a unique combination of hashtags (containing one or two hashtags) and then pooled together for CD8 T cell enrichment. To minimize the possibility of pMHC-dextramer binding by un-matched HLAs, we grouped patients and donors into two groups based on HLA types for cocktails of SCT and pCSCT dextramer staining. Group A individuals were stained only with A*02:01, A*11:01, B*07:02 pMHC-dextramers, while Group B individuals were also stained with A*24:02 and A*03:01 dextramers. Dextramer positive cells were sorted for single cell (sc) RNAseq to profile gene expression, TCR, surface protein, and antigen specificity (Figure 5.1A). As expected, both groups contain CMV-positive individuals, with cell frequencies of 0.16% and 0.66 (from all CD8 T cells) detected by the A*02:01 CMV-NLVP SCT (dextramer-PE and tetramer-APC, double positive population in Figure 5.1B). In addition, other antigen-specific CD8 T cell populations for two groups were detected at frequencies of 0.39% and 0.82%, respectively (PE positive population in Figure 5.1B).

HPV-vaccinated patients (N = 11) and Donors (N = 5)

FACS sample	Sample ID	Patient/Donor ID	HLA	timepoint	Dex Pool	Hashtag
A	1	CLT3071AS	A2	Wk9/10	A2,A11,B7	1
	2	CLT3104KT	A2, A11	Wk9/10	A2,A11,B7	2
	3			Rx (wk28)	A2,A11,B7	28
	4	CLT3126MN	A2	SCREEN	A2,A11,B7	3
	5			Wk9/10	A2,A11,B7	34
	6			Wk15	A2,A11,B7	35
	7	CLT3125GG	A2	SCREEN	A2,A11,B7	36
	8			Wk9/10	A2,A11,B7	37
	9	CLT3007BK	A2	Leuk	A2,A11,B7	38
	10	LP_627 (Donor)	A11	ePBMC	A2,A11,B7	4
	11	LP_501 (Donor)	A2, A11	ePBMC	A2,A11,B7	45
	12	CLT3047SH	A11,B7	Wk9/10	A2,A11,B7	46
	13			Wk15	A2,A11,B7	47
B	14	LP_650 (Donor)	A2, A24	ePBMC	A2, A11, B7, A24, A3	48
	15	LP_520 (Donor)	A2, A24	ePBMC	A2, A11, B7, A24, A3	5
	16	CLT3003LP	A2,A24	Wk15	A2, A11, B7, A24, A3	56
	17	CLT3055MC	A2,A3	Wk9/10	A2, A11, B7, A24, A3	57
	18			Wk15	A2, A11, B7, A24, A3	58
	19			Rx (wk28)	A2, A11, B7, A24, A3	6
	20	CLT3057KM	A2, A3, B7	Wk9/10	A2, A11, B7, A24, A3	67
	21	CLT3108JF	A11, A3, B7	SCREEN	A2, A11, B7, A24, A3	68
	22			Wk9/10	A2, A11, B7, A24, A3	7
	23			Wk15	A2, A11, B7, A24, A3	78
	24	CLT3062MH	A24,B7	Leuk	A2, A11, B7, A24, A3	8
	25	LP_512 (Donor)	A3	ePBMC	A2, A11, B7, A24, A3	CD8

Table 5.2 List of individuals for antigen-specific T cell staining

5.3.3 *Peptide-loaded CSCTs effectively captured virus-specific CD8 T cells across a broad range of epitopes*

Dextramer-positive (Dex^+), and dextramer-negative (Dex^-) CD8 T cells were pooled at a 1:1 ratio enabling the comparison between antigen-specific and pan-T cell populations. For each cell, the sample identity (which patient at certain timepoint, or which donor) was determined by hashtag-based demultiplexing (Method). Briefly, we applied a binary gate for each hashtag's UMI reads (Figure 5.2A), then sample identity was assigned by the unique hashtag combinations used during cell staining (Figure 5.2B).

Each cell's antigen assignment was determined by the DNA-barcoded dextramer. We considered a cell as dextramer-specific if more than 25% and more than 5 of the dextramer-associated UMIs corresponded to a certain dextramer's DNA barcode (Figure 5.3A). TCR-clonotype-based antigen assignment was applied to correct for outlier or false assignment (see methods). Ambiguous or unassigned cells were treated as non-antigen-specific cells. Notably, some well-recognized immunogenic epitopes failed for conventional SCT expression, but were feasible in the pCSCT format, and showed successful antigen-specific T cell detection from a wide range of individuals (Figure 5.3B). 2232 antigen-assigned cells were identified from 38 pMHC reagents, 71.6% (1598 cells) were captured via 19 pCSCT-only reagents (no cognate SCT used in the staining). Overall, we demonstrated the high-throughput application of pCSCT reagents as a complement to conventional SCTs. This combinatorial approach enables cost-effective screening of a broader range of epitopes.

5.3.4 *Phenotypes of antigen-specific CD8 T cells defined by gene expression and surface protein*

We then leveraged the single cell multiomic data to investigate the phenotypes of antigen-specific CD8 T cells in HPV-vaccinated patients and donors. Unsupervised clustering identified 16 transcriptional groups of CD8 T cells (Figure 5.4A). Differential gene expression analysis defined canonical CD8 T cell phenotypes[14,15] (Figure 5.4B, C). For instance, CCR7, LEF1, TCF7, IL7R, and SELL were up-regulated in naïve and memory cells, while GZMK were elevated in memory phenotypes. GZMB and GZMH were highly expressed in cytotoxic cells. We observed a mild expression of exhaustion markers (LAG2, TIGIT)[16] among memory and cytotoxic phenotypes, possibly related with persistent viral infection among these individuals. In additional to conventional CD8 T cell subsets, we also identified a mucosal-associated invariant

T (MAIT) CD8 cluster, characterized by high expression of KLRB1[17]. In addition to valuable biological functions in healthy individuals, MAIT cells have also been associated with COVID-19 disease severity[18], severe liver injury among patients with acute HAV infection[19], and airway inflammation in tuberculosis patients[17], etc. When integrated with scTCRseq, MAIT cells are characterized via semi-invariant TCRs comprised of an invariant TRAV1-2 gene, combining with limited β gene diversity[17] (Figure 5.4D). The cytotoxic cells exhibited a higher enrichment of clonally expanded T cell clonotypes, as anticipated (Figure 5.4E). The assignment of phenotypes was further validated by assessing the surface abundance of conventional markers (such as CCR7, CD45RA, CD45RO, CD25, etc.) using CITE-seq, which was consistent with gene expression profiles (Figure 5.4F). For instance, the upregulation or sustained expression of interleukin (IL)-7 receptor α (CD127) in memory cells suggests their likely differentiation from previously cytotoxic cells. Additionally, the surface expression of $CD45RO^{hi}CCR7^{low}$ further suggest a central memory-like phenotype for these cells.

5.3.5 *CD8 T cell response against immunodominant epitopes from multiple individuals*

Antigen-specific CD8 T cell responses were observed in both HPV16-positive patients and ePBMC donors (Figure 5.5A), displaying a heterogeneous T cell landscape within each individual (Figure 5.5B). Notably, HPV-specific CD8 T cells showed a generally low frequency, aligning with existing literature indicating their reduced frequencies in circulating blood compared to tissues[8]. We identified several immunodominant CD8 T cell epitopes that were consistently detectable across individuals. For instance, CD8 T cell responses against the A*02:01_EBV_GLCTLVAML epitope were detected in all 16 individuals (including patients and donors) (Figure 5.5C, Figure 5.6). Moreover, we identified 15 immunodominant epitopes that were recognized by at least 25% of the tested individuals, with seven epitopes originating

from HPV16 and predominantly restricted to HLA A*11:01(Table 5.3). This is also expected since we prioritized A*02:01 and A*11:01 individuals for cell staining.

Organism	HLA	Peptide	Protein	Detected/all individuals	Cell count	%Rank_EL	NetMHC Aff(nM)
HPV	A0201	LLMGTLGIV	E7	11/16	268	0.97	12.59
HPV	A1101	IILECVYCK	E6	4/16	5	1.369	56.89
HPV	A1101	IVYRDGNPY	E6	4/16	5	1.148	347.59
HPV	A1101	LLIRCINCQK	E6	14/16	123	4.273	233.96
HPV	A1101	RTAMFQDPQ	E6	10/16	13	6.205	3779.13
HPV	A1101	VQSTHVDIR	E7	4/16	24	4.277	6086.17
HPV	A2402	HYNIVTFCC	E7	8/16	170	3.974	8615.2

Table 5.3 List of immunodominant epitopes from HPV16 E6/E7 proteins

5.3.6 *Divergent phenotypic distribution for antigen-specific T cells against different viruses*

Upon antigen-exposure, antigen-specific CD8 T cells undergo rapid expansion for effector function, with a significantly lower frequency of memory T cells remaining after infection clearance. As expected, pMHC multimer-positive cells were enriched in cytotoxic and memory phenotypes, indicative of antigen experience, while multimer-negative cells were enriched in naive phenotypes (Data not shown). Mapping each cell's antigen origin onto UMAP revealed distinct density distributions for CMV, EBV, influenza, and HPV-specific cells (Figure 5.7A and B). Specifically, CMV-specific cells displayed an enrichment of the cytotoxic phenotype (Figure 5.7C), while influenza-specific cells predominantly exhibited memory phenotypes. In contrast, EBV-specific and HPV-specific cells were distributed across all phenotypes. Importantly, these phenotypic distributions were consistent in both ePBMC donors and HPV16-positive patients who underwent HPV vaccination (Figure 5.5A).

