High-throughput Identification and Analysis of Antigen-specific CD4+ T Cells

Rongyu Zhang

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Jame R. Heath, Chair

Evan W. Newell

Patrick S. Stayton

Program Authorized to Offer Degree:

Bioengineering

University of Washington

Abstract

High-throughput Identification and Analysis of Antigen-specific CD4+ T Cells

Rongyu Zhang

Chair of the Supervisory Committee:
James R Heath
Department of Bioengineering

We present a toolset for the high throughput detection and analysis of antigen-specific CD4+ T cells using DNA-barcoded, large libraries of single-chain trimers (SCTs) designed to mimic peptide-MHC multimers for class II human leukocyte antigen (HLA). Following platform validation, we executed an unbiased screen to capture and simultaneously analyze 2,188 CD4+ T cells with specificity to the Receptor Binding Domain (RBD) of SARS-CoV-2 Spike protein, from a longitudinal cohort of 24 HLA-DR1 matched participants. We tracked RBD-antigen-specific CD4+ T cell phenotypes out to over two years post-infection, and identified metrics for defining immunogenic class II-restricted viral antigens. We also identified human papilloma virus (HPV)-16 E6-specific CD4+ T cell receptors (TCRs) from HPV16+ patients with precancerous lesions. Those TCRs are analyzed for their therapeutic potential for treating HPV+ cancers. This platform enables detailed investigation of CD4+ T cell immune responses and can accelerate the discovery of both relevant epitopes and TCRs for immunotherapies.

Table of Contents

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to the individuals who mentored me, provided invaluable academic guidance, and supported me emotionally throughout my PhD journey.

First and foremost, I would like to express my deepest gratitude to my PhD advisor, Dr. James Heath. Jim has been an exceptional mentor, offering guidance while also granting me the freedom to explore the scientific world on my own. I truly appreciate the resources and unwavering support he provided throughout my PhD journey. I am deeply inspired by his passion for developing cutting-edge treatments to help patients with life-threatening diseases. I will always remember his words during the symposium celebrating his 60th birthday, where he encouraged us not to chase monetary success but to let compassion and a genuine desire to help others drive our work. Thanks to Jim, I never once doubted the significance of our work or its potential to make a positive impact on patients' lives. For this, I am sincerely grateful.

I also wish to acknowledge my committee members, Dr. Evan Newell, Dr. Joshua Veatch, Dr. Hao Yuan Kueh, Dr. Patrick Stayton, and Dr. Alexander Mamishev. Their guidance through various phases of my PhD journey has shaped me into a better scientist.

My gratitude extends to my mentors, Dr. William Chour, Dr. Jongchan Choi, and Dr. Yapeng Su, for their tremendous support throughout my academic journey. They have guided me in developing both wet lab and dry lab skills, discussed with me about the directions of my project, and provided advice on career development. Thank you for nurturing me into the scientist I aspired to become.

I am also grateful to my lab members. Dr. Jingyi Xie, Dr. Dan Yuan, Rachel Ng, Jingqi Qi, Conor Brennan, Sunga Hong, Dr. Andrew Webster, Yusuf Rasheed, Rick Edmark, Shuo Wang, Daniel Chen, Sarah Li, Dr. Kim Murray, Abby Anderson, Jimmi Hopkins, Dr. Lauren Jatt, Connor McDonald and Chong Xia. Your collaboration and support made this journey immensely enjoyable. I'd also like to thank former lab members who contributed to my thesis project: Michaela McKasson, Vanessa Gutierrez, Simran Sidhu, Megan van Meurs, Kelly Wang, and Nello Gu.

A special thanks to my friends, who always believed in me and encouraged me throughout my PhD. I wouldn't be able to do it without our weekend hangouts, hikes, skiing trips, long chats, and parties. I am especially grateful to Guan Wang, Dr. Qinghua He, Dr. Jingyi Xie, Dr. Dan Yuan, Caihong Huang, Dr. Qin Wang, Yu Jung Shin, Rachel Ng, Yunxuan Li, Jasmine Zhou, Ziyu Liu, Teng Liu, Yaping Shi, Xinming Liu, and so many others who supported and inspired me to keep pursuing my goals.

I am deeply thankful to my partner, Zach Cao, for his unconditional support, especially during the demanding final year of my PhD.

Finally, to my dearest parents—this achievement would not have been possible without your unwavering love and support. You have always been there for me, cheering me on and encouraging me to follow my dreams. My success is just as much yours as it is mine. I feel incredibly fortunate to have you as my parents.

DEDICATION

This dissertation is dedicated to my grandfather, whose unconditional love gave me the courage to explore the world. I know he will always be with me on every adventure I undertake.

Chapter 1: Introduction

1.1 CD4+ T cell

CD4+ helper T cells play multiple roles in the immune response to disease. They are activated following exposure to foreign or tumor antigens restricted to class II MHC and are known to enhance the functions of antigen-presenting cells, increase CD8+ T cell effector differentiation, and drive B cell activation and antibody affinity maturation[1]. In cancer settings, they can help sustain anti-tumor immune efficacy during cancer progression and even exhibit anti-tumor cytotoxic functions[2,3]. Understanding the specific and dynamic roles of antigen-specific CD4+ T cells is thus central to mapping immune responses within virtually any disease setting, and may help provide insights into epitopes most relevant for vaccine design, as well as TCRs with therapeutic potential for TCR-engineered adoptive cell cancer immunotherapies.

1.2 CD4+ T cell and class II pMHC

Although numerous methods for large-scale screening of antigen-specific CD8+ T cells have been established[4,5,6,7,8,9],[10], it has been more challenging to develop high-throughput platforms for the discovery and characterization of antigen-specific CD4+ T cells. This is largely related to class II MHC antigen presentation. Class II HLA molecules are composed of two polypeptide chains that assemble to form a peptide-binding groove. Unlike the closed cleft of class I MHC, which accommodates 9-12-mer peptides with dominant anchor residues, the open cleft of class II molecules binds longer peptides (13-24 amino acids) that may extend beyond the groove[11,12]. This structural difference means that class II peptide antigens consist of a core sequence within the binding groove with additional flanking residues, whose influence on TCR binding and CD4+ T

cell activation is not well-understood[13]. The more promiscuous nature of class II peptide binding and the limited availability of experimental binding data have hindered the development of accurate computational predictions.

1.3 High throughput identification of antigen-specific CD4+ T cells

One approach for the high-throughput identification of antigen-specific CD4+ T cells involves sorting cells based on their cytokine profile or activation markers following peptide-pulsed stimulation[14,15,16,17]. While this method is effective for obtaining TCR sequences, it depends on the functional capacity of T cells and results in the loss of phenotypic information. To preserve the phenotypic integrity of CD4+ T cells, direct capture of non-expanded T cells from blood or tissue using class II pMHC multimers is the gold standard, but is limited to relatively small libraries ($n \leq 15$) of pMHCs due to limited yield and finite stability[18,19,20,21,22,23,24]. Since computational predictions of antigen-MHC binding are not yet reliable, technologies capable of constructing larger-scale pMHC libraries—thereby enabling unbiased high-throughput screening of antigen-specific CD4+ T cells—would be of high value.

1.4 CD4+ T cells in cancer immunotherapy

T cell therapy represents a significant advancement in the field of immunotherapy, offering a highly specific approach to targeting cancer antigens[25]. By harnessing the antigen recognition capabilities of T cells, T cell therapy enables cancer-specific immune responses with minimal off-target effects. Strategic selection of target antigens holds the key to minimizing on-target off-cancer effects, thereby significantly enhancing the safety profile and reducing side-effects of the treatment. The ability of T cell therapy to specifically recognize cancer antigens expands its applicability to a wide range of cancers[26]. Unlike CAR-T cells, which rely on universal target

molecules present on cancer cells, antigen-specific T cell therapy can effectively target solid tumors. This specificity is achieved by the T cells' ability to distinguish between cancerous and healthy cells based on the recognition of small mutated peptides presented by HLA molecules[26]. Consequently, T cell therapy holds immense clinical promise for the treatment of solid tumors.

CD8+ T cells possess the ability to directly eliminate tumor cells through their cytotoxic effector functions. The potential of utilizing tumor-killing CD8+ T cells in clinical settings has been explored. Ongoing clinical trials are investigating the efficacy of CD8+ T cells targeting HPV-16+ cancers and other malignancies[27]. Despite promising results from some animal studies and clinical trials, the efficacy of single clone CD8+ T cell therapy has not met expectations. Cancer cells have developed various mechanisms to evade CD8+ T cell surveillance, including loss of HLA heterozygosity, downregulation of HLA molecules, and other immune evasion mechanisms[28]. Additionally, administered CD8+ T cells have been reported to have a short lifespan in patients, necessitating repeated cell infusions or resulting in diminished treatment outcomes[27]. CD8+ T cells rely on CD4+ T cell-mediated stimulation for prolonged proliferation and survival. Hence, combination therapies which include multiple clones of cancer-specific CD8+ and CD4+ T cells have been proposed as the next frontier in T cell therapy.

However, the implementation of CD4+ T cell therapy presents several challenges. CD4+ T cells exhibit a significantly larger repertoire compared to CD8+ T cells, making it more challenging to identify antigen-specific CD4+ T cells[29]. Furthermore, class II HLA alleles display greater heterogeneity, resulting in a more diverse array of alpha and beta combinations in the population[30]. This diversity poses a significant challenge in establishing CD4+ T cell treatments that are effective across a broad patient population. Moreover, current technologies in the field are limited in their ability to scale antigen-specific CD4+ T cell discovery. Cancer cells in patients present

thousands to millions of mutations or viral antigens, and targeting only a few of these antigens severely limits the efficacy of the therapy[31]. Cancer cells can downregulate the expression of target antigens to evade immune recognition. Incomplete elimination of cancer cells often leads to relapse[29]. Therefore, the development of an effective and comprehensive CD4+ T cell therapy requires the establishment of high-throughput screening assays to identify antigen-specific CD4+ T cells.

1.5 Dissertation overview

In this work, we discuss the development of a high-throughput platform for the detection and analysis of antigen-specific CD4+ T cells. Our approach starts with class II pMHCs expressed as libraries of single-chain trimers (SCTs). Class II SCT libraries support both unbiased proteome-wide screening and more targeted approaches, thus enhancing the breadth and precision of CD4+ T cell discovery. We validated the performance of class II SCTs across multiple HLA alleles based on their capacity to capture known antigen-specific CD4+ T cell clonotypes, as well as their ability to capture novel antigen-specific CD4+ T cells from healthy donor bloods. We then expanded the approach to screen longitudinal samples from 22 SARS-CoV-2 patients, using a high-throughput assay to identify CD4+ T cell clones specific to the receptor binding domain (RBD) of the Spike protein, and to analyze the phenotype dynamics of those same T cells from acute infection to out to 2 years post-infection. Finally, we constructed a class II SCT library targeting the E6 and E7 oncogenic proteins of HPV-16 and isolated HPV-16-specific CD4+ T cells from precancerous patients previously treated in therapeutic vaccine trials. Through a series of preclinical validation assays, we identified five HPV-16- E6 specific CD4+ TCRs. This high-throughput platform serves as a powerful tool for enabling the in-depth analysis of disease-specific CD4+ T cell immune responses, and for the discovery of immunogenic class II-restricted T cell epitopes, and for the

discovery of antigen-specific CD4+ T cells with potential applications in TCR-engineered adoptive cell immunotherapies.

Chapter 2: SCT technology and development

2.1 Introduction – pMHC and the Class II SCT technology

2.1.1 pMHC and CD4 T cell interaction

CD4 TCR interacts with peptide presented on the major histocompatibility complex (pMHC) molecule on the surface of professional antigen presenting cells (APCs) such as dendritic cells, macrophages, and B cells. Upon binding to the pMHC molecule, the intracellular domain of the TCRs on CD4 T cell is phosphorylated and a signaling cascade starts to activate the T cell[32]. The antigen presented by class II MHC are generally from extracellular sources, such as the degradation products of bacteria or viruses during an infection, and mutated peptides shed from cancer cells. CD4 TCR recognition to pMHC is unique and it makes pMHC molecular a great tool to identify and capture antigen-specific CD4+ TCRs.

pMHC is comprised of three components, the alpha and beta chains of the MHC molecule and the peptide that binds to the binding groove held by the alpha and beta chains. There are two types of pMHC – the class I and class II MHC. In class I MHC, the beta chain is constant across different alleles, and termed the beta-2M (B2M). Class I MHC holds a shorter peptide ranging from 8-12 amino acids. On the other hand, class II MHC alleles have variable alpha and beta chains, making the combination more complex. Class II MHC is also more promiscuous for the peptides that bind into the peptide binding groove. Class II MHC can bind peptides from 13-25 amino acids. Unlike class I MHC where the two ends (N- and C-terminus) of the binding groove are closed so that the peptides sit in a tight binding pocket, class II MHCs have an open binding pocket where the two ends of the peptide and flank freely. This unique property makes the recognition of CD4 TCRs

more promiscuous as they can bind a family of peptides of varying length as TCR can recognize the core peptide binding registrar shared in the peptide family[13].

2.1.2 Class II HLA alleles

MHC is a general term used across mammalian species. Human Leukocyte Antigen (HLA) is the specific terminology when referring to human MHC alleles. Class II HLAs are classified into three major subtypes: the DR, DP, and DQ alleles. The combinations of the alpha and beta alleles for each subtype are complex. We inherit an alpha and a beta allele from each of our parents, making it two alpha and beta chains for DR, DP, and DQ alleles. Hence, for each subtype, there are four possible combinations assuming none of the alpha or beta chains are homogeneous. The total possible combinations of all class II alleles in a person are twelve.

The DR alleles are unique as the alpha chain is a constant – DRA1*01:01. Based on the database, the top five common beta alleles for DR are DRB1*07:01, DRB1*15:01, DRB1*03:01, and DRB1*01:01[33,34]. The U.S. population has a dominant DP alpha allele DPA1*01:03, and the top three beta alleles, DPB1*04:01, DPB1*02:01 and DPB1*04:02 cover 80% of the population[34]. DQ alleles, on the other hand, are more complicated as the top alpha and beta chains are diverse.

2.1.3 Previous methods to express pMHCs

As class II pMHC molecule serves as a great tool to capture antigen-specific CD4 T cells, researchers have used various forms of pMHC to identify the target T cells. Previous studies have reported several ways to express soluble pMHCs to identify antigen-specific CD4+ T cells based on multimers. Methods evolved from the UV-liable peptide exchange from class I pMHC production are modified and applied to class II pMHC expressions where the peptide exchange occurs after enzymatic cleavage of the placeholder peptide, followed by targeted purification by

tag attached to the exchanged peptide[21,]. Since the class II MHC is more stable than class I MHC and hence could be expressed without the peptide, researchers have reported expressing the alpha and beta chains of class II MHC with an empty pocket followed by exogenous peptide loading[22,18,23,36]. Further engineering of pMHC led to linking the components together for easier manipulation of the design. Peptides are linked to the beta chain and co-expressed with the alpha chain with leucine zipper dimerization motifs to facilitate formation of heterodimer[24,37,38]. pMHC molecules were expressed in CHO cells, *E. Coli*, or insect cell lines.

The methods described above have low throughput (n<=15) and are time-consuming as the proteins need to be expressed separately and requires an extra step of synthesizing the peptides. For each new antigen, the whole expression, peptide synthesis and then the refolding process must be repeated. Furthermore, chemical synthesis omits post-translational modifications on the peptides.

2.1.4 Class II SCT

A new design that emerged was to express the class II pMHC molecule as a single construct connected by linkers. The design was initially introduced by Kozono et. al. to generate murine class II pMHC monomers[19]. Zhu et al. then adapted the approach to express SCTs encoding human HLA alleles in mouse cells[39]. Thayer et al. then established an alternative approach by inserting a fragment of the invariant chain between the alpha and the peptide to further stabilize the construct and improve yield[40]. They applied the method to express murine MHC alleles in monkey kidney tissue-derived COS cell line. We adapted the design to express pMHC proteins in mammalian cell lines in a high-throughput process to enable large-scale discovery of antigen-specific CD4+ T cells.

2.1.5 The motivation for building a high throughput class II SCT production platform

As discussed in Chapter 1.4, TCR-based T cell therapy targeting a single antigen often fail to result in a complete cancer elimination. To avoid downregulation of a particular set of antigens for cancer immune evasion, we need to perform high throughput and large-scale screening of CD4+ T cells responding to a large panel of antigen targets. Besides immune evasion by downregulating the antigens, patients also experienced loss of HLA heterozygosity (LOH) where cancer cells downregulate expression of a set of HLAs either coming from their paternal or maternal side to avoid immune surveillance[41]. Hence, administering patients with CD4+ T cells targeting antigens presented on more than one HLA alleles would be beneficial.

Majority of the TCR based immunotherapy has focused on the most prevalent HLA alleles in each HLA haplotype, such as A*02:01 for HLA-A alleles[42]. While this helps to cover a decent patient population as TCR immunotherapy first starts, many patients are turned away because their HLA haplotypes do not match with the TCRs.

