Molecular Technologies for Antigen-Based Immunity

Thesis by
William Chour

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2021
Defended March 5th, 2021

© 2021

William Chour
ORCID: 0000-0003-1817-0123

All rights reserved.

ACKNOWLEDGEMENTS

I would like to first thank my advisor, Professor James Heath, for supporting me throughout these five years of my PhD degree. Jim's unyielding passion for science has been the primary motivational force that enables his students to ask fundamental questions in our projects and to develop the means to answer them. My time spent working in his lab has shown me that there is no end in sight for this pursuit, both in breadth and depth. I would also like to thank the advisors on my committee, Professors Mikhail Shapiro, Ellen Rothenberg, Changhuei Yang, and Matt Thomson. From our exchanges, I came to realize that having the opportunity to present my science to an engaging, inquisitive audience is equally as important as having the opportunity to conduct science. I am incredibly grateful for their expertise and acuity in identifying in my project the key areas of weakness or focus that required emphasis along the way toward completion of my thesis.

To the members of the Heath lab, I am grateful to have had the opportunity to be surrounded by so much scientific passion. I would like to especially thank Songming Peng, Alphonsus Ng, and Alex Xu. These postdocs served as great mentors during the formative years of my graduate school training, and enabled me to become a more independent thinker. I would also like to thank the members of the antigen team of the lab, namely Jingyi Xie, Rongyu Zhang, Dan Yuan, Lesley Jones, Jongchan Choi, and Yapeng Su. I would not be here today without the immense amount of experimental support that they provided throughout our projects. I am grateful to Matt Idso, Sunga Hong, and Jimmi Hopkins, the lab experts on all matters peptide-related, for their assistance with peptide synthesis. Lastly, I would also like to give additional thanks to Jingxin Liang; while we never had any overlapping projects, I am grateful to her for bringing some much-needed camaraderie to the lab. The transition from Pasadena to Seattle was a chaotic chapter in our academic pursuits, so having some laughter and reprieve was much appreciated.

Coupled with this migration from California to Washington, I soon came to realize that having a local network of peers, mentors, and collaborators would be vital toward success in the lab. For this, I am thankful for all the friendships formed from my time spent with the labs at the Fred Hutch. From Dr. Roland Strong's lab, Kathryn Finton was extraordinarily helpful in easing me into the science behind single-chain trimers. Her expertise in all things structural made for enjoyable discussions into potential future projects and new ideas that neither of us had time for, but nevertheless were

entertaining to ponder upon. I am also grateful to her for helping me reach out to other key members of the Fred Hutch who would play a crucial role in assisting with the more biological components of my projects. From Dr. John Lee's lab, Diana DeLucia was incredibly helpful with protein validation experiments, and provided us with cell samples to conduct functional assays. From Dr. Phil Greenberg's lab, Tom Schmitt, Rachel Perret, Tijana Martinov, Mary Chaffee, and Jonah Butler, could not have made for a better team of immunology experts. Tom's guidance and discussion regarding antigen-specific T cells was extremely helpful. Our lab was primarily focused on technological matters of the project, so I appreciate the time taken by Tijana in showing me the biological side of things, from cell expansion to flow cytometry and everything in between. I would also like to note Mary's support for many of the experiments we conducted in this past year; her incredible willingness to spend long days (and nights) in the lab speaks to her character and her passion for science. Finally, I also wanted to express my appreciation to Dr. Jason Goldman, his team at the Swedish Medical Center, and the coordinators from ISB (Kim Murray, Lee Rowen, Rick Edmark, Yong Zhou, among many others), for helping to set into motion an extraordinarily ambitious series of projects. The COVID-19 pandemic made 2020 an insanely productive but crazy year in terms of what we had to accomplish given our window of time. Dr. Goldman's team was able to rapidly coordinate with ISB to put together a large set of COVID-19 blood samples for us to perform many of the analyses discussed in Chapter 3. Their involvement was crucial in allowing us to accomplish so much in our COVID projects, and their impact will undoubtedly be imprinted into the lab's future endeavors given the technological outcomes we achieved from the collaboration.

With that said, I would also like to thank our collaborators from California. Our early work with the Parker Institute on cancer-related projects required guidance from members of Dr. Antoni Ribas' lab, specifically Dr. Begonya Comin-Anduix, whose patience and willingness to accommodate us into her busy schedule enabled us to gather necessary data for the TESLA-related projects and papers. Last but not least, I would like to thank fellow graduate students Mao Zhiyuan and Pavlo Nesterenko from Dr. Owen Witte's lab for our collaboration together on both cancer and COVID projects. Our initial work came toward the end of my time as a graduate student, but I am sure it will open many doors for the two labs to coordinate on new projects. I am excited to see future developments of the Heath lab that arise from collaborations with all the aforementioned.

ABSTRACT

The presence and proliferation antigen-specific T cells is a defining characteristic of an adaptive immune response against various disease types (viral, fungal, bacterial, cancer, etc). The use of Class I and Class II peptide-major histocompatibility complex (pMHC) reagents to identify such cells, however, is technically difficult and expensive, and it has been challenging to refine synthesis protocols for higher yield and more efficient assembly to accommodate large-scale applications. This achievement would enable high-throughput capture of corresponding T cell receptors (TCR), which may be further used in clinical applications such as adoptive cell transfer therapies. Overcoming this hurdle requires the development and integration of various molecular technologies and analytical methods.

Toward this end, the bulk of my thesis work, covered in Chapter 2, introduces these developments in the context of pMHCs, where the three subunits of each reagent are covalent linked together and expressed as a single protein. These single-chain trimer (SCT) technologies primarily consist of traditional DNA cloning and protein production techniques which have been streamlined for applications requiring output on the scale of 10^2-10^3 reagents. This chapter serves as the foundation for much of the methodology discussed throughout the rest of my thesis, and thus should serve as a reference point. The generated constructs are also functionally validated here, and potential future research directions are outlined.

In Chapter 3, I explore the use of this technology in the context of COVID-19 to enumerate antigen specificity of the $CD8^+$ T cell immune response. Class I SCTs were constructed to present peptides across several SARS-CoV-2 protein domains, using various HLA alleles to match haplotyped participant blood samples. These reagents were then used to capture SARS-CoV-2-specific T cells through flow and nanoparticle cytometry to demonstrate HLA-dependent, domain-dependent immune responses. Identified TCRs were cloned into T cells for confirmation of antigen specificity and functional cytotoxicity.

In Chapters 4 and 5, I explore potential pMHC applications in cancer antigen contexts, covering both tumor-associated and tumor-specific antigens. Through various collaborations across the west coast (UCLA, Parker Institute, Fred Hutchinson Cancer Research Center), I make use of the SCT platform to showcase new assays to discover and rank key tumor targets (Chapter 4). Finally, Chapter 5 is a reproduction

of our lab's published work concerning identification of antigen-specific $CD8^+$ T cells from melanoma cancer patients.

In summary, the adaptation of SCTs in a high-throughput format allows for the rapid enumeration of antigen-specific T-cell receptor sequences. As demonstrated in the contexts of COVID-19 and cancer, this SCT platform enables subsequent downstream applications, such as single-cell, antigen-specific immunophenotypic mapping/analysis and target discovery for personalized immunotherapies.

PUBLISHED CONTENT AND CONTRIBUTIONS

Chour, W., Ng A.H.C., Peng, S., Zaretsky J.M., et al. Sensitive Detection and Analysis of Neoantigen-Specific T Cell Populations from Tumors and Blood. Cell Reports 28, 2728–2738 (2019). doi:10.1016/j.celrep.2019.07.106

- W.C. was involved in experimental design, data collection, data analysis, and manuscript preparation.

Ng, A.H.C., Peng, S., Xu A.M., **Chour, W.**, et al. MATE-Seq: microfluidic antigen-TCR engagement sequencing. Lab on a Chip 19, 3011–3021 (2019). doi:10.1039/C9LC00538B

- W.C. was involved in data collection.

Wells, D. K., **Chour, W.**, et al. Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction. Cell (2020). doi:10.1016/j.cell.2020.09.015

- W.C. was involved in data collection.

Chour, W. et al. Shared Antigen-specific CD8+ T cell Responses Against the SARS-COV-2 Spike Protein in HLA-A*02:01 COVID-19 Participants. medRxiv 2020.05.04.20085779 (2020) doi:10.1101/2020.05.04.20085779.

- W.C. was involved in experimental design, data collection, data analysis, and manuscript preparation.

Su, Y., Yuan D., Chen D.G., **Chour, W.**, et al. Heterogeneous immunological recovery trajectories revealed in post-acute COVID-19. medRxiv 2021.03.19.21254004 (2021) doi:10.1101/2021.03.19.21254004.

- W.C. was involved in manuscript review.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

NOMENCLATURE

***Organisms & Viruses*.**

APC. antigen-presenting cell.

CMV. cytomegalovirus.

DC. dendritic cell.

EBV. Epstein-Barr virus.

SARS-CoV-2. Severe acute respiratory syndrome coronavirus 2.

TIL. tumor-infiltrating lymphocyte.

***Peptides, Proteins, & Protein Complexes*.**

CDR. complementarity-determining region.

CLIP. class II-associated invariant chain peptide.

ERAAP. endoplasmic reticulum aminopeptidase associated with antigen processing.

ERp57. endoplasmic reticulum resident protein 57, a.k.a. PDIA3 (protein disulfide-isomerase A3).

HLA. human leukocyte antigen.

IFN. interferon.

Ii. invariant chain.

MHC. major histocompatibility complex.

pMHC. peptide-MHC.

RAG. recombination-activating genes.

SCT. single-chain trimer.

TAA. tumor-associated antigen.

TAP. transporters associated with antigen processing.

TCR. T-cell receptor.

TdT. terminal deoxynucleotidyl transferase.

TSA. tumor-specific antigen.

UPS. ubiquitin-proteasome system.

β**2m.** beta-2 microglobulin.

***Other Symbols*.**

ACT. adoptive cell transfer therapy.

COVID-19. Coronavirus Disease 2019.

ER. endoplasmid reticulum.

Chapter 1

INTRODUCTION

1.1 The Era of Immunotherapy

The concept of harnessing the power of the immune system to treat disease, termed immunotherapy, dates back to ancient tales of the Plague of Athens in 430 BC [1]. Greeks who survived an initial exposure to the plague were noted to have never succumbed to a second course of infection, implying a memory-like capability of the body to resist repeated infection. Throughout the following centuries, similar stories of retained immunity following an initial infection were recorded. Notably, one of the earliest scientifically recorded cancer immunotherapy experiments was conducted in the late 1800s, and served as a starting point for the basis of numerous immunology questions that ensued in the following decades. Dr. William Coley initially observed spontaneous remissions of advanced cancers in patients who had been administered live *Streptococcus pyogenes* cultures [2, 3]. Due to issues concerning risks and unpredictability of exposure to live bacteria, he later developed a vaccine consisting of proteins from dead bacteria, which would not cause an infection but could still trigger a fever in patients due to the presence of foreign protein material. This approach to cancer treatment was hailed to be remarkable; 5-year survival rates were up to 80% for certain malignancies, and prolonged regression was observed even beyond this timeline. More recent reports, however, cast doubt on the true efficacy of Coley's toxins, remarking that the early trials were limited by multiple factors, including patient sample size, selection bias, inaccurate staging of tumors, etc [4].

Nevertheless, Coley's studies were the catalyst for a number of hypotheses that emerged regarding immunity to cancer. The induction of a febrile response against bacterial toxins served as a starting point for numerous downstream immune reactions. First, endotoxins could cause release of inflammatory signals from damaged tissue, inducing activation of dormant cell types which might in turn present these immunogens to and activate T cells for a cell-mediated killing response. Cancer-targeting T cells, which might remain in an anergic state in the absence of infection, could be reactivated upon sensing signals that arise during tissue destruction. Usually dormant T cells with the capacity to target self-like tissue (termed allo-specific

T cells) might also be activated as part of this response, further supporting the notion of increased tumor-killing during bacterial infection. It was speculated that the immune bias for targeting cancer cells could be explained by the presence of weaker vasculature surrounding tumor tissue, which would enable more immune trafficking versus the non-necrotic stable vasculature of healthy tissue [5, 6].

In the following century, several works were conducted to confirm these hypotheses. While many studies did not necessarily focus on a tumor response in the context of bacterial infection, they were able to affirm some of the hypothesized mechanisms by which the immune system responds to infection or cancer, and served to inform some of the modern approaches to immunotherapy. In 1973, dendritic cells (DCs) were found to be the instigators of the cell-mediated response against infection [7, 8]. Initially overlooked due to their extreme rarity in the blood, the unusual characteristics of these cells, with their extensively branched, mitochondria-rich appendages and increased motility, intrigued scientists to re-examine the role they played in the immune response. It was found that they were not only capable of inducing cell-mediated immunity via interaction with T cells, but also capable of triggering humoral (antibody-mediated) immunity via interaction with B cells [9, 10]. The high level of expression of major histocompatibility complex (MHC) proteins by DCs was later shown to be crucial in the antigen presentation pathways that lead to T cell activation [11]. These discoveries led to numerous advances in the following decades. The characteristics of antigens involved in T cell activation were later unveiled to show that they consist of short peptides (8-25 aa) derived from non-self proteins [12, 13]. Crystal structures of peptide-loaded MHC (pMHC) proteins enabled an understanding of the molecular requirements dictating how peptides might be presented by DCs for T cell activation [14]. The mechanism by which T cells would be able to recognize pMHCs was subsequently revealed with the discoveries of the T-cell receptors [15] and landmark crystal structures of the ternary TCR-peptide-MHC complex [16]. Following these discoveries, it became clear that an antigen-specific form of immunity played a crucial role in establishing "memory" against future infections; patients whose immune systems could be "trained" to respond to these antigen targets would therefore have the capacity for what is now referred to as adaptive immunity.

More recently in the 1990s, the concept of adaptive immunity was clinically demonstrated by two key works in which virus-specific T cells were transplanted from allogeneic bone marrow donors to immuno-compromised recipients. After transfusion,

the recipients displayed reconstitution of cellular immunity against cytomegalovirus (CMV) or Epstein-Barr virus (EBV) [17, 18]. Both works demonstrated that this form of therapy, termed adoptive cell transfer (ACT), was immensely potent and highly specific, as no toxic side effects were encountered. Furthermore, Walter *et al.* noted that the beneficial cytotoxic activity of $CD8^+$ T cells against infected cells could not persist if patients were deficient in $CD4^+$ helper T cell count. This observation implicated the requirement of a network of interactions among the immune cells, particularly amongst the T cell subtypes, to orchestrate potent antigen-specific responses in a step-wise, cell-dependent manner. Almost two decades later, in an approach parallel to the two works discussed, Rosenberg *et al.* remarkably demonstrated that T cells within the tumor tissue of a patient, termed autologous tumor-infiltrating lymphocytes (TILs), could be extracted, cultured *in vitro* unsuppressed by the tumor micro-environment from which they originated, and then re-introduced to the same patient to achieve complete regression in up to 40% of metastatic melanoma patients [19]. This event marked an exciting new breakthrough in cancer therapeutics, as conventional drug treatments at the time (e.g. *BRAF* inhibitors) could only achieve 6% remission. It implied that regardless of the source of the foreign agent, whether it was infectious (viral or bacterial) or tumor-derived, T cells were capable of recognizing non-self peptides and mounting a strong response against cells harboring those peptides.

The concept of personalized medicine as a new treatment modality was now made more tangible with the emergent use of patient-derived cancer antigens as the key therapeutic targets. If tumor tissue could be screened to identify non-self peptides emerging from common oncogenic DNA mutations, then one could, in theory, develop a cocktail consisting of those peptides to serve as a preventative vaccine [20]. Alternatively, the identification of these peptides could enable identification and extraction of cancer-specific T cells, enabling one to re-constitute a formidable anti-cancer immune response after undergoing ACT, as described in earlier works [21]. Simultaneously, attempts to modulate the degree of the immune response, for instance through the use of antibodies against programmed death receptors and ligands (PD-1 & PD-L1) or against cytotoxic T-lymphocyte-associated protein 4 (CTLA4), also showed efficacy [20, 22]. By turning on the activity of cytotoxic T cells and dampening the activity of anti-inflammatory, regulatory T (T_{reg}) cells, researchers showed that T cell recognition of antigen and proliferation could be dramatically improved. These realizations brought forth the onset of engineered immunotherapies.

Up to this point, it was clear that the adaptive immune response needed to be the primary research focus to better understand and hone immunotherapeutics against infectious diseases and cancer. The manner in which T cells recognized peptides and how they could be activated to expand was well-elucidated. Briefly, intracellular or extracellular peptides could be taken up by cells, processed, and then loaded onto major histocompatibility complex (MHC) proteins, which could then be transported to the cell membrane for presentation to T cells [13]. A stabilized complex forms among the pMHC, the T-cell receptor (TCR). and the T cell co-receptor (CD8 or CD4). If the TCR's binding pocket is compatible with the epitope presented by the pMHC, then the T cell may be activated, triggering a downstream cascade of reactions to induce T cell proliferation and strengthening of the immune network response (**Fig. 1.1**) [23]. In particular, it was shown that the functions of T cells could be correlated with the expression of either the CD8 or CD4 co-receptor, dictating whether they predominantly played a cytotoxic role or a helper role, respectively [24]. These two categories of T cells were found to recognize two distinct classes of MHC proteins, class I and II, respectively, which bind and present two distinct types of peptides, generally classified by amino acid length as either 9-11mers or 13-25mers, respectively. Through decades of rigorous investigation, researchers were able to outline these general rules regarding peptide, MHC, and TCR interactions and subsequent downstream responses. However, following these rules and reverse engineering the immune system in a high-throughput manner such that one could induce a TCR-driven cure against any given disease was still the holy grail for many labs. The rules of the game were established, but the breadth of the variables to survey and understand was massive. In the following sections, pertinent biological details for pMHC-TCR interactions will be outlined to provide the reader with the necessary background to appreciate the biological complexity of antigen processing, presentation, recognition, and immune response. Following these details, an overview of the current progress and challenges in immunotherapeutic developments will be provided.

1.2 A Peptide-Based Overview of Adaptive Immunity

1.2.1 Peptide Processing and Presentation

To acquire immunological memory against a foreign pathogen, an adaptive immune response is required. Typically, this involves the recruitment of two main cell types: 1) B cells for humoral immunity via antibody production, and 2) T cells for

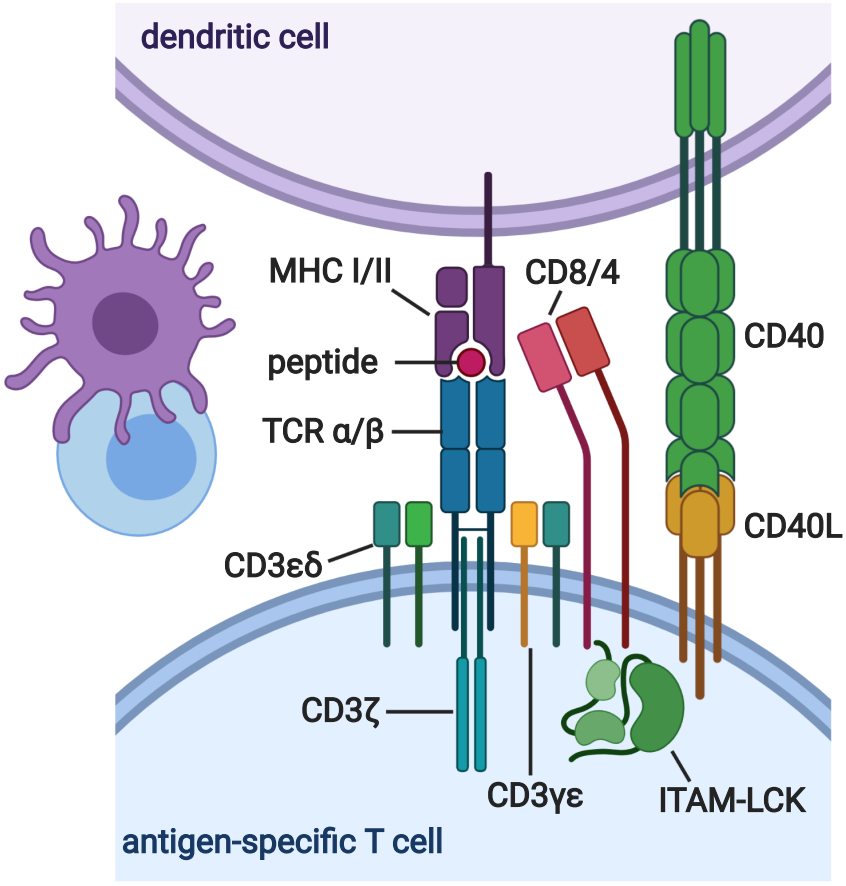


Figure 1.1: Antigen Presentation
A simplified model illustrating antigen presentation by an antigen-presenting cell (DC) to a T cell. The primary signal comes from the interaction of a peptide-loaded MHC from the APC with the TCR of the T cell. A second required signal comes from the CD8 or CD4 co-receptor interaction with the pMHC. Tertiary signals include B7/CD28, CD40/CD40L (depicted), among many others. Ultimately, these signals contribute to enforcement of T activation, which involves CD3-facilitated signal cascades with intracellular proteins (ITAM-LCK). *Created using BioRender.com (2020).*

cellular immunity via clonal activation and expansion of antigen-specific T cells. Cellular immunity requires the involvement of peptide, MHCs, and TCR, and will be the primary topic throughout this section. One of the earliest steps initiating cellular immunity involves dendritic cells (DCs), a form of antigen-presenting cell (APC), which circulate to the site of infection and can acquire non-self antigens. Upon maturation within lymphoid tissue, DCs are then capable of processing these antigens into short continuous peptides to be presented onto their MHC proteins (**Fig. 1.1**) [25].

The mechanisms behind antigen uptake, processing, and presentation are dependent upon the origins of the antigen [25, 26]. If the antigen arises from an intracellular agent (e.g. intracellular virus or mutated self-protein), then it may be directly processed in the cytosol, transported into the endoplasmic reticulum (ER), and loaded onto MHC class I proteins for recognition by $CD8^+$ T cells. This pathway, called direct presentation, results in priming of $CD8^+$ T cells whose TCRs are specific to the presented epitope, leading to clonal expansion and induction of cytotoxic activities. Direct presentation can be conducted by all nucleated somatic cells. If the antigen instead arises from an extracellular pathway (e.g. bacteria or protozoans resistant to macrophage digestion/uptake), DCs must first ingest and kill the pathogens in the low pH environment of endocytic vesicles. Subsequently, the remnant antigens are processed into peptides for presentation onto MHC class II proteins for recognition by $CD4^+$ T cells. This pathway, called indirect presentation, leads to priming of $CD4^+$ T cells to induce primarily helper functionalities (although some cytotoxic functionalities can exist), which, for instance, can lead to the recruitment of $CD8^+$ T cells and B cells. MHC class II peptide processing and presentation occurs only in APCs, which includes DCs, macrophages, and B cells.

These two classical presentation pathways are treated as the conventional routes, but there does exist some degree of overlap. Intracellular pathogens which evade traditional direct presentation pathways in their hosts, for instance, will avoid activation of $CD8^+$ T cells. To counteract this, DCs may take up antigens from these sources in an exogenous manner into phagolysosomes, and degrade the proteins via proteasomal pathways into the cytosol, whereby the peptides can then be loaded onto MHC class I. Thus, this phenomenon, called cross-presentation, makes use of the MHC class II conventional antigen uptake mechanisms for loading onto MHC class I molecules. In another non-conventional technique, endogenous proteins may circumvent traditional MHC class I presentation to be digested by lysosomes within endocytic vesicles. This process, called autophagy, enables endogenous proteins to be presented onto MHC class II molecules.

Antigen processing occurs after uptake, and may dictate the nature by which MHC class I peptide sequences may be immunogenic. Briefly, peptide generation is initiated by ubiquitinylation of proteins (**Fig.1.2**), in which they are tagged with ubiquitin for degradation by the ubiquitin-proteasome system (UPS). The proteasome consists of three proteolytic subunits, which recognize ubiquitinylated sites and degrade protein sequences into peptide fragments, typically without regard for

sequence specificity. In the presence of interferons (IFNs), which are signaling proteins released upon viral infection, three alternative proteasomal subunits may be expressed, giving rise to the immunoproteasome structure instead [27]. This complex digests proteins with an increased propensity for post-hydrophobic residue cleavage, and decreased propensity for post-acidic residue cleavage. The resultant peptides from immunoproteasomal processing have been found to have MHC class I anchor residues in higher abundance, implying that IFN skews the peptide repertoire in favor toward presentation. Another role played by IFNγ is that it can induce expression of the PA28 proteasome-activator complex, which accelerates proteasomal processing, and thus increases peptide production rate. After these peptides are produced in the cytosol, they enter the ER through hetero-dimerized ATP-binding cassette proteins, called transporters associated with antigen processing-1 and -2 (TAP1 and TAP2) [28]. Prior to peptide entry into the ER, MHC class I's alpha chain subunit, called the human leukocyte antigen (HLA) heavy chain, is stabilized by calnexin until its beta chain subunit (beta-2 microglobulin, or β2m) is present. After β2m binds to the HLA subunit and releases calnexin, two proteins (calreticulin, ERp57) bind onto the MHC, acting as a scaffold for stabilization. Tapasin then brings the dimerized TAP complex into proximity with the MHC, allowing for incoming peptides to be loaded into the MHC's peptide-binding groove. At this point, peptide editing may still occur, as tapasin ensures that the groove remains open until a peptide of sufficient affinity is bound. Furthermore, the peptides which enter the ER may be 8-16 aa in length, but the binding groove traditionally accommodates peptides up to 11 amino acids, thus the peptides must first be trimmed. For this purpose, a protein called the endoplasmic reticulum aminopeptidase associated with antigen processing (ERAAP) facilitates cleavage of amino acids at the carboxy terminus, reducing peptide length to accommodate the binding groove [29]. Finally, the stabilized pMHC is delivered from the ER membrane to the cellular membrane for presentation to TCRs.