A much lower frequency of MAIT cells was assigned with an MHC-I antigen SCT compared to conventional CD8 T cell phenotypes (Figure 5.7C), consistent with previous reports on their non-MHC-I restriction[18]. Further analysis of individual epitopes from each virus showed

consistent overall phenotype distributions, with certain epitopes showing a preference towards specific phenotypes (Figure 5.7D). For instance, two influenza epitopes consisted predominantly of memory cells. Certain HPV epitopes, such as A*11:01_HPV_VQSTHVDIR and A*11:01_HPV_RTAMFQDPQ, were enriched with memory phenotypes, while others, like A*02:01_HPV_LLMGTLGIV and A*24:02_HPV_HYNIVTFC, were consisted predominantly of cytotoxic and naive cells. Overall, we identified divergent phenotypic distributions of antigen-specific T cells against different viruses.

5.3.7 *Candidate HPV-specific TCRs for preclinical evaluation of adoptive cell therapy*

Identifying HPV-specific T cell receptors (TCRs) will significantly facilitate the development of TCR-based therapies for HPV-positive tumors. We aimed to generate a list of candidate TCRs identified in our study for functional validation, potentially advancing towards clinical trials for HPV-related cancers. The first consideration is to elucidate the ideal HPV target. In addition to HPV16, HPV18 is also prevalent; therefore, TCRs that could potentially recognize conserved epitopes between HPV16 and HPV18 would benefit a broader patient population. We mapped all immunogenic HPV16 epitopes that showed antigen-specific CD8 T cell capture in our study. Epitopes derived from the E7 protein share very few sequence similarities between HPV16 and HPV18 (Figure 5.8, top panel), indicating that TCRs targeting these epitopes are less likely to be functional in HPV18-positive patients. In contrast, the E6 protein is highly conserved between the two viral types (Figure 5.8, bottom panel). All HPV16 E6 epitopes share significant sequence similarities with their HPV18 counterparts, including immunodominant epitopes (brown rectangles in Figure 5.8, bottom panel). For example, A*11:01-presented epitope **I**ILE**CVYCK** (3 amino acids (AAs) different from HPV18, including

1 similar AA), I**VYRD**GN**P**Y (4 AAs different, including 2 similar), and **LLIRC**IN**CQK** (only 2 AAs different, with 1 similar); conserved AAs between HPV16 and HPV18 are shown in bold.

We subsequently identified clonally expanded TCRs assigned to various HPV16 epitopes for *in vitro* TCR cloning and functional validation. The selected potential HPV-specific clones are provided in Table 5.4. Cells within each clonotype tended to share similar UMAP distributions and gene expression profiles (Figure 5.9A). We observed HPV-specific T cell responses at multiple time points from some DNA-vaccinated patients (Figure 5.9B). Further studies are underway for an in-depth longitudinal analysis. The phenotype readout and TCR persistence from this experiment may significantly contribute to refining candidate TCRs. TCRs demonstrating potent cytotoxicity and long-term memory potential, or a combination of two types of TCRs for therapy, may enhance the longevity and efficacy of adoptive cell therapy.

For our ongoing research, we isolated and then cloned paired α and β chains into the TCR-KO Jurkat cell line for a pMHC-tetramer staining assay to validate specificity. Subsequently, candidate TCRs will be engineered into primary CD8 T cells, with endogenous TCRs being knocked out using CRISPR technology (to enhance the level of expression of our introduced TCR chains). These cells will then be stimulated with titrated concentrations of cognate peptide to assess the activation threshold and effector function based on measuring TNF-α and/or INF-γ production. Cytotoxicity against peptide-loaded APCs or HPV-transduced cell lines will be evaluated using a cell killing assay. Amino acid substitutions in individual residues of the peptide epitope, including Alanine and Glycine scanning, will be used to identify the TCR-pMHC interaction motif and to screen for potential cross-reactivity against similar "self-antigens." Finally, comprehensive preclinical evaluation and adherence to GMP procedures will be essential steps towards developing a clinical candidate[4,10,11].