To minimize cancer immune evasion and to maximize the patient population that we can treat, we aim to develop a high throughput antigen-specific CD4+ T cell screening platform that allows us to scan through and capture target T cells against a large panel of HLAs and antigens. We hope to use this technology to identify public TCRs against the top 10-20 common class II HLA alleles in the U.S. for off-the-shelf TCR-T cell immunotherapy products. This would greatly reduce the timeline to prepare the products to treat patients so that the patients are treated at the best clinical intervention timepoint. We also hope to apply this technology to develop personalized TCR-T cell

therapy for patients with rare HLA combinations, so all patients have equal chance of getting treatment.

2.1.6 The class II SCT design and the advantages of the class II SCT production platform

The class II SCT design starts with a signaling peptide exports the protein outside the cell so the SCT is a soluble monomer. The alpha chain comes after the signaling peptide, followed a short (GGGGS)2 linker 1. The partial invariant chain follows L1, adjacent to the antigen peptide, followed by Linker 2, another short GGSS(GGGGS)2 sequence, and the beta chain is right after L2. Linker 3 (L3) then connects the SCT with two tags for biotinylation and purification. The AviTag is a short sequence recognized by the BirA enzyme for the addition of a biotin molecule. The 6x histidine tag allows immobilized metal affinity chromatography (IMAC) based protein purification.

The SCT design offers several advantages:

1. Improved Production Efficiency: The SCT design allows the protein to be expressed as a single entity rather than three separate components, significantly enhancing overall production efficiency.
2. Versatile Antigen Modifications: The SCT design enables versatile modifications of the antigen sequence, facilitating rapid substitution with other antigens.
3. Easy Allele Switching: The alpha and beta allele can be easily switched out, allowing for adaptation to different combinations of Class II HLA alleles.

4. Mammalian Cell Expression with Natural Post-Translational Modifications: The SCTs are produced by mammalian cell lines, ensuring that they undergo post-translational modifications that closely mimic those of the natural pMHC protein.

2.2 Methods

2.2.1 Class II SCT template construction for various alleles

An SCT plasmid template encoding for a class II allele is assembled by a two fragment Gibson assembly each encoding for the alpha and beta chains of the allele. The FASTA amino acid sequences of the extracellular region of the alpha and beta chains are extracted from the IMGT and PDB websites and then reverse translated into DNA sequences followed by codon optimization. The DNA fragments are purchased from Twist Bioscience. The first fragment encodes for a partial MCS region that overlaps with the plasmid backbone, the secretion signal peptide, the alpha chain, L1, and a fraction of the pIi. The second fragment contains the second half of the pIi that overlaps 18 bp with the first fragment, a placeholder antigen peptide, L2, the beta chain, L3, followed by the AviTag and HisTag, the MCS region that overlaps with the plasmid backbone. The pcDNA3.1 plasmid backbone is linearized by enzyme double digestion. The first and second fragments are gibsoned together with the linearized backbone. The Gibson product is transformed into competent cells and go through standard molecular cloning to amplify the plasmids.

2.2.2 Molecular cloning to amplify plasmids

5-8 uL of a Gibson product or 10 ng of a pure plasmid are added to competent cells and incubated on ice for 25 min. The mixture is then heat shocked at exactly 30 seconds and incubated on ice for 2 minutes. 150 uL of SOC are added into each reaction and the culture is shaked horizontally at 37C 225 rpm for 1 hour. The culture is then evenly spread onto a LB with carbenicillin plate with

beads. Invert the plate and culture for 18-20 hours at 37C. Pick individual colonies and culture in 3 mL of LB with carbenicillin in 48-well deep culture plate for 18-20 hours at 250 rpm, 37C. Plasmids are extracted using the Qiagen miniprep kit and sequenced by next-generation sequencing services (Plasmidsaurus or in-house Nanopore sequencing).

2.2.3 Class II SCT plasmid construction

To build a library of SCTs of the same HLA allele but interrogating a list of antigens, a library of the DNA fragments encoding for the antigens need to be generated first. We have developed an algorithm that takes in the amino acid sequences of the antigens and outputs the list of forward and reverse primers that are checked through multiple checkpoints and optimized to remove hairpin structures and ensures ample hybridization region during PCR. The primers are purchased from IDT. Overlap extension PCR is used to generate the antigen fragments which extend at both ends to include a fraction of the pIi and L2. The antigen fragment and the linearized SCT backbone that omits the placeholder antigen are assembled through Gibson in a 5:1 insert to vector ratio. The Gibson plasmid is then amplified through the standard molecular cloning techniques described in 2.2.2.

2.2.4 Class II SCT protein production

The mammalian cell based Expi293 transfection system is used for SCT transfection and protein production according to the manufacturer protocol (Invitrogen). Briefly, the SCT plasmid and the Expi293 reagents are diluted with the OptiMEM media and mixed for 10-20 minutes at room temperature. Then the complex is added to Expi293 cells at 3M/mL in 1.25 mL cultures in 24-well plates at 225 rpm on a shaker in 37C incubator on Day 0. Enhancer 1 and 2 are mixed in a ratio of 10:1 and added to cells after 18-24 hours. The supernatant collected on Day 4. 30 uL of the

supernatant is denatured and run on unstained protein gels (BioRad) to visualize the expression of the SCT proteins. The rest of the supernatant undergoes 3 rounds of buffer exchange with bicine buffer through columns. The SCTs are then biotinylated by BirA enzymes at room temperature for 1.5 hours and then overnight in 4C. On Day 5, the SCTs are histag purified using the automatic Phynexus robot, and then desalted and buffer exchanged into PBS with the 7kbMW Zeba columns. The concentration of the SCT protein is measured through nanodrop and 20% v/v glycerol is added to the SCT in PBS for a more stable long-term storage.

2.2.5 TCR plasmid construction

We designed a high throughput approach to generate plasmids encoding for TCRs. The DNA sequences for TCRs are split into four fragments f1-f4. f1 encodes for the V, CDR3b, and J segment of TCRb, f2 encodes for the constant chain of TCRb. f3 and f4 are design in a similar way but encoding for TCRa. The f2 and f4 fragments are less diverse and hence reused while the f1 and f3 fragments are designed and purchased for each new TCR. Each of the fragment includes a front and end overlap region with either the backbone or the fragment next to it for Gibson assembly. We developed a TCR gene fragment auto-generation algorithm in-house[43] that takes in the amino acid sequences, performs codon optimization, balances GC content and outputs the DNA sequences. All TCR gene fragments are purchased from Twist Bioscience. The f1-f4 gene fragments are amplified, purified, and Gibson assembled into the pRRL plasmid backbone at a 4:1 insert to vector ratio, followed by standard molecular cloning procedure in 2.2.2 for plasmid amplification.

2.2.6 Lentiviral cloning for TCR expression

HEK293 cells are seeded at 0.5M in 3 mL R10 media in 6-well plates and cultured at 37C overnight. 0.24 ug, 0.48 ug, 0.48 ug, and 0.75 ug of pMD2-G VSVG, pMDL g/pRRE, pRSV-REV, and TCR-pRRL plasmids respectively are added to 150 uL EC buffer, followed by 8 uL of Enhancer, and incubated 5 min at r.t. 25 uL of effectene is then added incubated for 10 min at r.t. HEK293 media is aspirated and replenished with 2mL fresh R10 slowly. Then the transfection reaction mixture is combined with 1 mL media and applied to cells in a drop-wise fashion. After overnight incubation, the media is aspirated and 2-3 mL of the media that the target cells grew in is added. After another two days of incubation, lentivirus suspension in media is harvested by aspirating using a syringe and passed through a 45 um low protein binding filter. 1 mL of the viral supernatant is added to 500k target cells (Jurkats or primary T cells) in 1 mL media. Media is then changed the second day after transduction and TCR expression can be assessed via flow cytometry two days after the media change.

2.2.7 Cell line

To establish the NFAT-GFP reporter cell line, TCRb followed by TCRa knock-out was performed on the Jurkat E6-1 cell line (ATCC) using CRISPR gRNA and nucleofection, confirmed through flow cytometry, and sorted for purity. A plasmid encoding for NFAT-GFP reporter was then lentivirally transduced into the TCR- Jurkats and treated with blasticidin for selection.

The DR1+ K562 cells was established from WT K-562 (ATCC) by lentivirally transducing a plasmid encoding for the DRB1*01:01 allele with the transmembrane and cytoplasmic tail sections. Expression of DR1 on the surface was confirmed through flow cytometry and DR1+ K562 pure cell line was sorted.

2.2.8 Tetramer preparation

Tetramers are prepared by incubating the SCTs with a fluorophore-labeled streptavidin in a 4:1 ratio. 10 pmol of SCT protein and 2.5 pmol of streptavidin is usually used to generate a SCT tetramer reagent for four flow cytometry reactions. After incubating the SCT with streptavidin-PE or APC for 30 min at 4C protected from light, 1 uL of D-biotin at 20 uM is added to the mixture to block the free sites on streptavidin. The mixture is incubated at 4C for at least 30 min.

2.2.9 Antigen-specific CD4 T cells staining

Each vial of the target cells (usually PBMCs) was initially thawed in 5 mL of R10 media. Pre-enrichment of CD4+ T cells was performed using magnetic-activated cell sorting (MACS) according to the manufacturer's protocol. In the final step of CD4+ T cell enrichment, the cells were centrifuged and resuspended in 1 mL of media. The enriched CD4+ T cells were then incubated with PKI inhibitor (Dasatinib). The PKI buffer (concentration 100 nM) was added at a 1:1 ratio to the media volume. The cells were incubated at 37°C for 20 minutes, followed by centrifugation at 500 × g for 5 minutes without washing. A tetramer staining buffer mix was prepared in 50 nM PKI/SE buffer. For each antigen, 2 µL of PE and APC tetramers, along with 2 µL of HIV PE/Cy7 tetramer, were added to 100 µL of the buffer and mixed thoroughly. Each sample was resuspended in 100 µL of the tetramer staining buffer and incubated at 4°C for 20 minutes. Following incubation, 100 µL of SE buffer was added to wash the cells, which were then centrifuged at 500 × g for 5 minutes. A cell surface staining buffer was prepared by adding CD4-BV421 at a ratio of 1 µL per 100 µL of buffer per 50,000 CD4 cells, and Calcein Green (FITC) live stain at a final concentration of 0.2 µM. Specifically, 0.2 µL of Calcein Green (stock at 1mM) was added to 1 mL of PKI buffer, followed by 1 µL of CD4-BV421 in 100 µL of the PKI/live dye buffer. The cells were incubated for 10 minutes at 4°C. After staining, the cells were washed with

100 µL of SE buffer to remove excess reagents, centrifuged, and resuspended in 200 µL of PBS. The prepared cells were then subjected to flow cytometry analysis to assess antigen-specific CD4+ T cell populations.

2.2.10 Tetramer binding assay

This is a similar workflow as the antigen-specific T cell staining protocol. The TCR-transduced Jurkat cells were counted, and 50,000 cells from each sample were aliquoted. Additionally, 50,000 cells from the transduced population were set aside as the unstained control. Cells were incubated with PKI as described above and then 1.5 uL of the SCT PE-tetramer (target antigen) and 1.5 uL of HIV PE-Cy7-tetramer was used to stain the cells. The cells were further stained using a live dye and anti-TCR (APC) antibody in 50 nM PKI buffer. The staining was carried out for 10 minutes at 4°C. After staining, an additional wash with 100 µL of SE buffer was performed, followed by centrifugation. The cells were resuspended in 200 µL of PBS and analyzed using flow cytometry to assess antigen-specific responses of the TCR-transduced Jurkat cells.

2.2.11 Single cell plate sequencing for paired αβTCR

Cells were single cell sorted into 96-well plates containing 12 uL of 1x lysis buffer with RNAse inhibitor (RNAsin). The plates were spun down immediately after sorting and flash froze on dry ice. When ready for sequencing the TCRs, the lysate plate was removed from storage and thawed on ice for 3 minutes. The samples were then centrifuged at 1,000 × g for 1 minute to collect any residual liquid. A master mix of one step RT-PCR reaction was conducted according to the manufacturer's protocol (Qiagen). Briefly, the following reagents were mixed for one reaction: 3.2 uL 5x buffer, 0.64 uL dNTPs at 10 mM, 0.64 uL enzyme mix, 5.12 uL RNAse free water, 0.55 uL of a mixture of the alpha or beta variable primers (1.75 uM), and the alpha or beta constant primer

at 100 uM. Each lysate is split into two reactions for amplifying the alpha and beta sequences separately. The PCR protocol consisted of the following steps: an initial incubation at 50°C for 90 minutes, followed by 95°C for 15 minutes; The amplification step involved 40 cycles of a 3-step process: 94°C for 1 minute (denaturation), 68°C for 1 minute (annealing), and 72°C for 1 minute (extension). A final extension was performed at 72°C for 10 minutes, followed by a hold at 4°C. To verify the PCR products, 4 µL of the reaction mixture was loaded onto an agarose gel and run at 150V for 25 minutes. Following RT-PCR, cleanup was performed using 0.8x SPRI beads (Beckman Coulter PCR purification beads). The DNA was eluted in 30 µL of EB buffer for downstream applications. A second PCR further amplifies the TCR fragment and extends the fragment with overlap regions for the third PCR that adds in the barcoding regions for multiplexed sequencing through Illumina.

2.2.12 Peptide-pulsed activation assay

Peptides were reconstituted to a concentration of 4 mg/mL using 200 µL of DMSO per tube (with each tube containing 0.8 mg of peptide). Peptides were added directly to each well of cells at a final concentration of 2 µg/mL in a 100 µL media volume, equivalent to using 0.2 µg of peptide per well. Jurkat cells expressing the NFAT-GFP reporter were co-cultured with DR1-K562 cells at a 1:1 ratio, with 50,000 cells of each type per well in a round-bottom 96-well plate. A negative control was included, consisting of 50,000 DR1-K562 cells combined with an unmatched peptide, 50,000 TCR-transduced Jurkat cells, and 1 µL of anti-CD28. The co-cultured cells were incubated overnight at 37°C for 16-24 hours. On the next day, cells were stained with Calcein UV to identify live cells and assessed for GFP expression. All cells from each well were collected, and approximately 50,000 cells were aliquoted from the DR1+ K562 and TCR NFAT Jurkat populations to be used for separate unstained controls. This aided in distinguishing K562 from

Jurkat cells based on side scatter (SSC) and forward scatter (FSC) gating. Centrifuge all collected cells at 500 × g for 5 minutes, followed by a single wash with SE buffer. The cells were stained with the following antibodies: Live dye: 0.2 µL of Calcein UV, TCR: 1 µL of APC, HLA-DR: 1 µL of PE. The cells were incubated for 10 minutes at 4°C. A final wash by adding 100 µL of SE buffer, followed by centrifugation and the cells were resuspended in 200 µL of PBS.

2.3 Results and Discussion

2.3.1 Workflow for generating class II SCTs

We first start with building plasmid templates encoding for the target HLA allele (**Figure 2.1**). The amino acid sequence for the extracellular regions of HLA alpha and beta alleles were downloaded from IMGT database. The gene fragments encoding for the alpha and beta chains were codon optimized, ordered from Twist Bioscience, PCR amplified and purified. Both fragments were assembled into a linearized pcDNA3.1 backbone in the MCS region through Gibson assembly.

We then move to compile the list of SCT antigens. If we are probing for antigens from infectious pathogens, such as viruses or bacteria, we usually begin with either an unbiased screening to cover the entire protein or we select antigens based on prediction algorithms for likely binders to the MHC allele. On the other hand, if we are screening neoantigens from cancer patients, we would start with a list of mutations sequenced from exosome from both cancer cells and blood samples of the patient. With the list of the antigens, we generate a short DNA fragment encoding for the antigen with both ends extending to overlap with the partial invariant chain fragment and Linker 2 (**Figure 2.1**). We designed forward and reverse primers that hybridize at the 3' end. We generated the antigen fragment through overlap extension PCR. The purified fragment was Gibson assembled with a linearized pcDNA3.1 backbone with the targeted allele inserted in the template.

The Gibson product with the antigens inserted in the plasmid was amplified through the culture of *E.Coli* competent cells and the amplified plasmids were extracted and purified.

The plasmids were then transfected into Expi293F cells and after four days of culture, the secreted SCT proteins were first buffer exchanged for overnight biotinylation followed by his-tag purification (**Figure 2.1**). The purified SCTs were desalted and buffer exchanged into PBS, followed by addition of 20% glycerol and saved in -20C for long-term storage.

2.3.2 Automatic generation of primer designs encoding class II antigens

To develop a high throughput platform for generating class II SCTs, we wrote a computational pipeline to generate optimized sets of forward and reverse primers automatically. The algorithm takes in the one-letter amino acid sequence as the input, generates a gene fragment encoding the antigen and adds in the Gibson overlaps at both ends, confirms that the melting temperature of the 20bp hybridization region is within an optimized range of temperature for annealing (50-69C), checks the primers for secondary structures at the annealing and extension temperatures, confirms the length of hybridization, and if the primers can anneal at the PCR annealing temperature. If the primer fails at any checkpoint, the algorithm optimizes the primer by going through rounds of alternative codon replacement until a primer that passes through all checkpoints is generated.