As for MHC class II peptide processing, the endocytosed pathogens are first digested into peptide fragments, typically 13 to 25 amino acids in length, by acid proteases (cathepsins) within lysosomes (**Fig. 1.3**) [30]. In order for these peptides to bind onto MHC class II molecules, several steps must occur in the ER vesicles, where the MHCs are membrane-bound. First, MHCs synthesized in the ER must not bind to any self-peptides. To maintain this state, a membrane protein called the invariant chain (Ii) forms a trimerized complex, with each unit containing a domain called the class II-associated invariant chain peptide (CLIP), to occupy the peptide-binding

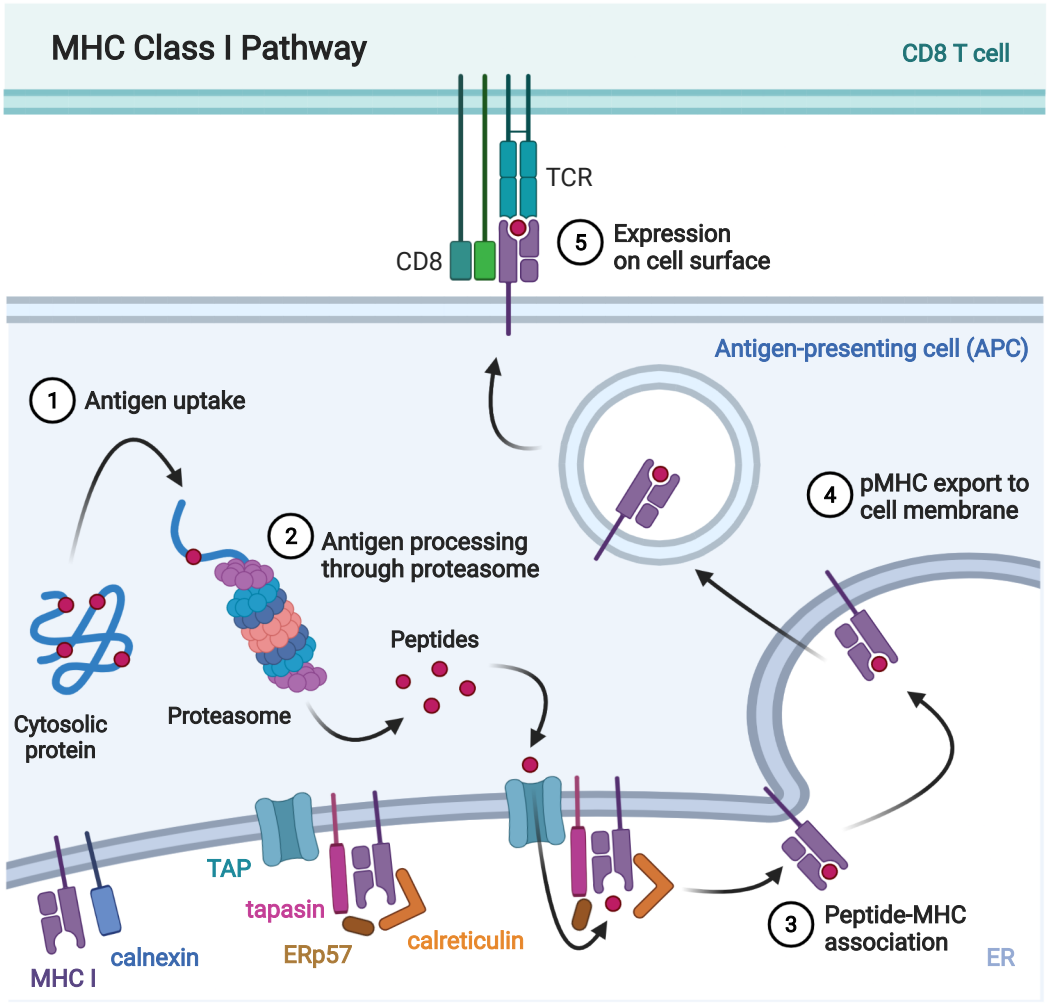


Figure 1.2: MHC Class I Peptide Processing and Presentation Pathway
In order to activate $CD8^+$ T cells, endogenous peptides (1) are ubiquitinylated for proteasomal processing (2) to produce short peptide fragments. A complex of proteins in the ER are simultaneously prepared. Calnexin serves to stabilize the class I heavy chain until β2m is able to bind, at which point the heterodimer is then stabilized by the complex of calreticulin and ERp57. Tapasin acts to bring the membrane porin TAP in proximity to the MHC. After peptides have been shuttled through TAP into the ER, tapasin facilitates peptide exchange with the MHC's binding groove until the pMHC is stabilized (3). The pMHC complex is then transported to the cell surface for presentation to TCRs of $CD8^+$ T cells. *Partially adapted from "MHC Class I and II Pathways", by BioRender.com (2020). Retrieved from https://app.biorender.com/biorender-templates*

groove of one MHC class II molecule [31]. This entire complex, consisting of three MHCs bound to trimerized Ii protein, is then released from the ER. Ii, now able to perform its second function, directs the entire complex toward the low-pH

environment of the endosomes, exposing the MHC class II heterodimers to the endosome's digested peptides. A series of endosomal cathepsins cleave Ii from the MHC proteins, leaving behind only the CLIP fragment. Next, HLA-DM acts as a chaperone protein to the CLIP-bound MHC. HLA-DM is similar in structure to other HLA class II complexes, consisting of an α and a β chain, but does not contain the prerequisite open groove for peptide binding. Instead, HLA-DM binds to other MHC class II molecules, stabilizing them during groove opening to allow for CLIP release. The binding groove is maintained in this open state until endosomal peptides can favorably bind. HLA-DM activity is inhibited by HLA-DO, which competes against a functional MHC class II molecule for binding to HLA-DM and is released from HLA-DM only in the presence of an acidic environment. Thus, HLA-DM serves to facilitate peptide editing by enabling affinity-based peptide selection within the acidic lysosomes before pMHC class II molecules are delivered to the cellular membrane [32].

So far we have discussed how peptides are processed and presented on both MHC class I and class II molecules. For class I, various immunoproteasomes will cleave proteins, subsequent length editing by ERAAP determines the ultimate MHC-compatible peptide repertoire, and tapasin acts as a sort of catalyst to reduce the energy barrier for peptide exchange. For class II, cathepsins engage in proteolytic activity to digest proteins, while HLA-DM acts as the final mediator in stabilizing peptide to binding groove. Beyond these processes, there is at least one more significant variable, the highly polymorphic HLA alleles comprising the MHC for the loaded peptide, that determine whether a peptide is a favorable binder.

1.2.2 The Diversity of Major Histocompatibility Complexes

A brief overview of HLA allele distributions is necessary to acclimate the reader to subsequent thesis chapters. MHC class I α chains are encoded by alleles from three HLA loci (A, B, C), and MHC class II α and β chains are encoded by alleles from three HLA loci (DR, DQ, DP). Both MHC protein classes are considered transmembrane, heterodimeric proteins, but MHC class I's β chain (β2m) is uniform throughout the human species, and while β2m appears to play a major role in stabilizing the MHC upon loading of a peptide, it has no direct interaction with the peptide itself. This is because the α heavy chain, typically referred to by its HLA class I allele (e.g. A*02:01), is the sole determinant of the amino acid structure of the peptide binding groove. The α heavy chain of MHC class I encodes not only the beta sheet

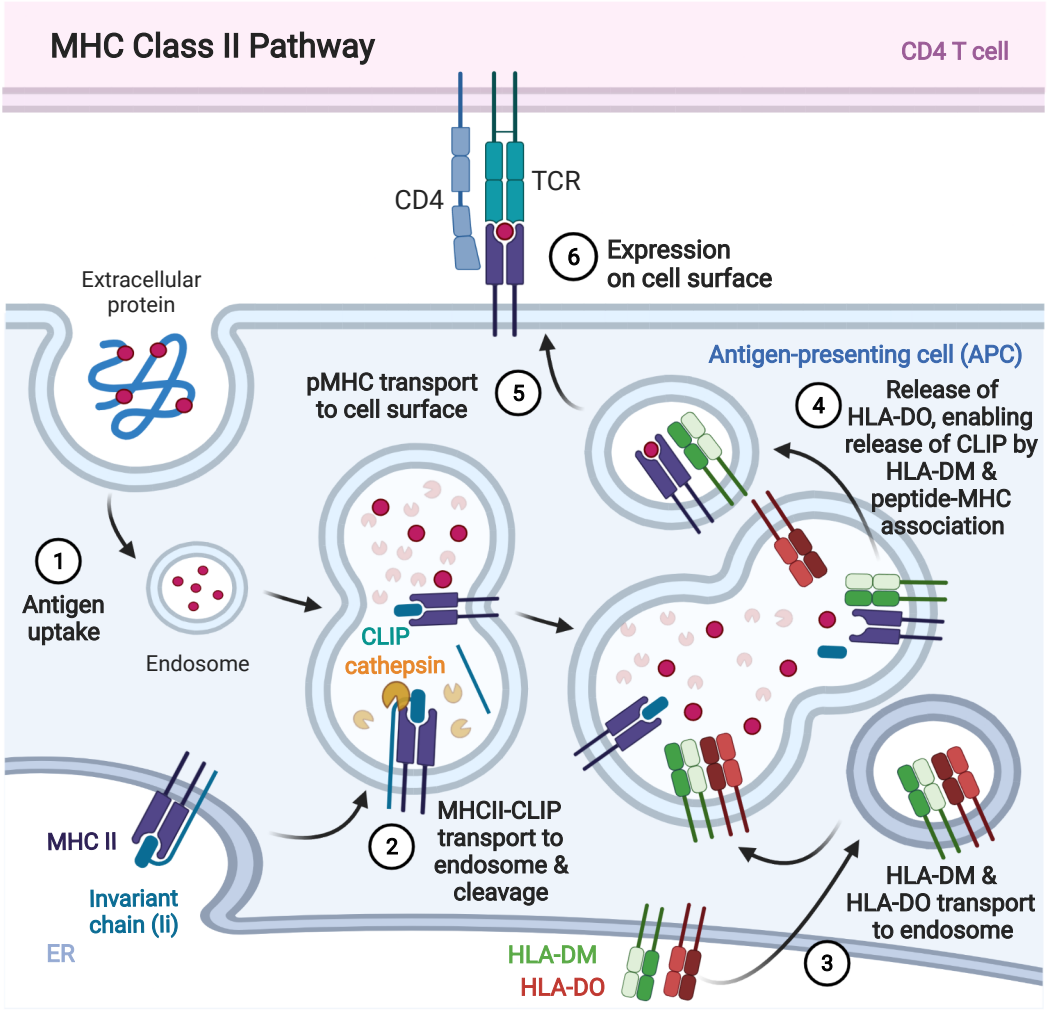


Figure 1.3: MHC Class II Peptide Processing and Presentation Pathway Activation of $CD4^+$ T cells begins with endocytosis of exogenous antigens (1). The acidic environment of lysosomes enables proteases (cathepsins) to cleave antigens into short peptide fragments. MHC class II proteins are also transported into this environment (2), where cathepsins cleave off Ii and leave behind CLIP in the MHC's binding groove. Peptide exchange with CLIP cannot occur until the HLA-DM/DO complex is able to join the endosome from the ER (3). Once HLA-DO is destabilized by the acidic environment, HLA-DM is able to serve its catalytic function to enable CLIP release and peptide exchange. the pMHC is then transported to the cell surface for presentation to TCRs of $CD4^+$ T cells. *Partially adapted from "MHC Class I and II Pathways", by BioRender.com (2020). Retrieved from https://app.biorender.com/biorender-templates*

of the groove (forming the groove's base), but also the two alpha helices (which encapsulate the peptide at the sides of the groove) (**Fig. 1.4 A-C**) [23].

MHC class II heterodimers, on the other hand, consist of unique α and β chains

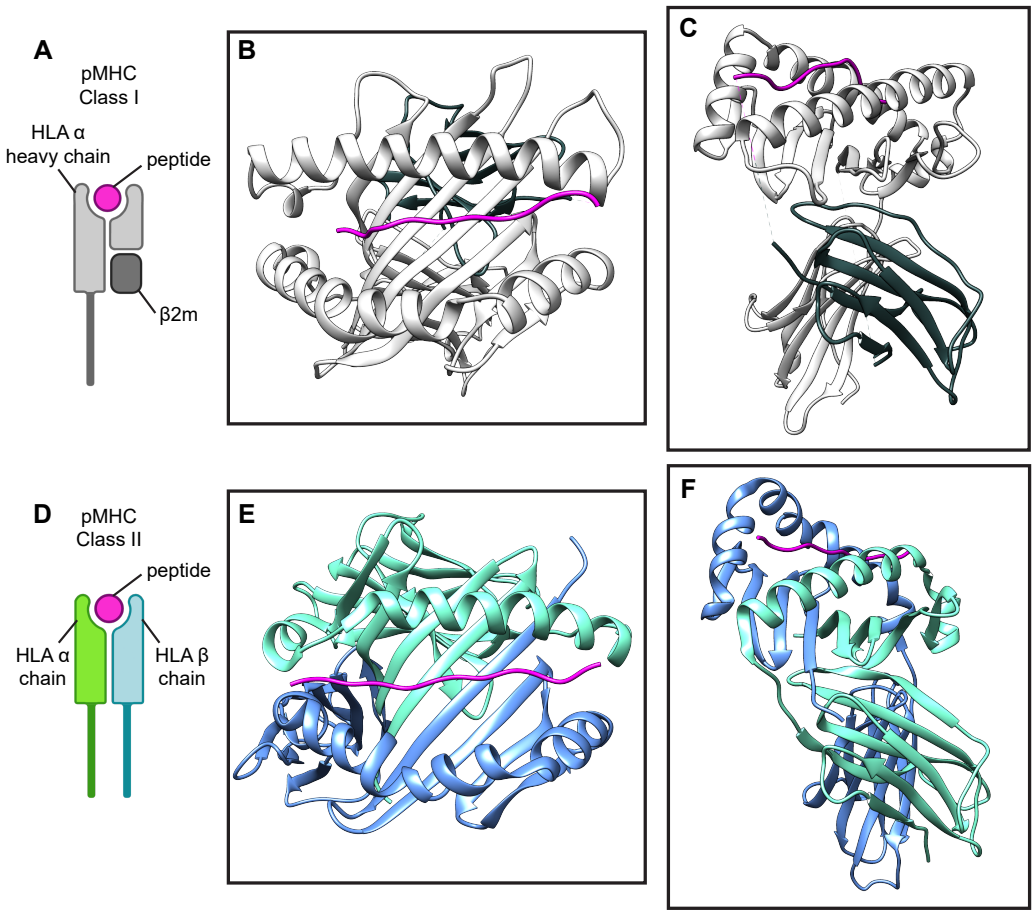


Figure 1.4: 3D Structures of pMHC Proteins
Peptide-loaded MHC class I cartoon is depicted (A). Axial views of pMHC-I, from the top (B) and the side (C). Peptide (pink), α heavy chain (light gray), and β2m (dark gray). Peptide-loaded MHC class II cartoon is depicted (D). Axial views of pMHC-II, from the top (E) and the side (F). Peptide (pink), α chain (cyan), and β chain (blue).

determined by separate class II HLA alleles per locus [23]. Because every individual may have up to two unique HLA alleles per chain type per class II locus, up to four uniquely paired α/β heterodimers may be expressed by each locus. One exception is the α chain of the HLA-DR locus, which is essentially invariant in the human species, and thus for HLA-DR, there are only two potential α/β combinations, which is analogous to the situation with β2m for MHC class I [the HLA-DR β chain alleles, termed HLA-DRB, actually consist of multiple paralogs (genes arising from duplication events), of which the most prevalent one is HLA-DRB1; HLA-DRB3/4/5 also exist, but are more rare, and for simplicity we ignore them here]. Structurally, both α and β chains of MHC class II proteins are involved in forming the β sheet and

α helices of the binding groove (**Fig. 1.4 D-F**). Taken together, the binding groove of MHC class II molecules is typically more diverse than that of MHC class I, given the various combinations by which α and β chains can come together to influence the molecular structure of the peptide binding groove.

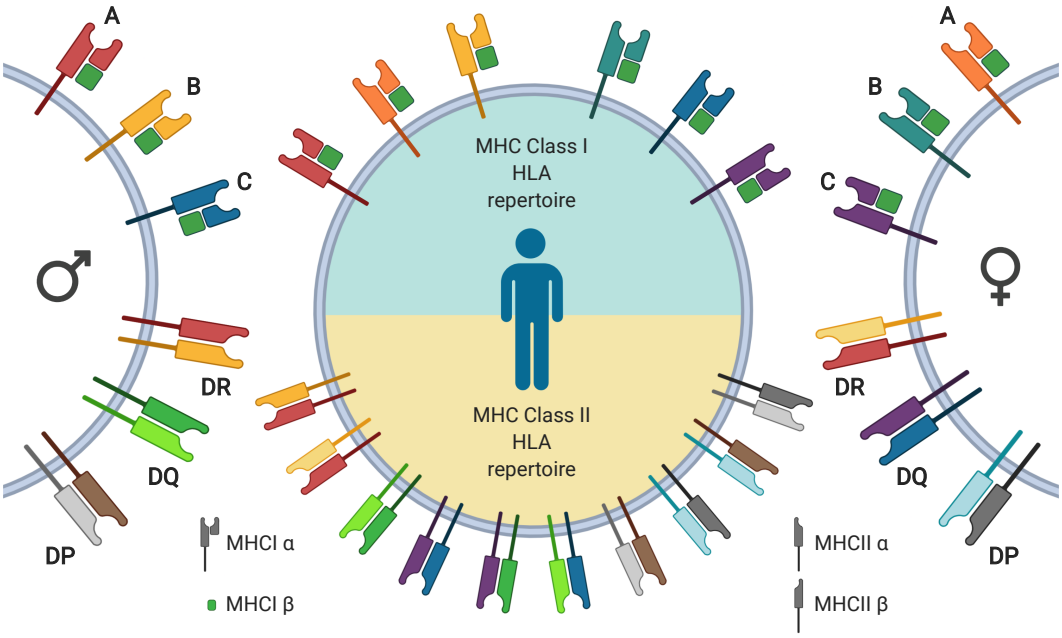


Figure 1.5: HLA Inheritance
An individual's HLA repertoire consists of inherited HLA alleles from each parent. Up to six unique HLA class I alleles may be inherited, and up to 12 unique HLA class II alleles may be inherited (however, due to HLA-DR α chain invariance, maximum diversity is usually 10, as depicted). *Created using BioRender.com (2020).*

Any given individual will inherit one copy of each HLA allele from each parent (one copy here means, for class I, one unique HLA α chain allele per locus; for class II, one unique HLA α chain AND one unique HLA β chain allele per locus). Therefore, an individual will express up to six unique MHC class I proteins and up to ten unique MHC class II proteins (we account here for the invariant HLA-DR α chain, as discussed above) (**Fig. 1.5**). In general, the closed nature of the N- and C-terminal regions of the peptide-binding groove for MHC class I result in shorter bound peptides (8-11 aa), while the open nature of these regions for MHC class II allow for longer peptides (13-25 aa). The presence of unique anchor pocket residues within binding grooves of HLA alleles from both MHC classes means that compatible peptides typically will have a canonical motif at these positions to facilitate favorable molecular interactions with the groove.

In summary, these six HLA loci are among the most polymorphic genes in the human species, resulting in tens or even hundreds of allelic variants for certain loci. Despite the combinatorial diversity that arises from these alleles, global HLA distribution studies indicate that approximately 8-12 alleles per loci will generally account for up to 90% of the human population. The mutational differences across HLAs are predominantly encoded within the peptide-binding regions of each MHC, which in turn determines the immense repertoire of compatible binding peptides per individual MHC variant. This suggests that there is significant evolutionary bias in these alleles, favoring prevalence of certain HLAs that are capable of covering most pathogenic peptides to confer immune protection [23].

1.2.3 The Diversity of T-Cell Receptors

Having discussed some of the details regarding pMHC assembly and function, we now turn our attention toward their cognate TCRs. The initial discovery of the TCR occurred in 1982 [15], followed by its sequencing from multiple labs two years later [33–35]. Somatic recombination of gene segments within the TCR was also discovered, which explains how the vast diversity of TCR sequences arises within any individual despite the relatively limited size of the human genome [36, 37]. Analogous to MHC class II proteins, TCRs are heterodimers consisting of an α and β chain subunit, where both chains play a role in interacting with the epitope presented by a bound pMHC. During lymphocytic development, the TCRα and TCRβ chain genes undergo numerous processing steps, whereby the gene segments that make up each chain recombine and undergo randomized modifications, to generate mRNA transcripts encoding the finalized TCR α/β paired sequence which define individual T cell clonotypes from each other.

The genetic information encoding each chain consists of two or three gene segments per chain. Each segment may contain numerous possible genes, but only one gene per segment will be selected for exon splicing and joined together as an mRNA strand to ultimately encode the TCR. This selection of gene segments is referred to as gene rearrangement, or somatic DNA recombination. The gene segments encoding the TCR α and β chains are categorically bifurcated into either the variable (V) regions or the constant (C) region. The C region plays no role in interacting with the pMHC surface and encodes the parts of the TCR that are distal to the pMHC and proximal to, and embedded within, the T cell's surface membrane. The C region is therefore much less diverse than the V region, with only one and two allelic variants for α and β

chains, respectively, resulting in C_α, $C_{\beta 1}$, and $C_{\beta 2}$ genes. The V region, meanwhile, encodes the proximal part of the TCR to the pMHC, and confers to each TCR its idiosyncratic peptide/HLA specificity. The V region of the α chain is split into two gene segments (V_α, J_α), and of the β chain is split into three gene segments (V_β, D_β, J_β). Therefore, the formation of a complete TCR chain consists of a V_α-J_α-C_α chain paired with a V_β-D_β-J_β-$C_{\beta 1,2}$ chain. While the V region does indeed encode for the regions which interact with the pMHC, there only exist a few select regions of the protein that actually come into contact with the pMHC (**Fig. 1.6**). These regions, called complementarity determining regions (CDRs), consist of three loops (CDR1, CDR2, CDR3) per chain. CDR1 and CDR2 loops are encoded by germline V gene segments and are therefore less variable. Furthermore, these two CDRs are not believed to play a significant role in terms of determining which peptide/HLA pairing will be a compatible binder. Numerous crystallography structures obtained for pMHC-TCR interactions show that CDR1 and CDR2 rest along the periphery of the peptide epitope, primarily in contact with the alpha helices of the binding groove, whose CDR contact sites are overall less variable. This leaves CDR3 as the primary determinant of the antigen binding site for the TCR chains. For virtually every known pMHC-TCR structure, regardless of the axial tilt of the TCR with respect to pMHC orientation, the CDR3 loops of both TCR chains will come into direct contact with the peptide, usually at the central location of the peptide [23]. Therefore, a TCR's peptide specificity is primarily defined by the CDR3 variations that arise from TCR α and β chains, and these CDR3 sequences act as the unique identifier to understand clonotype repertoires and relation to disease [38–40].