TCR_ID	Group	Full TCR
SCD10X_TCR_001	cef	TRAV13-2_CADLSSGGSYIPTF_TRAJ6_TRAC_TRBV4-3_CASSQRTGREQYF_TRBJ2-7_TRBC2
SCD10X_TCR_002	cef	TRAV17_CATNLGGSRLTF_TRAJ58_TRAC_TRBV14_CASSIGTVYNEQFF_TRBJ2-1_TRBC2
SCD10X_TCR_003	cef	TRAV39_CAAADSGGGADGLTF_TRAJ45_TRAC_TRBV4-3_CASSQRTGREQYF_TRBJ2-7_TRBC2
SCD10X_TCR_004	cef	TRAV25_CAGLLSNDYKLSF_TRAJ20_TRAC_TRBV5-5_CASSPDLNYGYTF_TRBJ1-2_TRBC1
SCD10X_TCR_005	cef	TRAV27_CAGGGSQGNLIF_TRAJ42_TRAC_TRBV19_CASSIRSSYEQYF_TRBJ2-7_TRBC2
SCD10X_TCR_006	cef	TRAV20_CAVQASNARLMF_TRAJ31_TRAC_TRBV9_CASSLQESSYNSPLHF_TRBJ1-6_TRBC1
SCD10X_TCR_007	cef	TRAV21_CAVLMDSNYQLIW_TRAJ33_TRAC_TRBV10-2_CASSGDGMNTEAFF_TRBJ1-1_TRBC1
SCD10X_TCR_008	cef	TRAV12-1_CVVNGLTGGGNKLTF_TRAJ10_TRAC_TRBV3-1_CASSPGLSYEQYF_TRBJ2-7_TRBC2
SCD10X_TCR_009	cef	TRAV12-1_CVVGNTGGFKTIF_TRAJ9_TRAC_TRBV7-2_CASSRDNGDEQFF_TRBJ2-1_TRBC2
SCD10X_TCR_010	cef	TRAV38-2/DV8_CACSNSGYALNF_TRAJ41_TRAC_TRBV30_CAWSVLDEKLFF_TRBJ1-4_TRBC1
SCD10X_TCR_011	cef	TRAV5_CAEGKAGGTSYGKLTF_TRAJ52_TRAC_TRBV5-6_CASSLKLWQKGSEQYF_TRBJ2-7_TRBC2
SCD10X_TCR_012	cef	TRAV35_CAGRVNNDMRF_TRAJ43_TRAC_TRBV7-2_CASSWGQGSNYGYTF_TRBJ1-2_TRBC1
SCD10X_TCR_013	cef	TRAV17_CATVPRMDSSYKLIF_TRAJ12_TRAC_TRBV7-9_CASSLIGTSERNEQFF_TRBJ2-1_TRBC2
SCD10X_TCR_014	cef	TRAV38-2/DV8_CAYRRDDKIIF_TRAJ30_TRAC_TRBV28_CASSLRLTSAGETQYF_TRBJ2-5_TRBC2
SCD10X_TCR_015	hpv1	TRAV12-3_CAMGGNDMRF_TRAJ43_TRAC_TRBV5-1_CARKETPHF_TRBJ2-7_TRBC2
SCD10X_TCR_016	hpv1	TRAV25_CAVDDNYGQNFVF_TRAJ26_TRAC_TRBV7-2_CASSEGPMYEQYF_TRBJ2-7_TRBC2
SCD10X_TCR_017	hpv1	TRAV21_CAVKDTDKLIF_TRAJ34_TRAC_TRBV6-2_CASSIAGGLVEQFF_TRBJ2-1_TRBC2
SCD10X_TCR_018	hpv1	TRAV19_CALSVNGMTTDSWGKFQF_TRAJ24_TRAC_TRBV27_CASSPSGTAYEQYF_TRBJ2-7_TRBC2
SCD10X_TCR_019	hpv1	TRAV27_CAGALGNQFYF_TRAJ49_TRAC_TRBV30_CAWGTSEAFF_TRBJ1-1_TRBC1
SCD10X_TCR_020	hpv1	TRAV17_CATDAESDGQKLLF_TRAJ16_TRAC_TRBV2_CASIGQGWTNEKLFF_TRBJ1-4_TRBC1
SCD10X_TCR_021	hpv1	TRAV4_CLVGDKGRNTDKLIF_TRAJ34_TRAC_TRBV5-1_CARKETPHF_TRBJ2-7_TRBC2
SCD10X_TCR_022	hpv1	TRAV35_CAGRRVAGGTSYGKLTF_TRAJ52_TRAC_TRBV4-1_CASSPSGLASETQYF_TRBJ2-5_TRBC2
SCD10X_TCR_023	hpv1	TRAV13-2_CAETSSGGSYIPTF_TRAJ6_TRAC_TRBV4-3_CASSQRTGREQYF_TRBJ2-7_TRBC2
SCD10X_TCR_024	hpv1	TRAV20_CAVQGADSWGKLQF_TRAJ24_TRAC_TRBV10-2_CASNPGPVGYEQYF_TRBJ2-7_TRBC2
SCD10X_TCR_025	hpv1	TRAV10_CVVSGRNSGYALNF_TRAJ41_TRAC_TRBV28_CASSQGWTSSYEQYF_TRBJ2-7_TRBC2
SCD10X_TCR_026	hpv1	TRAV8-3_CAVGDNFNKFYF_TRAJ21_TRAC_TRBV28_CASSLTSAAGELFF_TRBJ2-2_TRBC2
SCD10X_TCR_027	hpv1	TRAV38-2/DV8_CAWEGTSNTGKLIF_TRAJ37_TRAC_TRBV3-1_CASSQDSGDYGYTF_TRBJ1-2_TRBC1
SCD10X_TCR_028	hpv1	TRAV8-3_CAVGDNFNKFYF_TRAJ21_TRAC_TRBV28_CASSLSSAAAQAFF_TRBJ1-1_TRBC1
SCD10X_TCR_029	hpv1	TRAV29/DV5_CAASWGGGTSYGKLTF_TRAJ52_TRAC_TRBV20-1_CSASGQGSIYEQYF_TRBJ2-7_TRBC2
SCD10X_TCR_030	hpv1	TRAV38-2/DV8_CALGGEWGDMRF_TRAJ43_TRAC_TRBV5-8_CASRTDRDYEQYV_TRBJ2-7_TRBC2
SCD10X_TCR_031	hpv1	TRAV12-1_CVVNNYGQNFVF_TRAJ26_TRAC_TRBV6-1_CASSVRGAIETQYF_TRBJ2-5_TRBC2
SCD10X_TCR_032	hpv1	TRAV21_CAVAIKAAGNKLTF_TRAJ17_TRAC_TRBV12-3_CASSLNYAVEKLFF_TRBJ1-4_TRBC1
SCD10X_TCR_033	hpv1	TRAV8-3_CATNAGNNRKLIW_TRAJ38_TRAC_TRBV4-1_CASSHRGGYTF_TRBJ1-2_TRBC1
SCD10X_TCR_034	hpv1	TRAV29/DV5_CAASPIIQGAQKLVF_TRAJ54_TRAC_TRBV19_CASKVGPTGLDIQYF_TRBJ2-4_TRBC2
SCD10X_TCR_035	hpv1	TRAV29/DV5_CAASRADGLTF_TRAJ45_TRAC_TRBV3-1_CASSQTVTNQPQHF_TRBJ1-5_TRBC1
SCD10X_TCR_036	hpv1	TRAV4_CLVGDTNNFNKFYF_TRAJ21_TRAC_TRBV7-3_CASSLFPGGPHNEQFF_TRBJ2-1_TRBC2
SCD10X_TCR_037	hpv1	TRAV8-3_CAVGDNFNKFYF_TRAJ21_TRAC_TRBV28_CASSRTSATGELFF_TRBJ2-2_TRBC2
SCD10X_TCR_038	hpv1	TRAV6_CALGQGGKLIF_TRAJ23_TRAC_TRBV4-2_CASSPRLVSSSYEQYF_TRBJ2-7_TRBC2
SCD10X_TCR_039	hpv1	TRAV14/DV4_CAMSLSGAQKLVF_TRAJ54_TRAC_TRBV19_CASSISGRSYNEQFF_TRBJ2-1_TRBC2
SCD10X_TCR_040	hpv1	TRAV10_CVVSDRGSTLGRLYF_TRAJ18_TRAC_TRBV25-1_CASSEWGAGEELSYEQYF_TRBJ2-7_TRBC2
SCD10X_TCR_041	hpv1	TRAV14/DV4_CAMREVLDSNYQLIW_TRAJ33_TRAC_TRBV27_CASSINLGGRTDTQYF_TRBJ2-3_TRBC2
SCD10X_TCR_042	hpv1	TRAV19_CALSEAELVF_TRAJ54_TRAC_TRBV30_CAWSVLGTAYNEQFF_TRBJ2-1_TRBC2
SCD10X_TCR_043	hpv1	TRAV1-2_CAVRDSNYQLIW_TRAJ33_TRAC_TRBV6-1_CASSELVDRANEQFF_TRBJ2-1_TRBC2
SCD10X_TCR_044	hpv1	TRAV1-2_CAVRDSNYQLIW_TRAJ33_TRAC_TRBV20-1_CSARDPVTGEWTEAFF_TRBJ1-1_TRBC1
SCD10X_TCR_045	hpv1	TRAV5_CAEKSNTGKLIF_TRAJ37_TRAC_TRBV3-1_CASSQVAGEMRETQYF_TRBJ2-5_TRBC2
SCD10X_TCR_046	hpv1	TRAV3_CAVRDEGGATNKLIF_TRAJ32_TRAC_TRBV19_CASSIADGDEQYF_TRBJ2-7_TRBC2
SCD10X_TCR_047	hpv2	TRAV17_CATTLGGARLMF_TRAJ31_TRAC_TRBV14_CASSSSTSYNEQFF_TRBJ2-1_TRBC2
SCD10X_TCR_048	hpv2	TRAV2_CAVEGFSGGYQKVTF_TRAJ13_TRAC_TRBV5-4_CASRPGGEAEAFF_TRBJ1-1_TRBC1
SCD10X_TCR_049	hpv2	TRAV12-2_CAVNRSDMRF_TRAJ43_TRAC_TRBV6-2_CASSRQGSYNEQFF_TRBJ2-1_TRBC2
SCD10X_TCR_050	hpv2	TRAV10_CVVSELKRGSTLGRLYF_TRAJ18_TRAC_TRBV12-5_CASGWPGGTLDEQFF_TRBJ2-1_TRBC2
SCD10X_TCR_051	hpv2	TRAV17_CATALGNQFYF_TRAJ49_TRAC_TRBV10-3_CAISESNQPQHF_TRBJ1-5_TRBC1
SCD10X_TCR_052	hpv2	TRAV5_CAESTGKLIF_TRAJ37_TRAC_TRBV29-1_CSVGTGGTNEKLFF_TRBJ1-4_TRBC1
SCD10X_TCR_053	hpv2	TRAV17_CATVLGTGKLIF_TRAJ37_TRAC_TRBV2_CASSGLNQPQHF_TRBJ1-5_TRBC1
SCD10X_TCR_054	hpv2	TRAV5_CAEDRNARLMF_TRAJ31_TRAC_TRBV20-1_CSARDRGLGNTIYF_TRBJ1-3_TRBC1
SCD10X_TCR_055	hpv3	TRAV21_CAVTRTGTASKLTF_TRAJ44_TRAC_TRBV12-3_CASSPNYSNQPQHF_TRBJ1-5_TRBC1
SCD10X_TCR_056	hpv3	TRAV35_CAGRRGAQKLVF_TRAJ54_TRAC_TRBV2_CAKSTGGWSYEQYF_TRBJ2-7_TRBC2
SCD10X_TCR_057	hpv3	TRAV17_CATVLGTGKLIF_TRAJ37_TRAC_TRBV2_CASSGLNQPQHF_TRBJ1-5_TRBC1
SCD10X_TCR_058	hpv3	TRAV29/DV5_CASRGGAQKLVF_TRAJ54_TRAC_TRBV30_CAWSPDPRWAEAFF_TRBJ1-1_TRBC1
SCD10X_TCR_059	hpv3	TRAV3_CAVRDMFDNTDKLIF_TRAJ34_TRAC_TRBV6-1_CASSEAAGASTIQTQYF_TRBJ2-5_TRBC2
SCD10X_TCR_060	hpv3	TRAV13-1_CAASSNSSNTGKLIF_TRAJ37_TRAC_TRBV7-9_CASSSTLAGGYNEQFF_TRBJ2-1_TRBC2
SCD10X_TCR_061	hpv3	TRAV6_CALDRARLMF_TRAJ31_TRAC_TRBV29-1_CSVVSAQVDGYTF_TRBJ1-2_TRBC1
SCD10X_TCR_062	hpv3	TRAV23/DV6_CAATGANSKLTF_TRAJ56_TRAC_TRBV5-4_CASRTIGGTEAFF_TRBJ1-1_TRBC1

Table 5.4 List of TCR candidates for in-vitro cloning and validation.

5.4 CONCLUSION AND DISCUSSION

Overall, through the integration of the high-throughput peptide-loaded cleavable-SCT (pCSCT) platform with single-cell multi-omics analysis, we comprehensively profiled broad antigen-specific CD8 T cell responses against common viruses and oncogenic HPV16. This approach holds promise for broad applications in immune response phenotyping against diverse antigens, potentially facilitating comprehensive TCR discovery for therapeutic purposes.

We observed a diverse phenotypic distribution among antigen-specific CD8 T cells targeting CMV, influenza, EBV, and HPV, which may in part be influenced by the specific HLA restriction and epitope. Several observations align with existing literature. Notably, CMV infection often leads to memory T cell inflation, characterized by sustained high T cell responses and more limited contraction post-primary infection, resulting from restimulation by virus reactivation as well as bystander activation[20]. In our dataset, CMV-specific CD8 T cells were predominantly cytotoxic. Influenza-specific cells were primarily memory cells, suggestive of prior exposure and immune memory development. However, apart from the virus origin of the epitope, TCR sequence may also contribute to these observations. Follow-up analyses will focus on TCR repertoire analysis against these viral antigens.

While we successfully identified HPV-specific CD8 T cells from PBMCs, it's suggested that the majority of these cells reside within the tumor microenvironment (TME) in tissue, with their frequency in the blood being much lower[8,21]. By narrowing down the list of immunogenic epitopes, the peptide-loaded CSCT library can be easily applied for the detection of antigen-specific cells from any specimen type, in addition to PBMCs. Our ongoing direction involves applying the same approach to tumor-infiltrating lymphocytes (TILs) to map and characterize TCR clonotypes that are enriched in tissue. Additionally, we aim to discern the longitudinal

dynamics of HPV-specific immune response through screening samples collected from multiple treatment timepoints[12]. Paired immune profiling results, including Adaptive TCR repertoire analysis and ELISPOT assay, will be utilized for cross-validation of the observed results. Clinical observations of patients, such as viral titers, vaccine doses, responders versus non-responders, and tissue staining results, will need to be considered before drawing any clinical outcome conclusions.