2.3.3 Validation of the SCT against the known cognate TCRs

The pMHC complex for class II HLA alleles consists of three individual components: the α chain, the β chain, and the presented peptide antigen (**Figure 2.2**). While previous designs of class II pMHCs involved expressing the subunits separately, followed by peptide loading or exchange[18,20,23,24,44], it has been shown that the peptide may be covalently linked to either the α or β chain[21,24,37], or to both chains so that the full pMHC is expressed as a single, stable

protein[19,40,45,39] (**Figure 2.2**). These early demonstrations inform our basic approach, where a library of SCTs for a given HLA allele was prepared using a single plasmid template for the α and β chains, followed by insertion of each antigen fragment using Gibson assembly. Purified plasmids encoding the SCT construct were transfected into a mammalian cell line (Expi293F) for expression of soluble SCT proteins, which were his-tag purified, biotinylated and preserved in PBS with 20% glycerol for long-term storage.

To validate SCT templates across common class II HLA alleles, we identified antigen-TCR pairs for five HLA-DR and one HLA-DP alleles from the literature[46,47]. The TCRs were cloned into a TCR knock-out Jurkat cell line through lentiviral transduction (**Figure 2.3**), and the SCTs for each antigen pairing were expressed as described above. Each SCT specifically recognized its cognate TCR, and non-specific binding was not observed (**Figure 2.4**).

2.3.4 Comparing binding efficiency of SCT with conventional pMHC

We further evaluated the binding efficiency of an SCT encoding an influenza peptide ($HA_{306-318}$ PKYVKQNTLKLAT) relative to a conventional pMHC reagent. The SCT exhibited a similar recognition efficiency, binding 86.4% vs 94.5% of T cells (**Figure 2.5**), compared to the pMHC tetramer. Non-specific binding was not observed. These results suggest that class II SCTs may provide a robust approach for generating class II pMHC-like reagents for sensitive and selective antigen-specific CD4+ T cell capture and analysis.

2.3.5 Identify novel TCRs using an SCT library

We next moved towards validating SCT libraries for capturing novel CD4+ T cells from non-expanded PBMCs collected from healthy donors. A 23-element library was constructed from the DRB1*01:01 template, presenting antigens from cytomegalovirus (CMV), Epstein-Barr virus

(EBV), influenza viruses, and tetanus bacteria (CEFT) (**Figure 2.6, top**). Each of the five healthy donor PBMC samples was stained with individual SCT tetramers labeled with PE and APC fluorophores (**Figure 2.6**, **bottom**). Cells stained with a negative control SCT (HIV Gag_{41-56} SALSEGATPQDLNTML) were excluded, and the double-positive cells were sorted for single-cell plate sequencing of their TCRs.

2.3.6 Validation of the TCRs identified – Tetramer binding

The TCRαβ sequences from the captured CD4+ T cells were cloned into a TCR knock-out NFAT-GFP reporter Jurkat cell line. When these TCR-engineered cells were tested, varying levels of tetramer binding were observed (**Figure 2.7**). In **Figure 2.7**, the x-axis labels follow a *TCR-antigen* format, where "2-A" refers to TCR index 2 and an antigen index A. For example, TCRs 1-7 represent different clones that all bind to antigen A. TCR clones with validated tetramer binding were further assessed using a peptide-stimulation assay (**Figure 2.8, 2.9**).

2.3.7 Validation of the TCRs identified – peptide pulsed activation

K562 cells, which lack functional expression of wild-type class II pMHC on their surface[48], were engineered to serve as artificial antigen presenting cells (aAPCs) expressing the DRB1*01:01 allele with an empty peptide-binding pocket on their surface (DR1-K562)[49,50,51]. TCR NFAT-Jurkat cells were co-cultured overnight with their cognate peptide and the DR1-K562 cells to induce T cell activation (**Figure 2.8**). All TCR clones demonstrated NFAT-based T cell activation (**Figure 2.9**). TCR clones 9, 10, 11 and 15 showed comparable percentages of tetramer binding and activation in response to both peptides A and B, which share the same core sequence and differ by only one amino acid at the N-terminus. No activation from the negative control was observed. These results suggest that SCT libraries may be used to identify new TCRs, and that T cells

expressing those TCRs are functionally activated to varying degrees by the SCT-encoded peptide antigen when that antigen is presented by the native MHC molecule.

2.4 Figures

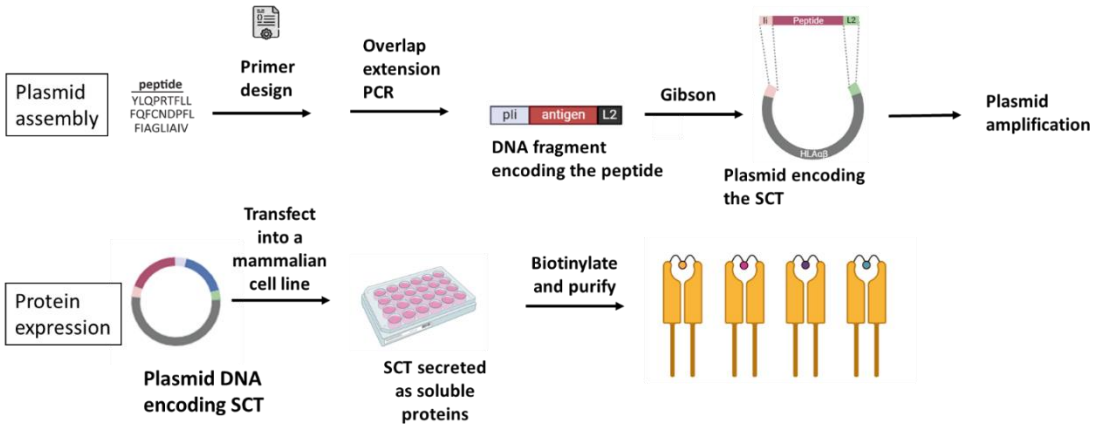


Figure 2.1 Experimental workflow of constructing the SCT plasmid and expression of the SCT protein.

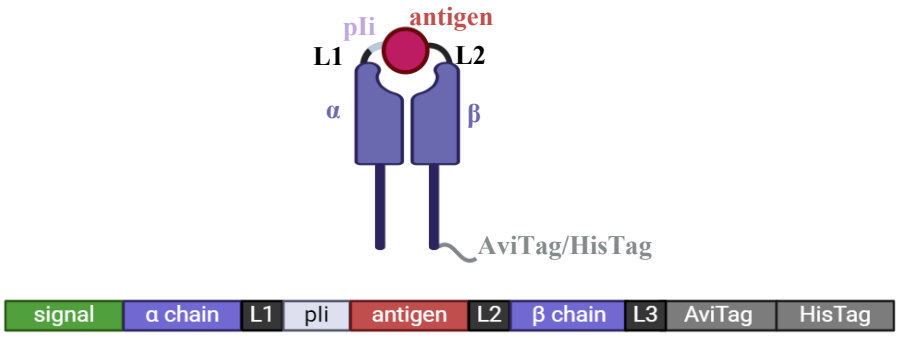


Figure 2.2 Plasmid design encoding the class II SCT protein. The α chain, antigen, and β chain are connected through two linkers (L1 and L2). A fragment of the invariant chain (pIi) stabilizes the SCT. A secretion signal is placed before the α chain, while purification tags are linked through L3.

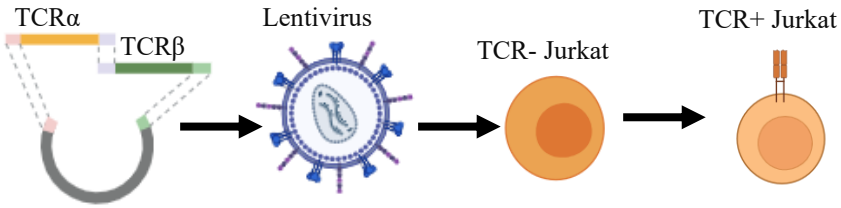


Figure 2.3 Lentiviral cloning of TCRs into a TCR knock-out Jurkat cell line.

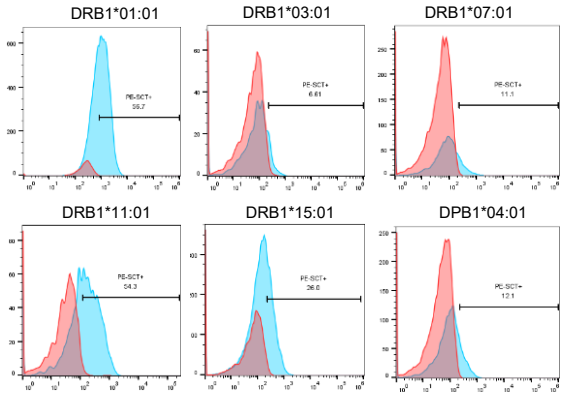


Figure 2.4 Validation of class II SCT templates for various HLA alleles through SCT tetramer binding to previously reported cognate TCRs.

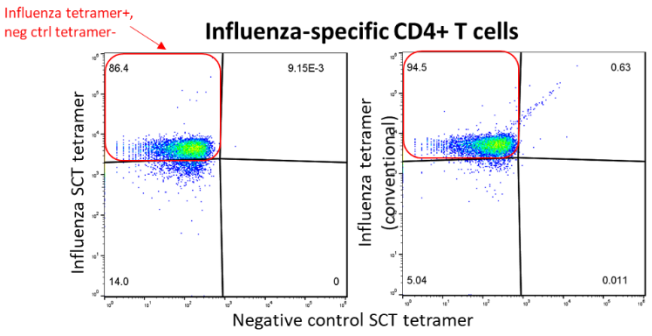


Figure 2.5 Comparison of the binding efficiency of class II SCT to pMHC multimers.

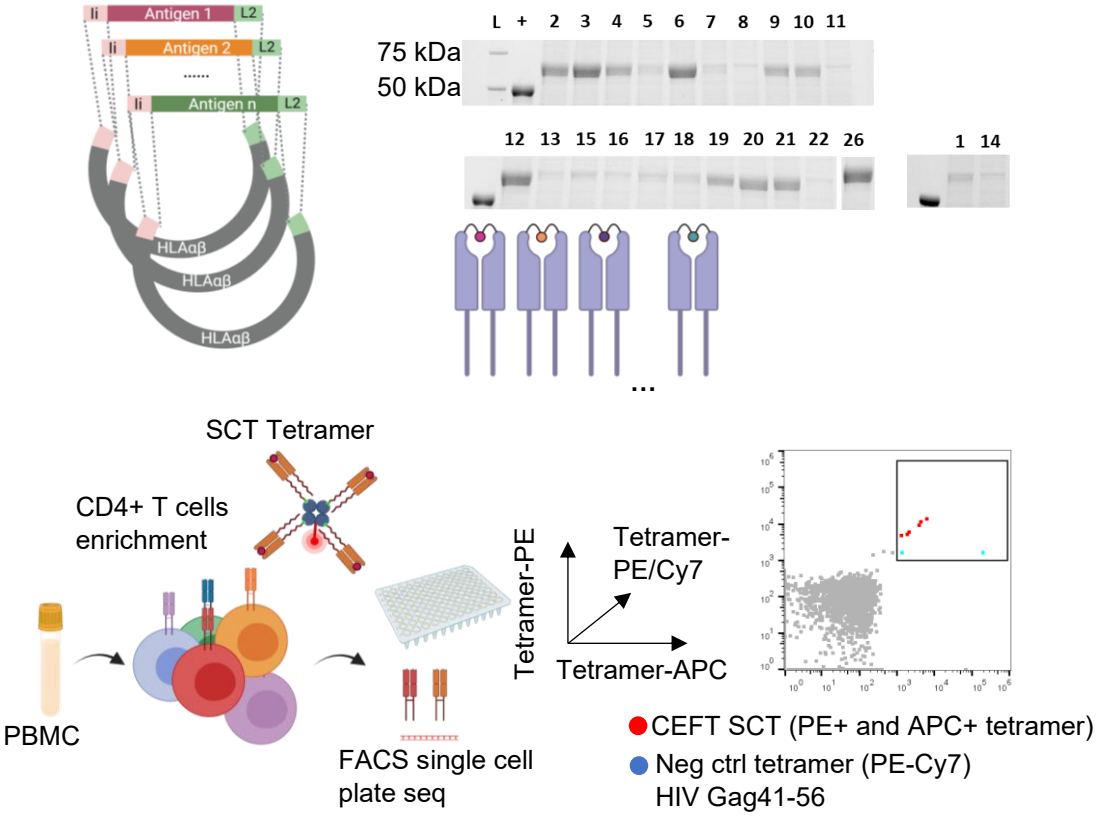


Figure 2.6 A 23-Element DRB1*01:01 CEFT library of SCTs used to identify novel antigen-specific CD4+ T cells. SCT expression was evaluated on SDS-PAGE protein gels (top right). PBMC samples from healthy donors were pre-enriched for CD4+ T cells. Antigen-specific T cells were identified through double-positive staining with PE and APC SCT tetramers and sorted for TCR-sequencing and subsequent cloning, with representative flow cytometry data shown at bottom right. An HIV Gag41-56 SCT tetramer was used to exclude non-specific binding cells.

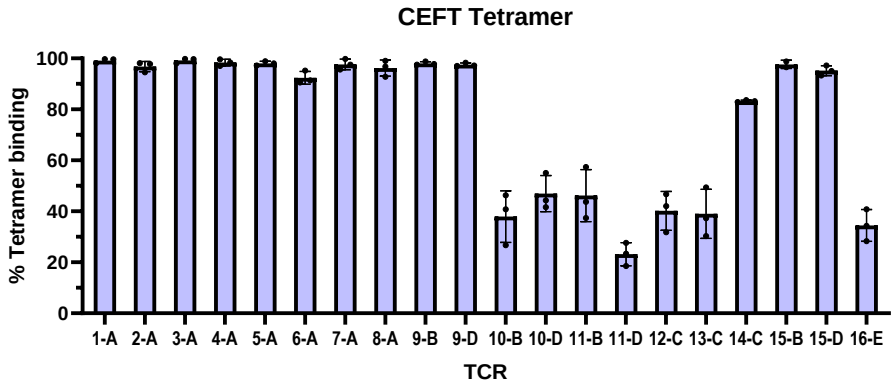


Figure 2.7 Tetramer binding validation of 16 cloned TCRs, each reactive to one of the five CEFT antigens (n=3). Positive tetramer binding read-outs were corrected for negative control tetramer staining.

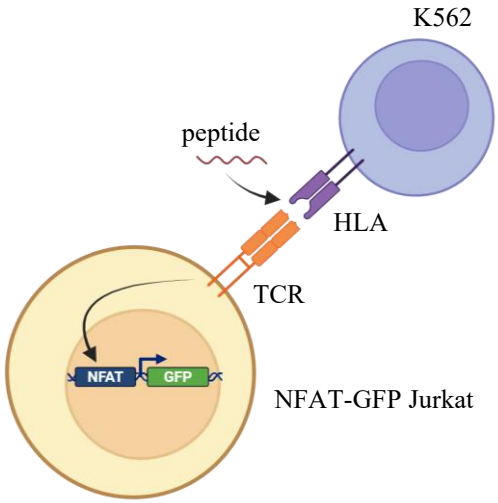


Figure 2.8 Peptide-pulsed activation assay. Illustration of peptide-pulsed activation assay. TCRs were transduced into TCR knock-out NFAT-GFP Jurkat cells. WT K562 cells were engineered to express the DRB1*01:01 (DR1) allele. TCR+ NFAT-GFP Jurkats were co-cultured with DR1+ K562 cells overnight in the presence of the peptide.

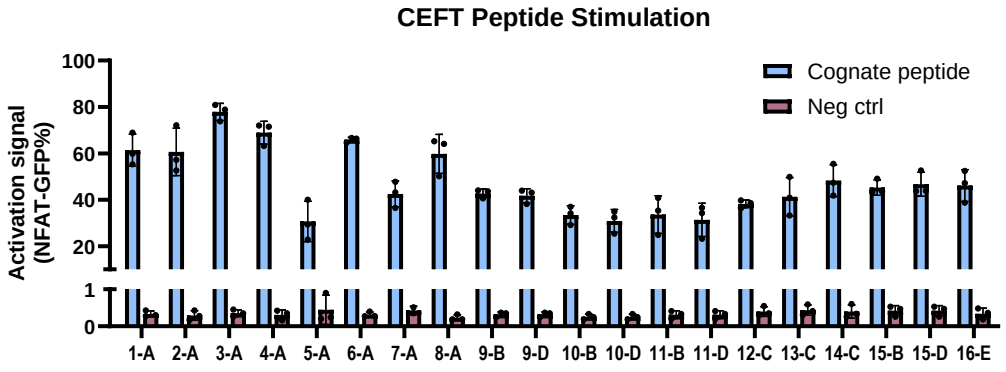


Figure 2.9 Peptide-pulsed activation of the TCR+ NFAT-GFP reporter cells. T cell activation response measured through the percentages of NFAT-GFP+ Jurkat cells upon stimulation with cognate peptides identified by SCT tetramer assays in d and e, in comparison to stimulation with a negative control peptide (n=3). **$P < 0.01$ for each group relative to the negative control peptide, determined by one-tailed independent t-test assuming equal variances.

Chapter 3: SARS-CoV-2 specific CD4+ T cell immune response

3.1 Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), also known as coronavirus disease 2019 (COVID-19), is a highly transmissible and pathogenic virus that caused a worldwide pandemic and resulted in millions of deaths. Patients showed symptoms of viral pneumonia, including fever, cough, chest discomfort, difficulty in breathing, and dyspnea or bilateral lung infiltration in severe cases[52].