With the importance of the CDR3 now established, we can now explore the question of how CDR3 diversity arises to generate individualized repertoire of TCR clonotypes. An examination of the CDR locations on the V genes (**Fig. 1.7**) reveals that CDR1 and CDR2 are located within the V gene itself, while CDR3 is at the 3'-terminus. As alluded to earlier, gene rearrangement governs the selection of one gene from each gene segment, enabling a V gene to combine with a J gene for TCRα, and a selected V gene to combine with a D gene and a J gene for TCRβ. Therefore, CDR3α's full amino acid sequence comprises the overlapping regions between Vα and Jα, while CDR3β's sequence encompasses the V_β-D_β-J_β overlapping regions, or more specifically, the entire junction from carboxy-terminal end of V_β to the amino-terminal end of J_β (D_β is an extremely short gene segment entirely embedded within CDR3β). Based on the total number of unique genes per segment (**Table 1.1**), this combinatorial diversity is expected to give rise to approximately 6 million potential

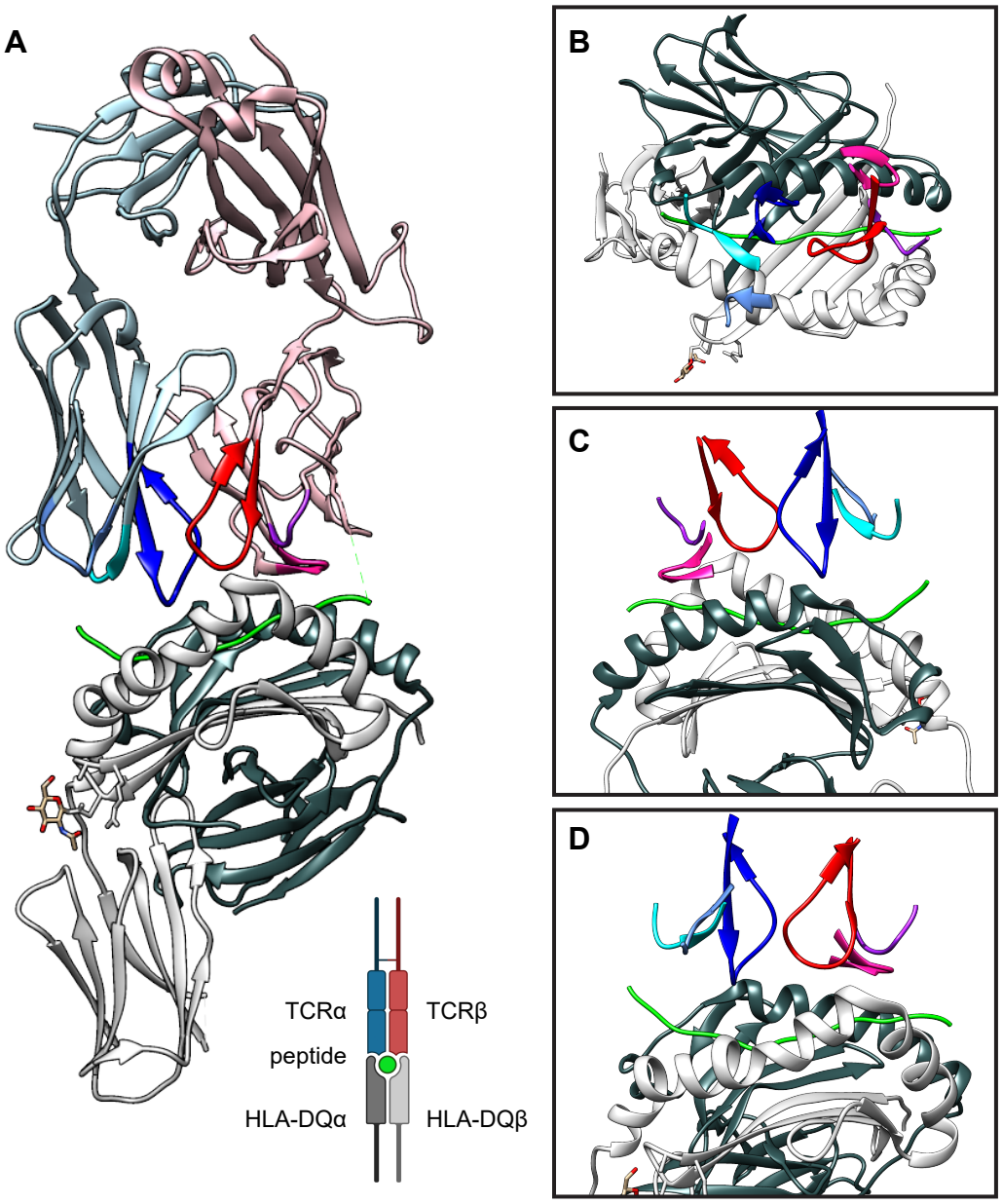


Figure 1.6: pMHC-TCR Structure
Complete structural view of the pMHC-TCR complex (A). Views of the complex with TCR FR components removed to highlight CDR interactions: top view over pMHC (B), side view (proximal to HLA-DQA) (C), and side view (proximal to HLA-DQB) (D). TCRα chain (light blue), TCRβ chain (light red). CDR1α (cyan), CDR2α (gray blue), CDR3α (dark blue), CDR1β (pink), CDR2β (purple), CDR3β (red), HLA-DQ α chain (dark gray), HLA-DQ β chain (light gray), and peptide (green).

unique paired combinations of TCR α/β chains [36]. In addition to combinatorial diversity, the phenomenon of junctional diversity comes into play to introduce an-

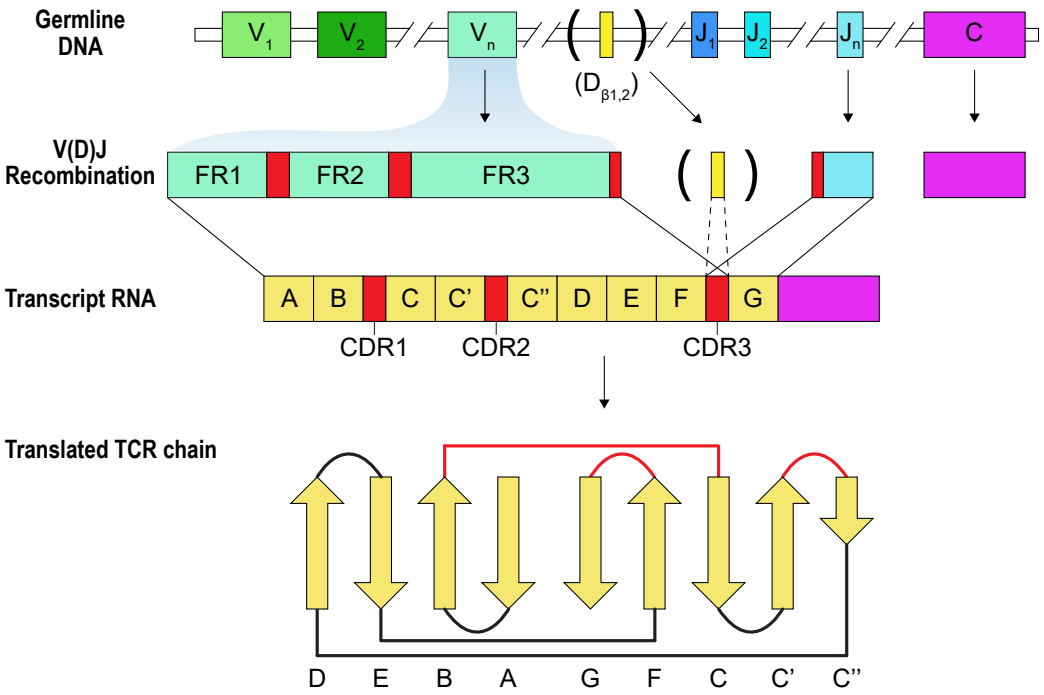


Figure 1.7: V(D)J Gene Recombination
TCR α and β chains consist of V, J, and C gene segments that are initially selected in the germline during lymphocytic development (first row). The selected genes from each segment are combined together to encode the entire TCR chain. For TCRβ, an additional D gene segment is present that is merged between V and J (second row). Most of the protein in the V(D)J product encodes the framework region (FR), which serves as a scaffold for the highly variable complementarity-determining regions (CDRs). V, D, and J gene segments all contribute to CDR3 (third row). A linearized map of the nine beta sheets comprising a TCR chain is laid out to depict the spatial 2-D relationships of the CDRs (fourth row).

other 200 billion potential combinations. Briefly, during each recombination event in which the V, D, and J genes are brought together, the recombination-activating genes (RAG) protein complex forms around the junction points, enabling nucleotide addition/deletion with the assistance of DNA hairpin formation from the template genes. In addition, the enzyme terminal deoxynucleotidyl transferase (TdT) drives non-template-informed nucleotide addition to randomly insert bases at the ends of the strands. When the single-stranded genes join together at the junctions, exonucleases cleave incompatible nucleotides, and ligation generates the double-stranded product. These combined mechanisms generate random edits to the CDR3-encoding junction points. Most of the chain products will be nonproductive due to frameshift insertions/deletions, and cells harboring these nonproductive chains will be elim-

inated. Altogether, combinatorial diversity and junctional diversity combined can generate a maximum diversity of 10^{18} possible TCR clonotypes.

TCR chain	α	β
variable (V) segments	70	52
diversity (D) segments	0	2
joining (J) segments	61	13
combinatorial diversity	5.8×10^6	
junctional diversity	2×10^{11}	
total diversity	10^{18}	

Table 1.1: TCR Diversity
The number of unique variants per TCR gene segment is depicted in this table. Recombinatorial diversity arises from random combinations of these variants to form the V(D)J arrangement of a mature TCR chain. Junctional diversity arises from further randomized mutational edits to the V(D)J junctions.

The TCR editing process explained above represents the primary diversification that occurs in any T cell during development in the thymus. Upon leaving the thymus, each T cell, now defined by its TCR clonotype, retains a fixed antigen specificity. Just before leaving the thymus, central tolerance eliminates T cells whose TCRs are reactive to self-antigens, although this process is not entirely comprehensive, and can result in the presence of self-reactive T cells which can lead to autoimmune diseases. Peripheral tolerance serves as a sort of backup method to central tolerance, whereby self-reactive T cells which have "escaped" deletion in the thymus may have their functions curtailed through a number of mechanisms (deletion, conversion to T_{reg}, suppression, anergy, etc.). Note that further TCR editing does not appear to play a role in this process, and it has been the conventional belief that TCR editing does not occur outside of the thymus and that a T cell's antigen specificity is fixed after thymic education. Recent developments, however, have introduced the concept of TCR revision, in which mature T cells in the periphery may demonstrate altered antigen specificity. This process was first identified by Vaitaitis *et al.*, when they found that the RAG complex could be induced to express and activate TCR editing [41]. Later studies showed that this phenomenon is more prevalent than originally believed, and is dependent on the interactions of TCRs with self-antigens [42, 43]. Further, the presence of dual receptor T cells (T cells containing two unique α/β pairs) was known for quite some time, and has been thought to also play a role during revision [44]. Additional pursuit of these concepts is beyond the scope of this introduction, and their mention here is intended to provide the reader with an appreciation of the biological origins of TCR diversity. In summary, attempts to

further our understanding of TCR diversity have generated a ever-growing web of questions.

1.3 Recent Progress and Developments Toward Personalized Medicine

Following the wake of major discoveries regarding structural properties of pMHCs and TCRs, many biological questions surrounding the interplay of peptide and T cell ensued. Which peptides of a foreign agent are most immunogenic? Are they biased for particular regions or proteins of a virus or a bacteria? What about the case of tumor tissue: which mutated regions are most likely to be presented to the immune system? Are there any immunodominant peptides which are abundantly presented from a particular foreign agent or shared across cancer patients? If so, are they also responsible for triggering the primary response that leads to recovery? Such questions have been explored on an individual or small-library basis for peptides arising from bacterial, viral, and cancer origin. However, major bottlenecks in these studies limited the scope of these works. As discussed above, the breadth of peptide diversity is immense, and even if scientists were to just focus on a single given bacterial protein (average length 320 aa) to explore the diversity of MHC class I peptides, for instance, they would generate on average close to 1,000 potential 9-11mers as candidates. Fortunately, at this time, peptide prediction software development is well underway. Researchers are now able to utilize neural networks and AI to predict the best peptide binders for popular HLA alleles based upon canonical sequence motifs reported for each allele's binding groove. These works, for example MHCflurry and NetMHC4.0 [45–47], operate on an allele-specific basis and may incorporate experimental stimulation assay data and mass spectroscopy results. These approaches were further generalized with NetMHCpan, an algorithm which extrapolates peptide/HLA interactions from a trained set to enable peptide binding affinity prediction against any given HLA class I structure [48]. Prediction software has also been developed for HLA class II proteins, but is less mature and remains a challenging problem due to the open-ended nature of the binding groove, allowing for presumably a broader range of affinities between peptide and HLA to result in a stable, viable compound [49]. Furthermore, the NetChop protein degradation prediction algorithm was also developed, which enabled researchers to further filter down peptide lists based on proteasomal activity [50].

Taken together, developments in peptide processing and binding prediction algorithms have been useful in assisting with *in silico* identification of immunogenic

epitopes. The overarching goal of these efforts have generally been to provide a pre-experimental filter of genomically processed antigen data to narrow down the list of peptide candidates for subsequent downstream experimental steps (e.g. vaccination testing or loading onto MHCs to identify TCRs). However, on the experimental side, verification of these predicted epitopes still remains an outstanding challenge. Given the vast space of TCR sequences, the identification of antigen-specific TCRs against any given peptide/HLA combination is akin to finding a needle in a haystack. Synthesizing pMHCs which mimic their biological membrane-bound counterparts in terms of TCR binding function would enable one to metaphorically use pMHCs as a magnet to comb through the haystack for such needles. One of the largest impediments is the difficulty of refolding MHC proteins. Traditionally, the MHC subunit chains are expressed in bacteria, which are then lysed to extract the unfolded components. These chains are then combined together in a purely chemical solution in the presence of a strong-binding peptide to enable refolding and stabilization of the pMHC product [51, 52]. This strategy is fraught with issues, however. Yield is typically low, as these subunits are bacterially expressed in their unfolded format and must be folded together without the assistance of their native, intracellular eukaryotic mechanisms. In addition, the folding efficiency has been found to be HLA-dependent, whereby different alleles have different proclivities for successfully binding. A strategy developed by Rodenko *et al.* in 2006 to facilitate library production of pMHCs incorporates the additional step of initially folding the pMHC complex using a UV-labile peptide, to allow for initial large-scale production of a "template" UV-labile pMHC [53, 54]. Subsequent exposure of the folded complex to UV light in the presence of a target peptide cleaves the UV-labile peptide and allows for exchange to generate a functional pMHC, which can then be mixed with streptavidin to form pMHC-tetramers, whose increased TCR avidity improve cognate TCR binding and capture. The primary issue with this strategy is that there is no guarantee of complete UV exchange; side products may occur, and the presence of these misfolded products may induce non-specific binding downstream. In addition, there is the additional difficulty of having to identify a UV-labile peptide per HLA allele that is compatible with the binding groove. A method to circumvent some of the challenges of this technique was recently introduced in which tapasin's ability to open the binding groove for peptide exchange was exploited to enable rapid, clean peptide insertion into refolded MHC proteins [55]. Unfortunately, this technique, similar to UV exchange, requires *in vitro* synthesis of every target peptide, which is either extremely laborious or expensive when operating on a library scale. Fur-

thermore, the reliance on synthetic peptides precludes the capacity of using pMHCs to present post-translationally modified peptides. Modifications such as phosphorylation and glycosylation at key amino acid residues can have a drastic impact on TCR recognition, but fall outside the scope of feasible studies when working with synthetic peptides.

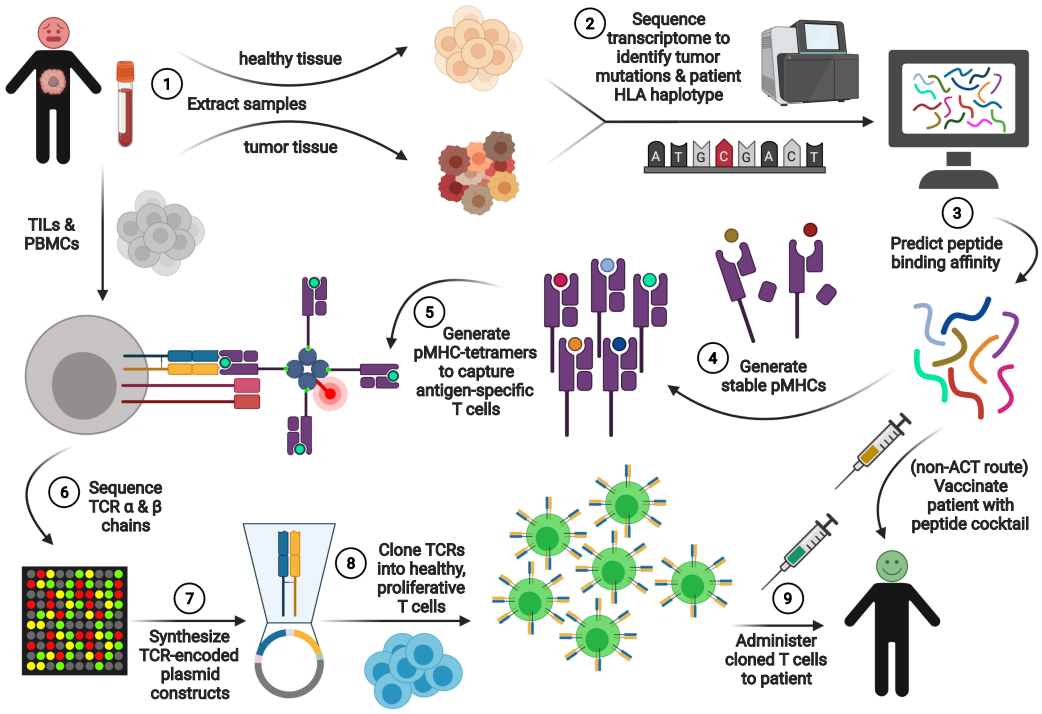


Figure 1.8: Adoptive Transfer Cell Therapy
ACT as a viable route of immunotherapy begins with the extraction of tissue (1) to identify neoantigens (2). Peptide-MHC binding affinity predictions are performed (3) to identify the best peptide candidates for pMHC generation (4). Stable pMHCs are then tetramerized and used to capture antigen-specific T cells (5), whose TCRs are subsequently sequenced (6), synthesized in plasmid format (7), cloned into healthy T cells (8), and delivered to patients (9). As therapeutic alternative to ACT, cancer vaccines may possibly be used, which circumvents the need to extract cells from the patient, requiring a similar pathway until peptide prediction, where peptides are formulated into a vaccine cocktail. *Created using BioRender.com (2020).*

There has been a significant amount of development in the field to further our understanding of pMHC-TCR pairing in spite of the bottlenecks described above. The preparation of pMHC formats in a library manner has been achieved by multiple groups, to assess not only the TCR sequences of antigen-specific T cells, but also their phenotypic and transcriptomic data. Emergent methods to creatively tackle

the questions of pMHC and TCR specificity have been devised, and have resulted in novel computational approaches. For example, Birnbaum *et al.* developed a yeast display assay to allow for a TCR of interest to be scanned against a pMHC library to identify reactivity [56]. Mark Davis' lab at Stanford University later developed technologies that could answer the inverse question. Rather than starting from one TCR to identify all potential binding pMHCs, could one instead start from one pMHC and capture all the compatible CDR3 sequences from TCRs? Their work led to a computational algorithm capable of learning from captured antigen-specific TCR sequences to predict the antigen specificity of novel TCRs [57, 58]. Around the same time, projects from David Baltimore's lab at Caltech generated massively multiplexed technologies using the single-chain trimer (SCT) format [59, 60] to encode cell surface-bound pMHCs [61, 62]. By relying on cell-cell interactions to facilitate pMHC-TCR binding, activation, and fluorescence readout, they could discover antigen specificities of cloned TCRs. Similar to the methods of Birnbaum *et al.*, this assay enables antigen discovery starting from one TCR, but is more facile in that it does not require synthesis of a soluble TCR construct. While the scope of antigens is high (up to 10^6 epitopes/assay), the authors noted limitations in capture sensitivity and TCR throughput.

1.4 Thesis Overview

The general course of my thesis work aims to address many of the bottlenecking problems that have hampered existing pMHC-TCR pairing technologies. This work encompasses assays intended for the discovery of multiple TCRs from multiple peptides, a.k.a. a "many-to-many" approach in contrast to the "one-to-many" assays described previously. Thus, Chapter 2 serves as a reference chapter, describing all of the molecular engineering technologies I developed to improve the robustness of pMHC class I reagents and the facility by which they may be synthesized. These advancements primarily hinge on the single-chain trimer approach to pMHC production introduced decades ago by others. The establishment of experimental protocols and automated scripts which adapt these techniques to enable high-throughput SCT template plasmid production, peptide-substituted SCT library production, DNA sequence alignment, protein expression, yield characterization, and thermal stability measurements are detailed within this chapter, and further discussed in the Appendix. Modulation of other components aside from the peptide, such as the identity of the covalent tethers linking peptide to MHC, peptide

groove-stabilizing mutations, and CD8-independent mutations, are also discussed within the context of Class I SCTs. In addition, a subsection is also devoted to the implementation of these techniques toward Class II SCT development, wherein I highlight functional validation of my initial conceptual work. A third component of this Chapter discusses similar endeavors to enable high-throughput TCR cloning, an essential component to validate the peptide-presenting integrity of the SCTs.

The introduced platforms for pMHC and TCR synthesis are implemented in the context of the recent COVID-19 pandemic to demonstrate a practical, immediate use case in Chapter 3. Immunodominant antigens for prevalent HLA alleles against CMV and EBV have been well-characterized and used as standard markers to identify immunity amongst patients. Here, I attempt to to identify similar immunogenic antigens from SARS-CoV-2. The initial section of this chapter details the earlier work conducted, in which SCT libraries were generated across three HLA alleles to encode predicted antigens derived from the structural spike protein. Additionally, an SCT library encoding antigens from the non-structural nsp3 protein was later developed for one HLA allele to comparatively assess for immunogenic differences based upon the origin and infective role of the protein source of the peptides. To confirm the antigen-specificity of detected T cells, cell sequencing technologies were utilized to extract TCR sequences, which were then cloned into healthy $CD8^+$ T cells to demonstrate antigen-specific elicitation of standard activation and proliferation markers. The latter section of this chapter addresses the expansion of this antigen discovery approach to cover more peptides and more patients, spanning the entire proteome of SARS-CoV-2 for one single HLA allele and surveying about 40 HLA-matched patients for reactivity to these peptides. Here we demonstrated the throughput power of the SCT platform, moving beyond technology maturation to highlight its applicability for biological discovery.

The same platforms are applied in the context of cancer in Chapter 4, showcasing the various use cases such as immunogenic peptide discovery and validation of epitope predictions in personalized medicine. Much of the work here is still in development, but is sufficient to demonstrate the potential for furthering tumor antigen discovery. Within this chapter, I highlight exemplary experiments of SCT libraries expressed for several HLA alleles to contain peptides from various origins, whether they are derived from tumor-associated antigens (TAAs), tumor-specific antigens (TSAs) (public and private), or even from spliced antigens. The SCT libraries developed in the context of TAAs and private TSAs illustrate how the SCT platform can

validate prediction software, while libraries developed for public TSAs illustrate the discovery potential of an unbiased approach to library production. Experiments from numerous cancer collaborators across the country are highlighted in this chapter to validate SCT integrity.

Finally, Chapter 5 will touch upon some earlier work conducted in my thesis, in which initial efforts to streamline existing lab technologies enabled the incorporation of pMHC-tetramers onto a magnetic nanoparticle scaffold to facilitate antigen-specific T cell enumeration. Subsequent work to augment this technology to allow for microfluidic extraction of such cells for sequence is described. These technologies were implemented in the Heath lab prior to the aforementioned SCT work, but are completely cross-compatible, enabling the improved robustness of SCT reagents to further bolster the quality of microfluidic assays.

References

[1] Jessica Eno. Immunotherapy Through the Years. *Journal of the Advanced Practitioner in Oncology*, 8(7):747–753, 2017. ISSN 2150-0878. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6188092/.

[2] M. A. Richardson, T. Ramirez, N. C. Russell, and L. A. Moye. Coley toxins immunotherapy: a retrospective review. *Alternative Therapies in Health and Medicine*, 5(3):42–47, May 1999. ISSN 1078-6791.

[3] DL Mager. Bacteria and cancer: cause, coincidence or cure? A review. *Journal of Translational Medicine*, 4:14, March 2006. ISSN 1479-5876. doi: 10.1186/1479-5876-4-14. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1479838/.

[4] S A Hoption Cann, J P van Netten, and C van Netten. Dr William Coley and tumour regression: a place in history or in the future. page 9.

[5] Uwe Hobohm. Fever and cancer in perspective. *Cancer Immunology, Immunotherapy*, 50(8):391–396, October 2001. ISSN 1432-0851. doi: 10.1007/s002620100216. URL https://doi.org/10.1007/s002620100216.

[6] M. H. Falk and R. D. Issels. Hyperthermia in oncology. *International Journal of Hyperthermia: The Official Journal of European Society for Hyperthermic Oncology, North American Hyperthermia Group*, 17(1):1–18, February 2001. ISSN 0265-6736. doi: 10.1080/02656730150201552.

[7] Ralph M. Steinman and Zanvil A. Cohn. IDENTIFICATION OF A NOVEL CELL TYPE IN PERIPHERAL LYMPHOID ORGANS OF MICE. *The*

Journal of Experimental Medicine, 137(5):1142–1162, May 1973. ISSN 0022-1007. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2139237/.

[8] R. M. Steinman and Z. A. Cohn. Identification of a novel cell type in peripheral lymphoid organs of mice. II. Functional properties in vitro. *The Journal of Experimental Medicine*, 139(2):380–397, February 1974. ISSN 0022-1007. doi: 10.1084/jem.139.2.380.

[9] Nussenzweig, M.C., Steinman, R.M., Gutchinov, B, and Cohn, Z.A. Dendritic cells are accessory cells for the development of anti- trinitrophenyl cytotoxic T lymphocytes. *The Journal of Experimental Medicine*, 152(4):1070–1084, October 1980. ISSN 0022-1007. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2185968/.

[10] Kayo Inaba, Ralph Steinman, and Wesley Voorhis. Dendritic cells are critical accessory cells for thymus-dependent antibody responses in mouse and in man. *Historical Scientific Reports*, January 1983. URL https://digitalcommons.rockefeller.edu/historical-scientific-reports/14.

[11] R. M. Steinman, G. Kaplan, M. D. Witmer, and Z. A. Cohn. Identification of a novel cell type in peripheral lymphoid organs of mice. V. Purification of spleen dendritic cells, new surface markers, and maintenance in vitro. *The Journal of Experimental Medicine*, 149(1):1–16, January 1979. ISSN 0022-1007. doi: 10.1084/jem.149.1.1.

[12] Clotilde Théry and Sebastian Amigorena. The cell biology of antigen presentation in dendritic cells. *Current Opinion in Immunology*, 13(1):45–51, February 2001. ISSN 09527915. doi: 10.1016/S0952-7915(00)00180-1. URL https://linkinghub.elsevier.com/retrieve/pii/S0952791500001801.

[13] Kenneth L. Rock, Eric Reits, and Jacques Neefjes. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends in Immunology*, 37(11):724–737, November 2016. ISSN 14714906. doi: 10.1016/j.it.2016.08.010. URL https://linkinghub.elsevier.com/retrieve/pii/S1471490616301004.

[14] P J Bjorkman, J L Strominger, and D C Wiley. Structure of the human class I histocompatibility antigen, HLA-A2. page 7, 1987.

[15] J. P. Allison, B. W. McIntyre, and D. Bloch. Tumor-specific antigen of murine T-lymphoma defined with monoclonal antibody. *Journal of Immunology (Baltimore, Md.: 1950)*, 129(5):2293–2300, November 1982. ISSN 0022-1767.