5.5 METHODS

5.5.1 *HPV-vaccinated individuals and samples*

Patients were enrolled in a clinical trial for the assessment of VGX-3100 in CIN2/3 associated with HPV-16 and HPV-18.[12] Women aged 18–55 years with histologically confirmed HPV-16-positive or HPV-18-positive CIN2/3 were recruited from seven countries. Enrolled patients were randomized to receive 6 mg VGX-3100[12,13] (Inovio Pharmaceuticals, 3 mg plasmid targeting HPV-16 E6 and E7, and 3 mg plasmid targeting HPV-18 E6 and E7) or placebo (1 mL sterile water), given intramuscularly using electroporation device CELLECTRA (Inovio Pharmaceuticals) at study weeks 0, 4, and 12. On week 12, interim colposcopy was done after the third dose. The primary endpoint visit was on week 36, where patients with colposcopic evidence of residual disease underwent standard therapeutic resection. Whole blood was collected at baseline, weeks 6, 14, 24, 36, and 88 for isolation of peripheral blood mononuclear cells (PBMCs). PBMCs were cryopreserved before use.

5.5.2 *Large-scale preparation of peptide-HLA complex libraries*

HPV peptide-MHC (pMHC) libraries were prepared in multiple methods. Single chain trimers (SCTs) for HPV antigens (HPV-SCTs) were provided by PACT Pharma[6]/7/24 2:12:00 PM.

Peptide-loaded cleavable SCTs (pCSCTs) for HPV, CMV, EBV, Influenza antigens were prepared following protocols described earlier in Chapter 4. Specifically, HLA-A*02:01, A*11:01 pCSCTs were prepared via 'two-step' protocol, while pCSCTs for other HLA alleles were prepared using 'one-step' protocol.

5.5.3 *Barcoded-dextramer generation and cell staining*

SCT or pCSCT pMHC dextramers were individually DNA barcoded using dCODE Klickmers (dCODE Klickmer, Immudex). Briefly, each pMHC monomer was mixed with barcoded dCODE-PE-dextramer at a ratio of 20 ligands per dextran and incubated for at least 1h on ice before adding biotin (100 uM) to block free binding sites. Dextramer cocktails were prepared by mixing 90 unique SCT and pCSCT pMHC dextramers freshly before cell staining. PBMCs from patients and donors (purchased from CTL ePBMC) were thawed for CD8 T cell enrichment (Human CD8 T cell Isolation Kit, Miltenyi Biotec) according to the manufacturer's protocol then incubated with Human TruStain FcX blocking reagent (422302, BioLegend) for 10 min at 4 °C before wash. Each tube of CD8 T cells were stained with a unique combination of DNA-barcoded TotalSeq-C anti-human hashtag antibody (Customized, Clone LNH-94 only, BioLegend) to label each sample's ID and timepoint for 15 mins on ice and washed 3 times. Then cells were pooled and stained with a cocktail of dextramers for 25 min on ice in the presence of herring sperm DNA according to the manufacturer's instructions, with individual dextramer concentration at 1.1 nM. Cells were washed three times before surface staining using a cocktail of CD8, Apotracker™ Green viability dye, and DNA-barcoded Biolegend Totalseq panels (CD45RA, CD45RO, CCR7, CTLA4, CD25, CD127, CD8_C0046) for 15 min on ice and washed 3 times before sorting.

5.5.4 *10X genomics single cell sequencing*

Cells were gated on Single/live/CD8-positive population and exclude populations from Uncleaved CSCT template. Three sub-populations were sorted into separate tubes containing FACS buffer (PBS, 2%FBS, 2mM EDTA and 10mM HEPES) using a Sony MA900 cell sorter: dextramer-PE single positive CD8 T cells, dextramer-PE and tetramer-APC (A*02:01 CMV-NLVP SCT tetramer) double-positive CD8 T cells (representing A*02:01 CMV-NLVP cells), dextramer negative CD8 T cells. A*02:01 CMV-NLVP cells were spiked into the final cell mixture for positive control purpose. Dextramer positive and dextramer negative cells were immediately were pelleted, resuspended and loaded into 10X Chromium reactions at approximately 1:1 ratio for single cell RNA sequencing (scRNA-seq). GEX, VDJ and Surface Protein libraries were generated using Chromium Next GEM Single-Cell 5′ kits v2 (10X Genomics) according to the manufacturer's protocol. Libraries were sequenced on an Illumina NovaSeq at a read length of 26x90 bp.

5.5.5 *Single-cell sequencing data processing*

Transcriptome, TCR, surface protein levels and antigen specificity were simultaneously analyzed for each cell. Raw data were processed via Cell Ranger Single-Cell Software Suite (v3.1.0, 10X Genomics) using GRCh38 as a reference. Cells that fit any of the following filters were excluded due to low quality: n-counts <1000 or >11,000, n-genes <300 or >3500, mitochondrial percentage >10%. After filter, there were 14,896 cells. Gene counts for each cell were normalized by total expression, multiplied by a scale factor of 10,000 and transformed to log scale. TCR sequencing result was integrated via Scirpy package. Unique clonotypes were defined by the combination of TCR α and β sequences including variable genes and CDR3 sequences. Cells with

orphan α or orphan β chains were treated as unique clones even when they share one chain with other fully paired clones.

5.5.6 *Single CD8 T cell phenotype assignment*

Single cells were assigned phenotypes by clusters determined through the leiden algorithm. Leiden group that highly expressed mitochondrial genes were further removed. Phenotype associated transcripts were acquired from literature as follows: Naïve/memory: LEF1, TCF7, CCR7. Memory: IL7R, GZMK. Cytotoxic: PRF1, GZMB. Exhaustion: TIGIT, PDCD1.

5.5.7 *Sample demultiplexing using hashtags*

The UMI reads for each hashtag were first normalized to determine the threshold (ranging from 20% to 50%) for considering a cell as positive for that particular hashtag. This positive gating threshold was calculated as the percentage of reads for a given hashtag relative to the total hashtag reads for that cell. Cells were then assigned to one or more specific hashtags if they expressed that hashtag above the positive gating threshold level. By deconvoluting the combination of hashtags for each cell, the cells could be mapped back to their corresponding individual sample and timepoint. Any cells that had an ambiguous or indeterminate hashtag combination, not clearly matching one sample, were left "unassigned" (265 cells were unassigned). After sample assignment, batch correction (based on bbKNN) was performed between two sample sources (HPV-vaccinated clinical samples, and the purchased donor ePBMC samples).

5.5.8 *Antigen demultiplexing using dextramers*

For each cell, we calculated the number of unique molecular identifiers (UMIs) for each dextramer, and the percentage of each dextramer. We assigned each cell an antigen only if their UMI count was > 5 and the UMIs specific for that dextramer occupied > 25% of that cells' dextramer reads.

Antigens were then assigned by the maximally mapped dextramer for each cell. Ambiguous cells that didn't assigned with any dextramer were left "unassigned". To correct for potential dextramer assignment outlier or missing dextramer assignments for cells within a clonotype, we made clonotype-level antigen assignment adjustment. Specifically, we assigned all cells within one clonotype the antigen specificity of the majority of cells from that clone, with the criteria: if more than 25% of the cells within a clonotype were assigned to the same antigen, then all cells in that clone were adjusted to be specific for that antigen. Otherwise, the final antigen assignment was defined by the original single-cell dextramer-based assignment without adjustment. For example, for clone A with a clone size of 30, if only 2 cells were initially assigned antigen A by dextramer, we would keep only those 2 cells assigned to antigen A without assigning other cells in the clone, since the percentage of assigned cells is less than 25%. However, for clone B with a size of 20, if 17 cells were initially assigned antigen B by dextramer, passing our >25% filter, we would assign all 20 cells within this clonotype to antigen B.

Additional single cell analysis methods refer to Chapter 3.