SARS-CoV-2 is a betacoronavirus that shares 79% genome sequence identity with SARS-CoV and 50% with MERS-CoV. It contains 14 open reading frames (ORF) which encode 29 proteins. Four of the ORFs encode a set of structural proteins, which include the spike protein (S), the nucleocapsid (N), membrane protein (M), and the envelop protein (E), that are essential for viral assembly, interaction with the target cells in host, entering the host cells, and suppression of the immune responses in host[53,54]. The S protein on the surface of the virus binds to the human angiotensin-converting enzyme 2 (hACE2) receptor on the host cellular membrane, followed by fusion of the viral and host cellular membrane. Viral RNA is then released into the host cells and viral proteins are produced through host cellular mechanisms. Translation of 16 non-structural proteins (NSPs) are then initiated, whose function is to replicate and transcribe viral RNA. The viral genomic RNA and the NSPs then assemble into mature virions which exit the current cell through exocytosis to infect surrounding host cells.

Upon binding to epithelial cells in the upper respiratory tract, SARS-CoV-2 replicates and migrates down to the airways and enters alveolar epithelial cells in the lungs. Severe immune responses might be triggers by the rapid replication of the viruses, leading to cytokine storm syndrome which

causes acute respiratory distress (ARD) symptom, the main cause of death due to COVID-19. Histopathology analyses showed bilateral diffused alveolar damage (DAD), alveolar epithelial damage, desquamation of pneumocytes and fibrin deposits in lungs of patients with severe COVID-19[52,55]. DAD is the predominant pattern of lung injury that is observed in deceased patients infected with SARS-CoV-2[55].

Humoral immunity has been considered a major player in anti-SASR-CoV-2 responses. Vaccines have focused primarily on inducing SARS-CoV-2 specific antibodies to neutralize and prevent spread of viral infection in host. However, it has been noted that T cell dependent immune responses are also critical in disease progression and recovery. Studies have shown that SARS-CoV-2 induced significant CD4 and CD8 responses in addition to the IgG and IgA humoral responses[56]. In fact, early T cell responses indicate faster and viral clearance and less disease severity whereas early antibodies do not correlate with better disease outcomes[57,58]. More specifically, early effective CD4+ T cell responses have the strongest association with mild disease severity compared to CD8+ T cells and antibody responses. Viral specific CD4+ T cell responses are predominantly from the Th1 effector types to antigens from the S, M, N, and several NSPs[57,58].

A large effort has been made to understand the repertoire of epitopes recognized by T cell responses against SARS-CoV-2. Over 2,000 epitopes have been identified and cataloged by the Immune Epitope Database (IEDB) with repeated reporting of some of the dominant epitopes such as spike 269–277 (YLQPRTFLL) and the N 105–113 (SPRWYFYYL) for CD8+ T cells, and M 176–190 (LSYYKLGASQRVAGD) and spike 166–180 (CTFEYVSQPFLMDLE) epitopes for CD4+ T cells. However, it has not been clear for what constitutes an immunogenic antigen and how they would have affected the T cell immune responses over time.

As variants of SARS-CoV-2 virus emerge and repeated infections haunt the global community, understanding the durability of the immune responses after previous infections is critical. Memory CD4 and CD8 T cell responses are observed in the majority of infected individuals for over eight months which indicates that T cell immunity may persistent years after recovery[59]. Although CD4 T cell responses in COVID-19 patients have been assessed, antigen-specific CD4+ T cell responses have not been well evaluated longitudinally.

Here we evaluate the properties of immunogenic antigens and the longitudinal immune responses to SARS-CoV-2 specific CD4+ T cells via the high throughput screening platform using a library of class II SCTs for the receptor binding domain (RBD). SARS-CoV-2 serves as an effective model to study human immune responses to infectious disease.

3.2 Methods

3.2.1 Building an unbiased library of the RBD SCTs

The unbiased RBD library starts at amino acid position 317 on the Spike protein of the SARS-CoV-2 virus. Each SCT encodes for a 15-mer antigen that shifts down the RBD domain by four amino acids. In this case, every two consecutive antigens overlap by 11 amino acids. A total of 54 DRB1*01:01 SCTs spanning the entire RBD domain (317-543) was included in the design. The SCT library was constructed and purified as described in the Methods section in Chapter 2. Additionally, 19 literature reported antigens from structural proteins such as the Membrane protein, Nucleocapsid protein, Envelope protein, and the other regions of the Spike protein were also expressed.

3.2.2 Preparation of DNA-barcode dextramers and negative control tetramers

6 pmol of each SCT was combined with 0.269 pmol (molar ratio 22:1) of a dextramer labeled with a unique DNA barcode to constitute for one reaction and PBS was added to make up the volume to 3.5 uL. The mixture was incubated for 30 minutes at 4°C. Following this incubation, 0.5 uL of 20 uM biotin was added, and the mixture was incubated for an additional 30 minutes at 4°C. For subsequent experiments, 1.67 µL of each prepared dextramer was utilized per reaction. HIV SCT tetramer was prepared as described in 2.2.8 as a PE-Cy7 tetramer.

3.2.3 High throughput screening of SARS-CoV-2-specific CD4+ T cells

Cell Preparation

In the morning of the experiments, 1.67 uL of each dextramer was pooled together. Cells were thawed in batches in pre-warmed R10 media to prevent prolonged exposure to the freezing medium. Cells from multiple timepoints were then combined based on their hashtag grouping. A maximum of 5 mL of thawed cells was added per tube with 40 mL media to achieve a 1:10 dilution. Each batch consisted of thawing five vials, with each vial containing approximately 1 mL of cells. Cells could be temporarily stored at 4°C after centrifuge and resuspension in the Selection Buffer if a pause was necessary. For cell counting, a 5 µL cell suspension mixed with an equal volume of trypan blue was used

Hashtag Staining Round 1

Samples from Other Timepoints (OT, i.e. PD1 and 06M-36M) were stained with unique DNA-barcode hashtag antibodies based on their timepoint grouping. Staining was performed using 2 µL of hashtag per million cells, with incubation at 4°C for 10 minutes. The cells were washed with 300 µL of SE buffer and centrifuged.

MACS Sorting of CD4+ T Cells

Next, the combined samples from AC, CV, and OT were each enriched for CD4+ T cells through a negative selection through MACS according to manufacturer's protocol.

Hashtag Staining Round 2

The enriched CD4+ T cells were then treated with PKI buffer as described in 2.2.9. The enriched CD4+ T cells from AC and CV timepoints then went through a second round of hashtag staining. After washing with 300 μL SE buffer and centrifuged, the final cells from AC, CV and OT were combined and counted.

Multimer Staining

For multimer staining, the dextramer pool and HIV PE-Cy7 tetramer were combined which was used to stain the combined cell sample after the addition of PKI. Samples were incubated for 20 minutes at 4°C, and cells were washed with 4 mL SE buffer and centrifuged at 500 g for 5 minutes.

Dead Cell Staining and CD4 Staining

To stain for dead cells and CD4, 0.5 uL of Apotracker Green (FITC) and 5 uL of CD4-BV421 was added in 100 uL SE buffer to stain up to 1M cells. Cells were incubated for 10 minutes at 4°C, washed with 500 μL of SE buffer, and centrifuged.

FACS sorting

The final cell suspension was adjusted to 2 mL in FACS buffer, with a target concentration of 1-10 million cells/mL. Before FACS sorting, cells were filtered through a 40 μm cell strainer into flow cytometry tubes to ensure a single-cell suspension. Dextramer+ and HIV tetramer- samples were tube-sorted and stored on ice until ready for 10x single cell sequencing.

Preparation for 10x Genomics

The PBS buffer with 0.04% BSA was used to coat all the tubes to minimize cell loss during processing. Once the tubes are coated, add 2 mL of the wash buffer directly to the FACS-sorted cells and centrifuge them at 500g for 5 minutes. After centrifugation, carefully aspirate the supernatant, leaving less than 0.5 mL of liquid behind. Next, without mixing the cells, top up the volume with an additional 2 mL of fresh wash buffer and spin down again at 500g for 5 minutes. Aspirate the supernatant once more, leaving about 200 µL remaining. Mix the cells thoroughly at this point, and transfer the remaining cells into a pre-coated DNA LoBind tube to reduce cell adhesion, then top up the tube with 1 mL of wash buffer and spin down. Resuspend the cells to a volume and concentration based on the protocol outlined in the 10x Genomics manual.

Single cell multi-omics assay

The cells were loaded onto a Chromium Next GEM chip. Cells were lysed for reverse transcription and cDNA amplification using the Chromium Controller (10X Genomics). Polyadenylated transcripts underwent reverse transcription within each gel bead-in-emulsion. Full-length cDNA, along with unique cell barcode identifiers, was then amplified via PCR. Specifically, the following three libraries for each cell were constructed: mRNA transcriptome, TCR sequences, and surface barcoded proteins. The resulting sequencing libraries were prepared, normalized, and sequenced on the Illumina Novaseq platform.

3.2.4 Computational analysis of single cell sequencing data

The transcriptome, TCR sequences, surface protein levels, and antigen specificity were simultaneously assessed for each cell. Raw data were processed using the Cell Ranger Single-Cell Software Suite (v3.1.0, 10X Genomics) with GRCh38 as the reference genome. Cells that did not

meet quality thresholds were filtered out if they met any of the following criteria: n-counts below 1,000 or above 10,000, n-genes below 250 or above 2,500, or a mitochondrial content above 10%. Gene expression counts for each cell were normalized by the total expression, scaled by a factor of 10,000, and transformed to a log scale. Phenotypes were assigned to single cells based on clusters identified using the Leiden algorithm.

Raw reads for each Hashtag and Dextramer were normalized, and cells were classified based on their highest expressed Hashtag. Cells showing high expression of multiple Hashtags were excluded as potential doublets. The identified Hashtag labels were then used to determine the corresponding SCT-dextramer cocktail for each cell. For each cell, the unique molecular identifiers (UMIs) for each dextramer were counted, and the proportion of UMIs for each dextramer was calculated. Cells were assigned a specific antigen only if they had a UMI count greater than 25, with those UMIs representing more than 25% of the cell's total dextramer reads. Antigens were assigned according to the dextramer with the highest mapping for each cell. Cells that could not be definitively assigned to a specific dextramer were discarded.

3.3 Results and Discussion

3.3.1 Unbiased screening of the SARS-CoV-2 RBD domain

We explored the use of class II SCT libraries as high throughput screening tools for capturing and characterizing antigen-specific CD4+ T cells in an HLA-matched longitudinal cohort of patients infected with SARS-CoV-2 (**Figure 3.1**).

As discussed before, computational predictions for antigens that bind class II HLAs are not as accurate as class I antigen predictions. Hence, we decided to construct an SCT library for unbiased screening of the SARS-CoV-2 RBD domain. We curated an inventory of DRB1*01:01 PBMC

samples from a longitudinal cohort of patients and designed an unbiased 54-element library presenting epitopes from the RBD domain from the Spike protein using DRB1*01:01 SCT template. The library was designed to contain 15-mer epitopes spanning the entire RBD domain in 4-amino acid increments (**Figure 3.2**), ensuring unbiased coverage of all potential core epitopes. Of the 54 SCTs, 52 were successfully expressed, with 46 reaching sufficient concentrations for downstream applications (**Figure 3.3**). Additionally, we included 18 SCTs representing previously reported epitopes from other structural domains (e.g., membrane protein (M), envelope protein (E), nucleocapsid protein (N), and regions from the Spike protein outside the RBD)[60,17,61,62] (**Figure 3.3**).

3.3.2 Experimental design to identify SARS-CoV-2 specific CD4+ T cells

Each SCT was converted into a multimer by appending it onto a fluorophore-labeled-dextramer, with each dextramer also containing a unique DNA barcode (**Figure 3.1**). The SCT-dextramers were then pooled and used to stain CD4+ T cells enriched from PBMC samples of 22 SARS-CoV-2 participants. Samples were collected at multiple timepoints: acute (AC, within 1 week of infection), convalescent (CV, 2-3 months post-acute), and at 6, 12, 18, 24, and 36 months (6M/12M/18M/24M/36M) post-AC, with all time points included in the study. Samples from the same timepoint were pooled and tagged with barcoded antibodies for sequencing-based demultiplexing. Antigen-specific CD4+ T cells were sorted and pooled for 10x single-cell sequencing to resolve single cell gene expression profiles, TCR sequences, and barcode sequences for epitope specificity and blood collection time points (**Figure 3.1**). After data processing (see Methods), a total of 2,188 antigen-specific CD4+ T cells were identified.

3.3.3 SCT expression and cells captured versus prediction

SCT expression levels did not correlate with predicted binding affinity or the percentile rank of eluted ligand prediction scores (**Figure 3.2**). Similarly, an SCT's ability to capture cells did not correlate with its expression level or predicted binding affinity, which contrasts with findings for class I SCT libraries and antigen predictions for common HLA alleles[63].

3.3.4 Clonotype characterization of SARS-CoV-2 CD4+ T cells

In **Figure 3.4**, we present the diversity of T cell clones captured by each SCT (top half of graph), as well as the distribution of antigen-specific CD4+ T cells detected across individual participants. Clonal expansion was observed in most antigen-specific CD4+ T cells, with at least two copies of a T cell clonotype detected for 44 out of 64 antigens and in 49-58% of cells across various timepoints (**Figure 3.5**). A subset of the CD4+ T cells was selected for cloning, and their antigen specificity was validated through tetramer binding and peptide-pulsed activation assays (**Figure 3.6**).

3.3.5 Phenotypic characterization of SARS-CoV-2 CD4+ T cells

The SCT library-based approach in **Figure 3.1** permits an in-depth analysis of the antigen-specific CD4+ T cells. Demultiplexing the sequencing results reveals the patient and time-point origin of each cell, along with its antigen-specificity, TCRα/β clonotype, and single cell transcriptome (**Figure 3.7**).

The 2,188 antigen-specific CD4+ T cells were clustered based on similarity in gene expression and projected onto a Uniform Manifold Approximation and Projection (UMAP) (**Figure 3.8**). SELL, CCR7, TCF7, SATB1, and LEF1 serve as signature gene markers for naïve or central memory CD4+ T cells (Naïve/T_{CM}), while KLRG1, S100A4, ANXA1, AHNAK, IL32, and CLIC1

indicate effector memory T cells (T_{EM}) (**Figure 3.9**). Th17 cells exhibit high expression in RORC, KLRB1 and CCR6 genes, whereas Treg cells show upregulated expression of FOXP3. The Th1 cluster demonstrates elevated expression of TBX21, IFNG, STAT4, RUNX3, PRF1, CCL5 and HOPX. A subset of Th1 cells upregulates cytotoxic CD4+ T cell markers, including NKG7, CST7, GNLY, PRF1, TNF, and members of the granzyme family. The exhausted phenotype is characterized by high expression of exhaustion markers such as PDCD1, TIGIT, LAG3, and CTLA4. Additionally, a small cluster of cells displays high expression of proliferation markers, including MKI67, MYBL2, BUB1, PLK1, and CCNE1. The UMAP analysis reveals distinct clusters representing phenotypic variations among SARS-CoV-2-specific CD4+ T cells.

3.3.6 Phenotypic evolution of SARS-CoV-2 CD4+ T cells

We clustered the antigens into three groups based upon the immune phenotypes they induced during acute disease, including naïve/T_{CM}, exhausted, and T_{EM} immune responses (**Figure 3.10**). At later time points, we observed that all of these phenotypes tended to shift towards (or remain as) T_{EM} phenotypes. This observation is consistent with existing literature, which indicates that CD4+ T cell responses often transition towards effector memory phenotypes following antigen exposure[64,65].

We also observed three distinct clonal properties: clonal expansion, clonal persistence, and public TCR clonotypes (**Figure 3.13**). The naïve/T_{CM} cell cluster exhibited the lowest percentage of clonally expanded cells, whereas the Th1 and proliferative CD4+ T cell clusters displayed a high degree of clonal expansion. Additionally, we identified clonally persistent CD4+ T cells enriched in the exhausted, T_{EM}, cytotoxic and Th17 phenotypes. Surprisingly, the Treg population was also enriched with clonally persistent SARS-CoV-2 specific CD4+ T cells. Moreover, we identified TCR clonotypes that were shared across participants.

3.3.7 Discovery of immunogenic antigens

To assess immunogenicity of the T cell epitopes, we scored each antigen based on several properties: the number of unique CD4+ T cell clonotypes induced by the antigen, the number of donors that showed antigenic CD4+ T cell immune responses, the percentage of antigen-specific CD4+ T cell clonotypes that persisted over time, and the percentage of clonal expansion among the antigen-specific CD4+ T cell clonotypes. Each property was categorized into groups based on a $\log_2$ scale and assigned scores from 0 to 5. The antigens were then ranked into three groups based on their overall immunogenicity scores by a $\log_2$ scale (**Figure 3.11**). In **Figure 3.12**, we plot the phenotypic evolution of each of these groups. The most highly immunogenic antigens induced more exhausted CD4+ T cell responses during acute disease and exhibited the largest fractional changes towards T_{EM} phenotypes over time. In contrast, the least immunogenic antigens had a higher fraction of naïve cells during acute disease and showed relatively stable phenotypic distributions over time.