[16] David N. Garboczi, Partho Ghosh, Ursula Utz, Qing R. Fan, William E. Biddison, and Don C. Wiley. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature*, 384(6605):134–141, November 1996. ISSN 1476-4687. doi: 10.1038/384134a0. URL https:

//www.nature.com/articles/384134a0. Number: 6605 Publisher: Nature Publishing Group.

[17] Elizabeth A Walter, Rosalynde J Finch, Käthe S Watanabe, E Donnall Thomas, and Stanley R Riddell. Reconstitution of Cellular Immunity against Cytomegalovirus in Recipients of Allogeneic Bone Marrow by Transfer of T-Cell Clones from the Donor. *THE NEW ENGLAND JOURNAL OF MEDICINE*, page 7, 1995.

[18] Cliona M Rooney, Colton A Smith, Catherine Y C Ng, Susan K Loftin, John W Sixbey, Yanjun Gan, Deo-Kumar Srivastava, Laura C Bowman, Robert A Krance, Malcolm K Brenner, and Helen E Heslop. Infusion of Cytotoxic T Cells for the Prevention and Treatment of Epstein-Barr Virus–Induced Lymphoma in Allogeneic Transplant Recipients. page 7.

[19] Steven A. Rosenberg, James C. Yang, Richard M. Sherry, Udai S. Kammula, Marybeth S. Hughes, Giao Q. Phan, Deborah E. Citrin, Nicholas P. Restifo, Paul F. Robbins, John R. Wunderlich, Kathleen E. Morton, Carolyn M. Laurencot, Seth M. Steinberg, Donald E. White, and Mark E. Dudley. Durable Complete Responses in Heavily Pretreated Patients with Metastatic Melanoma Using T-Cell Transfer Immunotherapy. *Clinical Cancer Research*, 17(13):4550–4557, July 2011. ISSN 1078-0432, 1557-3265. doi: 10.1158/1078-0432.CCR-11-0116. URL https://clincancerres.aacrjournals.org/content/17/13/4550. Publisher: American Association for Cancer Research Section: Cancer Therapy: Clinical.

[20] Lu Zhang, Yanyu Huang, Aaron Raymond Lindstrom, Tzu-Yin Lin, Kit S Lam, and Yuanpei Li. Peptide-based materials for cancer immunotherapy. *Theranostics*, 9(25):7807–7825, October 2019. ISSN 1838-7640. doi: 10.7150/thno.37194. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6831480/.

[21] Alex D. Waldman, Jill M. Fritz, and Michael J. Lenardo. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nature Reviews Immunology*, 20(11):651–668, November 2020. ISSN 1474-1741. doi: 10.1038/s41577-020-0306-5. URL https://www.nature.com/articles/s41577-020-0306-5. Number: 11 Publisher: Nature Publishing Group.

[22] Judith A. Seidel, Atsushi Otsuka, and Kenji Kabashima. Anti-PD-1 and Anti-CTLA-4 Therapies in Cancer: Mechanisms of Action, Efficacy, and Limitations. *Frontiers in Oncology*, 8, March 2018. ISSN 2234-943X. doi: 10.3389/fonc.2018.00086. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5883082/.

[23] Jr Charles A Janeway, Paul Travers, Mark Walport, and Mark J. Shlomchik. Antigen recognition by T cells. *Immunobiology: The Immune System in Health and Disease. 5th edition*, 2001. URL https://www.ncbi.nlm.nih.gov/books/NBK27098/. Publisher: Garland Science.

[24] E. K. Barber, J. D. Dasgupta, S. F. Schlossman, J. M. Trevillyan, and C. E. Rudd. The CD4 and CD8 antigens are coupled to a protein-tyrosine kinase (p56lck) that phosphorylates the CD3 complex. *Proceedings of the National Academy of Sciences of the United States of America*, 86(9):3277–3281, May 1989. ISSN 0027-8424. doi: 10.1073/pnas.86.9.3277.

[25] Pierre Guermonprez, Jenny Valladeau, Laurence Zitvogel, Clotilde Théry, and Sebastian Amigorena. Antigen Presentation and T Cell Stimulation by Dendritic Cells. *Annual Review of Immunology*, 20(1):621–667, 2002. doi: 10.1146/annurev.immunol.20.100301.064828. URL https://doi.org/10.1146/annurev.immunol.20.100301.064828. _eprint: https://doi.org/10.1146/annurev.immunol.20.100301.064828.

[26] Jatin M. Vyas, Annemarthe G. Van der Veen, and Hidde L. Ploegh. The known unknowns of antigen processing and presentation. *Nature Reviews Immunology*, 8(8):607–618, August 2008. ISSN 1474-1741. doi: 10.1038/nri2368. URL https://www.nature.com/articles/nri2368. Number: 8 Publisher: Nature Publishing Group.

[27] Michael Basler, Christopher J. Kirk, and Marcus Groettrup. The immunoproteasome in antigen processing and other immunological functions. *Current Opinion in Immunology*, 25(1):74–80, February 2013. ISSN 1879-0372. doi: 10.1016/j.coi.2012.11.004.

[28] Brigitte Lankat-Buttgereit and Robert Tampé. The transporter associated with antigen processing: function and implications in human diseases. *Physiological Reviews*, 82(1):187–204, January 2002. ISSN 0031-9333. doi: 10.1152/physrev.00025.2001.

[29] Thomas Serwold, Federico Gonzalez, Jennifer Kim, Richard Jacob, and Nilabh Shastri. ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature*, 419(6906):480–483, October 2002. ISSN 0028-0836. doi: 10.1038/nature01074.

[30] Richard J Riese, Paula R Wolf, Dieter Brömme, Lisa R Natkin, José A Villadangos, Hidde L Ploegh, and Harold A Chapman. Essential Role for Cathepsin S in MHC Class II–Associated Invariant Chain Processing and Peptide Loading. *Immunity*, 4(4):357–366, April 1996. ISSN 1074-7613. doi: 10.1016/S1074-7613(00)80249-6. URL http://www.sciencedirect.com/science/article/pii/S1074761300802496.

[31] Peter Cresswell. Invariant Chain Structure and MHC Class II Function. *Cell*, 84(4):505–507, February 1996. ISSN 0092-8674, 1097-4172. doi: 10.1016/S0092-8674(00)81025-9. URL https://www.cell.com/cell/abstract/S0092-8674(00)81025-9. Publisher: Elsevier.

[32] Elizabeth D. Mellins and Lawrence J. Stern. HLA-DM and HLA-DO, key regulators of MHC-II processing and presentation. *Current opinion in immunology*,

0:115–122, February 2014. ISSN 0952-7915. doi: 10.1016/j.coi.2013.11.005. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3944065/.

[33] Stephen M. Hedrick, David I. Cohen, Ellen A. Nielsen, and Mark M. Davis. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature*, 308(5955):149–153, March 1984. ISSN 1476-4687. doi: 10.1038/308149a0. URL https://www.nature.com/articles/308149a0. Number: 5955 Publisher: Nature Publishing Group.

[34] Stephen M. Hedrick, Ellen A. Nielsen, Joshua Kavaler, David I. Cohen, and Mark M. Davis. Sequence relationships between putative T-cell receptor polypeptides and immunoglobulins. *Nature*, 308(5955):153–158, March 1984. ISSN 1476-4687. doi: 10.1038/308153a0. URL https://www.nature.com/articles/308153a0. Number: 5955 Publisher: Nature Publishing Group.

[35] Yusuke Yanagi, Yasunobu Yoshikai, Kathleen Leggett, Stephen P. Clark, Ingrid Aleksander, and Tak W. Mak. A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. *Nature*, 308 (5955):145–149, March 1984. ISSN 1476-4687. doi: 10.1038/308145a0. URL https://www.nature.com/articles/308145a0. Number: 5955 Publisher: Nature Publishing Group.

[36] Susumu Tonegawa. Somatic generation of antibody diversity. *Nature*, 302 (5909):575–581, April 1983. ISSN 1476-4687. doi: 10.1038/302575a0. URL https://www.nature.com/articles/302575a0. Number: 5909 Publisher: Nature Publishing Group.

[37] Yueh-hsiu Chien, Nicholas R. J. Gascoigne, Joshua Kavaler, Nadine E. Lee, and Mark M. Davis. Somatic recombination in a murine T-cell receptor gene. *Nature*, 309(5966):322–326, May 1984. ISSN 1476-4687. doi: 10.1038/309322a0. URL https://www.nature.com/articles/309322a0. Number: 5966 Publisher: Nature Publishing Group.

[38] Kun Yu, Ji Shi, Dan Lu, and Qiong Yang. Comparative analysis of CDR3 regions in paired human $\alpha\beta$ CD8 T cells. *FEBS Open Bio*, 9(8):1450–1459, July 2019. ISSN 2211-5463. doi: 10.1002/2211-5463.12690. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6668380/.

[39] Ming Zheng, Xin Zhang, Yinghui Zhou, Juan Tang, Qing Han, Yang Zhang, Qingshan Ni, Gang Chen, Qingzhu Jia, Haili Yu, Siqi Liu, Elizabeth Robins, Ning Jenny Jiang, Ying Wan, Qi-Jing Li, Zhi-Nan Chen, and Ping Zhu. TCR repertoire and CDR3 motif analyses depict the role of $\alpha\beta$ T cells in Ankylosing spondylitis. *EBioMedicine*, 47:414–426, September 2019. ISSN 2352-3964. doi: 10.1016/j.ebiom.2019.07.032. URL http://www.sciencedirect.com/science/article/pii/S2352396419304748.

[40] Nadezhda N. Logunova, Valeriia V. Kriukova, Pavel V. Shelyakin, Evgeny S. Egorov, Alina Pereverzeva, Nina G. Bozhanova, Mikhail Shugay, Dmitrii S. Shcherbinin, Mikhail V. Pogorelyy, Ekaterina M. Merzlyak, Vasiliy N. Zubov, Jens Meiler, Dmitriy M. Chudakov, Alexander S. Apt, and Olga V. Britanova. MHC-II alleles shape the CDR3 repertoires of conventional and regulatory naïve CD4+ T cells. *Proceedings of the National Academy of Sciences*, 117(24):13659–13669, June 2020. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2003170117. URL https://www.pnas.org/content/117/24/13659. Publisher: National Academy of Sciences Section: Biological Sciences.

[41] Gisela M. Vaitaitis, Michelle Poulin, Richard J. Sanderson, Kathryn Haskins, and David H. Wagner. Cutting Edge: CD40-Induced Expression of Recombination Activating Gene (RAG) 1 and RAG2: A Mechanism for the Generation of Autoaggressive T Cells in the Periphery. *The Journal of Immunology*, 170(7):3455–3459, April 2003. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.170.7.3455. URL https://www.jimmunol.org/content/170/7/3455. Publisher: American Association of Immunologists Section: CUTTING EDGE.

[42] Erica Lantelme, Luca Orlando, Paola Porcedda, Valentina Turinetto, Mario De Marchi, Antonio Amoroso, Stefania Mantovani, and Claudia Giachino. An in vitro model of T cell receptor revision in mature human CD8+ T cells. *Molecular Immunology*, 45(2):328–337, January 2008. ISSN 0161-5890. doi: 10.1016/j.molimm.2007.06.153. URL http://www.sciencedirect.com/science/article/pii/S0161589007004075.

[43] J Scott Hale and Pamela J Fink. T-cell receptor revision: friend or foe? *Immunology*, 129(4):467–473, April 2010. ISSN 0019-2805. doi: 10.1111/j.1365-2567.2010.03250.x. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2842493/.

[44] Xin He, Charles A. Janeway, Matthew Levine, Eve Robinson, Paula Preston-Hurlburt, Christophe Viret, and Kim Bottomly. Dual receptor T cells extend the immune repertoire for foreign antigens. *Nature Immunology*, 3(2):127–134, February 2002. ISSN 1529-2916. doi: 10.1038/ni751. URL https://www.nature.com/articles/ni751. Number: 2 Publisher: Nature Publishing Group.

[45] Timothy J. O'Donnell, Alex Rubinsteyn, Maria Bonsack, Angelika B. Riemer, Uri Laserson, and Jeff Hammerbacher. MHCflurry: Open-Source Class I MHC Binding Affinity Prediction. *Cell Systems*, 7(1):129–132.e4, July 2018. ISSN 2405-4712. doi: 10.1016/j.cels.2018.05.014. URL http://www.sciencedirect.com/science/article/pii/S2405471218302321.

[46] Morten Nielsen, Claus Lundegaard, Peder Worning, Sanne Lise Lauemøller, Kasper Lamberth, Søren Buus, Søren Brunak, and Ole Lund. Reliable pre-

diction of T-cell epitopes using neural networks with novel sequence representations. *Protein Science : A Publication of the Protein Society*, 12(5): 1007–1017, May 2003. ISSN 0961-8368. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2323871/.

[47] Massimo Andreatta and Morten Nielsen. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics*, 32(4):511–517, February 2016. ISSN 1367-4803. doi: 10.1093/bioinformatics/btv639. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6402319/.

[48] Birkir Reynisson, Bruno Alvarez, Sinu Paul, Bjoern Peters, and Morten Nielsen. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Research*, 48(W1):W449–W454, July 2020. ISSN 0305-1048. doi: 10.1093/nar/gkaa379. URL https://doi.org/10.1093/nar/gkaa379.

[49] Kamilla Kjærgaard Jensen, Massimo Andreatta, Paolo Marcatili, Søren Buus, Jason A. Greenbaum, Zhen Yan, Alessandro Sette, Bjoern Peters, and Morten Nielsen. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology*, 154(3):394–406, July 2018. ISSN 0019-2805. doi: 10.1111/imm.12889. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6002223/.

[50] Can Keşmir, Alexander K. Nussbaum, Hansjörg Schild, Vincent Detours, and Søren Brunak. Prediction of proteasome cleavage motifs by neural networks. *Protein Engineering, Design and Selection*, 15(4):287–296, April 2002. ISSN 1741-0126. doi: 10.1093/protein/15.4.287. URL https://doi.org/10.1093/protein/15.4.287.

[51] Henrik Ferré, Emmanuel Ruffet, Thomas Blicher, Christina Sylvester-Hvid, Lise Lotte B. Nielsen, Timothy J. Hobley, Owen R.T. Thomas, and Søren Buus. Purification of correctly oxidized MHC class I heavy-chain molecules under denaturing conditions: A novel strategy exploiting disulfide assisted protein folding. *Protein Science : A Publication of the Protein Society*, 12(3): 551–559, March 2003. ISSN 0961-8368. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2312438/.

[52] Tatyana Sandalova, Jakob Michaëlsson, Robert A. Harris, Hans-Gustaf Ljunggren, Klas Kärre, Gunter Schneider, and Adnane Achour. Expression, refolding and crystallization of murine MHC class I H-2Db in complex with human β2-microglobulin. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 61(Pt 12):1090–1093, November 2005. ISSN 1744-3091. doi: 10.1107/S1744309105037942. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1978157/.

[53] Arnold H. Bakker, Rieuwert Hoppes, Carsten Linnemann, Mireille Toebes, Boris Rodenko, Celia R. Berkers, Sine Reker Hadrup, Wim J. E. van Esch, Mirjam H. M. Heemskerk, Huib Ovaa, and Ton N. M. Schumacher. Conditional MHC class I ligands and peptide exchange technology for the human MHC gene products HLA-A1, -A3, -A11, and -B7. *Proceedings of the National Academy of Sciences*, 105(10):3825–3830, March 2008. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0709717105. URL https://www.pnas.org/content/105/10/3825. ISBN: 9780709717102 Publisher: National Academy of Sciences Section: Biological Sciences.

[54] Boris Rodenko, Mireille Toebes, Sine Reker Hadrup, Wim J. E. van Esch, Annemieke M. Molenaar, Ton N. M. Schumacher, and Huib Ovaa. Generation of peptide-MHC class I complexes through UV-mediated ligand exchange. *Nature Protocols*, 1(3):1120–1132, 2006. ISSN 1750-2799. doi: 10.1038/nprot.2006.121.

[55] Sarah A. Overall, Jugmohit S. Toor, Stephanie Hao, Mark Yarmarkovich, Sara M. O'Rourke, Giora I. Morozov, Son Nguyen, Alberto Sada Japp, Nicolas Gonzalez, Danai Moschidi, Michael R. Betts, John M. Maris, Peter Smibert, and Nikolaos G. Sgourakis. High throughput pMHC-I tetramer library production using chaperone-mediated peptide exchange. *Nature Communications*, 11(1):1909, April 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-15710-1. URL https://www.nature.com/articles/s41467-020-15710-1. Number: 1 Publisher: Nature Publishing Group.

[56] Michael E. Birnbaum, Juan L. Mendoza, Dhruv K. Sethi, Shen Dong, Jacob Glanville, Jessica Dobbins, Engin Özkan, Mark M. Davis, Kai W. Wucherpfennig, and K. Christopher Garcia. Deconstructing the Peptide-MHC Specificity of T Cell Recognition. *Cell*, 157(5):1073–1087, May 2014. ISSN 00928674. doi: 10.1016/j.cell.2014.03.047. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867414004760.

[57] Jacob Glanville, Huang Huang, Allison Nau, Olivia Hatton, Lisa E. Wagar, Florian Rubelt, Xuhuai Ji, Arnold Han, Sheri M. Krams, Christina Pettus, Nikhil Haas, Cecilia S. Lindestam Arlehamn, Alessandro Sette, Scott D. Boyd, Thomas J. Scriba, Olivia M. Martinez, and Mark M. Davis. Identifying specificity groups in the T cell receptor repertoire. *Nature*, 547(7661):94–98, July 2017. ISSN 0028-0836. doi: 10.1038/nature22976. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5794212/.

[58] Huang Huang, Chunlin Wang, Florian Rubelt, Thomas J. Scriba, and Mark M. Davis. Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nature Biotechnology*, 38(10):1194–1202, October 2020. ISSN 1546-1696. doi: 10.1038/s41587-020-0505-4.

[59] L Ignatowicz, G Winslow, J Bill, J Kappler, and P Marrack. Cell surface expression of class II MHC proteins bound by a single peptide. page 12, January 1995.

[60] Y. Y. L. Yu, N. Netuschil, L. Lybarger, J. M. Connolly, and T. H. Hansen. Cutting Edge: Single-Chain Trimers of MHC Class I Molecules Form Stable Structures That Potently Stimulate Antigen-Specific T Cells and B Cells. *The Journal of Immunology*, 168(7):3145–3149, April 2002. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.168.7.3145. URL http://www.jimmunol.org/cgi/doi/10.4049/jimmunol.168.7.3145.

[61] Guideng Li, Michael T. Bethune, Stephanie Wong, Alok V. Joglekar, Michael T. Leonard, Jessica K. Wang, Jocelyn T. Kim, Donghui Cheng, Songming Peng, Jesse M. Zaretsky, Yapeng Su, Yicheng Luo, James R. Heath, Antoni Ribas, Owen N. Witte, and David Baltimore. T cell antigen discovery via trogocytosis. *Nature Methods*, 16(2):183–190, February 2019. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-018-0305-7. URL http://www.nature.com/articles/s41592-018-0305-7.

[62] Alok V. Joglekar, Michael T. Leonard, John D. Jeppson, Margaret Swift, Guideng Li, Stephanie Wong, Songming Peng, Jesse M. Zaretsky, James R. Heath, Antoni Ribas, Michael T. Bethune, and David Baltimore. T cell antigen discovery via signaling and antigen-presenting bifunctional receptors. *Nature Methods*, 16(2):191–198, February 2019. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-018-0304-8. URL http://www.nature.com/articles/s41592-018-0304-8.

Chapter 2

MOLECULAR ENGINEERING TECHNOLOGIES FOR IMMUNOTHERAPEUTICS

2.1 Class I SCTs

2.1.1 Abstract

A major obstacle to the discovery of immunogenic viral or tumor-derived CD8 T cell epitopes is the labor-intensive synthesis and construction of peptide-major histocompatibility complex (pMHC) tetramers. In this section, we adopt single-chain trimer technologies to develop a high-throughput platform for Class I pMHC generation, showing that hundreds can be rapidly prepared at ease and characterized to show peptide-dependent or HLA template-dependent trends in SCT expression and thermal stability. Functional characterization of some SCTs are further highlighted, and we demonstrate biological relevance of SCTs targeting common viral epitopes. The technologies described herein will enable rapid analyses of peptide-based T cell responses for therapeutic development across several contexts, including autoimmune, cancer, or infectious disease.

2.1.2 Introduction

The emergence of novel, pathogenic virus strains (and the predicted acceleration of such events) has driven the need for high-throughput approaches to epitope-based reagent production [1]. In particular, the use of peptide-MHC reagents to capture antigen-specific T cells will enable identification of relevant TCR sequences and shed light on the role played by immunodominant epitopes in the host immune response. Toward this end, vaccine therapies must involve assessment of HLA haplotypes and HLA-based epitope landscapes to predict and identify the most prominent immunogenic viral peptides. The number of compatible epitopes per HLA allele may differ vastly, ranging from only a handful up to hundreds or thousands based on the desired scope of inclusion, the natural receptivity of each HLA allele's binding pocket to peptide motifs, and the accuracy of existing peptide binding prediction algorithms. To accommodate this scale, soluble pMHC reagents must be produced on a per-peptide, per-HLA basis in a high-throughput manner to identify and rank immuno-responsive TCRs from peripheral blood mononuclear cells

(PBMCs). Soluble pMHCs are conventionally produced by individual expression of the subunits of the MHC within E. coli, followed by subsequent *in vitro* refolding of the HLA heavy chain and β2-microglobulin (β2m) subunit inclusion bodies in the presence of a target peptide [2]. A modified version to produce the refolded pMHC complex makes use of a UV-cleavable peptide during the reaction [3–5]. This peptide serves as a placeholder, enabling rapid production of UV-exchanged pMHCs (UV-pMHCs) where UV light exposure facilitates exchange of the cleavable peptide for target peptide. As reported by others, the production of refolded pMHCs and UV-pMHCs is prone to several technical problems. Overall protein yield from refolding is HLA-dependent, and the success of UV exchange is highly dependent upon chemico-physical properties of the individual peptide.

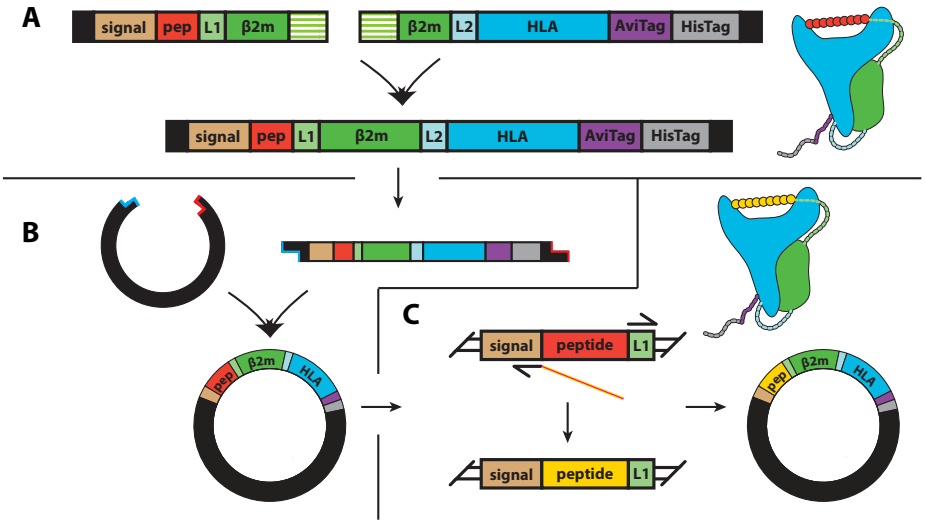


Figure 2.1: SCT design for Class I pMHC constructs
A. SCTs encoding Class I pMHC molecules are constructed by Gibson assembly from two fragments, enabling modular insertion of any desired Class I HLA subunit to design a template plasmid for peptide insertion. B. Template SCT constructs are ligated into pcDNA3.1 vector by restriction digest and ligation. C. An SCT library containing various peptide elements can be constructed from an initial template plasmid by inverse PCR and ligation.

Single-chain trimers (SCTs) are an alternative approach to construct pMHCs that may address the issues posed by refolding and UV exchange [6–8]. Briefly, the SCT format consists of a pcDNA3.1 plasmid construct encoding a secretion signal, peptide, peptide-β2m linker (L1), β2m, β2m-HLA linker (L2), HLA, glycine linker, AviTag, and 6x HisTag (**Fig. 2.1A**). The three primary units, joined by two linkers to give a single chain, are secreted as a single protein unit. Initially expressed in

bacterial cells, SCTs were adopted into mammalian expression systems, enabling significant improvements to overall protein yield, presumably due to the use of a native mammalian system's internal protein folding and quality control mechanisms to regulate SCT output [8]. SCTs were also engineered with several peptide binding pocket mutations to minimize interference of the peptide linker with pMHC function and to improve immunogenicity of the pMHC reagents [9–12]. Recently, the SCT expression system was adopted into Expi293 cells to maximize SCT expression quantity, and peptide modularity was introduced using homologous recombination of peptide-encoded DNA fragments into the plasmid to enable production and functional characterization of two SCTs encoding HLA-A*24:02 viral peptides [13]. The works outlined above are key studies which point to the potential use of SCTs as functionally similar alternatives to UV-pMHCs. To our knowledge, however, we have not seen works that simultaneously assess soluble SCTs on an antigen library scale and encompass the engineered mutations as outlined above.