5.6 FIGURES

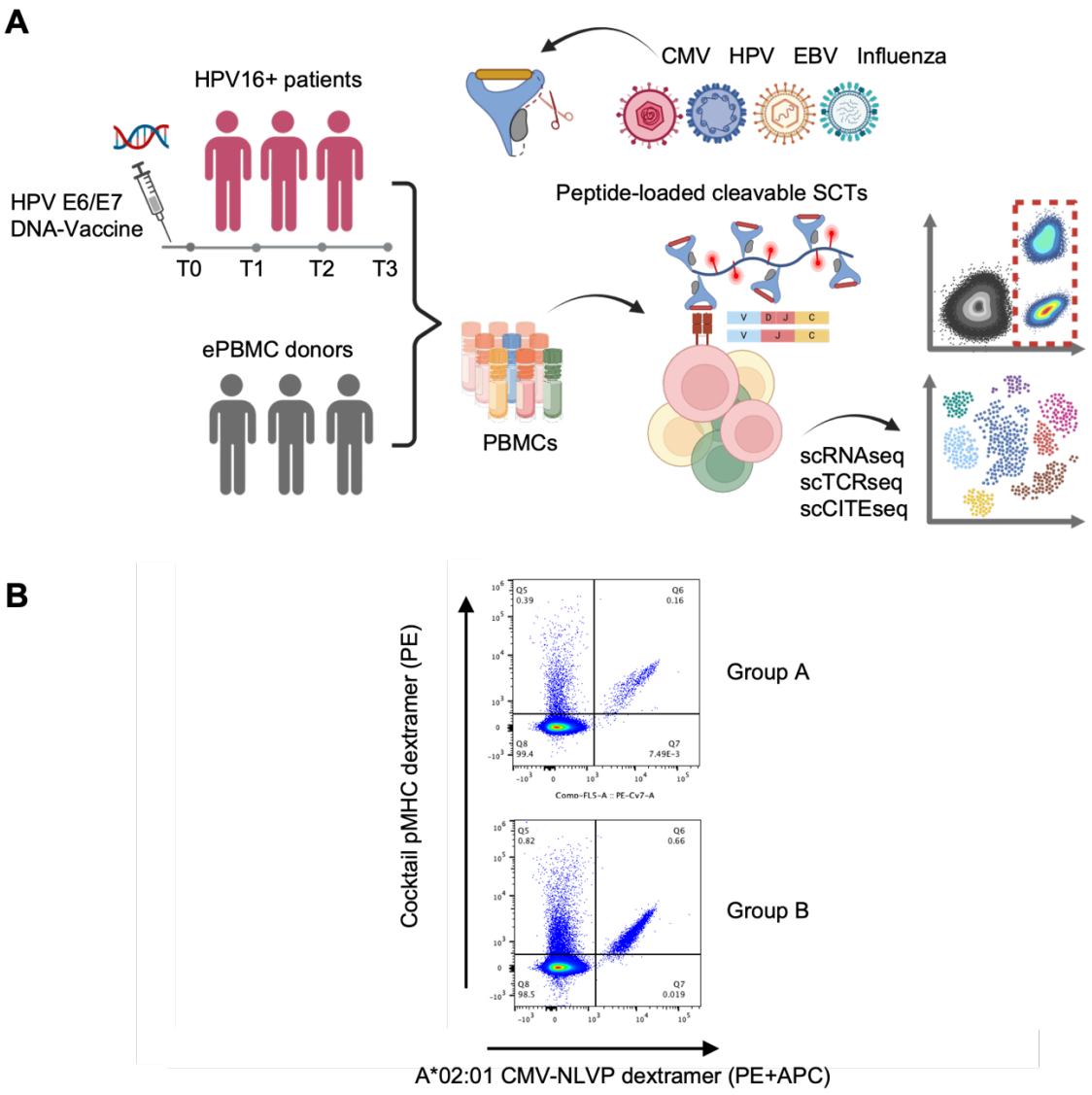


Figure 5.1 Experimental design

A. Schematic illustration. PBMCs from HPV16-positive, vaccinated patients and irrelevant donors were screened. DNA-barcoded, peptide-loaded cleavable SCTs (pCSCTs) representing immunogenic epitopes from common virus (CMV, EBV, influenza, and HPV) were prepared for antigen-specific CD8 T cell staining. Dextramer positive and negative CD8 T cells were both sorted for single cell RNAseq, TCRseq, and CITEseq.

B. Flow cytometry results of antigen-specific CD8 T cell staining for two groups of PBMCs. Cells were gated from lymphocytes/single/live/CD8+/CSCT_template_CMV- populations. X-axis: HLA A*02:01 CMV-NLVP tetramer (APC) (positive control). Y-axis: pCSCT-dextramer (PE) cocktail.

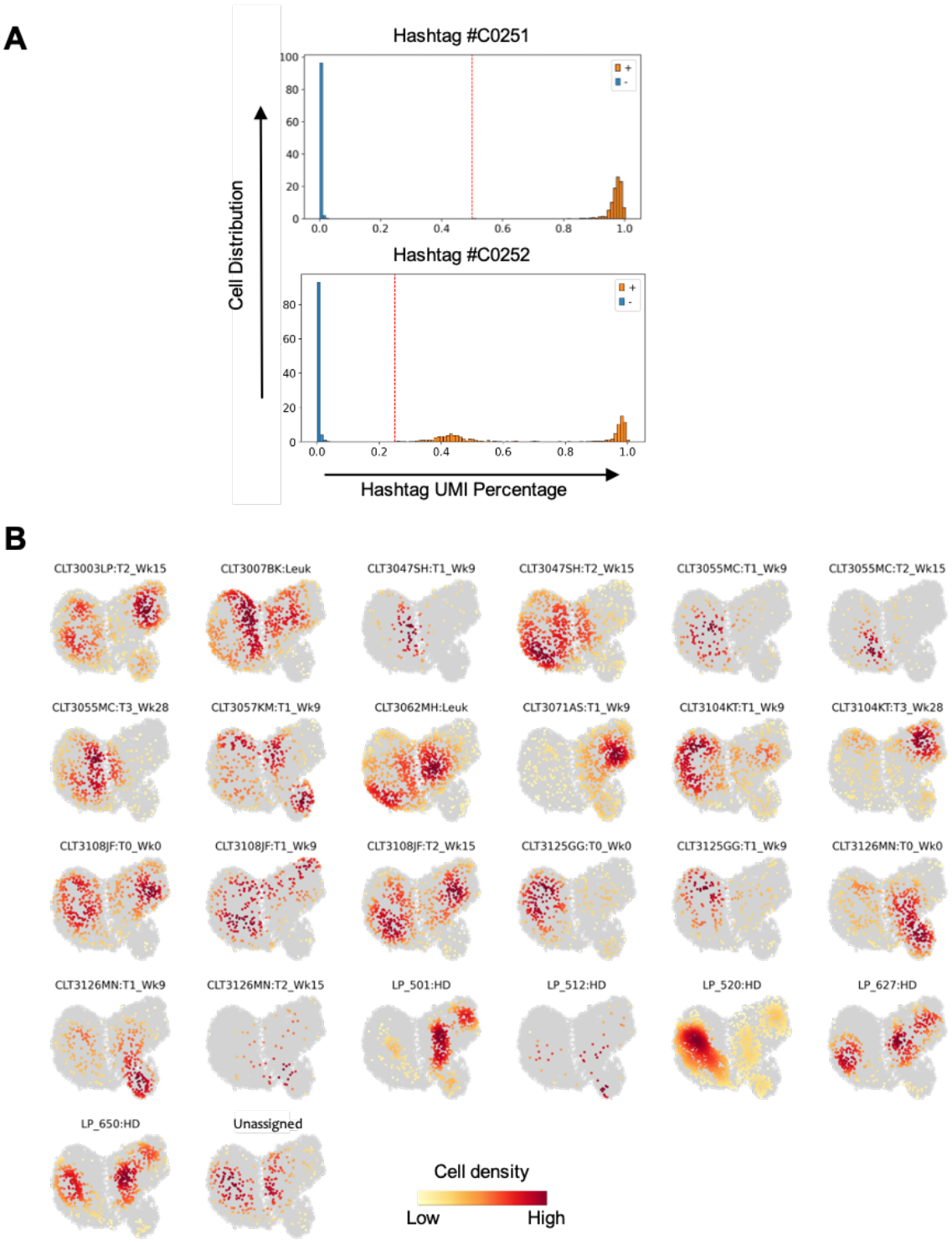


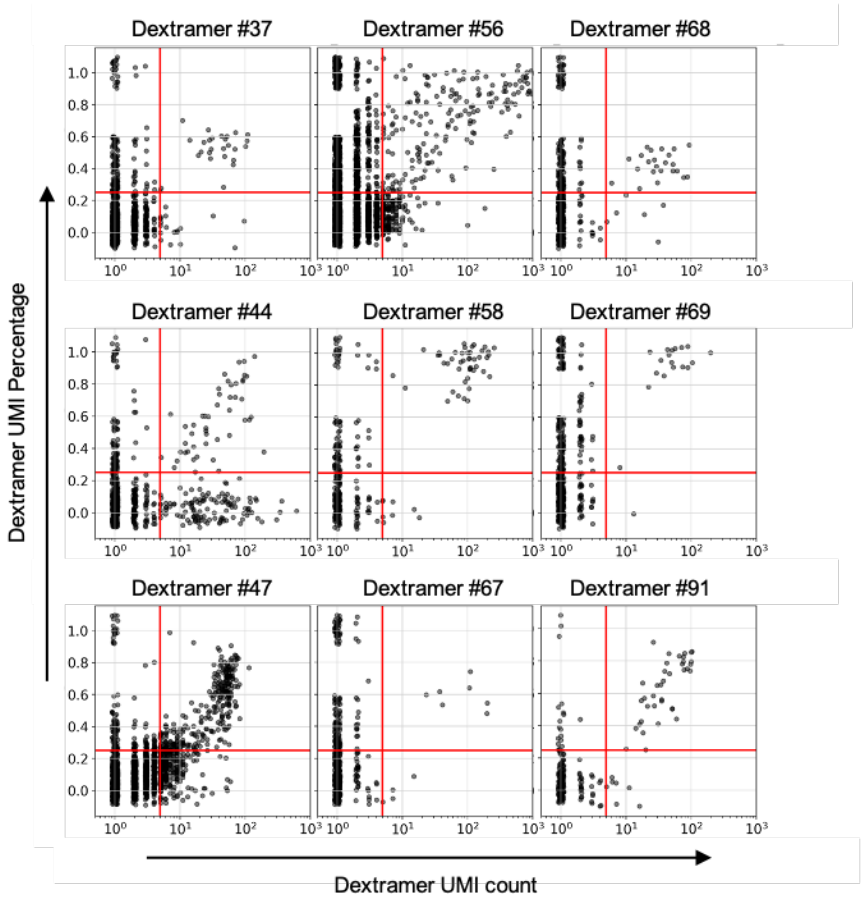
Figure 5.2 Technical methods for sample assignment based on hashtags

A. Histograms show representative hashtag signals for sample demultiplexing. The red dashed line indicates the positive threshold for each hashtag. Orange bars represent hashtag-

positive cells, and blue bars represent hashtag-negative cells. The hashtag #C0251 was used alone for one patient sample, showing a single peak for positive signal. The threshold of 0.5 indicates that the unique molecular identifier (UMI) associated with hashtag #C0251 must occupy >50% of a cell's total hashtag reads to be assigned as #C0251-positive. For hashtag #C0252, which was used alone and in combination with another hashtag for cell labeling, two positive signal peaks are observed with a threshold of 0.25.