These results demonstrate that the SCT library-based experimental approach shown in **Figure 3.1** enables the multi-omic discovery and in-depth characterization of large numbers of antigen-specific CD4+ T cells from a cohort of HLA haplotype matched patients within the context of a viral infection. The analysis enables a quantitative assessment of the immunogenicity of the viral antigens and an evaluation of the phenotypic evolution of viral-specific CD4+ T cell clonotypes over a disease course.

3.4 Figures

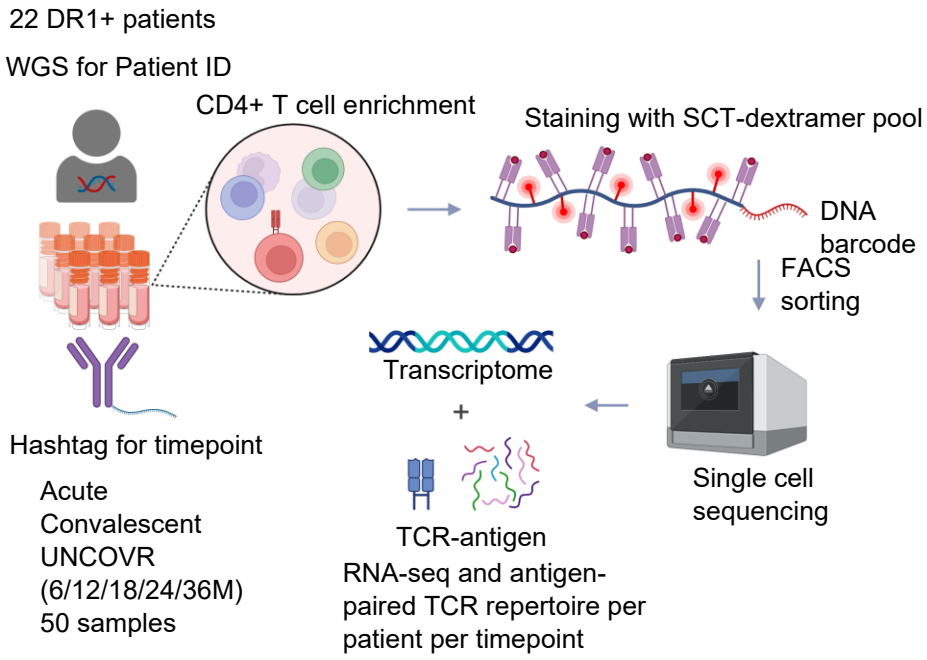


Figure 3.1 Experimental workflow of the high throughout identification of SARS-CoV-2 specific CD4+ T cells. PBMC samples from 22 HLA-DR1+ patients were barcoded by timepoints and pooled for CD4+ T cell enrichment, followed by SCT-dextramer staining and FACS sorting, and subjected to single cell sequencing for RNA expression, TCRαβ pairs, and SCT IDs through barcode. Whole genome sequencing (WGS) for each participant was used to demultiplex patient identity.

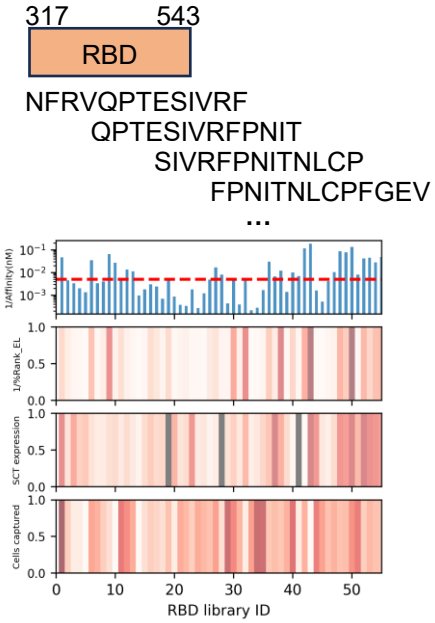


Figure 3.2 Strategy for unbiased screening of the RBD domain (top), and the correlation between SCT expression, cells captured and class II antigen-presentation prediction algorithms.

Library ID	Domain	Peptide	Antigen
1	RBD	NFRVQPTESIVRFPN	S_317-331
2	RBD	QPTESIVRFPNITNL	S_321-335
3	RBD	SIVRFPNITNLCPFG	S_325-339
4	RBD	FPNITNLCPFGEVFN	S_329-343
5	RBD	TNLCPFGEVFNATRF	S_333-347
6	RBD	TRFASVYAWNRKRIS	S_345-359
7	RBD	SVYAWNRKRISNCVA	S_349-363
8	RBD	WNRKRISNCVADYSV	S_353-367
9	RBD	CVADYSVLYNSASFS	S_361-375
10	RBD	YSVLYNSASFSTFKC	S_365-379
11	RBD	YNSASFSTFKCYGVS	S_369-383
12	RBD	SFSTFKCYGVSPTKL	S_373-387
13	RBD	FKCYGVSPTKLNDLC	S_377-391
14	RBD	GVSPTKLNDLCFTNV	S_381-395
15	RBD	TKLNDLCFTNVYADS	S_385-399
16	RBD	TNVYADSFVIRGDEV	S_393-407
17	RBD	ADSFVIRGDEVRQIA	S_397-411
18	RBD	VIRGDEVRQIAPGQT	S_401-415
19	RBD	DEVRQIAPGQTGKIA	S_405-419
20	RBD	QIAPGQTGKIADYNY	S_409-423
21	RBD	GQTGKIADYNYKLPD	S_413-427
22	RBD	KIADYNYKLPDDFTG	S_417-431
23	RBD	VIAWNSNNLDSKVGG	S_433-447
24	RBD	NSNNLDSKVGGNYNY	S_437-451
25	RBD	LDSKVGGNYNYLYRL	S_441-455
26	RBD	VGGNYNYLYRLFRKS	S_445-459
27	RBD	YNYLYRLFRKSNLKP	S_449-463
28	RBD	YRLFRKSNLKPFERD	S_453-467
29	RBD	RKSNLKPFERDISTE	S_457-471
30	RBD	LKPFERDISTEIYQA	S_461-475
31	RBD	ERDISTEIYQAGSTP	S_465-479
32	RBD	STEIYQAGSTPCNGV	S_469-483
33	RBD	YQAGSTPCNGVEGFN	S_473-487
34	RBD	STPCNGVEGFNCYFP	S_477-491
35	RBD	NGVEGFNCYFPLQSY	S_481-495
36	RBD	GFNCYFPLQSYGFQP	S_485-499
37	RBD	YFPLQSYGFQPTNGV	S_489-503
38	RBD	QSYGFQPTNGVGYQP	S_493-507
39	RBD	FQPTNGVGYQPYRVV	S_497-511
40	RBD	NGVGYQPYRVVVLSF	S_501-515
41	RBD	YQPYRVVVLSFELLH	S_505-519
42	RBD	RVVVLSFELLHAPAT	S_509-523
43	RBD	LSFELLHAPATVCGP	S_513-527
44	RBD	LLHAPATVCGPKKST	S_517-531
45	RBD	PATVCGPKKSTNLVK	S_521-535
46	RBD	KSTNLVKNKCVNFNF	S_529-543
47	S	CTFEYVSQPFLMDLE	S_166-180
48	S	ITRFQTLLALHRSYL	S_235-249
49	S	TRFQTLLALHRSYLT	S_236-250
50	S	FNFNGLTGTGVLTES	S_541-555
51	S	NLLLQYGSFCTQLNR	S_751-765
52	S	TQLNRALTGIAVEQD	S_761-775
53	S	CAQKFNGLTVLPPLL	S_851-865
54	S	TDEMIAQYTSALLAG	S_866-880
55	S	WTFGAGAALQIPFAM	S_886-900
56	S	IPFAMQMAYRFNGIG	S_896-910
57	S	TLVKQLSSNFGAISS	S_961-975
58	S	VQIDRLITGRLQSLQ	S_991-1005
59	S	LITGRLQSLQTYVTQ	S_996-1010
60	N	SWFTALTQHGKEDLK	N_51-65
61	N	KDGIIWVATEGALNT	N_127-141
62	N	AIVLQLPQGTTLPKG	N_156-170
63	M	GAVILRGHLRIAGHHLGR	M_141-158
64	M	PKEITVATSRTLSYYKL	M_165-181

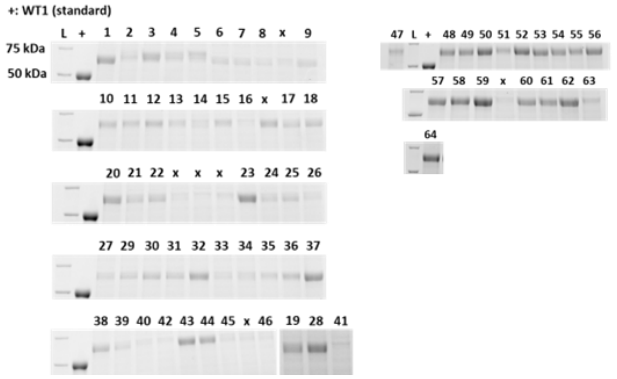


Figure 3.3 SARS-CoV-2 SCT library. top, the list of the RBD antigens and epitopes from other structural domains – non-RBD spike (S), membrane (M), nucleocapsid (N). bottom, SCT expression on SDS-PAGE gel. +, standard; L, ladder.

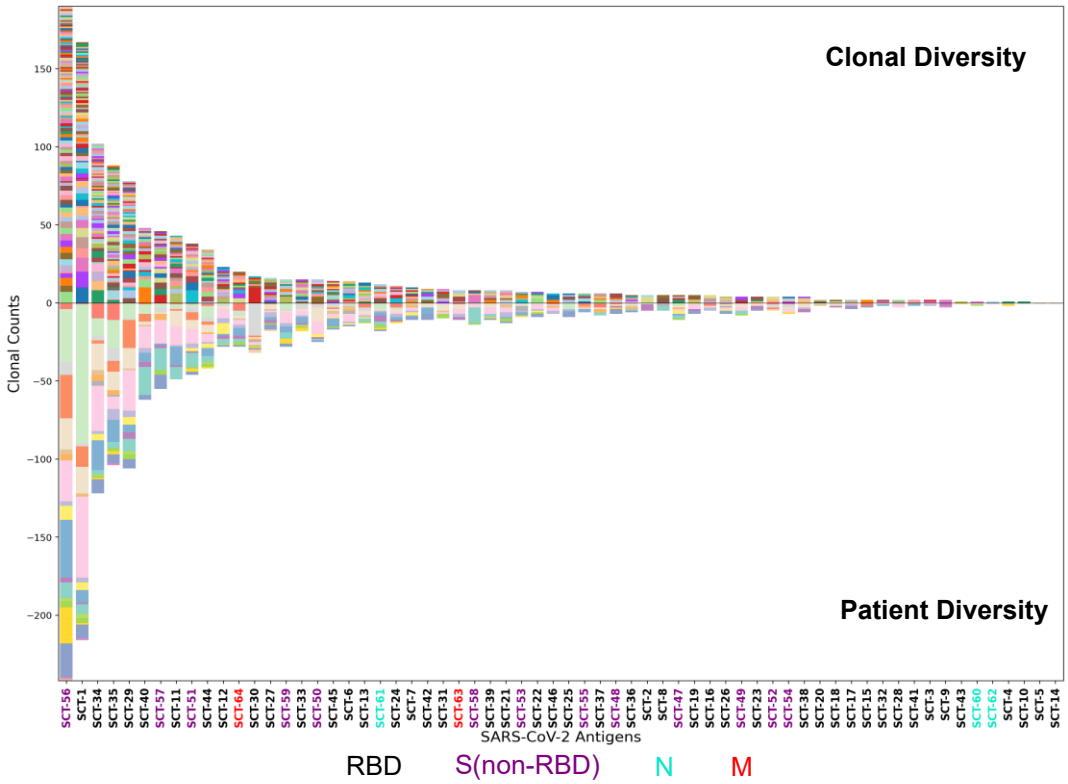


Figure 3.4 Clonal and patient diversity of CD4+ responses against each antigen. Each column represents an antigen from the RBD (black), non-RBD spike domain (purple), nuclear protein (teal), or membrane protein (red). top, TCR clonotypes identified reacting to an antigen. Bottom, patients that showed antigen-specific CD4+ T cell responses against each antigen.

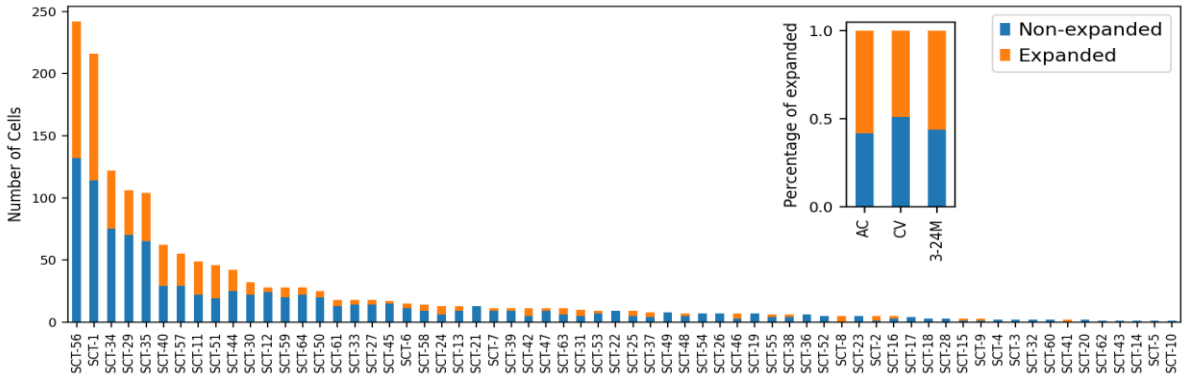


Figure 3.5 Clonal expansion of CD4+ T cell responses across antigen and timepoints.

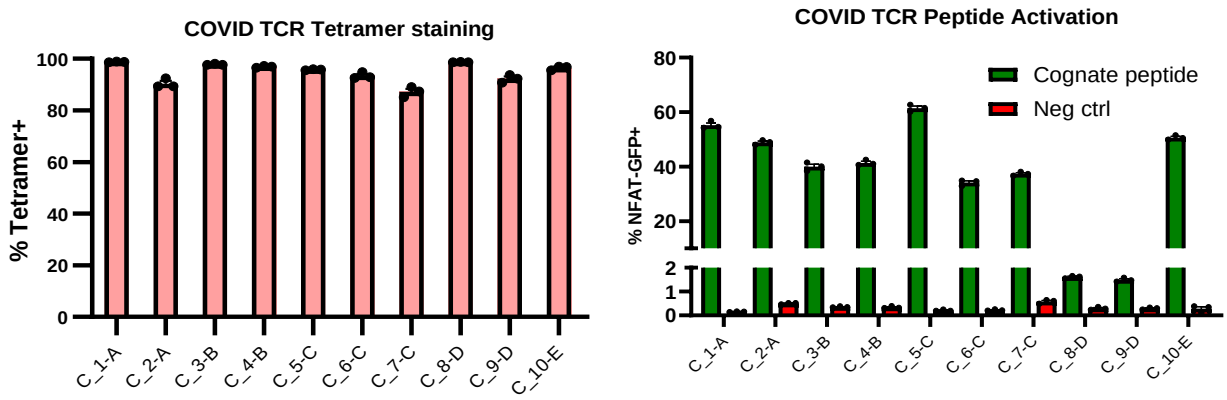


Figure 3.6 SARS-CoV-2 TCR validation through tetramer binding and peptide-pulse activation assay (n=3). ****$P < 0.0001$ for each group relative to the negative control peptide, determined by one-tailed independent t-test assuming equal variances.

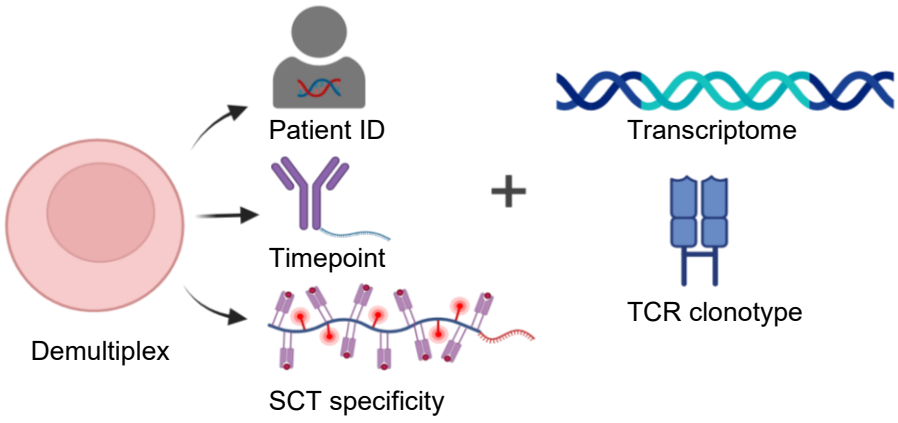


Figure 3.7 Coupling multiple information layers for each SARS-CoV-2-specific CD4+ T cell. Patient identity was resolved by comparing single nucleotide polymorphisms (SNPs) in transcriptome data with WGS patient profiles. DNA barcodes and hashtags permit demultiplexing for timepoint and SCT identity.