Herein we introduce a high-throughput SCT expression platform enabling production of SCTs for any pairing of peptide and Class I HLA allele. Whereas with traditional pMHC folding, epitope and HLA modularity are determined by peptide synthesis and refolded MHC subunits, respectively, the SCT platform utilizes a primer and a PCR template plasmid to determine these two variables. The facile nature of handling and scaling up these PCR reagents enables a mix-and-match approach that allows one to rapidly screen across a peptide library and list of HLA template variants to optimize pMHCs. We initially apply this system for a test case of 18 tumor-associated antigens (TAAs) for HLA-A*02:01, utilizing nine different L1/HLA templates as previously published, in order to two-dimensionally assess the impact of peptide identity and L1/HLA templates on SCT protein expression and thermal stability. Next, we highlight the functionality of these SCTs in a disease context by assembling HLA-A*02:01 and A*24:02 SCTs loaded with epitopes derived from common viral strains, demonstrating that they can bind to healthy donor T cells stimulated against the synthesized forms of these epitopes.

2.1.3 Methods

2.1.3.1 SCT template production

Class I SCT-encoded plasmids were constructed using a combination of Gibson assembly and restriction enzyme digest methods for insertion into pcDNA3.1 Zeo(+)

plasmid (Thermo Fisher Scientific) (**Fig. 2.1A**). Briefly, the SCT inserts were designed to be modular to allow for any choice of L1 to be paired with any choice of HLA allele. Because β2m has no allelic variation in the human species, the SCT was split into two Gibson assembly fragments within this region to allow for decoupling of L1 from HLA. Fragments were purchased from Twist Bioscience, PCR-amplified with KOD HotStart Hi-Fi polymerase (MilliporeSigma), and joined together by Gibson assembly using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). The PCR-amplified Gibson product's flanking regions were digested by EcoRI and XhoI (New England Biolabs) to be ligated into pcDNA3.1's MCS region at the same enzyme recognition sites (**Fig. 2.1B**). Codon optimization was applied to the designed fragments under three considerations: 1) selection of only highly prevalent codons in the human species, 2) avoidance of continuous gene segments (24+ bp) where GC content is above 60% (to avoid manufacturer error rates during synthesis), and 3) avoidance of key recognition cut sites within the fragments, which must only exist at the flanks of the Gibson product for insertion into pcDNA vector. This strategy was initially used successfully across three HLA alleles (A*01:01, A*02:01, and A*03:01). Subsequently, the design of the second fragment (encoding HLA allele) was automated with a Python script, encompassing all aforementioned design criteria and accounting for all alleles from Class I HLA-A, B, C loci. The protein sequences of each HLA allele were obtained from an FTP server hosted by The Immuno Polymorphism Database (ftp://ftp.ebi.ac.uk/pub/databases/ipd/imgt/hla/fasta/). A detailed protocol and discussion of this design can be found in **Appendix A**. To date, all existing Class I HLA sequences from the IMGT database (linked above) have been converted in this manner into ready-to-order DNA sequences. From these sequences, we have constructed at least 40 unique plasmid templates, encompassing 24 HLA-A, HLA-B, and HLA-C alleles.

2.1.3.2 SCT peptide library production

A PCR-facilitated approach was implemented to enable high-throughput substitution of peptides into SCT-encoded plasmids. We opted to use extension PCR methods among other potential approaches after consideration of cost, ease-of-use, and flexibility for various L1 choices coupled next to the plasmid (**Fig. 2.1C**). Briefly, for any given peptide substitution, a peptide-encoded reverse primer (binding to the signal sequence upstream of peptide region) and a forward primer (binding to

L1 downstream of peptide region) is required. The peptide-encoded primer varies for any given peptide, while the forward primer remains fixed across all peptide elements (unless one chooses to use a different L1/HLA template plasmid). In this manner, an SCT plasmid library, encompassing *n* peptides and *m* templates, requires the purchase of *n* + *m* total primers. Extension PCR was conducted with KOD Hot Start polymerase (MilliporeSigma). The product was phosphorylated and ligated with a mixture of T4 Polynucleotide Kinase and T4 DNA Ligase, and then template DNA was digested with DpnI (New England Biolabs). The peptide-substituted plasmids were then transformed into One Shot TOP10 Chemically Competent E. coli (Thermo Fisher Scientific). Plasmids were verified by Sanger sequencing using a Python script prior to use in transfection. A detailed discussion and protocol of high-throughput implementations for each of these steps can be found in **Appendix B & C**.

2.1.3.3 SCT expression

Purified SCT plasmids were transfected into Expi293 cells (Thermo Fisher Scientific) within 24-well (2.5 ml capacity) plates. Briefly, 1.25 µg of plasmid was mixed with 75 µl Opti-MEM reduced serum media. 7.5 µl of ExpiFectamine Reagent was mixed with 70 µl Opti-MEM reduced serum media, incubated at room temperature for 5 minutes, and combined with the plasmid mixture. After a 15-minute room temperature incubation, the solution was added to 1.25 ml of Expi293 cells at 3 million cells/ml into a 24-well plate, which was then shaken at 225 RPM at 37 °C in 8% CO_2 overnight. Twenty hours later, a solution containing 7.5 µl of ExpiFectamine Transfection Enhancer 1 and 75 µl of ExpiFectamine Transfection Enhancer 2 was added to each well. The plate was kept on the shaker using aforementioned settings for a total of 4 days from start of transfection. The supernatant of the transfection solution was collected and filtered through 0.22 µm PVFD membrane syringe filters (MilliporeSigma) prior to yield analysis via SDS-PAGE. The supernatant solutions of SCTs which expressed at high yield were concentrated down to 200 µl PBS using 30 kDa centrifugal filter units (Amicon) and subsequently biotinylated with BirA enzyme kit (Avidity) overnight. The biotinylated SCTs were then purified with HisTag resin tips (Phynexus) and desalted back into PBS buffer with Zeba 7K MWCO spin desalting columns (Thermo Fisher Scientific). For long-term storage, the SCTs were re-suspended into 20% glycerol w/v prior to storage in -20 °C. The implementation of the aforementioned methods discussed in this section in

a high-throughput manner is discussed in more detail in **Appendix D**.

2.1.3.4 SCT yield characterization

After 4 days of transfection, a 15 µl solution containing 3:1 mix of transfection supernatant and Laemmli buffer with 10% β-mercaptoethanol was denatured at 100 °C for 10 minutes, and subsequently loaded into Bio-Rad Stain-Free gels for SDS-PAGE (200V, 30 minutes). A reduced, purified WT1 (RMFPNAPYL) A*02:01 SCT sample in 20% glycerol PBS solution (containing approximately 2 µg) was run in each gel to serve as a positive control and intensity reference for relative protein yield calculation. Images were obtained using a Bio-Rad ChemiDoc MP gel imaging system (manual settings: 45 seconds UV activation, 0.5 second exposure). Because of our intention to express SCTs in a high-throughput manner for peptide libraries on the scale of over one hundred samples, and our desire to be able to uniformly assess emergent trends based on protein expression, it was necessary to identify a consistent approach for analyzing SCT expression. Toward this end, a custom Python script was developed specifically for the analysis of SCT proteins run on Stain-Free gels (Bio-Rad). The script allows for user-defined selection of protein bands of interest, and provides background reduction and uniform normalization of SCT yield across all gels given the consistent use of a control protein lane. The accuracy of this approach was measured by SDS-PAGE of titrated, pre-quantified samples of purified SCTs to demonstrate a 99% correlation between true protein A280 concentration (as measured by NanoDrop 8000 Spectrophotometer) and quantified relative band intensity (**Fig. E.2**). SCTs which expressed above an established cutoff for yield were selected for subsequent biotinylation and purification steps. A more detailed discussion of the drawbacks behind conventional or available approaches and a protocol of how to use our script can be found in **Appendix E**.

2.1.3.5 Thermal stability characterization

SYPRO™ Orange Protein Gel Stain was purchased from ThermoFisher Scientific and diluted with H_2O to give a 100X working solution. To each 19 µl aliquot of Class II SCT protein solution (diluted to 10 µM, if possible), 1 µl of the 100X dye solution was added. A Bio-Rad thermal cycler equipped with a CFX96 real-time PCR detection system was used in combination with Precision Melt Analysis

software to obtain melting curves of each SCT sample. Thermal ramp settings were 25 °C to 95 °C, 0.2 °C per 30 seconds.

2.1.3.6 Peptide stimulation

The thawed PBMCs were incubated in complete R10 media (500 ml of RPMI 1640; 50mL Heat-inactivated FBS; 5 ml of Pen/strep (100 U/mL penicillin and 100 ug/mL streptomycin); 1x GlutaMAX) by adding 1 μM of peptide and anti-CD40 antibody (1 μg/mL) for 16 hrs. On the next day, the PBMCs were washed and stained with Annexin V-BV421 (1 μg/mL), CD8-FITC antibody (1 μg/mL) and CD137-PE antibody (1 μg/mL) for 10 mins at 4 °C. Activation-induced expression of CD137 by peptide stimulation permits the sorting of antigen specific T-cells into tubes using FACS sorter equipment.

2.1.3.7 SCT multimer formation

Biotinylated SCT monomers have been successfully used in at least three different formats. First, they have been tetramerized with Streptavidin-Phycoerythin (PE) (BioLegend) for use as conventional flow cytometry staining reagents. Second, they have been tetramerized with a custom-made streptavidin-DNA conjugate to allow for subsequent binding onto complementary ssDNA-biotin molecules affixed on streptavidin-coated magnetic Dynabeads (Thermo Fisher Scientific). These reagents can be utilized in our previously published work describing the nanoparticle-nucleic acid cell sorting platform (NP-NACS) [14], which allows for enhanced pMHC-TCR avidity and microfluidic-guided extraction and analysis of antigen-specific T cells. Third, we have also been able to conjugate our SCT monomers onto 10X-compatible DNA barcoded dextramers (Immudex). These reagents enable coupling of the antigen-specific identity (DNA barcoded onto dextramers) of a captured CD8 T cell and its corresponding TCR α and β chain sequences (single-cell mRNA sequencing). SCTs used as conventional tetramers are exemplified throughout various figures (e.g. **Fig. 4.11**). SCTs used within NP-NACS were the primary focus of TESLA A*01:01 and A*03:01 libraries (**Fig. 4.20 & Table 4.7**), and of initial COVID assays (**Figs. 3.4 & 3.6**). Finally, SCT-dextramers were incorporated into 10x single-cell sequencing for SARS-CoV-2 proteome analysis (**Figs. 3.8 & J.1**) to enable large-scale multiplexing of single-cell antigen-TCR analysis.

2.1.4 Results

2.1.4.1 Expression of SCT library

Our initial SCT library consisted of 18 HLA-A*02:01 antigens derived from various sources (**Table F.1**). To identify candidate L1/HLA mutations to introduce into the SCT, we surveyed the literature for engineered improvements made to SCT design. Three generations of L1-HLA combinations (closed groove (wild-type HLA Y84), open groove (HLA Y84A), and thiol linker (HLA Y84C)) have been previously explored and shown to demonstrate gradual improvements in pMHC stability. We implemented these three generations into five unique designs, abbreviated D1 (L1 = $(GGGGS)_3$; closed groove), D2 (L1 = $(GGGGS)_3$; open groove), D3 (L1 = $GCGGS(GGGGS)_2$; thiol linker), D4 (L1 = $GGCGS(GGGGS)_2$; thiol linker), and D5 (L1 = $GCGAS(GGGGS)_2$; thiol linker) (**Fig. 2.2A-B**) [9–11]. Designs which contained a cysteine in the linker (D3-D5) also incorporated the Y84C mutation in the HLA subunit to enable dithiol linkage. Next, we implemented an orthogonal HLA mutation, H74L, into three of the templates (D6-D8) [12]. The H74L mutation forms a portion of the C pocket in the peptide binding groove of the HLA subunit and has been reported to facilitate peptide loading and pMHC immunogenicity, so we rationalized that its inclusion may improve overall pMHC stability and function. Our final design (D9, termed DS-SCT) accounts for a recent report which stated that the paired Y84C-A139C mutation to the HLA binding pocket could introduce further stabilization to refolded pMHC construct [15–18]. This 162-element plasmid library, encompassing nine HLA templates and 18 peptides, was transfected into Expi293 cells (**Fig. 2.2B**). Reduced SDS-PAGE analysis of the SCT protein bands revealed significant variations in protein yield that was dependent on peptide and template (**Fig. 2.2B**). To decouple the effect of transfection efficiency on SCT yield, a subset of the library under design D3 was further modified to incorporate an IRES-GFP sequence, such that regardless of peptide identity or degree of SCT expression, transfected cells would be induced to express intracellular GFP (**Fig. 2.3**) [19]. Flow cytometry-based detection of GFP-positive cells indicated that the degree of transfection efficiency was approximately uniform (70%) across all tested SCT constructs (**Fig. 2.3A**). A biological triplicate of this subset, with and without the IRES-GFP insert, was conducted to demonstrate that the peptide-dependent SCT yield variations are consistent (**Fig. 2.3B**). The three H74L mutation templates among the library generally demonstrated improved protein expression relative to their wild-type counterparts, and the templates making use of thiol linkers produced

the highest overall yields of SCTs (**Fig. 2.2B**). In some cases, such as the peptide AIQDLCLAV, SCT expression could only be obtained with the D8 template, which incorporates both H74L and thiol linker features, or with the D9 template, possibly due to stability at the F pocket conferred by the dithiol mutation. Interestingly, we observed a slight upward shift of the SCT band for VLQELNVTV, indicating increased mass due to the NXT glycosylation consensus sequence in the peptide region (**Fig. 2.2B**). This phenomenon is absent in assembly methods which require exogenous introduction of peptide and shows that SCTs undergo biological protein-processing pathways prior to secretion. Thus, SDS-PAGE analysis of this library revealed that SCT expression is dependent on the choice of peptide and backbone template, and produces protein containing post-translational modifications.

Next, we wanted to see if the peptides or design templates played any discernible role in determining the SCT's thermal stability. SCTs which expressed above a yield threshold were subsequently HisTag-purified into PBS buffer at pH 7.4 for thermal shift assays. The measured T_m values were within expected values of reported SCTs compared to native pMHC counterparts, providing a trend of increased stability for the same peptide from wild-type groove (D1 & D6) to open groove (D2 & D7) to thiolated linker/groove (D3, D4, D5, D8, & D9) (**Fig. 2.2C**). SCT thermal stability for each peptide was also higher for H74L variants than wild-type counterparts. Interestingly, for some peptides (such as AIQDLCLAV or FLKANLPLL) in which SCTs expressed only for some templates, we detected two distinct T_m values, the lower of which may indicate an improperly folded SCT species.

2.1.4.2 SCT functional assay against tumor-associated antigen

To validate the functionality of the SCT constructs, we needed to assess SCT binding efficiencies across various designs against known TCRs. For the Wilms Tumor 1 (WT1) peptide (RMFNAPYL) [20], we assessed the binding of this series of six SCTs (D1, D2, and D7 yields were too low for use) against the WT1-specific C4 TCR, which has been characterized by others for reactivity to the peptide *in vivo* (**Fig. 2.4**). Expressed WT1 SCTs were purified and used in binding assays against a 95/5 mixed population of C4 TCR-transduced and MART-1-specific F5 TCR-transduced Jurkat cells. Significant differences in the degree of binding by WT1 SCTs to WT1-specific Jurkat cells was observed. The H74L SCT variants (D6 and D8) displayed the poorest performance, capturing approximately two-fold fewer cells within the gates compared to the wild-type H74 counterparts. The DS-SCT variant

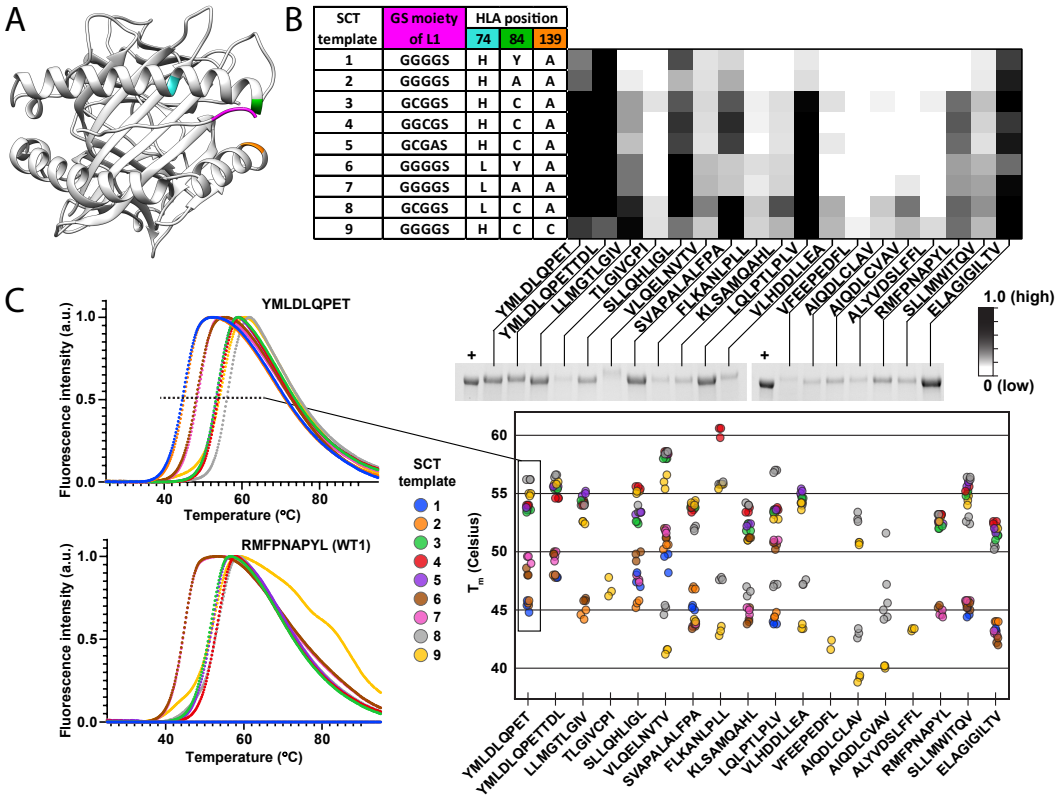


Figure 2.2: SCT design and optimization

A. Axial view of crystal structure of HLA-A*02:01 SCT (RDB ID: 6APN). Highlighted regions of interest: H74 (blue), Y84 (green), A139 (orange), first three amino acids of L1 linker (purple). Peptide is loaded into pocket in N-to-C direction (left-to-right). B. Table: Summary of L1 and HLA amino acid modifications for each of the nine SCT templates tested. Heatmap: Relative expression of each SCT combination, as designated by template (row) and peptide (column). Relative expression is quantified by automated measurement of protein band intensities, as exemplified by reduced SDS-PAGE image of 18 SCTs constructed using design template D9 (bottom). Previously expressed and purified aliquot of WT1 (RMFPNAPYL) SCT is used as positive control (+) for band intensity quantification. C. Thermal shift assay measurements of SCTs. Tm measurements of two peptides designed using the nine SCT templates are depicted (left). Their T_m values are plotted in the scatterplot (right) to show relative changes in stability based on template and peptide. Individual thermal shift curves (left) are representative of a biological triplicate measurement, with all individual T_ms plotted (right).

for WT1 demonstrated the best binding efficiency in the same assay against C4 TCR-transduced Jurkat cells, capturing 97.3% of the WT1-specific cell population. A similar assay was performed for the MART-1 epitope against a pure population of F5 TCR-transduced TCR Jurkat cells to produce similar results. Consequently,

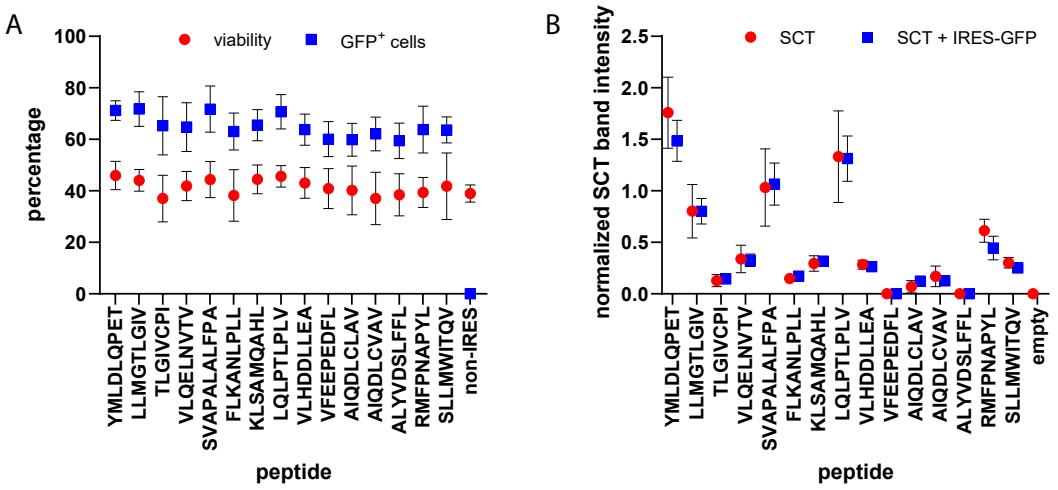


Figure 2.3: SCT transfection quality control
SCT transfection efficiency is uniform and expression is peptide-dependent. A. Expi293 cells transfected with an SCT library consisting of 15 different peptide elements (x-axis) with or without an IRES-GFP indicator are measured for viability and GFP fluorescence after 4 days of transfection. B. Measurement of SCT protein band intensity in SDS-PAGE is performed after transfection using the same plasmid library elements. A negative control ("empty") consists of Expi293 cells transfected with all standardized reagents except SCT plasmid.

the DS-SCT template was used for peptide libraries in future experiments.

2.1.4.3 SCT functional assay against viral antigens

To extend our platform toward use cases in infectious disease, we expressed a small SCT library targeting common epitopes. We initially constructed plasmid templates against 66 total A*02:01 or A*24:02 epitopes commonly reported in the literature (**Fig. 2.5A**). Similar to the previous library, all plasmids displayed peptide-dependent SCT expression (**Fig. 2.5B**). The SCTs were ranked by protein expression, and ten epitopes derived from common viral strains (CMV, EBV, influenza, and rotavirus) from each of two HLA types and resulting in the highest SCT expression were selected for further use in identification of antigen-specific specific T cells. For each of the two HLA types, the ten selected SCTs were pooled together and used to sort for antigen-specific $CD8^+$ T cells from HLA-matched healthy donor PBMCs. One donor was used per HLA type, and the cells were split across ten lines to demonstrate reproducibility. The sorted cells from each line were subsequently expanded over several weeks, and then sorted by individual tetramers from the ten-element SCT pool. The results of this flow cytometry experiment are displayed

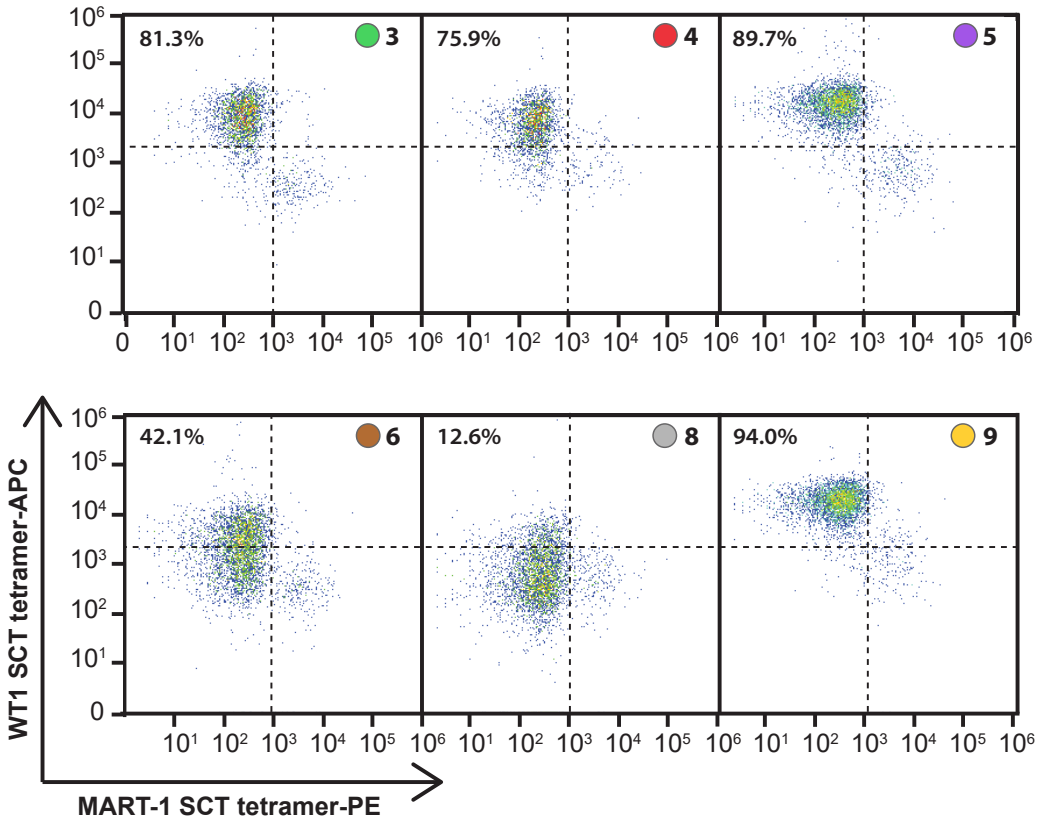


Figure 2.4: Flow cytometry assay to optimize WT1 SCT-TCR capture. WT1 (RMFPNAPYL) SCTs constructed according to each of six template designs in **Figure 2.2B** were paired with a MART-1 (ELAGIGILTV) SCT (D3 template) to identify their cognate TCR-transduced cells in a 95/5 mixture of C4 TCR-transduced primary T cells and MART-1 Jurkat T cells. Number/color at top right of each plot indicates the SCT template used for WT1 SCT in the flow assay. Percentages indicate the proportion of total cell population captured in the WT1 SCT-positive quadrant by each of the six WT1 SCT designs.

in **Figure 2.6A** for the A*02:01 SCTs, and in **Figure 2.6B** for the A*24:02 SCTs. For each peptide, the general consistency in cell capture percentage across each of the ten lines indicates that expansion was uniformly expanding all cells captured by the pooled tetramers. Across each peptide, however, we observed a broad range of capture frequency. This indicates that antigen-specific T cells against particular epitopes might be relatively rare in the captured cells, or that those particular cells may not be as susceptible to expansion, and therefore are detected more rarely. To further demonstrate that the captured cells are not only tetramer-specific, but actually can be activated by biological presentation of the tetramer's associated peptide, we measured IFNγ production after peptide stimulation against the expanded cell lines

or the original PBMCs (**Fig. 2.6C**). In both HLA types, we generally observed agreement between PBMCs and cell lines for IFNγ production after stimulation with each peptide. Scenarios where the PBMCs showed a greater number of IFNγ+ count might indicate that there were other T cells responsive against the epitope that were not detected by our tetramers. These cells might respond to a different epitope conformation not presented our SCT construct, or perhaps contain antigen-responsive TCRs with HLA restrictions not covered by our SCT library. For most peptides which resulted in low capture frequency (e.g. Rota-VP2 & Rota-VP6 of **Fig. 2.6A**; Influenza PB1 216, EBV EBNA3B, & EBV BMLF1 of **Fig. 2.6B**), a very low IFNγ count was obtained after peptide stimulation of either the cell lines or PBMCs. This indicates that the SCT tetramers for these peptides were performing at maximal capacity in terms of capturing all potential antigen-specific T cells. In summary, the individual tetramer binding assay and peptide stimulation results indicate that SCTs can capture cognate TCRs which recognize and are activated by the same epitope bound onto native, surface-bound MHCs.