B. UMAP density map for sample demultiplexing result. Single cells assigned to each sample (patient with timepoint, or donor) were highlighted in each panel. Legend on the bottom.

A



B

Epitope	HLA	antigen	NetMHC (nM)	SCT expression	Peptide-loaded CSCT		
					pCSCT successful	Cell count	Positive individuals
GILGFVFTL	A0201	Flu-M1	7.27	No	Yes	519	16/16 (100%)
CLGGLLTMV	A0201	EBV-LMP2A	75.6	No	Yes	226	15/16 (93.8%)
GLCTLVAML	A0201	EBV-BLMF1	62.23	No	Yes	46	8/16 (50%)

Figure 5.3 Technical methods for antigen assignment based on DNA-barcoded dextramers

A. Scatter plots show antigen assignment for representative dextramer signals. The x-axis represents the number of UMIs mapped to the specific dextramer (density). The y-axis represents the dextramer dominance, defined as the percentage of the cell's dextramer-associated UMIs that mapped to the antigen's specific dextramer. The red lines indicate the positive thresholds for

UMI count (>5) and percentage (>25%). Cells positive for both criteria are assigned to the corresponding antigens.

B. List of epitopes that are not expressible by conventional SCT, but succeed in pCSCT format and identified antigen-specific T cells from multiple individuals.

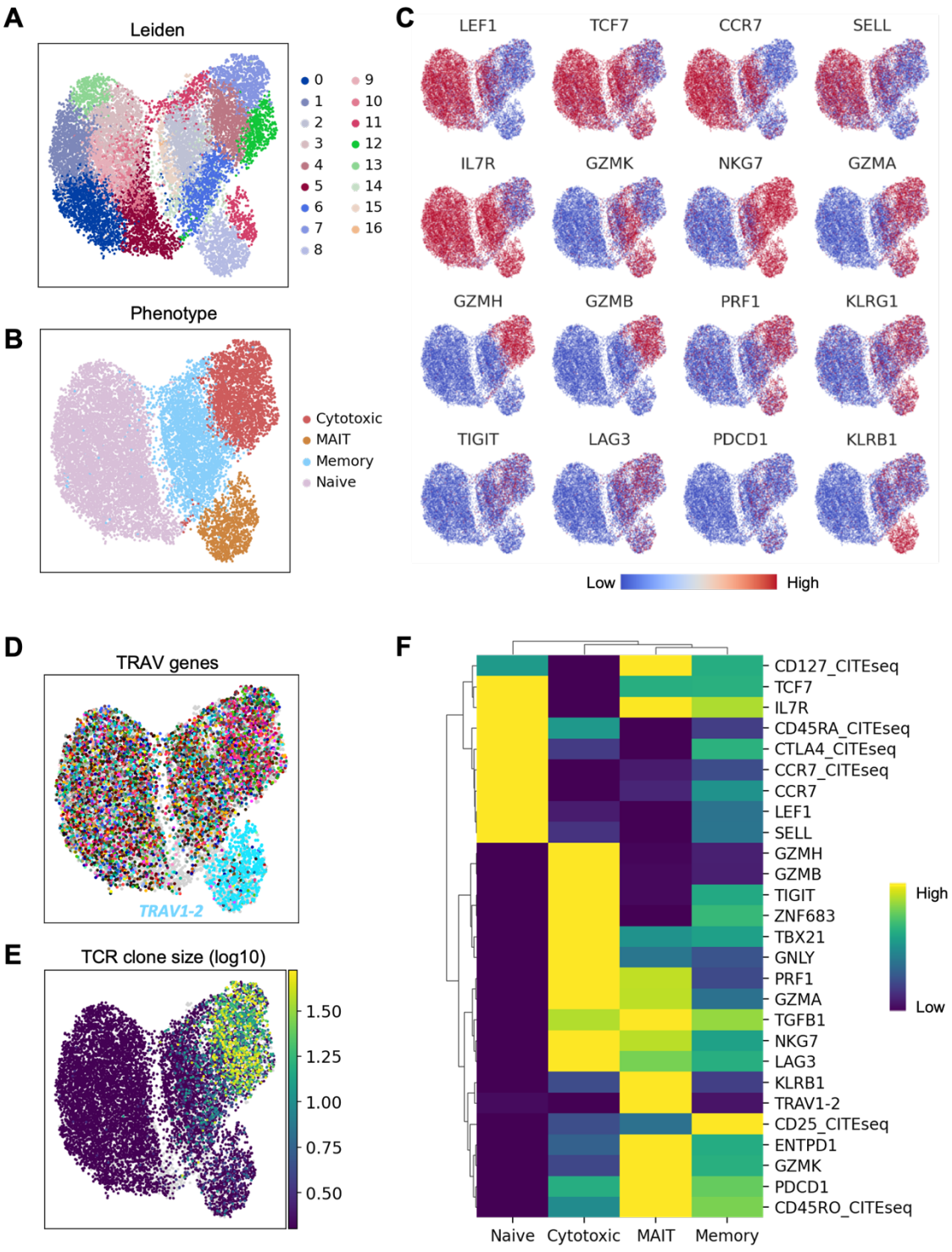


Figure 5.4 UMAP of antigen-specific CD8 T cells

A**.** Dex+ and Dex- CD8 T cells from HPV16-positive patients and ePMBC donors were sorted for scRNAseq and TCR sequencing. Gene expression UMAP was generated and color-coded with leiden groups. Legend on the right.

B. UMAP of CD8 T cells with different phenotypes color-encoded (legend on the right).

C. Normalized expression levels of selected mRNA transcripts (legend on the bottom).

D. UMAP embedding for TRAV gene usage. Each color represents an unique TRAV gene, with light blue cells were enriched with TRAV 1-2.

E. UMAP embedding of clone size for each TCR, defined by the combination of alpha and beta V(D)J genes and CDR3 sequences.

F. Clustermap of phenotype-related genes and surface protein measured via scCITEseq. Values are row-normalized, legend on the left.

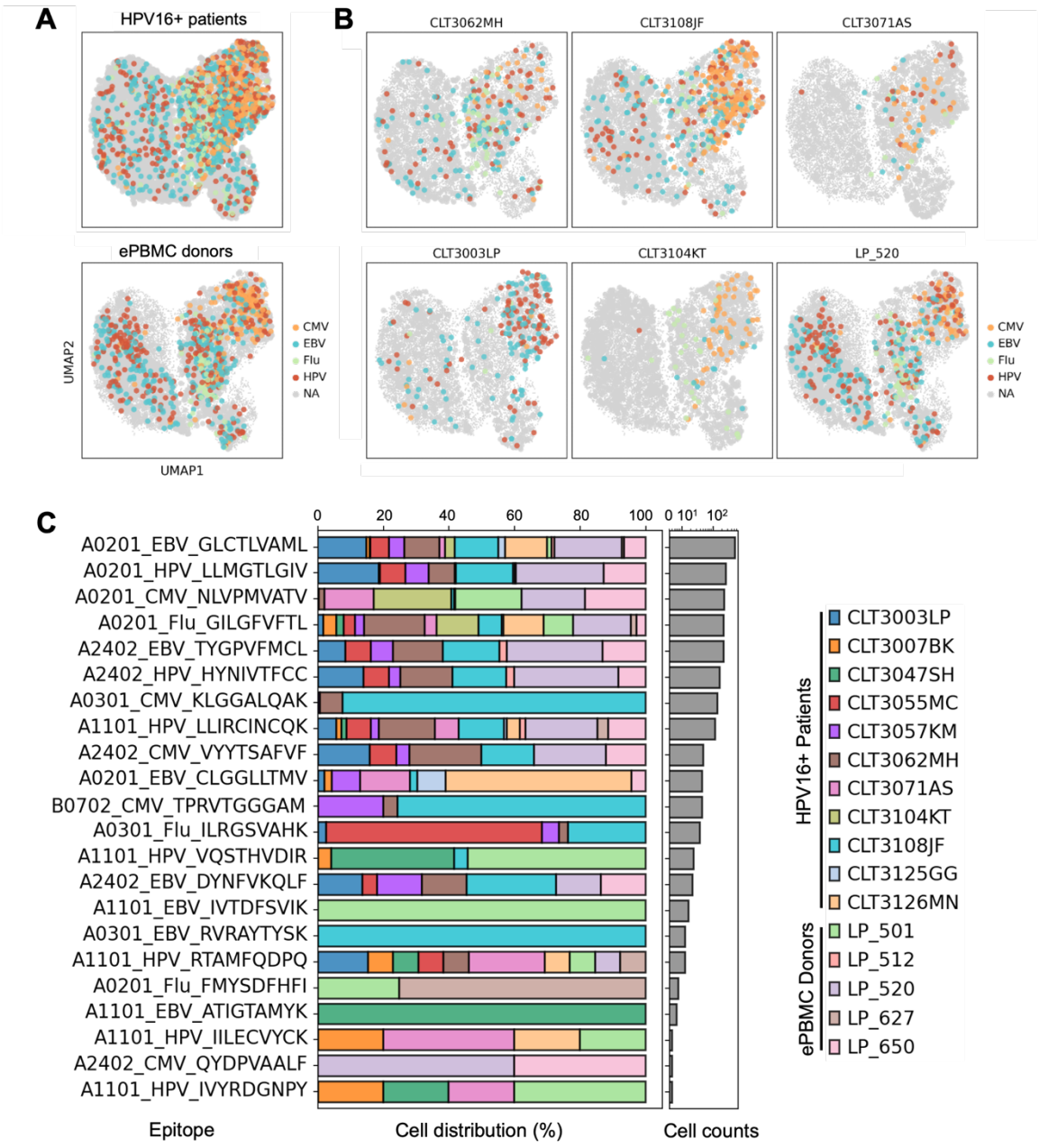


Figure 5.5 Antigen-specific CD8 T cells identified from each individual

A. UMAP of antigen-assigned CD8 T cells with each virus color-coded from HPV16-POSITIVE patients, and ePBMC donors, respectively (legend on the right).