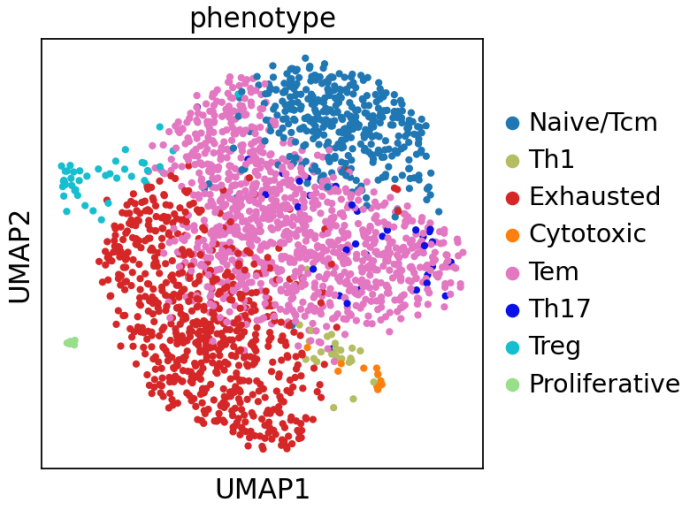


Figure 3.8 UMAP projection of CD4+ T cells.

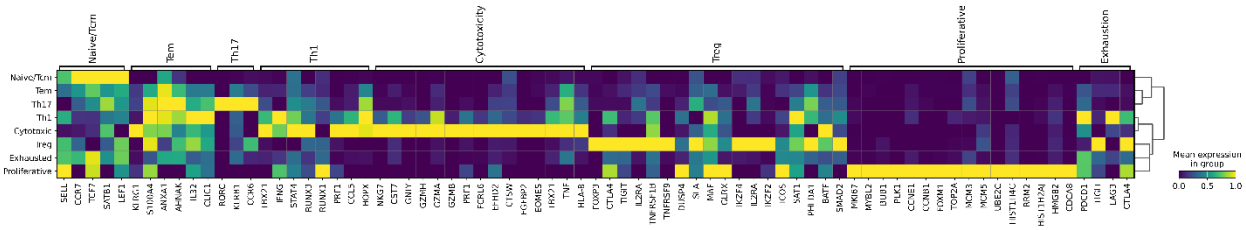


Figure 3.9 Signature gene markers for annotation of Leiden groups to T cell phenotypes.

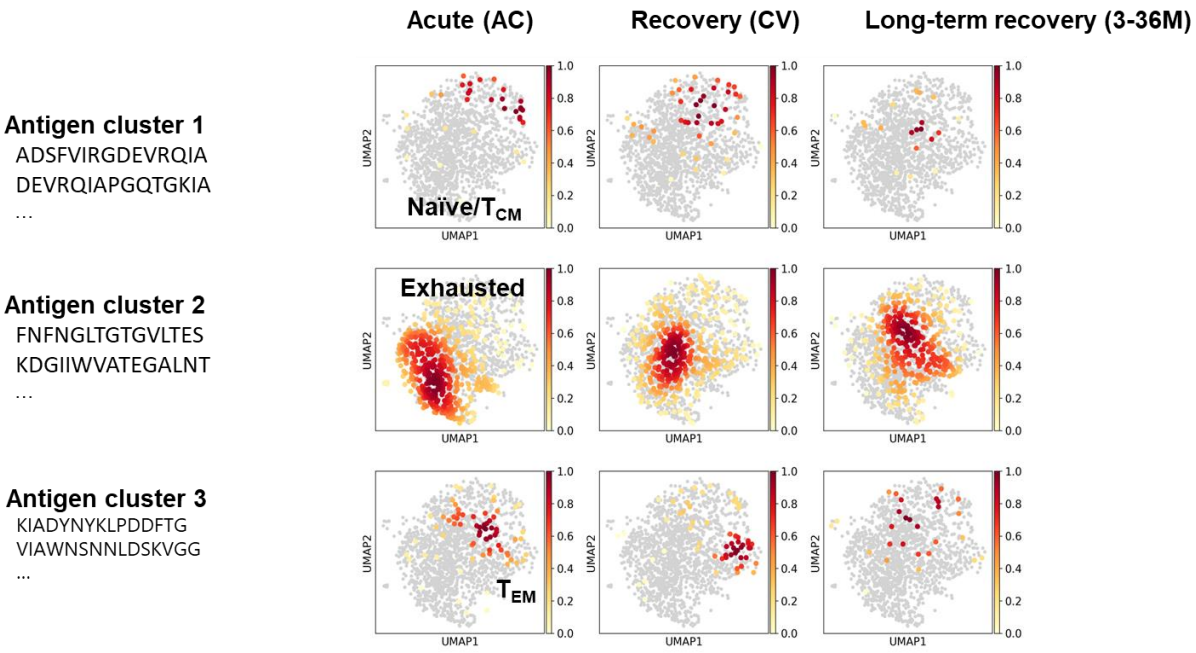


Figure 3.10 Trajectories of antigen-associated T cell phenotypes over time. Three clusters of antigens were identified based on the initial immune response induced during acute infection: the antigen cluster that predominantly induced naïve/T_{CM}, exhausted, or T_{EM} responses. CD4+ T cell responses against the antigen clusters were tracked over time into convalescent phase (CV), and long-term recovery (3-36 months).

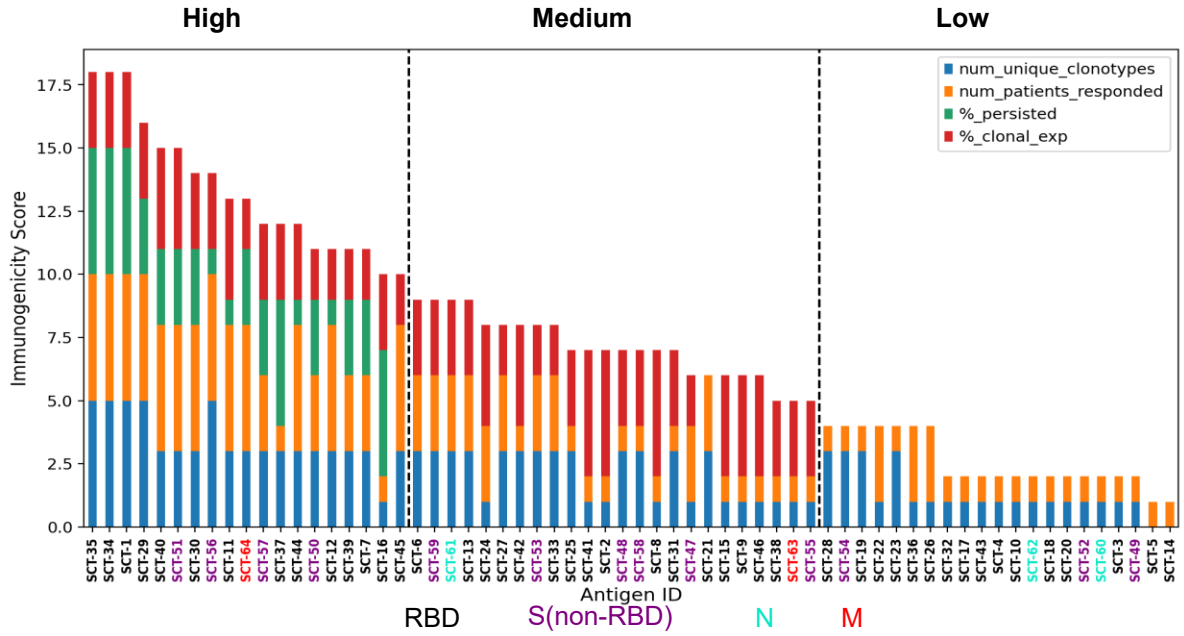


Figure 3.11 Ranking of DRB1*01:01 restricted SARS-CoV-2 antigens by immunogenicity.

Each property was scored from 0 to 5 based on a log2 scale, and then the properties summed.

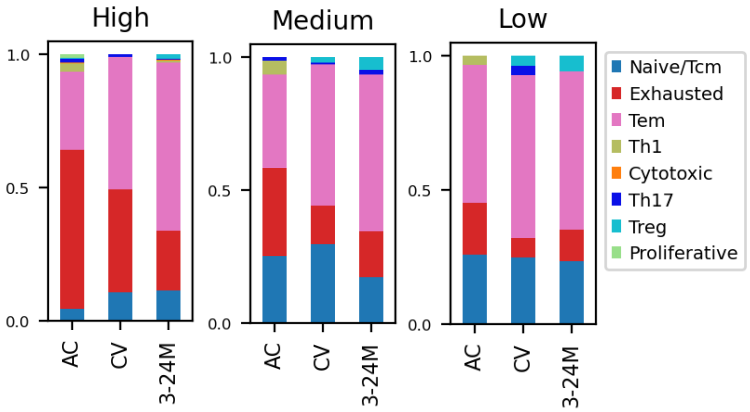


Figure 3.12 Kinetic evolution of CD4 T cell phenotypes induced by antigens of differing immunogenicity.

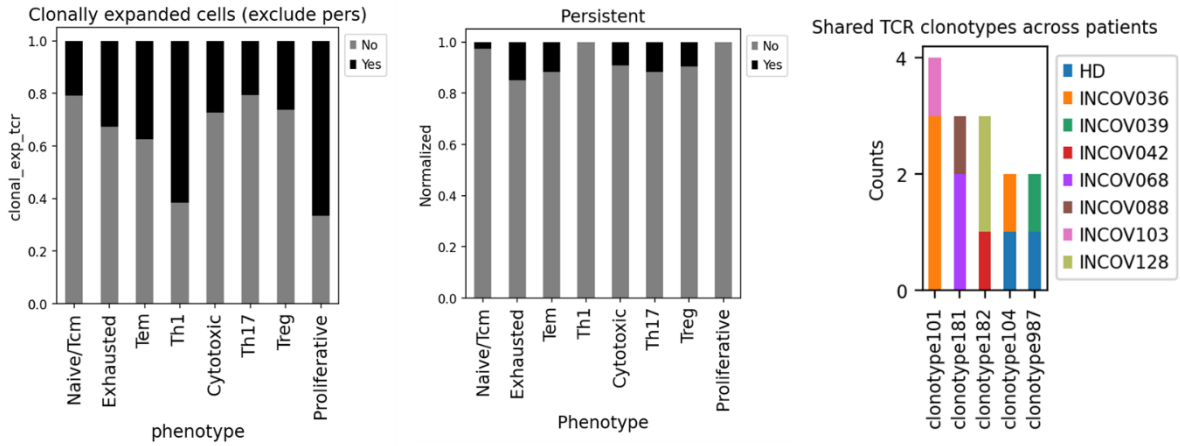


Figure 3.13 Clonal expansion and persistence across phenotypes and public TCRs.

Chapter 4: Identifying HPV-16 specific CD4 T cells for TCR T cell cancer immunotherapy

4.1 Introduction

Human papillomaviruses (HPVs) are one of the most prevalent sexually transmitted viruses affecting a large worldwide population. Over a hundred HPV subtypes have been identified but only a few of those are considered high-risk subtypes that may lead to cancer, which constitutes ~5% of human cancer incidences. HPV-16 has been identified as the most malignant variant and the main cause for cervical, oropharyngeal, and anogenital cancers[66].

HPV-16 is a small, non-enveloped DNA virus which interacts with and enters the undifferentiated cells in the squamous epithelium. The first stages of virion replication induce immune responses which usually clear out the infection spontaneously. However, in hosts with weakened immunity, viral infection persists and replicated virions accumulate and spread to deeper layers of cells under the epithelium through formation for microlesions, and if left untreated, will eventually lead to cancer.

The HPV-16 genome encodes two late-stage structural proteins and six early-stage non-structural proteins. E6 and E7 are two of the early-stage proteins which are critical for malignancy. E6 protein recruits ubiquitin ligase (E6AP) and p53 molecules to degrade p53, hence disrupting cell regulations. The E7 protein binds to the retinoblastoma protein, leading to its dissociation from E2F transcription factors. Proteins necessary for DNA replication are expressed again and the cell progresses into the S phase, resulting in uncontrolled cell division[34].

While the preventative HPV vaccine has done justice, patients infected with HPV-16 virus with late-stage cancer do not respond well to traditional treatments. Besides therapeutic vaccines for treating solid tumor, T cell-based therapy has been a focus of cancer immunotherapy. Despite the initial clinical efficacy of CD8+ T cells, persistence of the administered cells for prolonged anti-cancer activity is limited[67,68]. It has been reported that CD4+ T cells are required to facilitate CD8+ T cells in their cytotoxic effects and hence, the first in-human CD4+ TCR T-cell targeting cancer germline antigen MAGE-4 has been applied in clinical trial[69].

Here, we use the class II SCT library to discover E6 or E7-specific CD4+ T cells from HPV16+ precancerous patients. With the SCT library approach, we aim to identify multiple CD4 TCRs for combinatorial therapy to treat patients with a cocktail of CD4 and CD8 TCRs.

4.2 Methods

4.2.1 Building the HPV E6 and E7 SCT library using a liberal approach based on computational prediction

The class II SCT library encoding the E6 and E7 proteins encompasses 92 antigens varying from 13-25 amino acids. The antigens were selected based on the NetMHCIIpan4.0 and MixMHC2pred predictions. The top 15% of the antigens predicted to bind the DRB1*01:01 allele were prioritized. The SCTs were constructed, expressed, and purified as described in 2.2.3 and 2.2.4.

4.2.2 High throughput screening of HPV16-specific CD4+ T cells

The workflow is similar to the method described in 3.2.3. Out of the 92 SCTs, 85 of the SCTs were successfully expressed with a concentration applicable for downstream applications. In addition to the 85 HPV-16 SCTs, a panel of 11 SCTs encoding the common antigens such as influenza, EBV,

and CMV were also included. This 96-SCT panel was constructed into a pool of dextramers where each DNA barcode corresponds to a unique SCT. Four DRB1*01:01 patients at various timepoints were each labeled with unique combinations of the hashtag antibodies. After hashtagging, all samples were combined together for CD4+ T cell enrichment through MACS sorting, followed PKI incubation, staining with the dextramer pool together with the HIV negative tetramer control. Cells were then stained with anti-CD4 and anti-TCR surface antibodies, together with Total-seq antibodies for CD45RA, CD45RO, CCR7, CTLA4, CD25, CD127 tagged with DNA barcodes. Cells were then prepared for FACS and single cell sequencing for multiomics analysis using 10x Genomics kit.

4.2.3 TCR transduction into primary T cells

CD4+ T cells or CD8+ T cells were extracted via MACS sorting and resuspended at a density of 1 M/mL in Prime-XV media with 2% Physiologix serum replacement supplemented with IL-7 (581904, Biolegend) and IL-15 (570304, Biolegend), both at a final concentration of 12.5 ng/mL. Each 1M cells were activated by adding 10 µL of TransAct human CD3/CD28 activator (#130-111-160, Miltenyi) and incubated for 48 hours. For this experiment, 5M T cells were incubated in 5 mL of Prime-XV media in 6-well plates.

The sgRNA was reconstituted to a concentration of 120 µM by adding IDTE buffer. To prepare the ribonucleoprotein (RNP) complex, 0.75 uL of 62 uM Cas9 (Alt-R™ S.p. HiFi Cas9 Nuclease V3, IDT) was combined with 0.75 uL of 120 uM sgRNA (IDT) and incubated at room temperature for 10-20 minutes, ensuring thorough mixing as Cas9 contains glycerol. Separate RNP complexes were prepared for targeting TRAC and TRBC loci. While the RNP complex was incubating, the T cells were processed. T cells were activated beforehand and washed with media and then resuspended in 50 uL of T buffer.

Next, 60 uL of T buffer was added first, followed by 1 uL of 100 uM electroporation enhancer (EH), and then the RNP complexes for both α and β chains. The cell suspension was transferred into the RNP mixture and mixed thoroughly. Afterward, 3 mL of electrolytic buffer was added to the Neon electroporation tube, which was then inserted into the Neon system. The prepared cell/RNP mixture was transferred to the electroporation tip using the Neon pipet. Electroporation was conducted using the settings optimized for editing primary T cells: 1700V, 10 ms pulse duration, with 3 pulses. Post-electroporation, the cells were evenly distributed into pre-warmed media and allowed to recover in the incubator for 30 minutes.

For lentiviral transduction of TCR, the 5 million electroporated T cells were rested in warm media for 30 minutes before being divided into 10 wells, with each well containing 500,000 cells. Subsequently, 500 µL of PRIME-XV media and 1 mL of viral particles were added to each well for the transduction process.

4.2.4 ELISA assay

For peptide pre-loading, 100,000 K562 cells were plated per well in R10 media. 20 µg of peptide was added to each well containing the K562 cells, which equates to a final concentration of 0.1 mg/mL. The cells were incubated at 37°C for 1-2 hours to allow sufficient peptide loading. Following incubation, the cells were spun down, washed once with fresh media, spun down again, and resuspended in 100 µL of R10. For preparing the primary T cells, the required amount of cells was aliquoted, spun down, and resuspended in R10 media. To set up the co-culture, 100,000 peptide-loaded K562 cells were combined with 100,000 primary CD4 T cells in flat-bottom plates, ensuring thorough mixing. The co-cultures were then incubated for 16-20 hours. ELISA assays were performed according to manufacturer's manual.