To further assess functional capacity of the SCTs, we queried the sequences of the CDR3 regions from TCR α and β chains captured by SCT dextramers. A healthy A*02:01 donor was identified to have positive reactivity against the peptide NLVPMVATV, which is derived from human cytomegalovirus (CMV) pp65 protein. We used this SCT element and its folded pMHC counterpart to sort for CMV-specific T cells from the donor PBMCs (**Fig. 2.7A**). 10X single-cell sequencing of the sorted population revealed a similar distribution of antigen-specific clones captured by the two reagents. As seen in the table in **Figure 2.5B**, Levenshtein distances (LD) of the CDR3α and CDR3β chains against a public database were low, indicating high similarity between our detected CMV-specific TCR chains and those previously reported [21]. Two paired clones (red and light orange wedges in (**Fig. 2.7B**) contained CDR3α chains exactly matching literature results (LD = 0) [22, 23]. An additional clone (light green wedge) contained an α/β pair for which both chains have been reported as CMV-specific [22, 24], and was captured by the SCT at a ten-fold higher frequency. These results indicate that SCT tetramers have at least similar flow cytometry performance to the gold standard of folded pMHCs.

2.1.5 Discussion

The technologies introduced herein showcase an extension of SCT expression methods into a high-throughput platform for assembly of hundreds of elements in a

A

A*02:01 viral antigens

ID	peptide	antigen source
1	LLFGYPVYV	HTLV-1 Tax
2	KLVALGINAV	HCV
7	GLCTLVAML	EBV-BLMF1
11	WLSLLVPFV	HBV-SAg
14	YVLDHLIVV	EBV-BRLF1
19	SITEVECFL	Human polyomavirus 2
23	FLLSLGIHL	HBV
24	GILGFVFTL	Flu-M1
31	SLFNTVATL	HIV gag
41	YLLFEVFDV	AdV 11 Hexon
42	LLFEVFDVV	AdV 11 Hexon
43	YVLFEVFDV	AdV 11 Hexon
44	FLDKGTYTL	EBV BALF4
45	YLQQNWWTL	EBV-LMP1-2
46	YLLEMLWRL	EBV-LMP1-1
49	FLYALALLL	EBV-LMP1-2
57	VLEETSVML	CMV-IE1
62	TLNAWVKVV	HIV-gag
70	AIMDKNIIL	Influenza NS1
76	KLIANNTRV	M. tuberculosis Ag85A
84	ALWALPHAA	varicella-zoster IE62 593-601
86	NLVPMVATV	CMV-pp65
87	FMYSDFHFI	Influenza A
88	YLLPGWKL	Rota-VP3
89	NMLSTVLGV	Flu-PB1
90	SLMDPAILTSL	Rota-VP1
91	TLLANVTAV	Rota-VP6
92	FMDILTTCVET	CMV-IE1-2
93	QMWQARLTV	CMV-pp65-2
94	SLISGMWLL	Rota-VP2-1
95	LLNYILKSV	Rota-VP7-1
96	LMNGQQIFL	CMV-pp65-3
97	FLDSEPHLL	Rota-NSP1
98	ALWGPDPAAA	Proinsulin precursor 15-24
99	TLDYKPLSV	EBV BMRF1
100	CLGGLLTMV	EBV-LMP2A

A*24:02 viral antigens

ID	peptide	antigen source
1	TYFNLGNKF	AdV 11 Hexon (37-45)
2	VYSGSIPYL	AdV 11 Hexon (696-704)
3	TYFSLNNKF	AdV 5 Hexon (37-45)
4	DYNFVKQLF	EBV BMLF1 (320-328)
5	TYPVLEEMF	EBV BRLF1 (198-206)
6	RYSIFFDYM	EBV EBNA3A (246-254)
7	TYSAGIVQI	EBV EBNA3B (217-225)
8	IYVLVMLVL	EBV LMP2 (222-230)
9	PYLFWLAAI	EBV LMP2 (131-139)
10	TYGPVFMSL	EBV LMP2 (419-427)
11	TYGPVFMCL	EBV LMP2 (419-427)
12	EYLVSFGVW	HBV core (117-125)
13	KYTSFPWLL	HBV pol (756-764)
14	QYDPVAALF	HCMV pp65 (341-349)
15	EYVLLLFLL	HCV E2 (717-725)
16	PFHCSFHTI	HHV-6B U54 (267-275)
17	RYLRDQQLL	HIV env gp160 (584-592)
18	RYLKDQQLL	HIV env (67-75)
19	RYPLTFGW	HIV nef (134-141)
20	VYDFAFRDL	HPV16 E6 (49-57)
21	FFQFCPLIF	HTLV-1 Env (43788)
22	LFGYPVYVF	HTLV-1 Tax (43819)
23	PYKRIEELL	HTLV-1 Tax (187-195)
24	SFHSLHLLF	HTLV-1 Tax (301-309)
25	YYLEKANKI	Influenza PA (130-138)
26	SYLIRALTL	Influenza PB1 (216-224)
27	RYTKTTYWW	Influenza PB1 (430-438)
28	SYINRTGTF	Influenza PB1 (482-490)
29	RYGFVANF	Influenza PB1 (498-505)
30	TYQWIIRNW	Influenza PB2 (549-557)

B

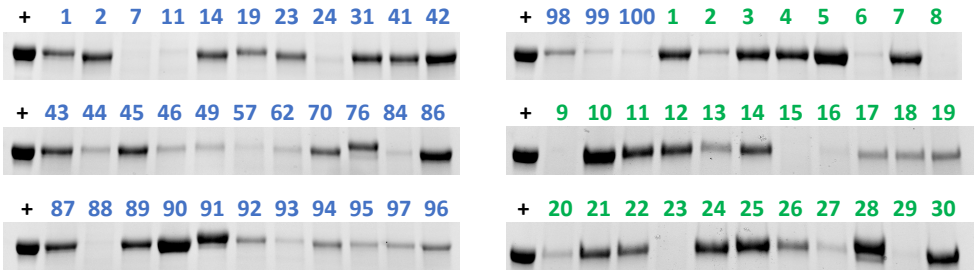
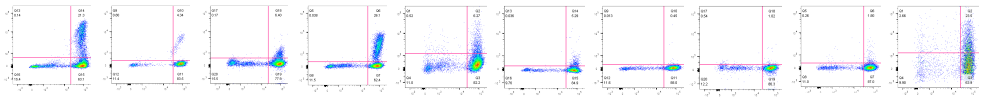


Figure 2.5: Expression of A*02:01 & A*24:02 SCT viral libraries. A. Table of A*02:01 and A*24:02 viral antigens used to construct SCT library. B. Reduced SDS-PAGE of SCT expression for each of the peptide elements highlighted in Table A. Lane number corresponds to peptide ID in Table A. Additional data on expression yield of selected SCTs for downstream stimulation assay can be found in **Table F.2**.

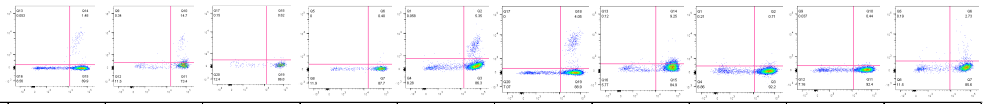
relatively rapid and facile manner. Subsequent characterization of the expressed elements reveals distinctive peptide-dependent trends in SCT expression, thermal

A



peptide	YVLDHLIVV	YLQQNWWTL	AIMDKNIIL	NLVPMVATV	FMYSDFHFI	NMLSTVLGV	SLMDPAILTSL	TLLANVTAV	SLISGMWLL	LMNGQQIFL
pep ID	14	45	70	86	87	89	90	91	94	96
antigen	EBV-BRLF1+	EBV-LMP1-2+	Influenza NS1+	CMV-pp65+	Influenza A+	Flu-PB1+	Rota-VP1+	Rota-VP6+	Rota-VP2-1+	CMV-pp65-3+
Line 1	17.00%	2.60%	4.50%	26.10%	6.30%	0.30%	0.10%	0.30%	1.10%	17.30%
Line 2	11.90%	2.00%	4.20%	12.80%	4.80%	5.30%	0.20%	0.40%	0.90%	8.80%
Line 3	21.30%	4.10%	6.40%	18.50%	3.80%	0.50%	0.20%	0.90%	0.30%	3.50%
Line 4	14.80%	2.40%	4.40%	16.50%	3.30%	0.80%	0.50%	0.50%	1.60%	23.50%
Line 5	13.60%	1.70%	1.90%	14.60%	1.80%	1.00%	0.30%	0.50%	0.80%	7.30%
Line 6	7.20%	1.20%	1.20%	8.20%	1.20%	1.20%	0.30%	0.80%	1.80%	21.00%
Line 7	14.20%	1.20%	0.90%	10.10%	0.80%	0.20%	0.00%	0.20%	0.10%	0.20%
Line 8	6.40%	2.70%	1.10%	10.30%	4.40%	0.70%	0.20%	1.00%	0.50%	10.20%
Line 9	11.60%	4.40%	1.10%	15.10%	0.90%	0.70%	0.10%	0.30%	0.70%	13.60%
Line 10	12.50%	1.10%	2.00%	19.80%	6.60%	0.60%	0.10%	0.70%	0.60%	5.10%

B



peptide	DYNFVKQLF	TYPVLEEMF	TYSAGIVQI	TYGPVFMSL	TYGPVFMCL	QYDPVAALF	YYLEKANKI	SYLIRALTL	SYINRTGTF	TYQWIIRNW
pep ID	4	5	7	10	11	14	25	26	28	30
	EBV BMLF1 320+	EBV BRLF1 198+	EBV EBNA3B+	EBV LMP2-SL+	EBV LMP2-CL+	HCMV pp65+	Influenza PA+	Influenza PB1 216+	Influenza PB1 482+	Influenza PB2 549+
Line 1	0.88	6.1	0	0.22	5.35	4.05	0.55	0.71	0.44	2.31
Line 2	0.56	3.64	0.17	0.32	3.34	0.94	0.04	0.2	0.15	0.68
Line 3	1.48	5.2	0.24	0.35	3.31	2.13	0.05	0.22	0.17	1.78
Line 4	0.54	14.7	0.2	0.4	1.94	1.65	0.09	0.5	0.21	0.81
Line 5	0.93	3.97	0.21	0.12	3.33	0.58	0	0.3	0.02	0.43
Line 6	0.35	3.18	0.3	0	1.46	0.41	0.27	0.23	0.27	0.99
Line 7	0.16	3.15	0.28	0.04	2.51	1.28	0.08	0.19	0.03	2.73
Line 8	0.93	10.1	0.62	0	4.64	1.86	9.25	0.37	0.17	2.15
Line 9	1.11	5.34	0.53	0.2	2.97	1.75	4.29	0.17	0.15	2.26
Line 10	0.72	3.09	0.41	0.07	3.14	1.45	0.18	0.48	0.12	0.85

C

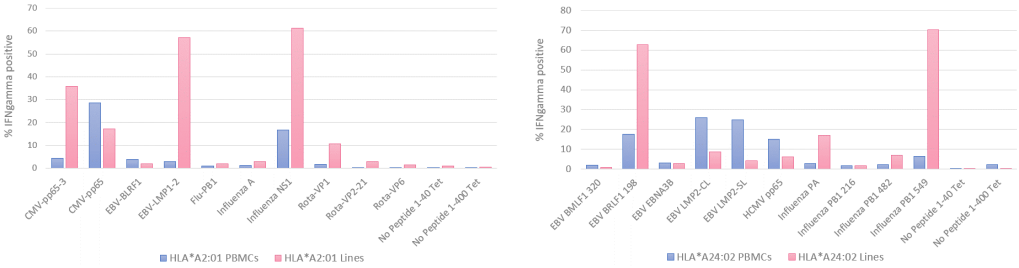
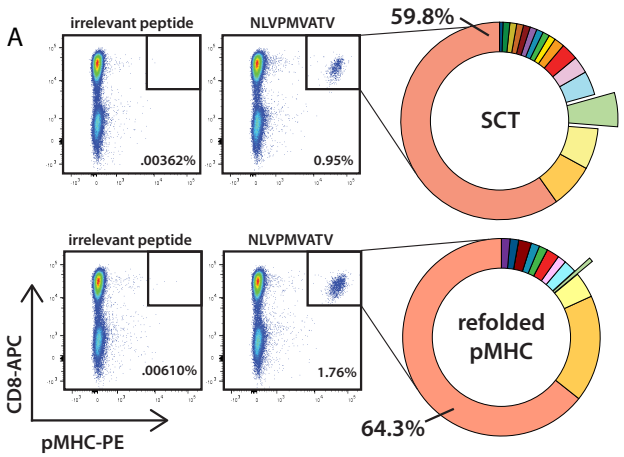


Figure 2.6: A*02:01 & A*24:02 SCTs can capture cognate TCRs that recognize and are activated by the same epitope bound onto native, surface-bound MHCs. A & B. Antigen-specific T cell populations against select SCT elements of **Figure 2.5** were sorted with SCT tetramer pools from healthy donor PBMCs (ten lines each). The sorted cells were expanded and re-stained with individual SCT tetramers, as depicted in the flow cytometry plots. Tables represent percentage of tetramer-positive $CD8^+$ T cells from expanded sample against each SCT element, and flow cytometry plots above each peptide are representative of the tetramer flow assay which generated the highest count among all ten lines [A*02:01 SCTs are depicted in (A), A*24:02 SCTs depicted in (B)]. C. PBMCs or expanded antigen-specific T cell lines from (A) & (B) were expanded and stimulated with individual peptides to demonstrate immunogenic response via IFNγ assay (left: A*02:01 peptides/SCTs/cells, right: A*24:02 peptides/SCTs/cells).

stability, and functionality. For peptides that are difficult to express in the SCT format, it is possible that our expression platform reflects to some extent the natural binding affinities of these peptides for the HLA binding groove, as noted by others [8], and thus may be used as a proxy to validate peptide binding algorithms. The assessment of peptide yield in a systematic manner using the gel analysis methods



B

	CDR3α	LD	CDR3β	LD
	CATVGTASKLTF	5	CASSLWLNEQFF	2
	CARNTGNQFYF	0	CASSPKTGASYGYTF	2
	CVVGYGQFYF	4	CASSFVSFDEQFF	4
	CAGPMKTSYDKVIF	0	CASSSAYYGYTF	0
	CAASRKGSNYKLTF	5	CASSADSYGANVLTF	4
	CAVRWGGKLSF	5	CSVDPGHTGEKLFF	6
	CAEIPNYGGSQGNLIF	0	CASSLVGGRHGYTF	2
	CAESSASKIIF	5	CASSHDPTWGPGNTIYF	6
	CAVRDRWSGGYQKVTF	8	CASSFGQGSSPLHF	4
	CAVRVSGGYNKLIF	5	CASSLETVNTEAFF	3
	CAVTLNNNAGNMLTF	6	CASSSFYDSNEKLFF	4
	CALSPRTQGGSEKLVF	4	CASSLASPGHFTGELFF	4

Figure 2.7: Functional comparison of CMV pMHC reagents. A. Flow cytometry assays of tetramers prepared using SCT or refolded format (left). Pie charts (right) depict the unique clonotypes identified by 10X single-cell sequencing of tetramer-positive cells. B. TCR CDR3α/β gene sequencing analysis of captured CMV-specific T cells. Table contains CDR3α and CDR3β sequences of the twelve most frequently captured clonotypes from SCT tetramer along with their Levenshtein distance to publicly reported CMV-specific clonotypes from VDJdb [16]. The light green clonotype (offset wedge in both pie charts) corresponds to a published pair of CMV-specific CDR3α and CDR3β chains, indicating an exact match (LD=0).

introduced herein should enable one to generate cumulative libraries comparing SCT yield with peptide identity, which in turn should provide a robust and growing dataset for training machine learning algorithms.

Aside from changing the identity of the loaded peptide in each SCT element, our initial library also demonstrated that SCT expression and function can be compatible with various mutations inserted into L1 and HLA domains. The spectrum of binding efficiencies in our flow assay against the C4 TCR indicates that perhaps each WT1 SCT variant is displaying a different epitope conformation to the TCR as a result of the mutations. Whether this phenomenon is peptide-dependent, HLA-dependent, or TCR-dependent remains to be seen, as there may not exist a single L1/HLA template solution that can capture the "best" TCRs for any given peptide. Future work in this direction will consist of performing similar assays against other WT1-specific TCRs, utilizing different peptide-loaded SCTs in a similar manner against cognate TCRs, and performing such assays on peptide/TCR pairings for different HLA alleles.

Prior works have implied that accessibility of the peptide to the terminal ends of the HLA binding groove plays a crucial role in stabilizing the MHC. Our identification

of the DS-SCT template as being among the best of the tested templates for binding to the C4 TCR, therefore, was an interesting result, as we speculated that a dithiolated closure of the groove at the C-terminus coupled with the presence of a peptide linker would sterically hinder the peptide from being presented in a manner capable of binding to cognate TCRs. To our surprise, the DS-SCT template did not negatively impact binding functionality of the WT1 SCT against the C4 TCR, and instead was among the best designs in terms of binding efficiency. Further elucidation of this phenomenon is currently being undertaken to obtain crystal structures of the WT1 DS-SCT. Additionally, crystal structures of the YML series of peptides (from 9mer to 14mer length) have been obtained for various template designs from this work. These structures, to be published in a separate manuscript[1], should elucidate the impact that peptide length and L1 identity have on the peptide-binding groove interactions at the C-terminal pocket region.

Following template optimizations, our use of a viral SCT library to capture functional TCRs, verified either by means of peptide stimulation or by TCR sequencing, indicates that the epitope conformations presented by the SCTs to the TCR interface are biologically relevant. The two libraries discussed in this section showcase the versatility of pMHC production and applications available under our SCT production platform, enabling modular analysis of different template mutations, different peptides, or different HLA alleles. The characterization profiles of these library elements should provide a foundation for further expansion into other infectious disease or cancer contexts.

2.1.6 Future Directions

2.1.6.1 CD8-inhibiting mutations

While pMHC reagents have enabled researchers to capture T cells at the level of antigen specificity, studying T cells on this basis fulfills just one of the numerous requirements necessary to better understand the immunogenic potential behind their TCRs. Numerous works have assessed the binding affinity of antigen-specific TCRs to their target pMHCs, showing that there is a broad affinity range which may drastically impact downstream immunogenic potential of each T cell [25]. It would be desirable then, to also distinguish all TCRs specific to one antigen on the basis of their binding affinity to the epitope. Such a function has been built

[1] Kathryn AK Finton*, Peter Rupert*, William Chour, Matthew Buerger, Ana Dinca, Erica Lovelace, Roland K Strong. *Manuscript in preparation.*

into existing pMHC constructs by others via the introduction of HLA mutations which inhibit pMHC binding interactions with the CD8 co-receptor of T cells [26–30]. The resultant loss of this secondary interaction during pMHC-TCR coupling is hypothesized to weaken the overall affinity and skew successful binding interactions toward those TCRs with high affinity to pMHC. Therefore, pMHC tetramer flow cytometry reagents with this mutation can be utilized as a filter to remove low-affinity TCRs from the sorted, antigen-specific T population.

The engineering of pMHCs with this capacity has been conducted throughout the past two decades. Luescher et al. illustrated that a single HLA mutation (D227K) has the capacity to reduce binding to low affinity TCRs [28]. Pittet *et al*. improved this development by implementing a pair of mutations (D227K and T228A) which completely abolished binding to low affinity TCRs [30]. Bodinier *et al*. later demonstrated that A245V can also perform a similar function [26]. Taken together, these works demonstrate that there are variable degrees to which the CD8-pMHC interaction can be tuned, such that the resultant T cell population pool captured might be different in terms of cytotoxic potential. Because these works occurred prior to the onset of single-cell sequencing technologies that democratized TCR sequencing and phenotypic characterization, their output metrics primarily relied on traditional lysis activation assays. By combining these CD8-inhibiting mutations with SCTs and a modern single-cell sequencing protocols, we are now in an era where TCR sequences of high and low affinity may be easily distinguished and potentially used to define downstream functional behavior.

The exploration of CD8-inhibiting mutations in our lab is still in its infancy; much of the SCT work has been done, but was conducted prior to the integration of TCR sequencing technologies in our lab such that we did not have a chance to properly generate output TCRs representative of high- and low-affinity populations. Herein, I briefly demonstrate the impact of CD8-inhibiting mutations on SCT function in three experimental contexts.

Impact of D227K & T228A mutations on WT1 SCTs

The D227K mutation has been reported as a lone mutation capable of blocking CD8 interaction with pMHCs [28]. This mutation's CD8 inhibition has also been reported to be improved when paired together with the T228A mutation [27]. We explore the function of these mutations when implemented into our SCT platform to generate a small library of A*02:01 SCT variants loaded with the WT1 peptide (RMFPNAPYL). We have previously shown the WT1 SCTs to be capable of expres-

sion only for certain template variations (**Fig. 2.2B & 2.8A**). The plasmid templates which successfully led to expression were subsequently mutated to introduce either D227K or D227K+T228A together across all templates . Standard transfection of this library, encompassing seven core templates across three CD8-interaction variants [wild-type (no HLA mutation), D227K, or D227K+T228A)], was performed over four days, and SDS-PAGE was conducted to characterize the yield (**Fig. 2.8A**; NOTE: Lane 8's cells were found to be low viability, so no transfection occurred, leading to no detectable SCT output for this plasmid). From Dr. Philip Greenberg's lab at Fred Hutch, we obtained T cells transduced with a WT1-specific A*02:01-restricted TCR (CD4ba). This TCR was transduced either into a $CD8^+$ or a $CD4^+$ cell line, enabling us to assess the impact of various HLA mutations on their capacity to interact with the CD8 co-receptor.

Transfection of these SCTs led to template-dependent yields across each HLA mutation type. Templates D6 and D7, which do not contain a cysteine-modified L1 linker and which also do not have the pocket-stabilizing dithiol mutation seen in D9, gave consistently lower yields relative to other templates. We also observed a double-banding pattern of SCTs in a non-reduced SDS-PAGE environment. Because this pattern was only observed in templates which implemented a cysteine linker, we suspect this is the cause of double-banding, but do not suspect that it has any impact on function (see top left plot of **Fig. 2.8B**). We currently do not understand the structural manifestations of this double-banding pattern; dimerization is ruled out as these bands would be expected to be significantly upward-shifted relative to the positive control SCT band.