B. UMAP of cells from selected patients and donor. Large grey dots represent pan-CD8 T cells assigned to that individual, colored dots represent antigen-assigned cells with each virus color-coded (legend on the right).

C. Cell distribution for each antigen specificity identified from each individual, with total cell count on the right (log10 scale). Only antigens assigned with more than 5 cells were plotted (legend on the right).

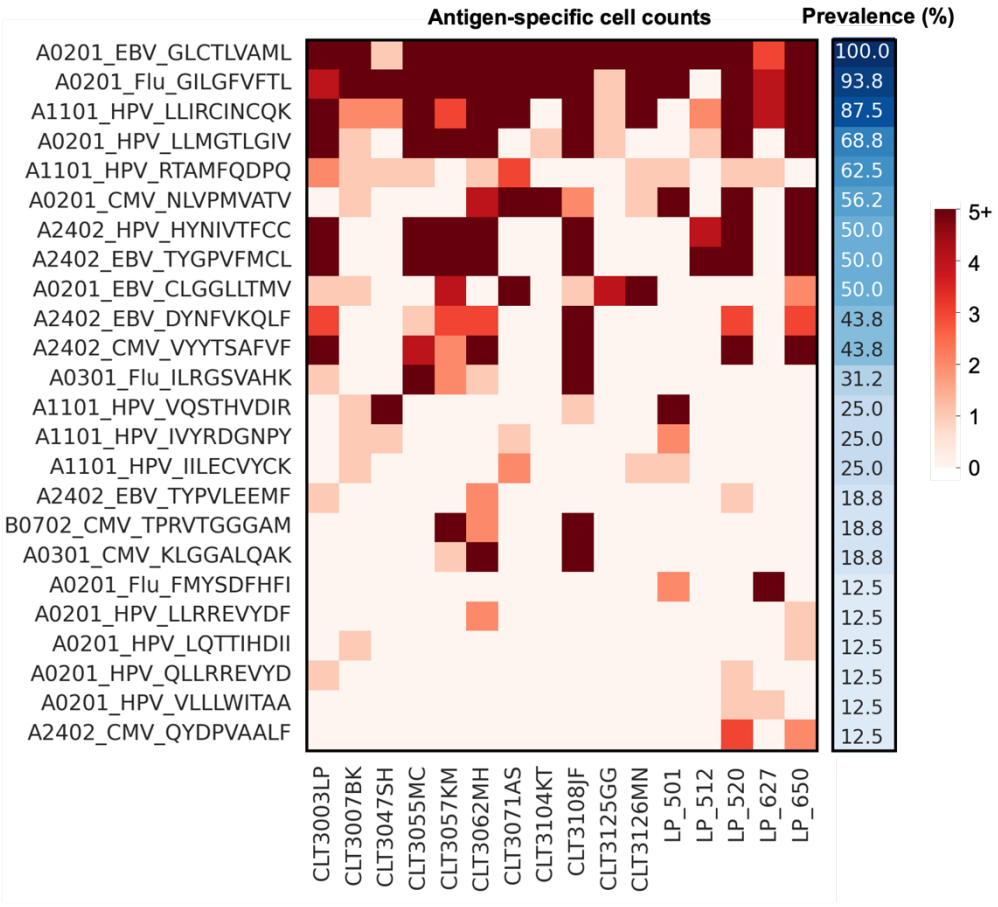


Figure 5.6 Immunodominant viral epitopes among individuals

Heatmap showing cell counts for each antigen identified from individual samples. Cell counts exceeding 5 are colored as dark red (legend on the far right). Only antigens those are detected from 2 or more individuals are included. “Prevalence %” was calculated by dividing the number of positive individuals by the total number of individuals (N = 16). For example, 50% indicates the antigen was detected in 8 individuals.

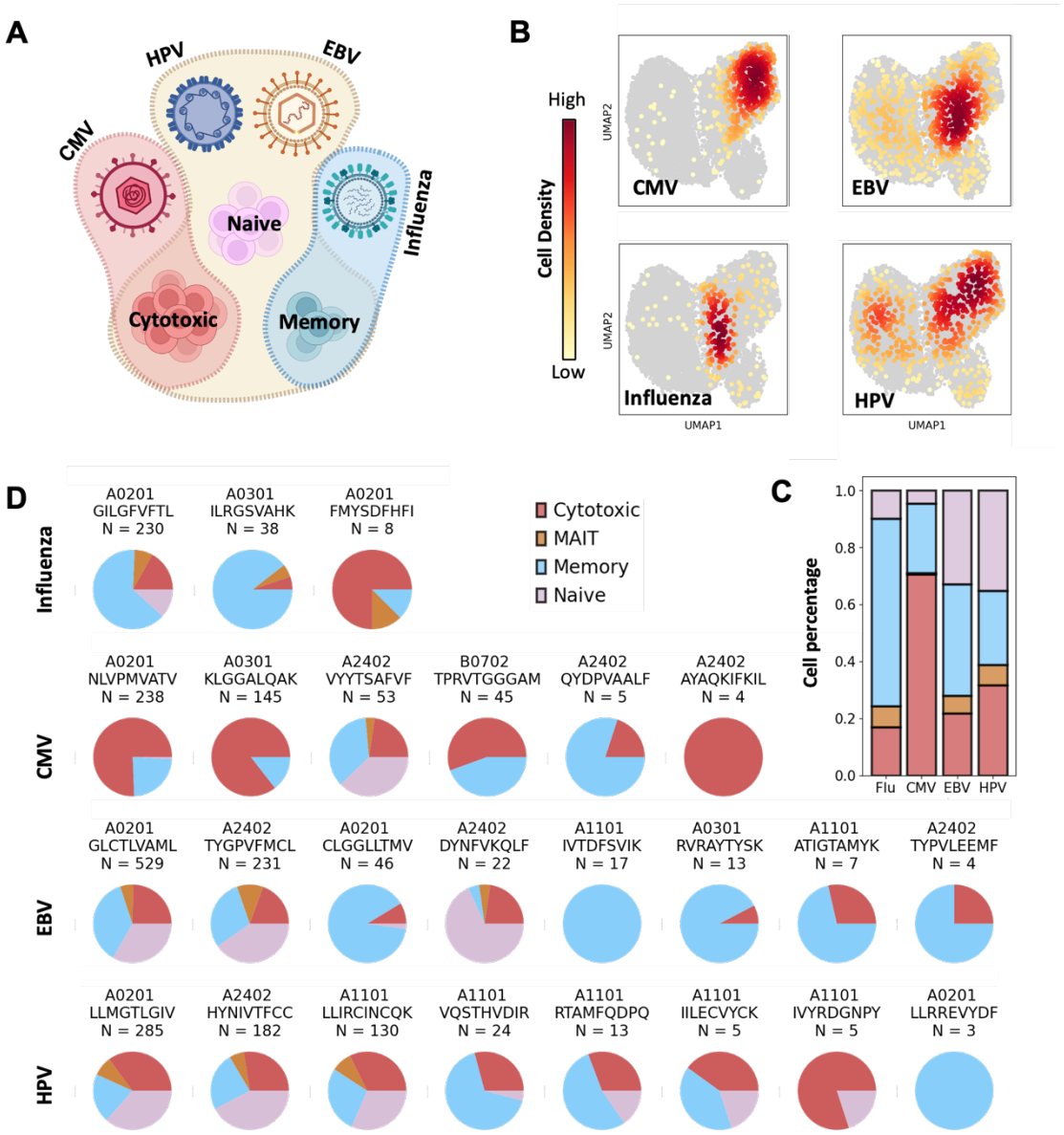


Figure 5.7 Phenotype distribution for antigen-specific T cells against different viruses

A. Schematic illustration.

B. UMAP density map for antigen-specific CD8 T cells against each virus. Legend on the left.

C. Cell distribution of phenotypes for antigen-specific CD8 T cells against each virus.

D. Cell distribution for antigen-specific CD8 T cells against each epitope. Epitopes derived from the same virus are plotted in same row with cell counts labeled. Only antigens assigned with 3 or more cells are included. Legend on the top right.