4.2.5 Cytotoxicity assay

The cytotoxic assay was conducted through IncuCyte. DR1-K562 cells were preloaded with target peptide as described in 4.2.4 and seeded at a final cell density of 1M/mL. 10 µL of the 100X working solution of Cytolight Rapid Dye to every 1 mL of K562 cell suspension. The cells were incubated for 20 minutes at 37°C, and gently mixed every 10 minutes to allow the dye to fully bind. Next, any excess dye was washed out by adding 6 mL of R10 media to every 1 mL of cell/dye suspension. The suspension was centrifuged at 300g for 3 minutes, and the supernatant was completely aspirated to remove residual fluorescence debris. The cells were then resuspended to a concentration of 100,000-500,000 cells/mL in R10 media. Subsequently, 10,000-50,000 K562 cells were seeded per well (100 µL per well) and allowed to settle at room temperature for 30 minutes. The apoptosis reagent was prepared by diluting Caspase-3/7 green to a final concentration of 20 µM using a 1:250 dilution (4x) in culture media, ensuring enough reagent for 50 µL per well. 50 µL of this 4x apoptosis reagent was added to each well. Next, the immune effector cells (primary TCR-T cells) were added to the DR1-K562 cells at an effector-to-target (E) ratios of 1:1. The assay plate was placed in the IncuCyte Live-Cell Analysis System for 44-hour repeat scanning.

4.3 Results and Discussion – IND-enabling validation

4.3.1 Experimental design to identify HPV-specific CD4+ T cells

Although CD8+ T cell immunotherapies have been explored for many years, CD4+ T cell-based therapies remain largely unexplored. We subsequently applied the SCT library approach to identify and characterize HPV16-specific CD4+ T cells for potential application in cancer immunotherapy. Advanced-stage HPV+ cancers have shown poor prognosis with traditional treatment[70,71]. The oncogenic HPV E6 and E7 proteins are universally expressed in HPV+ cancer cells, and E6 and

E7-reactive CD8+ T cells have demonstrated clinical efficacy in TCR-engineered cancer immunotherapies (TCR-T)[67,68]. To identify HPV-specific CD4+ T cells, we constructed an 85-element DR1-restricted SCT library encoding 13-25-mer antigens from E6 and E7 (see Methods, Table S2). PBMCs were collected from patients enrolled in either an HPV-16 specific DNA vaccine trial or an artesunate treatment trial[72]. Patients in these trials presented prolonged HPV-16 infection with Cervical Intraepithelial Neoplasia (CIN) II/III lesions. Patients received vaccines covering E6 and E7 proteins at Week 0, 4, and 9/10, and PBMC samples were collected before vaccination, at Week 9, and at Week 15 after completing the entire vaccine series (**Figure 4.1**). PBMC samples were prelabeled with hashtag antibodies for patient and timepoint identification. Similarly to the approach in **Figure 3.1**, CD4+ T cells were enriched and stained with DNA-barcoded SCT dextramers. Antigen-specific CD4+ T cells were sorted and subjected to single-cell sequencing for gene expression analysis, TCR repertoire profiling, and antigen identification.

4.3.2 Validation of the identified TCRs

We identified HPV-16 specific CD4+ T cells and selected an initial subset of clones for validations (**Table S3**). We discovered five CD4+ TCRs (H1-H5) specific to E6 epitopes, and validated their specificity through tetramer binding and peptide-pulsed activation in TCR-transduced NFAT-GFP Jurkat cells (**Figure 4.2**). H1, H2, and H4 displayed cross-reactivity against a family of peptides (**Table S3**) that share the same 9-mer core antigen, which is likely the key recognition motif for the TCRs[73]. Surprisingly, we found that, according to the NFAT-GFP Jurkat cell assay, binding does not necessarily lead to downstream activation. While H1 and H4 TCRs showed medium to high tetramer binding signal, this was accompanied by little evidence of T cell activation (**Figure 4.2**).

4.3.3 Functional validation in primary CD4 T cells

To further confirm the therapeutic potential of the HPV CD4 TCRs, we transduced the TCRs into primary healthy donor TCR knock-out CD4 T cells (**Figure 4.3 and Figure 4.5**). We compared the knock-in efficiency of the CD4 TCRs into primary CD4+ and CD8+ T cells. Notably, primary CD8+ T cells consistently showed significantly lower knock-in efficiency than primary CD4+ T cells (**Figure 4.4**), perhaps reflecting a mechanistic preference for certain TCRs to be expressed in CD4+ T cells versus CD8+ T cells.

Next, we assessed a suite of antigen-stimulated effector functions of the CD4 TCRs. We first exposed the cells to peptide-pulsed antigen-presenting DR1+ K562 cells and measured T cell activation through the secretion of IFNγ, TNFα, IL2, and GZMB (**Figure 4.6**). Cells engineered with H2, H3, and H5 TCRs displayed on-target secretions of all cytokines. In contrast, the H1 TCR showed low secretion of IFNγ and GZMB secretion, with no detectable secretion of TNFα or IL2. The H4 TCR showed no cytokine secretion for four of its cognate peptides; however, when stimulated with the F-7 peptide, it secreted high levels of IFNγ and GZMB. Notably, the H4 TCR showed tetramer binding to F-7, although, as described above, NFAT-based activation signal was not observed, suggesting that these standard T cell functional assays each yield an incomplete picture.

4.3.4 Cytotoxicity of the CD4 TCRs

Cytotoxic CD4+ T cells are increasingly recognized for their role in anti-tumor immunity[3]. Therefore, we further investigated the cytotoxic functions of the five HPV-specific TCRs. Live DR1+ K562 cells were preloaded with peptides, labeled with the Cytolight Rapid Dye, and co-cultured with the HPV TCR+ primary CD4 T cells (**Figure 4.7**). Apoptotic activity was measured

over a 42-hour incubation. The TCR:epitope combination of H3:F-2 showed the highest cytotoxic effector function, while H5:F-2, H2:L-1 and H4:F-7 demonstrated medium cytotoxic activity. All TCR:epitope pairs showed measurable apoptosis signals relative to the negative controls. These results suggest that this assay, although not commonly used to validate CD4+ T cells activated by class II antigens, could be valuable for screening CD4+ T cell TCRs for cancer immunotherapies.

4.4 Figures

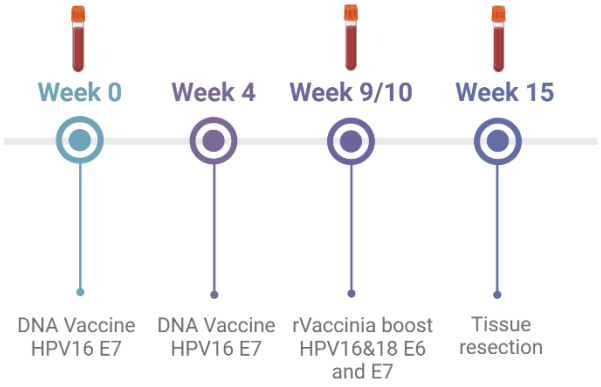


Figure 4.1 HPV DNA vaccine clinical trial scheme.

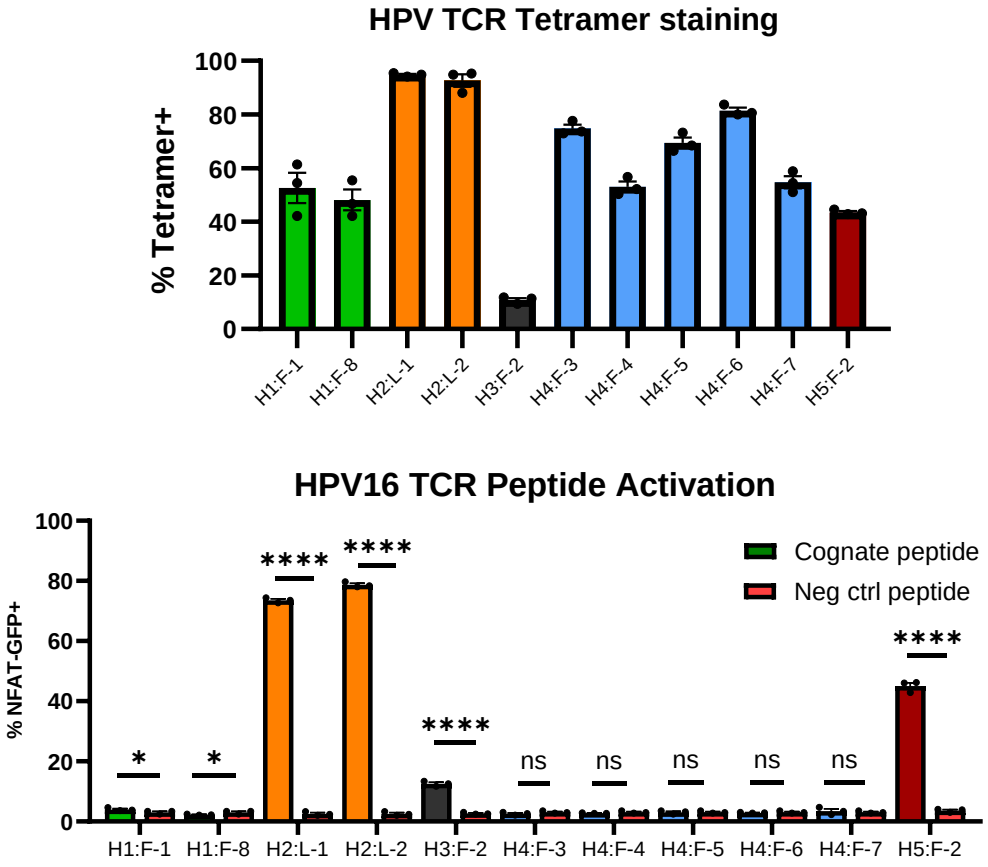


Figure 4.2 Validation of HPV CD4 TCRs H1-H5 against antigens 1-8 through tetramer binding and peptide-pulsed activation assays (n=3). ****P < 0.0001, *P < 0.05, ns P >0.05 labeled for each group relative to the negative control peptide, determined by one-tailed independent t-test assuming equal variances.

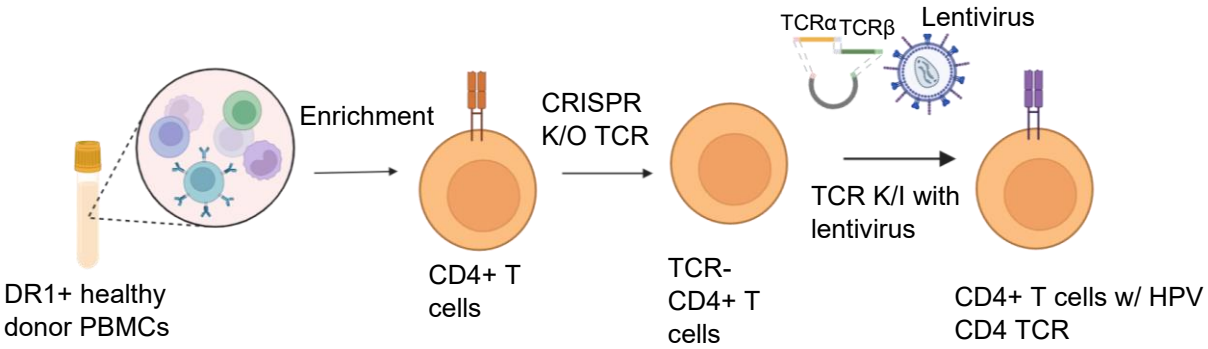


Figure 4.3 Engineering of HPV TCR+ primary CD4 T cells. Haplotype matched PBMC samples were enriched for CD4+ T cells and the endogenous TCRs were knocked out using CRISPR gene editing. HPV TCRs were then transduced through lentivirus. TCR expression and tetramer binding were confirmed (n=3).

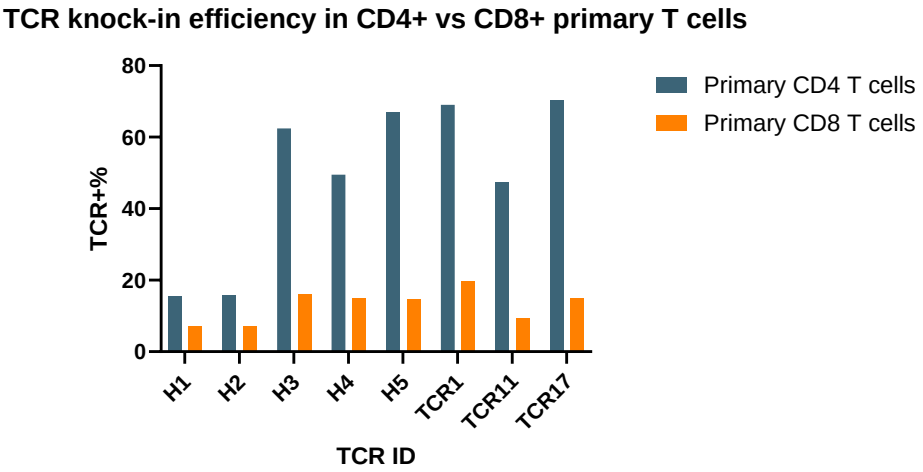


Figure 4.4 HPV CD4 TCR transduction efficiency into primary CD4+ T cells versus primary CD8+ T cells.

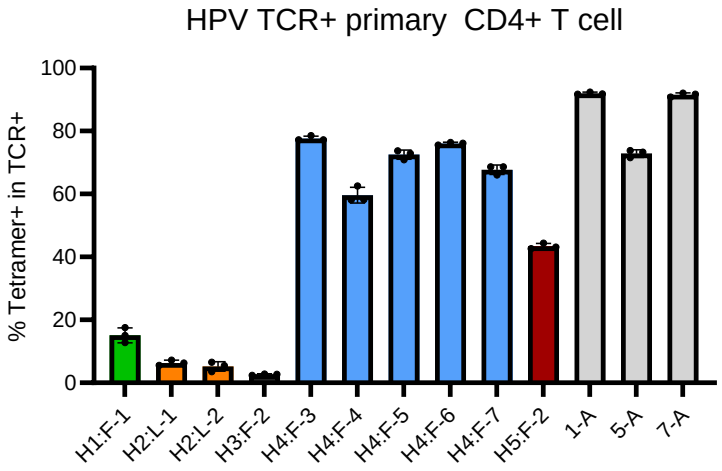
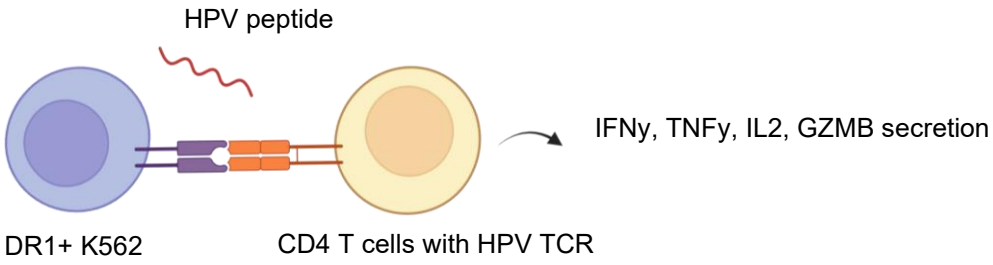
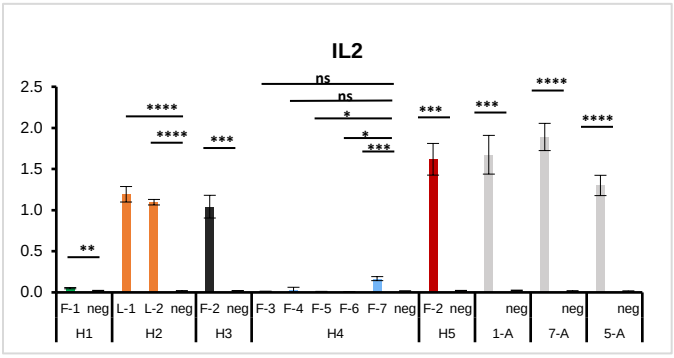
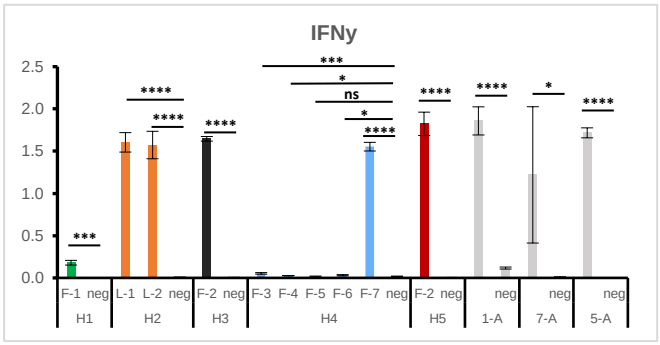
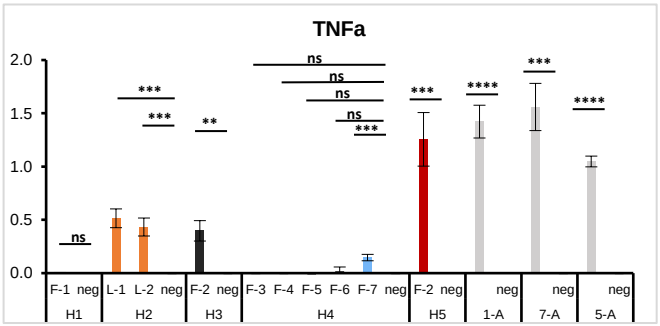


Figure 4.5 Tetramer binding in HPV TCR+ CD4+ T cells (n=3).