The tetramer binding assays against TCR-transduced cell lines showed distinctive binding patterns for each HLA variant. The wild-type SCTs, when used to stain TCR-transduced $CD8^+$ T cells, displayed variable degrees of successful binding to the cognate TCR (**Fig. 2.8B & 2.4**). Among this wild-type subset, D3 and D9 templates showed remarkably high binding efficiency, capturing at least 90% of all cells. When either D227K or D227K_T228A mutations were introduced, we observed essentially complete abolishment of TCR binding across all SCT variants (**Fig. 2.8B**, middle-left and bottom-left plots) except for D227K_T228A D9 variant, which still showed some degree of binding capability. We also assessed TCR-transduced $CD4^+$ T cells to see if the absence of CD8 on these T cells might still result in binding by any SCT variant. If binding does occur, one interpretation is that the particular SCT variant paired with the C4 TCR results in a high-affinity

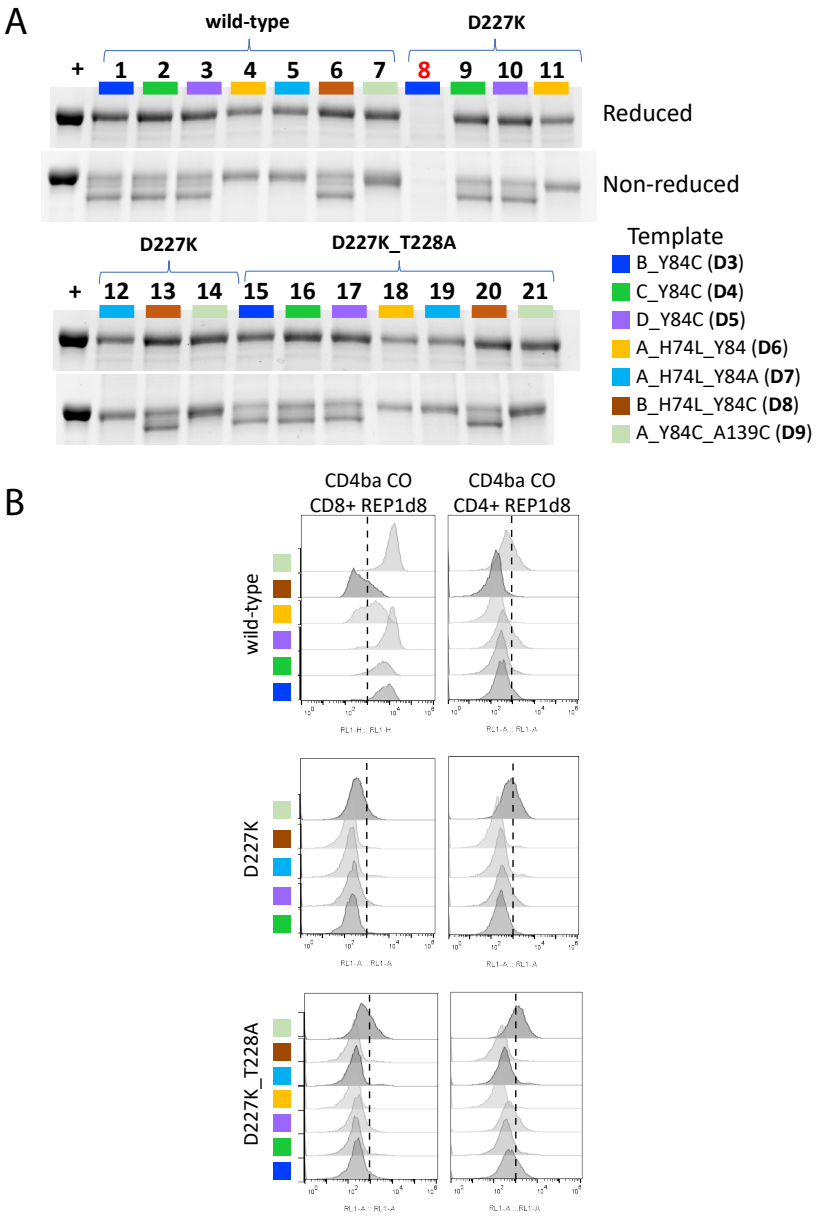


Figure 2.8: D227K and T228A mutations inhibit CD8 interaction with pMHCs. A. SDS-PAGE of A*02:01 SCTs expressed with the WT1 epitope (RMFPNAPYL) for various templates (**Fig. 2.2**). Labels above each bracket indicate the CD8-inhibiting mutation applied to each set of SCTs ("wild-type" refers to no mutation against CD8 interaction). +, purified WT1 SCT. B. Flow cytometry intensity plots of tetramer binding interaction between expressed WT1 SCTs and TCR-transduced T cells. Y-axis denotes SCT type (colors correspond with legend in (A)). Binding experiments were performed with $CD8^+$ T cells (left column) and $CD4^+$ T cells (right column). In each plot, the dashed line indicates the positive signal threshold of 10^3 mean fluorescence intensity units (right of line = positive).

interaction exclusive of CD8 co-binding. As seen in the top-row plots of **Figure 2.8B**, the wild-type SCTs showed a drastic reduction in binding against $CD4^+$ T cells compared to binding against $CD8^+$ T cells, indicating that most of these SCT variants relied on the CD8 co-receptor to facilitate pMHC-TCR affinity. For all SCT variants which contained a CD8-inhibiting mutation, binding efficiencies against $CD8^+$ or $CD4^+$ T cells were virtually unchanged.

Across all cell lines and HLA mutations, the D9 SCT template appeared to be the best binder in terms of signal retention beyond the 10^3 MFI threshold. Indeed, as seen across all cases where CD8 interaction is removed (either with introduction of CD8-inhibitory mutation or substitution of CD8 with CD4), the D9 tetramer was capable of still generating some signal beyond noise. While we have not completely ruled out the possibility that this could be due to WT1 D9 SCTs non-specifically binding all cells by testing binding to an off-target TCR, we have been able to confirm that D9 SCTs for other peptides and other HLAs do not non-specifically bind (**Fig. 4.11A,D**). Thus, we find the results of **Figure 2.8** to indicate that the D9 template is superior to other designs in terms of epitope presentation for enhanced affinity against TCR.

Impact of A245V mutation on neoantigen-encoded SCTs

Another HLA mutation, A245V, has been previously demonstrated to reduce CD8 interaction with pMHCs during TCR activation [26]. Herein, we implemented such a mutation for a private neoantigen-encoded library of SCTs, showing its capacity to significantly reduce background noise from binding of non-specific T cells. In our collaborative work with the TESLA consortium (**Section 4.3.3.2**), we pursued a parallel direction of generating A*03:01 SCTs (D3 template) with the A245V mutation for A*03:01-restricted peptides of a melanoma patient. The expression results of this library (data not shown) matched in terms of expressed protein band intensity per peptide-encoded SCT against its wild-type (no A245V mutation) variant, indicating that the mutation had no significant impact on protein expression capabilities of transfected cells. Subsequently, the biotinylated, purified SCTs were tetramerized for use against PBMC samples from the melanoma patient to detect antigen-specific T cells. The tetramers were utilized in groups of three to assess for three antigen specificities per flow experiment, where one antigen specificity was tetramerized with streptavidin-PE while the other two specificities were tetramerized with streptavidin-APC. In this manner, detection of double-positive fluorescence signal would indicate non-specific cross-binding of SCT tetramers. Cells which

exhibit significant PE signal but not APC would be truly specific T cells.

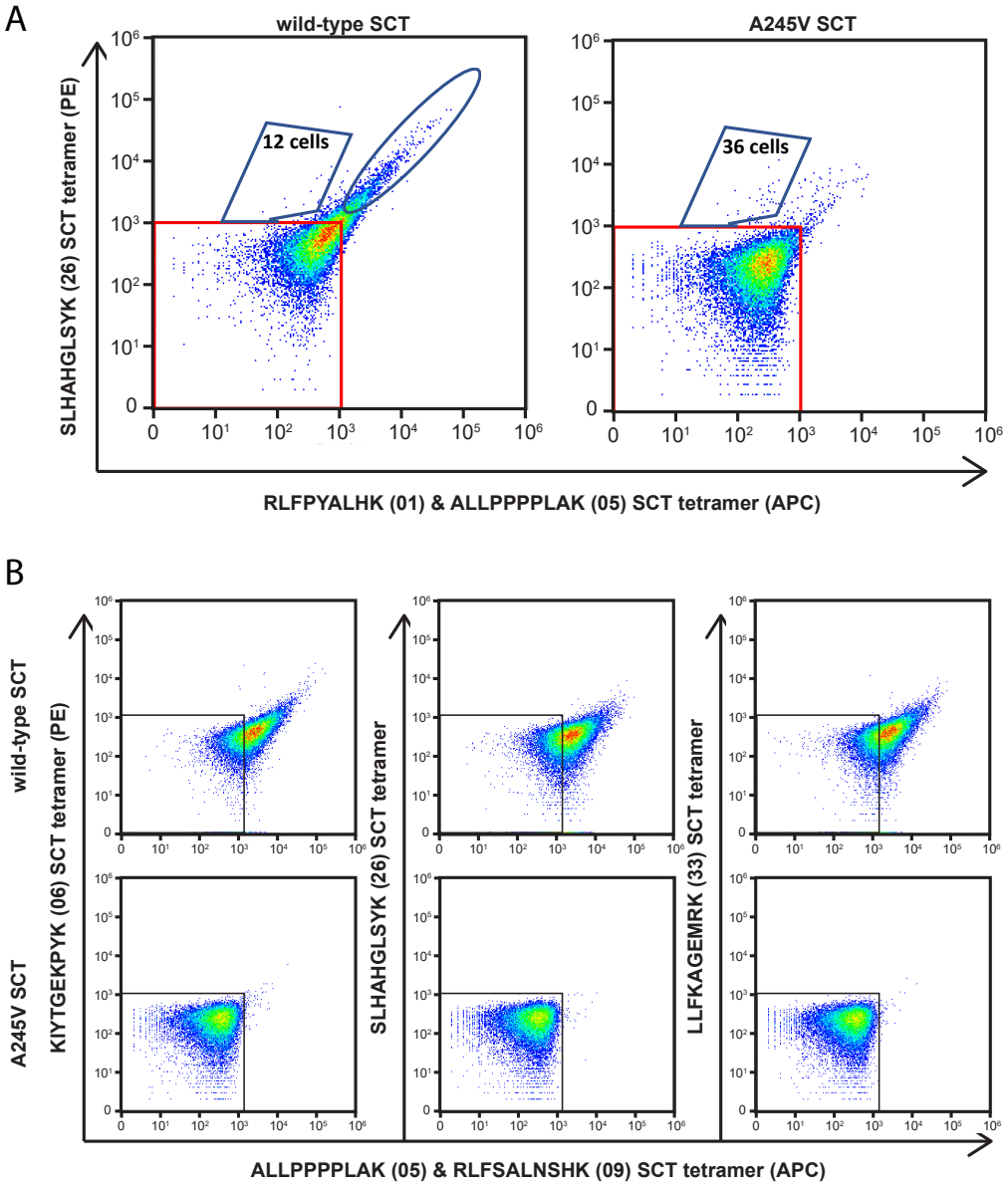


Figure 2.9: A245V mutation inhibit CD8 interaction with pMHCs loaded with neoantigens.

A. Flow cytometry profiles of neoantigen-loaded A*03:01 SCT tetramers incubated with PBMCs from a melanoma patient. Lower left quadrant (red) indicates non-binding. B. Experiment in (A) was expanded to cover various other combinations of SCT tetramers. Lower left quadrant (black) indicates non-binding.

When this experiment was performed for a set of three SCTs without using the A245V mutation (**Fig. 2.9A**), we observed significant cross-binding, strongly skew-

ing the tetramer-bound populations into a diagonal on the flow plot. However, when the A245V mutation was implemented for SCTs of the same antigen specificities, this cross-bound population was essentially removed. Furthermore, the counts of SLHAHGLSYK-specific T cells based on PE-specific signal (polygonally bound region) increased. Our interpretation here is that without the A245V mutation, non-specifically bound cells will overwhelm the tetramer-positive population, in essence masking the true positive reads from being properly detected. Once the A245V mutation is inserted to inhibit CD8 interaction, however, some of the truly PE-specific population (found in the oval-bound region of left in **Fig. 2.9A**) will only bind to the peptide-associated PE tetramer, thus increasing PE-specific binding counts.

This experimental setup was repeated three additional times (**Fig. 2.9B**), with each case having a unique arrangement of one SCT tetramerized with streptavidin-PE and two SCTs tetramerized with streptavidin-APC. In all cases, when comparing binding results of wild-type SCTs versus A245V SCTs, there are two major observations. First, the overall signal intensity decreases such that most cells give below 10^3 MFI (our cutoff threshold for establishing specific binding). Second, for A245V SCT tetramer staining, cells which generate a signal beyond 10^3 MFI tend to only do so in one axis, indicating that perhaps they only have specificity to one of the three SCT tetramers assessed. This is a strong contrast to what is observed with the wild-type SCTs, where again, similar to Figure (**Fig. 2.9A**), there is clearly a strong inclination for non-specific binding events to occur to generate a skewed diagonal.

Impact of A245V mutation on viral SCTs

Similar to the SCT library production design for the A*03:01 neoantigens, an A*02:01 SCT library (D8 design) containing the A245V mutation was generated to encode various A*02:01 viral and bacterial peptides. We selected four of these elements to be utilized in tetramer binding assays against PBMCs obtained from a healthy A*02:01 donor sample, where for each assay, one of three viral peptide SCT elements (tetramerized with streptavidin-PE) was mixed with the bacterial SCT element (tetramerized with streptavidin-APC) prior to staining. The viral SCTs encode peptides derived from EBV, CMV, and influenza viruses which have been reported in the literature to have cognate TCRs in virtually all A*02:01 individuals, whereas the bacterial SCT encodes a peptide from *M. tuberculosis*, for which we do not expect there to be much reactivity given the low prevalence of this disease. Therefore, the former elements serve essentially as positive controls in the staining

assay, while the latter element serves as a negative control.

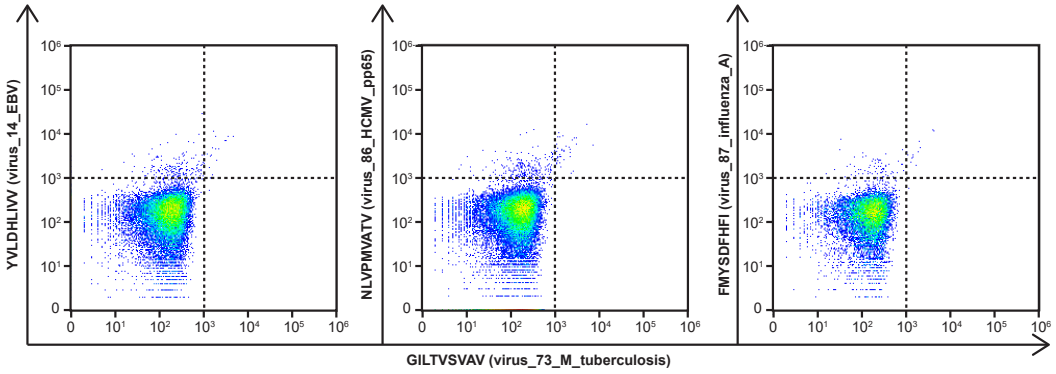


Figure 2.10: A245V SCTs loaded with viral antigens also give reduced non-specific binding profile.

PBMCs from an A*02:01-positive healthy donor were incubated with SCT tetramers encoding positive control peptides (from EBV, CMV, and influenza) and negative control peptide (from M. Tuberculosis).

As seen in (**Fig. 2.10**), the flow cytometry results for these A245V SCTs displayed a remarkably similar profile to that of the A*03:01 A245V SCTs (**Fig. 2.9B**), where most of the staining signal was contained within the lower left quadrant and no diagonal skew was present. This is highly suggestive of a strong reduction in non-specific binding compared to wild-type SCTs. Furthermore, for cells which do generate a positive signal in this experiment, we find that in all three assays, this was only observed for tetramer-PE, indicating specific binding only by tetramers designed to present a common viral epitope. The lack of any binding by the *M. tuberculosis* antigen SCT tetramers is in alignment with our expectations for negative control results.

Conclusion

Taken altogether, the data presented here in this section is a snapshot of our current progress in the design of SCTs tailored to avoid interaction with CD8. Our initial work with the WT1 SCTs against TCR-transduced cell lines provided a completely controlled experiment where we could alter SCT, TCR, and CD8 presence. This enabled us to demonstrate that D227K and D227K+T228A can completely abolish tetramer binding, at least in the context of our selected peptide and TCR system. With the two A245V experiments, which were used in human PBMCs across two HLA types, we showed that a similar reduction in TCR binding can occur due to removal of low-affinity TCR interactions that relied on CD8 co-binding. The tetramer binding profiles with the A245V mutation across A*02:01 and A*03:01

libraries were remarkably similar, highlighting the strong reduction in non-specific binding.

There appears to be a sort of "catch-22" when it comes to the incorporation of these CD8-inhibiting mutations. A very obvious advantage when using these SCT reagents is their capability to significantly reduce noise (**Fig. 2.9 & 2.10**). Unfortunately, the inhibition of CD8 binding, while conducive to elimination of non-specific events, may be unexpectedly counter-productive in some cases, such that we may end up capturing very little positive signal whatsoever. As seen with the case of the WT1 SCTs, CD8-inhibiting mutations blocked most SCTs from binding to the C4 TCR cell line. Given that this TCR has been shown to be cytotoxic and has clinical potential for therapy, TCR discovery using CD8-inhibiting SCTs would have completely missed this clonotype.

What is missing from our dataset so far then, is the demonstration of a scenario where high-affinity TCRs are captured by these reagents and properly sequenced and characterized to show antigen specificity. **Figures 2.9 & 2.10** show early steps toward this objective, where we have clearly captured specific binding signal using A245V SCTs. A follow-up to these results using the TCR sequencing, cloning, and validation methods discussed later (**Sections 2.3 & 3.1.3.4**) should be pursued. It would enable us to ask deeper biological questions and experimentally assess the functional difference between TCRs of low or high affinity against particular pMHCs.

2.1.6.2 Analysis of the impact of peptide length on SCT expression

Introduction & Methods

During our initial analysis of SCT expression across various templates (**Fig. 2.2**), we were surprised to see that the 12mer YML peptide was capable of expression. The high-yield expression of SCTs for both the 9mer and 12mer peptides prompted us to examine whether additional YML peptides of various length extensions toward the C terminus might also enable SCT expression. This peptide sequence of the HPV E7 protein (YMLDLQPETTDLYC) was adapted into lengths of 8 to 14 amino acids. Primers encoding these peptides were utilized in inverse PCR reactions to insert these codons into the peptide region of A*02:01 SCT templates (eight designs total). The plasmids were transfected into Expi293 cells, incubated for four days,

and the SCT expression was measured by SDS-PAGE analysis. The SCTs were further assessed for thermal stability by performing thermal shift assays.

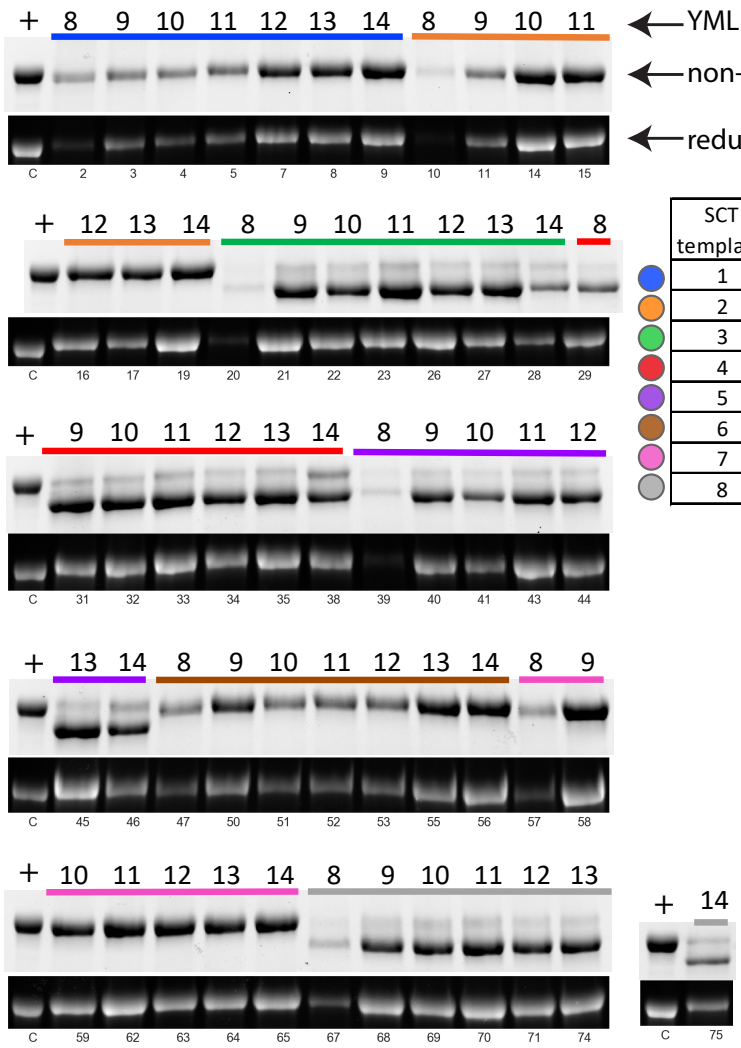
Results & Discussion

All SCTs containing the YML 8mer produced the weakest expression in general (**Fig. 2.11**). The highest expression yields across all design templates for the 8mer peptide were for those which used template designs without a cysteine linker (D1, D2, D6, D7). Our current hypothesis is that the cysteine linkers force the 8mer into a configuration within the HLA's binding pocket that is not amenable to stabilization and expression. Amongst the 8mer SCTs which had high yield, we observed that the D1 variant produced higher expression than D2, indicating that the Y84A mutation might be slightly worse at stabilization. The expression difference between these two templates is moderately reduced when the H74L mutation is added (D6 vs D7), where now the yields appear comparable.

SCTs with 9mer to 13mer peptides showed consistent expression levels across all templates. 9mer, 10mer, 11mer SCTs have relatively worse expression than other peptides for D1, while the 9mer is relatively worse for D2. Similar to the 8mer, the 14mer seems to experience significantly reduced expression when using a cysteine linker template. For D3, D4, D5, & D8, the 14mer shows significantly lower expression compared to 9-13mers of the same templates. However, for non-cysteine linker templates (D1, D2, D6, & D7), the 14mer ranks among the high-expressing SCTs compared to peptides of the same template. Our interpretation here is that the 14mer may be constrained by the presence of a cysteine linker. By forcing all amino acids upstream of the cysteine link to fit in the binding groove ahead of the C-terminal pocket enclosure, there is a high likelihood that the steric hindrance introduced will not enable the epitope to remain as stably bound to the groove. This issue becomes more apparent as the peptide length increases, explaining the expression differences of the 14mer versus other lengths.

All templates with cysteine linker display a double-banding pattern in non-reduced SDS-PAGE. This is a template-dependent phenomenon, similar to what we previously observed when expressing WT1 SCTs.

To further assess the stability of these SCTs, melting temperatures of the proteins were performed. As seen in **Figure 2.12**, T_m values across the peptide series showed dependencies on SCT template and peptide length. Across all the peptides, the most stable constructs consisted of templates which made use of a cysteine linker



SCT template	GS moiety of L1	HLA position 74	84	139
1	GGGGS	H	Y	A
2	GGGGS	H	A	A
3	GCGGS	H	C	A
4	GGCGS	H	C	A
5	GCGAS	H	C	A
6	GGGGS	L	Y	A
7	GGGGS	L	A	A
8	GCGGS	L	C	A

Figure 2.11: SCT platform can generate library matrices to analyze peptide length and template.
SDS-PAGE analysis of transfected SCT plasmids modified with combinations of various peptide lengths (8-14mer: from YMLDLQPE to YMLDLQPETTDLYC) and various template designs. +, purified WT1 SCT.

template. Templates without a cysteine mutation experience a drastic reduction in melting temperature, dropping by approximately 6 °C.

The H74L mutation was also another significant factor which increased protein stability. When comparing templates which are identical except for the presence of this mutation (D1 vs. D6, D2 vs. D7, and D3 vs. D8), the template with the H74L mutation is typically more stable. When we examined T_m values on the basis of peptide length, there is a clear drop in stability for 8mer SCTs. Beyond this length, all SCTs experience substantial improvement in stability, but there is

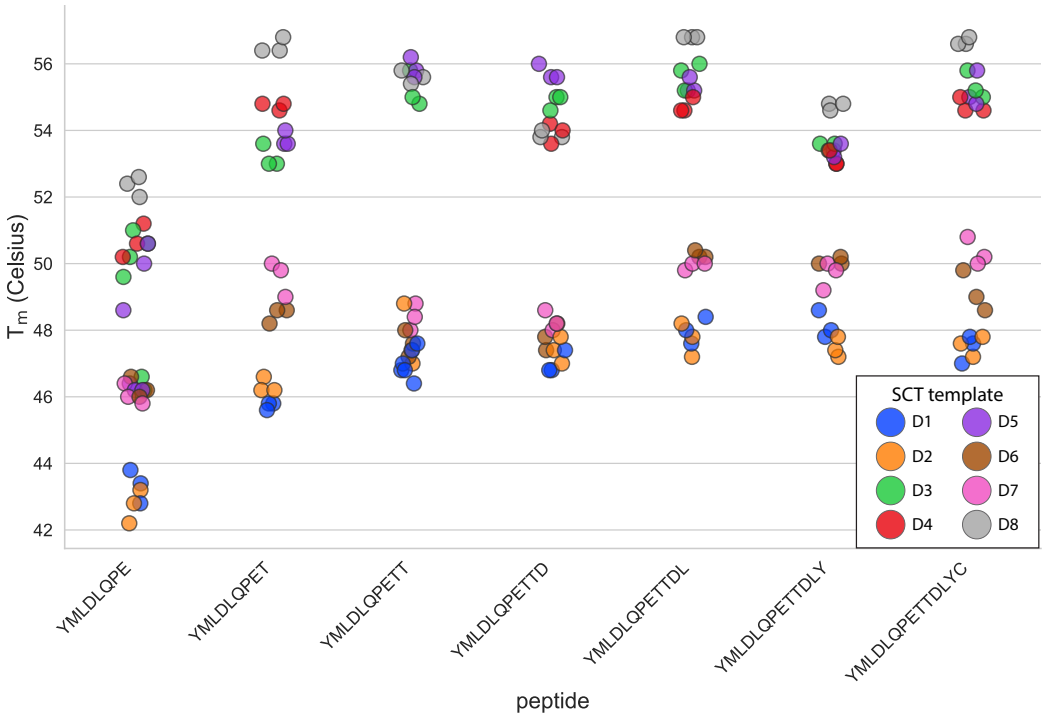


Figure 2.12: Thermal shift assay measurements of YML series SCTs. Tm values of YML SCTs are shown in the scatterplot, color-coded by design template and arranged left-to-right by peptide length. Biological triplicate measurements were performed for each peptide/template SCT combination. NOTE: One plasmid failed to express during transfection due to human error (D4 SCT loaded with 10mer), so no measurements could be performed for that sample.

no clear T_m difference per template for 9mer to 14mer, with the exception of the 9mers, for which D1 and D2 appear to afford slightly less stability than what would be expected of their counterparts for 10mers or longer. The most stable template across all templates was consistently the D8 template. The H74L mutation most likely explains the improved stabilization, given that D3 SCTs (which do not contain the H74L mutation) are always less stable than D8.