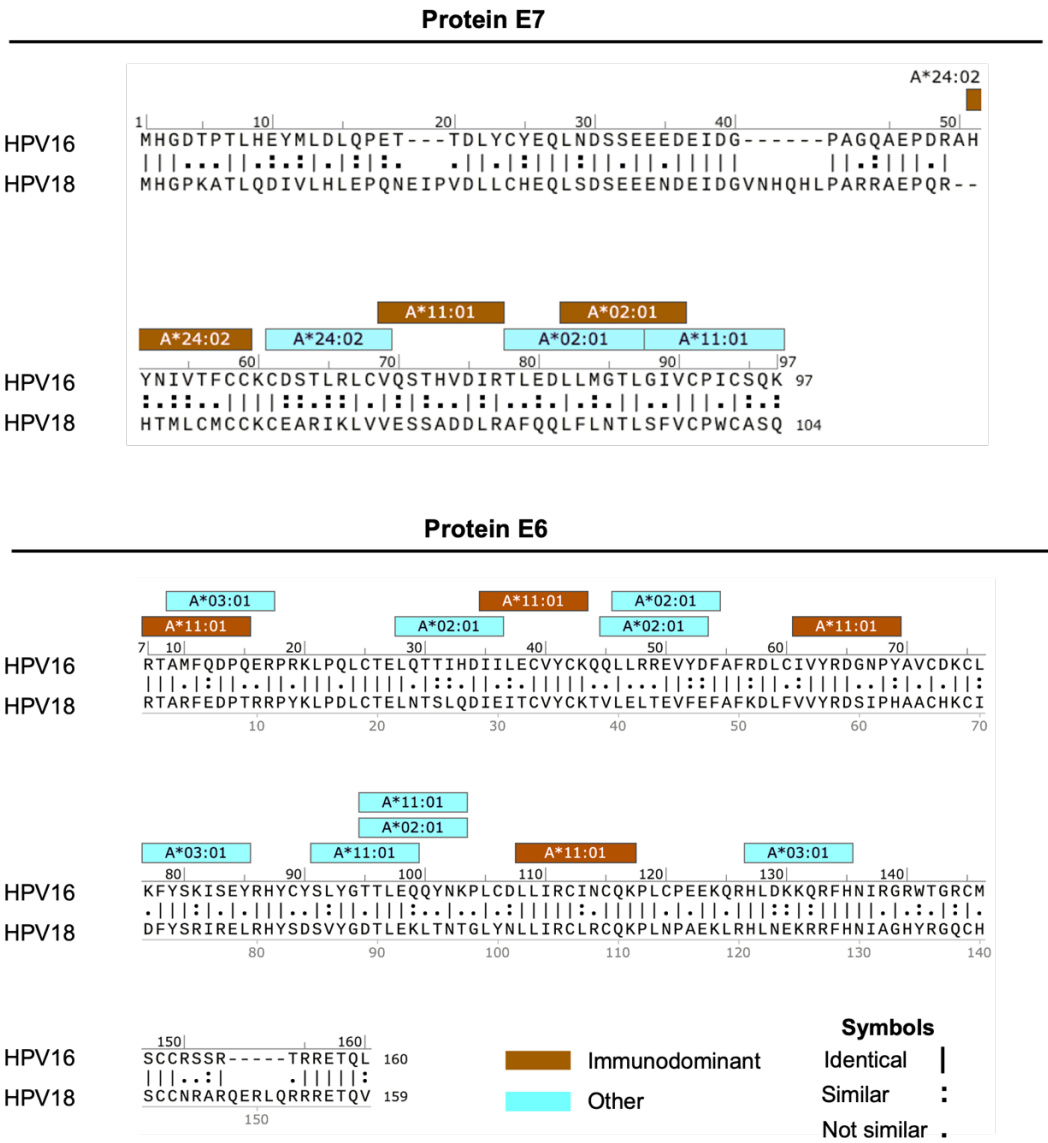


Figure 5.8 Conserved immunogenic epitopes from HPV16 E6 protein

Protein sequence alignment comparing HPV16 and HPV18 for proteins E6 (Bottom panel) and E7 (Top panel). Sequences were aligned using Snapgene's Smith-Waterman method, with symbols denoting amino acid similarity between HPV16 and HPV18 (legend located in the bottom right). Immunogenic epitopes from HPV16 are delineated by colored rectangles, each annotated with its corresponding HLA restriction. Brown rectangles denote seven

immunodominant epitopes from Table 5.3 (identified in at least 25% of tested individuals), while light blue rectangles represent all other epitopes.


A
clone_29
A0201_CMV_NLVPMVATV
clone_55
A0201_EBV_CLGGLLTMV
clone_51
A0201_Flu_GILGFVFTL
clone_94
A1101_HPV_VQSTHVDIR
clone_105
A2402_HPV_HYNIVTFCC
clone_12
A0201_HPV_LLMGTLGIV
clone_124
A1101_HPV_VQSTHVDIR
clone_144
A2402_HPV_HYNIVTFCC
clone_145
A0201_HPV_LLMGTLGIV
clone_148
A0201_HPV_LLMGTLGIV
clone_130
A0201_HPV_LLMGTLGIV
clone_119
A0201_HPV_LLMGTLGIV
clone_88
A1101_HPV_VQSTHVDIR
clone_139
A1101_HPV_LLIRCINCQK
clone_141
A1101_HPV_LLIRCINCQK
UMAP2
UMAP2
UMAP1
B
CLT3055MC
CLT3108JF
UMAP2
UMAP1
HPV-assigned cells
T0_Week0
T1_Week9/10
T2_Week15

Figure 5.9 Examples of HPV-specific CD8 T cell clonotypes for validation

A**.** UMAP of selected CD8 T cell clonotypes assigned with HPV antigens. In each panel, colored cells represent HPV-antigen assigned CD8 T cells share same TCR.

B. UMAP of HPV-specific CD8 T cells (including multiple clonotypes) from selected patients. Colored dots represent HPV antigen-assigned cells with each timepoint color-coded (legend on the right).

5.7 REFERENCES

1. Skolnik, J. M. & Morrow, M. P. Vaccines for HPV-associated diseases. Mol. Aspects Med. 94, 101224 (2023).
2. de Moura, I. A. et al. Enhancing the Effect of Nucleic Acid Vaccines in the Treatment of HPV-Related Cancers: An Overview of Delivery Systems. Pathogens 11, 1444 (2022).
3. IJMS | Free Full-Text | Therapeutic Vaccines for HPV-Associated Oropharyngeal and Cervical Cancer: The Next De-Intensification Strategy? https://www.mdpi.com/1422-0067/23/15/8395.
4. Nagarsheth, N. B. et al. TCR-engineered T cells targeting E7 for patients with metastatic HPV-associated epithelial cancers. Nat. Med. 27, 419–425 (2021).
5. Potent, Selective CARs as Potential T-Cell Therapeutics for HPV-positive Cancers - PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8415731/.
6. Yim, E.-K. & Park, J.-S. The Role of HPV E6 and E7 Oncoproteins in HPV-associated Cervical Carcinogenesis. Cancer Res. Treat. Off. J. Korean Cancer Assoc. 37, 319–324 (2005).
7. Mo, Y. et al. Prophylactic and Therapeutic HPV Vaccines: Current Scenario and Perspectives. Front. Cell. Infect. Microbiol. 12, 909223 (2022).
8. Eberhardt, C. S. et al. Functional HPV-specific PD-1+ stem-like CD8 T cells in head and neck cancer. Nature 597, 279–284 (2021).
9. Chour, W. et al. Large libraries of single-chain trimer peptide-MHCs enable antigen-specific CD8 T cell discovery and analysis. Commun. Biol. 6, 1–13 (2023).
10. Foy, S. P. et al. Non-viral precision T cell receptor replacement for personalized cell therapy. Nature 1–10 (2022) doi:10.1038/s41586-022-05531-1.
11. Puig-Saus, C. et al. Neoantigen-targeted CD8 T cell responses with PD-1 blockade therapy. Nature 615, 697–704 (2023).
12. Trimble, C. L. et al. Safety, efficacy, and immunogenicity of VGX-3100, a therapeutic synthetic DNA vaccine targeting human papillomavirus 16 and 18 E6 and E7 proteins for cervical intraepithelial neoplasia 2/3: a randomised, double-blind, placebo-controlled phase 2b trial. The Lancet 386, 2078–2088 (2015).
13. Bhuyan, P. K. et al. Durability of response to VGX-3100 treatment of HPV16/18 positive cervical HSIL. Hum. Vaccines Immunother. 17, 1288–1293 (2021).
14. Adamo, S. et al. Signature of long-lived memory CD8 T cells in acute SARS-CoV-2 infection. Nature 602, 148–155 (2022).
15. Su, Y. et al. Multi-Omics Resolves a Sharp Disease-State Shift between Mild and Moderate COVID-19. Cell 183, 1479-1495.e20 (2020).
16. Daniel, B. et al. Divergent clonal differentiation trajectories of T cell exhaustion. Nat. Immunol. 23, 1614–1627 (2022).
17. Wong, E. B. et al. TRAV1-2+ CD8 T-cells including oligoconal expansions of MAIT cells are enriched in the airways in human tuberculosis. Commun. Biol. 2, 1–13 (2019).
18. Parrot, T. et al. MAIT cell activation and dynamics associated with COVID-19 disease severity. Sci. Immunol. 5, eabe1670 (2020).
19. Lee, H., Jeong, S. & Shin, E.-C. Significance of bystander T cell activation in microbial infection. Nat. Immunol. 23, 13–22 (2022).
20. van den Berg, S. P. H. et al. The hallmarks of CMV-specific CD8 T-cell differentiation. Med. Microbiol. Immunol. (Berl.) 208, 365–373 (2019).

21. Sun, Y. et al. Evolution of CD8 T Cell Receptor (TCR) Engineered Therapies for the Treatment of Cancer. Cells 10, 2379 (2021).