HPV peptide
IFNy, TNFy, IL2, GZMB secretion
DR1+ K562
CD4 T cells with HPV TCR


IL2
F-1 neg L-1 L-2 neg F-2 neg F-3 F-4 F-5 F-6 F-7 neg F-2 neg neg neg neg
H1 H2 H3 H4 H5 1-A 7-A 5-A


IFNy
F-1 neg L-1 L-2 neg F-2 neg F-3 F-4 F-5 F-6 F-7 neg F-2 neg neg neg neg
H1 H2 H3 H4 H5 1-A 7-A 5-A


TNFa
F-1 neg L-1 L-2 neg F-2 neg F-3 F-4 F-5 F-6 F-7 neg F-2 neg neg neg neg
H1 H2 H3 H4 H5 1-A 7-A 5-A

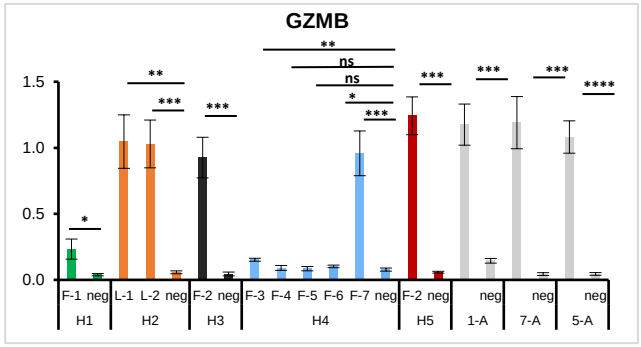

GZMB
F-1 neg L-1 L-2 neg F-2 neg F-3 F-4 F-5 F-6 F-7 neg F-2 neg neg neg neg
H1 H2 H3 H4 H5 1-A 7-A 5-A

Figure 4.6 Functional performance of the TCRs evaluated through antigen-induced production of IFNγ, TNFα, IL2, and GZMB in an ELISA assay (n=3). ****P < 0.0001, ***P < 0.001, **P < 0.01, *P < 0.05, ns P ≥ 0.05 labeled for each group relative to the negative control peptide, determined by one-tailed independent t-test assuming equal variances.

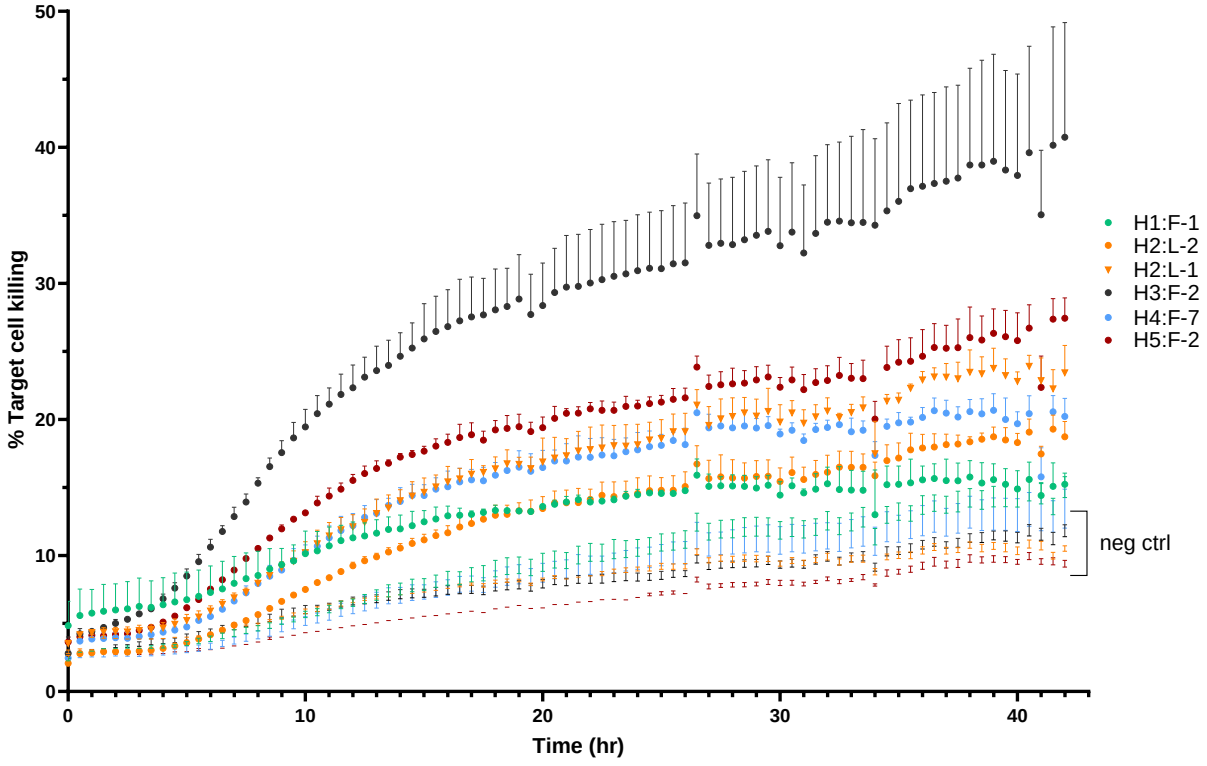


Figure 4.7 Cytotoxicity of TCR-engineered T cells toward DR1-K562 cells pulsed with cognate HPV-16 E6 antigens and tracked over 44 hours (n=2).

Reference

1. Luckheeram, R. V., Zhou, R., Verma, A. D. & Xia, B. CD4+T Cells: Differentiation and Functions. *Clinical and Developmental Immunology* **2012**, 925135 (2012).
2. Tay, R. E., Richardson, E. K. & Toh, H. C. Revisiting the role of CD4+ T cells in cancer immunotherapy—new insights into old paradigms. *Cancer Gene Ther* **28**, 5–17 (2021).
3. Oh, D. Y. & Fong, L. Cytotoxic CD4+ T cells in cancer: Expanding the immune effector toolbox. *Immunity* **54**, 2701–2711 (2021).
4. Gangaev, A. *et al.* Identification and characterization of a SARS-CoV-2 specific CD8+ T cell response with immunodominant features. *Nat Commun* **12**, 2593 (2021).
5. Bentzen, A. K. *et al.* Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nat Biotechnol* **34**, 1037–1045 (2016).
6. 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).
7. 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).
8. Puig-Saus, C. *et al.* Neoantigen-targeted CD8+ T cell responses with PD-1 blockade therapy. *Nature* **615**, 697–704 (2023).
9. Foy, S. P. *et al.* Non-viral precision T cell receptor replacement for personalized cell therapy. *Nature* **615**, 687–696 (2023).
10. Newell, E. W., Klein, L. O., Yu, W. & Davis, M. M. Simultaneous detection of many T-cell specificities using combinatorial tetramer staining. *Nat Methods* **6**, 497–499 (2009).
11. Bjorkman, P. J. *et al.* Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**, 506–512 (1987).

12. Brown, J. H. *et al.* Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364**, 33–39 (1993).
13. Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. *The Journal of Experimental Medicine* **178**, 27 (1993).
14. Moravec, Z. *et al.* Discovery of tumor-reactive T cell receptors by massively parallel library synthesis and screening. *Nat Biotechnol* 1–9 (2024) doi:10.1038/s41587-024-02210-6.
15. Frentsch, M. *et al.* Direct access to CD4+ T cells specific for defined antigens according to CD154 expression. *Nat Med* **11**, 1118–1124 (2005).
16. Poloni, C. *et al.* T-cell activation–induced marker assays in health and disease. *Immunology & Cell Biology* **101**, 491–503 (2023).
17. Le Bert, N. *et al.* SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature* **584**, 457–462 (2020).
18. Uchtenhagen, H. *et al.* Efficient ex vivo analysis of CD4+ T-cell responses using combinatorial HLA class II tetramer staining. *Nat Commun* **7**, 12614 (2016).
19. Kozono, H., White, J., Clements, J., Marrack, P. & Kappler, J. Production of soluble MHC class II proteins with covalently bound single peptides. *Nature* **369**, 151–154 (1994).
20. Day, C. L. *et al.* Ex vivo analysis of human memory CD4 T cells specific for hepatitis C virus using MHC class II tetramers. *J Clin Invest* **112**, 831–842 (2003).
21. Moon, J. J. *et al.* Naive CD4(+) T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity* **27**, 203–213 (2007).
22. Scriba, T. J. *et al.* Ultrasensitive detection and phenotyping of CD4+ T cells with optimized HLA class II tetramer staining. *J Immunol* **175**, 6334–6343 (2005).

23. Kwok, W. W. *et al.* Frequency of epitope-specific naive CD4(+) T cells correlates with immunodominance in the human memory repertoire. *J Immunol* **188**, 2537–2544 (2012).
24. Christophersen, A. *et al.* Distinct phenotype of CD4+ T cells driving celiac disease identified in multiple autoimmune conditions. *Nat Med* **25**, 734–737 (2019).
25. Garber, K. Driving T-cell immunotherapy to solid tumors. *Nature Biotechnology* **36**, 215–219 (2018).
26. Perica, K., Varela, J. C., Oelke, M. & Schneck, J. Adoptive T cell immunotherapy for cancer. *Rambam Maimonides Med J* **6**, e0004 (2015).
27. Zhang, J. & Wang, L. The Emerging World of TCR-T Cell Trials Against Cancer: A Systematic Review. *Technol Cancer Res Treat* **18**, 1533033819831068 (2019).
28. Sun, Y. *et al.* Evolution of CD8+ T Cell Receptor (TCR) Engineered Therapies for the Treatment of Cancer. *Cells* **10**, 2379 (2021).
29. Liu, Y. *et al.* TCR-T Immunotherapy: The Challenges and Solutions. *Front. Oncol.* **11**, (2022).
30. Neefjes, J., Jongsma, M. L. M., Paul, P. & Bakke, O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol* **11**, 823–836 (2011).
31. Martincorena, I. & Campbell, P. J. Somatic mutation in cancer and normal cells. *Science* **349**, 1483–1489 (2015).
32. Rudolph, M. G. & Wilson, I. A. The specificity of TCR/pMHC interaction. *Current Opinion in Immunology* **14**, 52–65 (2002).
33. Gragert, L., Madbouly, A., Freeman, J. & Maiers, M. Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. *Human Immunology* **74**, 1313–1320 (2013).

34. The Allele Frequency Net Database - Allele, haplotype and genotype frequencies in Worldwide Populations. https://www.allelefrequencies.net/.
35. Ng, R. H., Lee, J. W., Zhang, R. & Heath, J. R. Class I and II HLA allele coverage of the US population accounting for demographics. *Unpublished manuscript* (2024).
36. Dolton, G. *et al.* More tricks with tetramers: a practical guide to staining T cells with peptide-MHC multimers. *Immunology* **146**, 11–22 (2015).
37. Dahal-Koirala, S. *et al.* Discriminative T-cell receptor recognition of highly homologous HLA-DQ2-bound gluten epitopes. *J Biol Chem* **294**, 941–952 (2019).
38. Bodd, M. *et al.* Direct cloning and tetramer staining to measure the frequency of intestinal gluten-reactive T cells in celiac disease. *Eur J Immunol* **43**, 2605–2612 (2013).
39. Zhu, X. *et al.* A recombinant single-chain human class II MHC molecule (HLA-DR1) as a covalently linked heterotrimer of α chain, β chain, and antigenic peptide, with immunogenicity in vitro and reduced affinity for bacterial superantigens. *European Journal of Immunology* **27**, 1933–1941 (1997).
40. Thayer, W. P., Dao, C. T., Ignatowicz, L. & Jensen, P. E. A novel single chain I-Ab molecule can stimulate and stain antigen-specific T cells. *Molecular Immunology* **39**, 861–870 (2003).
41. Zhang, X. & Sjöblom, T. Targeting Loss of Heterozygosity: A Novel Paradigm for Cancer Therapy. *Pharmaceuticals* **14**, 57 (2021).
42. Zhang, Y., Liu, Z., Wei, W. & Li, Y. TCR engineered T cells for solid tumor immunotherapy. *Exp Hematol Oncol* **11**, 38 (2022).
43. Chour, W. Molecular Technologies for Antigen-Based Immunity. (California Institute of Technology, 2021). doi:10.7907/z20t-nq62.

44. Ayyoub, M. *et al.* Monitoring of NY-ESO-1 specific CD4+ T cells using molecularly defined MHC class II/His-tag-peptide tetramers. *Proceedings of the National Academy of Sciences* **107**, 7437–7442 (2010).

45. Rhode, P. R. *et al.* Single-chain MHC class II molecules induce T cell activation and apoptosis. *J Immunol* **157**, 4885–4891 (1996).

46. Deng, L., Langley, R. J., Wang, Q., Topalian, S. L. & Mariuzza, R. A. Structural insights into the editing of germ-line–encoded interactions between T-cell receptor and MHC class II by Vα CDR3. *Proceedings of the National Academy of Sciences* **109**, 14960–14965 (2012).

47. Galperin, M. *et al.* CD4+ T cell–mediated HLA class II cross-restriction in HIV controllers. *Science Immunology* **3**, eaat0687 (2018).

48. Anczurowski, M. *et al.* Mechanisms underlying the lack of endogenous processing and CLIP-mediated binding of the invariant chain by HLA-DP84Gly. *Sci Rep* **8**, 4804 (2018).

49. Glanville, J. *et al.* Identifying specificity groups in the T cell receptor repertoire. *Nature* **547**, 94–98 (2017).

50. Rosskopf, S. *et al.* Creation of an engineered APC system to explore and optimize the presentation of immunodominant peptides of major allergens. *Sci Rep* **6**, 31580 (2016).

51. 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).

52. Hu, B., Guo, H., Zhou, P. & Shi, Z.-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* **19**, 141–154 (2021).

53. Hardenbrook, N. J. & Zhang, P. A structural view of the SARS-CoV-2 virus and its assembly. *Current Opinion in Virology* **52**, 123–134 (2022).

54. Yang, H. & Rao, Z. Structural biology of SARS-CoV-2 and implications for therapeutic development. *Nat Rev Microbiol* **19**, 685–700 (2021).

55. Lamers, M. M. & Haagmans, B. L. SARS-CoV-2 pathogenesis. *Nat Rev Microbiol* **20**, 270–284 (2022).

56. Grifoni, A. *et al.* Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell* **181**, 1489-1501.e15 (2020).

57. Sette, A., Sidney, J. & Crotty, S. T Cell Responses to SARS-CoV-2. *Annu Rev Immunol* **41**, 343–373 (2023).

58. Moderbacher, C. R. *et al.* Antigen-Specific Adaptive Immunity to SARS-CoV-2 in Acute COVID-19 and Associations with Age and Disease Severity. *Cell* **183**, 996-1012.e19 (2020).

59. Dan, J. M. *et al.* Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* **371**, eabf4063 (2021).

60. Peng, Y. *et al.* Broad and strong memory CD4+ and CD8+ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. *Nat Immunol* **21**, 1336–1345 (2020).

61. Mateus, J. *et al.* Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. *Science* **370**, 89–94 (2020).

62. Nelde, A. *et al.* SARS-CoV-2-derived peptides define heterologous and COVID-19-induced T cell recognition. *Nat Immunol* **22**, 74–85 (2021).

63. Wells, D. K. *et al.* Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction. *Cell* **183**, 818-834.e13 (2020).

64. Harrington, L. E., Janowski, K. M., Oliver, J. R., Zajac, A. J. & Weaver, C. T. Memory CD4 T cells emerge from effector T-cell progenitors. *Nature* **452**, 356–360 (2008).

65. Harrington, L. E., Janowski, K. M., Oliver, J. R., Zajac, A. J. & Weaver, C. T. Memory CD4 T cells emerge from effector T-cell progenitors. *Nature* **452**, 356–360 (2008).

66. Malik, S., Sah, R., Muhammad, K. & Waheed, Y. Tracking HPV Infection, Associated Cancer Development, and Recent Treatment Efforts—A Comprehensive Review. *Vaccines* **11**, 102 (2023).

67. Doran, S. L. *et al.* T-Cell Receptor Gene Therapy for Human Papillomavirus–Associated Epithelial Cancers: A First-in-Human, Phase I/II Study. *JCO* **37**, 2759–2768 (2019).

68. 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).

69. Lu, Y.-C. *et al.* Treatment of Patients With Metastatic Cancer Using a Major Histocompatibility Complex Class II–Restricted T-Cell Receptor Targeting the Cancer Germline Antigen MAGE-A3. *JCO* **35**, 3322–3329 (2017).

70. Trimble, C. L. & Frazer, I. H. Development of therapeutic HPV vaccines. *The Lancet Oncology* **10**, 975–980 (2009).

71. Ramondetta, L. What Is the Appropriate Approach to Treating Women With Incurable Cervical Cancer? *Journal of the National Comprehensive Cancer Network* **11**, 348–355 (2013).

72. 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. *Lancet* **386**, 2078–2088 (2015).

73. Sant'Angelo, D. B. *et al.* The Specificity and Orientation of a TCR to its Peptide–MHC Class II Ligands. *Immunity* **4**, 367–376 (1996).