In order to more deeply investigate the relationship between peptide length and SCT template, the Strong lab (Fred Hutch) have obtained crystal structures to determine how YML peptides of various lengths are structurally configured in the binding groove of each SCT.[2] The following is a series of key observations highlighted from the preliminary crystal structure data (**Fig. 2.13**):

- The first four residues of the N-terminus of all lengths of YML peptides are in

[2]Kathryn AK Finton*, Peter Rupert*, William Chour, Matthew Buerger, Ana Dinca, Erica Lovelace, Roland K Strong. *Manuscript in preparation.*

	Peptide	SCT Template	L1	74	84
	YMLDLQPET	2	GGGG	H	A
	YMLDLQPET	8	GCGG	L	C
	YMLDLQPETTDL	2	GGGG	H	A
	YMLDLQPETTDL	8	GCGG	L	C
	YMLDLQPETTDLYC	2	GGGG	H	A
	YMLDLQPETTDLYC	8	GCGG	L	C

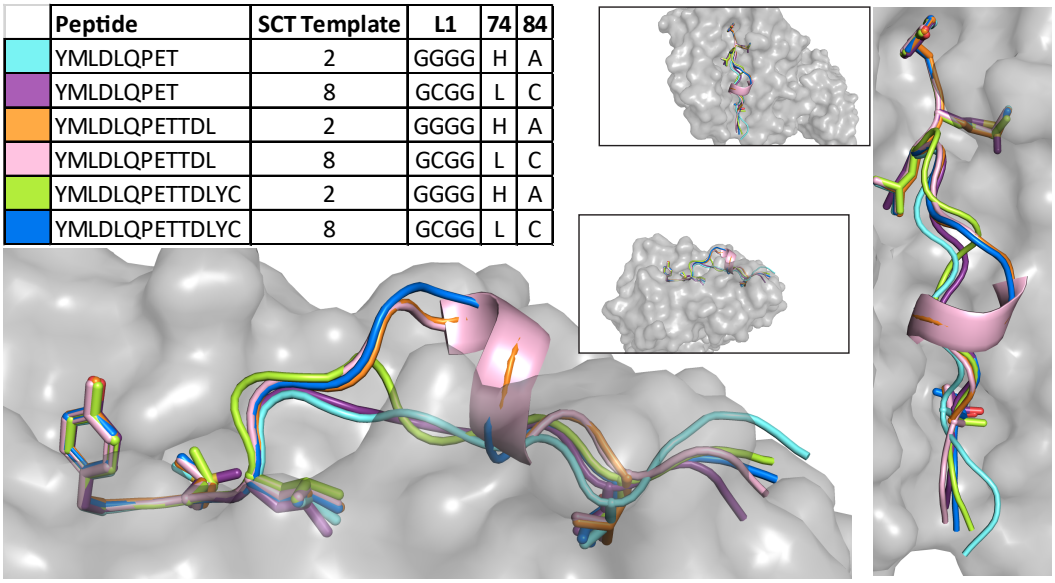


Figure 2.13: Superposition of SCT crystal structures.
All structures were superposed in Pymol; only one SCT is shown as a surface while the superposed peptides are all shown and are color coded. Side view of SCT displayed at bottom left, top view of SCT displayed on right. Residues 1, 2, 3, and the C-term pocket residue are shown in licorice stick. Other residues are only shown as main chain for clarity. Peptide configurations are colored according to the SCT template utilized (table, top left).

the same conformation with 2M binding in the P2 pocket. the conformation of the peptide from residues 5 through 9, 12, or 14 then change both depending on length and, in the case of the 14mer, on choice of Y84 mutation.

- The YML 9mer Y84A SCT structure was solved previously and uses the murine H2K alpha 3 domain with human alpha 1/2 and B2M. Despite minor structural differences in the binding cleft due to the chimeric SCT, the YML 9mer from the Y84A and Y84C constructs are virtually identical.

- The YML 12mer peptide protrudes outwards from the center of binding cleft compared to the 9mer. This changes the residue at the C-terminus anchor from 9T in the 9mer to 12L in the 12mer.

- The addition of the Y84C mutation does not affect the position of the 12mer peptide in the binding cleft or alter the structure of the peptide or HLA Heavy chain

- The addition of the Y84C mutation does affect the 14mer. With the Y84C mutation, the Cys present at position 14 of the peptide (not the Cys in the linker)

forms a disulfide with 84C. This results in an overall peptide conformation most like the 12mer with 12L as the C-terminus anchor and the center of the peptide bulging up in the middle with only a slightly different conformation from the 12mer.

- The 14mer peptide in the Y84A construct extrudes out the binding cleft and lays somewhere in between the 9 and 12mer in the center. The C-terminus anchor position is occupied by 10T in this case. However, the electron density for the center part of the peptide is poorly defined indicating that more than one conformation may be sampled.

- The 14mer YML Y84A had almost identical T_m as the other Y84A YML Xmers indicating that in this context, with the groove opening mutation, they all bind with similar affinity. Likewise, all YML Xmers in the Y84C platform had similar T_ms, gaining stability over the Y84A platform.

We are currently obtaining the structure of the 14mer in the DS-SCT (D9 of table in **Fig. 2.2B**) platform because we would like to see what happens when the groove is closed in the presence of a longer peptide that appears to want its C-term anchor at position 10 instead of 14, and whether the engineered DS-SCT disulfide will form, or, if like in the case of Y84C, the disulfide will form with the peptide Cys. This may help answer how the YML 14mer peptide is presented in a more native context and, if a Cys is present at the C-term of a peptide, will a DS-SCT work (because the other Y84C variants do not).

Conclusion

In summary, we demonstrated in these experiments the capacity to utilize our SCT platform for the generation of large SCT libraries, tuning peptide length and template design. In terms of SCT expression, the yield was found to be strongly dependent upon these two variables. For peptides of length 9-13 aa, templates containing a cysteine linker often did not reduce expression yield, possibly because the tethering impact of the dithiol linkage between linker and binding groove did not significantly deter the peptide from arriving at an optimal, stabilizing conformation within the peptide groove. For 8mers and 14mers, however, this dithiol linkage proved to be too cumbersome, either reducing a short peptide's capacity to properly fit in the groove (e.g. 8mer D3) or possibly generating too much steric burden for a long peptide within the binding groove (e.g. 14mer D3). Finally, the H74L mutation appeared to

provide an additional stabilizing impact on all template designs regardless of peptide identity.

While the initial goal during these experiments was simply to assess expression patterns as a function of SCT template and peptide length, the fact that some of these epitopes have been previously reported to be capable of eliciting T cell cytotoxicity enables us to explore immunogenicity with these reagents. TCRs captured with this library can be assessed for cross-reactivity to SCTs of various lengths or of different templates. Positive responses across multiple reagents per TCR would indicate shared epitope presentation by the SCTs, and might shed light on optimal SCT design for TCR capture.

2.1.6.3 Generation of soluble single-chain dimers from cleavable SCTs

Saini *et al*. in 2019 reported the development of a peptide-free pMHC construct [17]. Typically, MHCs can only be stabilized and folded together in the presence of a peptide; otherwise, the two subunits will not stably bind together. This requirement has posed a challenge for applications such as high-throughput pMHC production, which have traditionally necessitated individual three component refolding, and more recently have moved onto UV exchange as the main standard for library generation. By introducing two HLA mutations (Y84C and A139C), the authors were able to refold bacterially expressed MHC in the absence of a peptide, therefore generating peptide-free stabilized MHC constructs. They proceeded to demonstrate that these reagents can essentially be used as off-the-shelf reagents, in which the user can mix the empty MHCs with any peptide desired to generate functional pMHCs. This alternative approach to generate high-throughput libraries was applied in their concomitant paper, in which they analyzed binding affinities of an alanine scan library of the SIINFEKL peptide, demonstrating the potential of their dual mutation system [16]. We were initially interested in seeing whether this pair of mutations could also be used in the context of SCTs to provide additional stability to the MHC and therefore improve expression yield or protein function. As shown in **Figure 2.4**, the inclusion of these mutations (D9 template, a.k.a. DS-SCT) in the presence of $(GGGGS)_3$ L1 linker for the WT1 peptide resulted in the best tetramer binding performance against the C4 TCR.

One potentially useful innovation we would like to test using this DS-SCT template is to introduce a TEV protease cleavage site into the L1 linker. As previously described,

Moritz *et al*. were able to produce a large amount of the refolded pMHC using bacterial expression [16], but ideally, one would aim to study a human-origin protein utilizing mammalian-derived post-translational modifications, especially when the protein expression system has the added capability of generating already-folded product in high yield. The DS-SCTs we generated fulfill these criteria, but require the presence of a peptide upstream of L1 to stabilize the construct before the SCT can be secreted into media. We take advantage of this requirement to maximize protein yield of the DS-SCT by selecting a peptide which generally gives high yield (MART-1). Then, to create a peptide-free DS-SCD (single-chain dimer), we introduce the TEV protease cleavage site into the L1 linker upstream of each repeating GGGGS unit (**Fig. 2.14**). These MART-1-loaded TEV DS-SCTs (termed cleavable SCTs, or cSCTs) have been shown to express at relatively similar yields to the unmodified MART-1 DS-SCT. We have also shown that SCTs can easily be transfected at large volumes and purified for larger yields via size-exclusion FPLC (**Fig. 2.15**), and anticipate that this can also be performed for cSCTs should the need arise for downstream experiments.

At this point, we have so far demonstrated expression of cSCTs. Future work to demonstrate their functionality will consist of experiments that will study peptide cleavage and elution, peptide binding, stability assays, and functional binding, as outlined below:

- Linker cleavage: TEV protease and various mild acid elution steps will be performed to remove the MART-1 peptide from the cSCT to generate a SCD. This construct is expected to remain stable due to the dithiol mutations introduced into the HLA chain (Y84C and A139C), which has been previously reported for native MHCs.

- Peptide binding: peptides of interest may subsequently be loaded into the peptide-free SCDs via simple incubation methods, similar to those used for UV exchange protocols.

- Stability assay: ELISA assays making use of anti-β2m antibodies may be used to assess the stability of peptide-bound SCDs. The robustness of this technique has not yet been completely resolved, as there remains doubt about whether this antibody-binding approach will work for a single-chain variant of pMHCs, given that the β2m should remain covalently connected to the rest

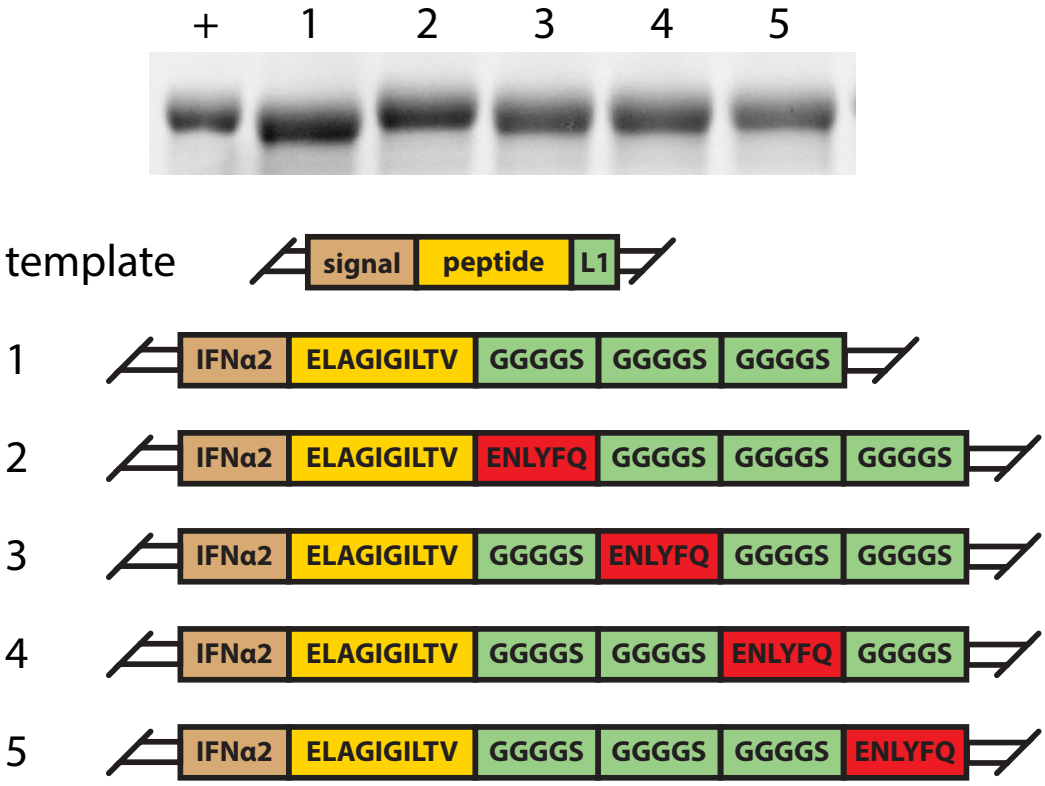


Figure 2.14: Expression of cleavable SCTs.
A*02:01 SCT plasmids (template D9) loaded with MART-1 peptide (ELAGILGILTV) were modified with a TEV protease cleavage site insertion within the L1 linker. Shown above is a reduced SDS-PAGE of transfection results from various designs.

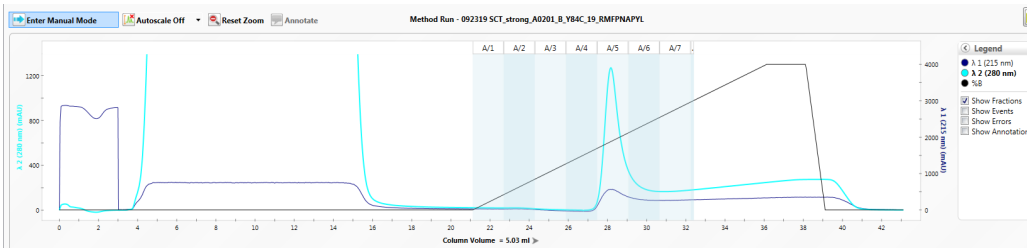


Figure 2.15: SCTs may be purified at larger scale to obtain milligram-scale yields.
A*02:01 SCT plasmid (D3 template) loaded with the WT1 peptide (RMFPNAPYL) was transfected at large scale (30 ml) for four days. The secreted protein was collected and purified by size-exclusion FPLC. Shown above is the absorbance analysis of eluent, where the desired purified SCT (column fraction A/5) appears as a singular peak.

of the MHC regardless of whether a peptide is present or not, so further testing is warranted.

- Functional binding: The peptide-bound SCDs will be tetramerized and used to bind against a TCR-transduced T cell line (for example, any one of the cell lines generated against SARS-CoV-2 peptides from **Fig. 3.9**). In order to assess purity of the SCDs after undergoing MART-1 cleavage, purification, and target peptide loading, one can mix in MART-1 TCR-transduced Jurkat cells to assess the relative proportions of tetramer-bound MART-1 Jurkat cells versus tetramer-bound target peptide-specific T cells.

Another significant application of the peptide-free SCD approach outlined here is that it may provide additional insight into the mechanism of peptide loading and presentation. As outlined by Anjanappa *et al.*, peptide-free refolded A*02:01 MHC was needed to study how the A*02:01 peptide-binding groove reacts to the presence of a peptide [31]. Prior to this work, no crystal structure of a peptide-free MHC was available; the DS mutation introduced to stabilize MHCs therefore allow for a peptide-free MHC construct to be crystallized, which enabled this insight into peptide-binding dynamics. The introduction of the TEV linker into our high-throughput platform enables one to generate DS-SCDs on a larger scale across multiple HLA alleles, therefore allowing for the same type of crystallography and molecular dynamics simulation work to be applied across other highly prevalent HLA alleles. Other potentially interesting outcomes from this work include the use of cSCTs for ELISA assays to measure peptide binding, and the use of cSCTs to facilitate discovery of HLA-compatible peptides by mass spectrometry.

2.2 Class II SCTs

2.2.1 Introduction

Class II Major Histocompatibility Complexes (MHCs) are antigen-presenting constructs recognized by $CD4^+$ T cells. As a general rule, Class II MHCs are presented on cell surfaces and draw their antigens from extracellular materials, such as the degradation products of bacteria, viruses, or mutated protein fragments (neoantigens) that are shed by cancer cells. Relative to Class I peptide (p)MHCs, which are recognized by $CD8^+$ T cells, Class II pMHCs are more diverse, due to the construction of the MHC itself, and due to the nature of the peptide antigen. For example, Class I pMHCs are typically comprised of a 9- to 11-mer peptide antigen, the HLA heavy chain, and an invariant β-2-microglobulin subunit, with the MHC allele providing a well-defined binding pocket for the peptide antigen. The nature of

this binding pocket, coupled with the relatively short length of the antigen, means that computational algorithms, such as NetMHC4.0, can provide good predictions for which candidate antigens bind most strongly to MHC, and thus are most likely to be presented in a biological context. By contrast, Class II pMHCs are constructed from two protein subunits that combine to form the binding pocket for a peptide antigen that can contain a highly variable number of amino acid residues, with a mean antigen size of around a 20-mer peptide (**Fig. 2.16A**). As a result, compared to Class I, the haplotype diversity of Class II MHCs is much higher for a given individual, the binding pocket of the MHC is less defined, and antigen prediction algorithms [32–34] are less accurate. Because of this, and because of the critical role of $CD4^+$ T cells as master immune regulators [35, 36], there has been a need for high-throughput Class II pMHC reagent production for the capture and analysis of antigen-specific $CD4^+$ T cells.

Soluble Class II pMHC tetramers have been used in prior reports to enumerate antigen-specific $CD4^+$ T cells. A brief scan of the literature shows that while numerous protocols exist outlining their assembly, generally there are two methods by which to synthesize them. The α and β chain subunits of the MHC complex may be individually encoded in plasmids for protein expression, typically in *E. coli* or baculovirus cells, and refolded together in the presence of a peptide to form stable pMHC monomers [37–43]. Alternatively, the two chain subunits may be encoded either in a single or dual plasmid system, with the peptide covalently linked to the N-terminus of the β chain, and utilized to express the pMHC monomers [6, 7, 44–46]. The primary advantage of the latter method is that peptide synthesis and purification may be avoided because all three components are genetically encoded. Furthermore, because the peptide is biologically loaded onto the MHC immediately after protein expression, there is no need to conduct an *in vitro* refolding step, obviating the need for tedious lower yield steps of protein refolding and pH optimization to facilitate peptide loading. The single-plasmid approach to generate single-chain trimers (SCT) was initially conducted in 1996 to generate murine class II pMHC molecules [7]. This strategy was then adapted by Zhu *et al.* to express SCTs encoding human MHC alleles in mouse cells [44] (**Fig. 2.16B**). An alternative approach to this SCT design was developed by Thayer *et al.*, in which a fragment of the invariant chain (proximal to the class II invariant chain-associated peptide (CLIP)) was attached between α chain and peptide to bind onto the HLA and position the peptide in proximity to the binding pocket [45] (**Fig. 2.16C**). This method was used in the context of murine MHC alleles for expression in the monkey kidney tissue-derived,

COS cell line. In both designs, the authors demonstrated that their SCTs are capable of binding to antigen-specific T cells to stimulate T cell responses such as cytokine release and apoptosis.

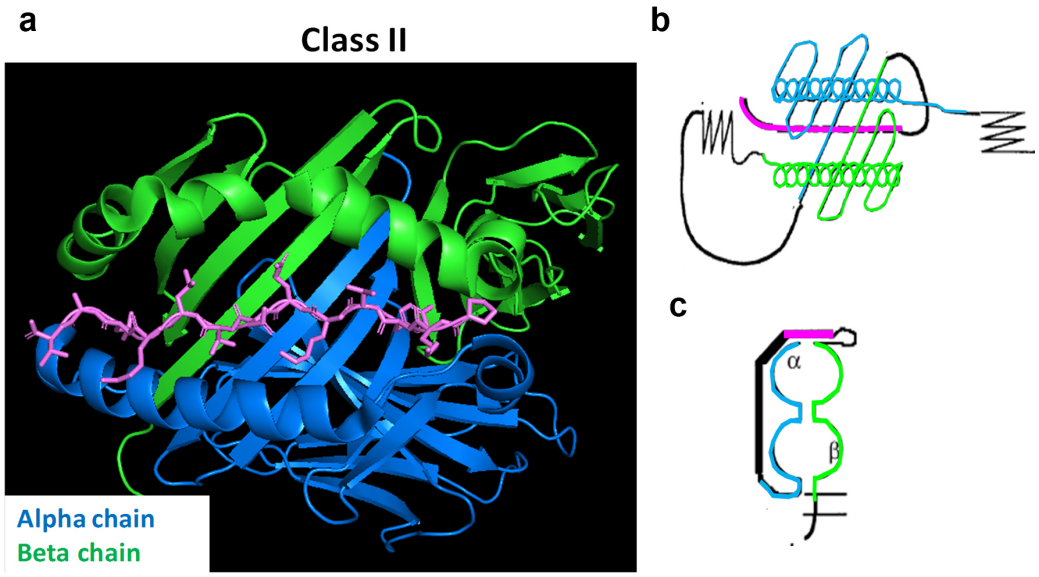


Figure 2.16: Class II pMHC structures and SCT constructs
The blue (alpha chain), green (beta chain), and pink (antigen) regions all correspond across the three structures. A. Crystal structure of a Class II pMHC, showing the peptide antigen residing in a binding pocket that is formed from both the alpha and beta chains. B. Class II SCT construct as reported by Zhu *et al* (1997). The black regions are linkers, purification tags, etc., that are engineered designs to promote stability and to facility purification. C. Class II SCT construct as reported by Thayer *et al* (2003). The primary difference between b and c is that the peptide antigen is linked to both the α and β MHC subunits in c, while it is only linked to the β subunit in b.

As reported by others, one demerit of the single-chain method is that a large number of unique pMHC proteins may generally be slower to prepare, as the mentioned works generally made use of large DNA fragments encoding the entirety of the pMHC construct. Additionally, both of the SCT expression protocols utilize non-human (insect and mammalian) systems, which can alter certain parameters away from human cell expression, such as glycosylation patterns. Thus, these methods are low-throughput, and yield expressed protein products that may not contain the types of post-translational modifications that would be representative of a pMHC expressed by human-derived cells. Here, we overcome these bottlenecks by genetically encoding HLA and peptide modularity into SCT Class II pMHC platforms, enabling plasmid production for any desired Class II HLA/peptide pMHC combi-

nation. We first show that the two different SCT designs used by Zhu *et al.* and Thayer *et al.* can lead to protein expression in a modern HEK293 cell line system, although average yields may differ. We validate the functionality of the Thayer *et al.* design, which gives higher protein yield, by showing binding and capture against a cell line of known antigen specificity. Finally, we demonstrate the feasibility of a library approach toward Class II pMHC construction by generating SCTs targeted against epitopes derived from various structural proteins of SARS-CoV-2, which may allow for capture of antigen-specific $CD4^+$ T cells from patient blood samples.

2.2.2 Methods

2.2.2.1 SCT template production

The construction of plasmids was initiated by designing Class II SCT-encoded fragments to be inserted into a pcDNA3.1 vector for subsequent protein expression using the Expi293 transfection kit (Thermo Fisher Scientific). All ordered fragments (Twist Bioscience) were codon-optimized for human species protein expression according to Expi293 expression guidelines. The Zhu *et al.* fragment design (SCT-Z) consists of protein subunits encoded in the following order: secretion signal, peptide, peptide-β chain linker (L1), β chain, β-α chain linker (L2), α chain, AviTag, 6xHisTag (**Fig. 2.17A**). The Thayer *et al.* fragment design (SCT-T) consists of protein subunits encoded in the following order: α chain secretion signal, α chain, α chain-invariant chain linker (L1), invariant chain fragment, peptide, peptide-β chain linker (L2), β chain, AviTag, 6xHisTag (**Fig. 2.17B**). For the SCT-Z designs, Gibson assembly overlaps (40 bp) were introduced within the initial 30 aa region of the post-signal sequence alpha chain for HLA alleles from each of the three loci (**Fig. 2.18**). HLA protein sequences at this region are invariant within each locus, which enables the ability to swap β and α chain of the first and second fragments of each construct, respectively, to generate new α/β pairs. For the SCT-T designs, the invariant chain region is a fixed sequence across all α/β pairs, so this part of the fragment encodes a Gibson assembly overlap (40 bp) to allow for HLA modularity (**Fig. 2.19**). Alternatively, one can make use of the Bsu36I recognition site embedded within to generate new α/β pairs by restriction enzyme digest (**Fig. 2.19**). The protein sequences of each HLA allele were obtained from an FTP server hosted by The Immuno Polymorphism Database (ftp://ftp.ebi.ac.uk/pub/databases/ipd/imgt/hla/fasta/). The peptide sequence derived from the invariant chain spans position 74-101, obtained from uniprot.org (P04233). The mouse MHC alleles for I-Ab α and β chains were ob-

tained from uniprot.org (P14434 and P14483, respectively). The ordered fragments were PCR-amplified by KOD Hot Start DNA Polymerase (MilliporeSigma) and paired together as described above using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). The complete SCT-encoded fragment was subsequently double digested at the flanking regions by EcoRI and XhoI (New England Biolabs), and ligated into the MCS region of pcDNA3.1 vector. This process can be iterated for every unique α/β pairing under either Z or T designs to generate a template plasmid upon which additional molecular engineering steps are performed to substitute the encoded peptide for library production.

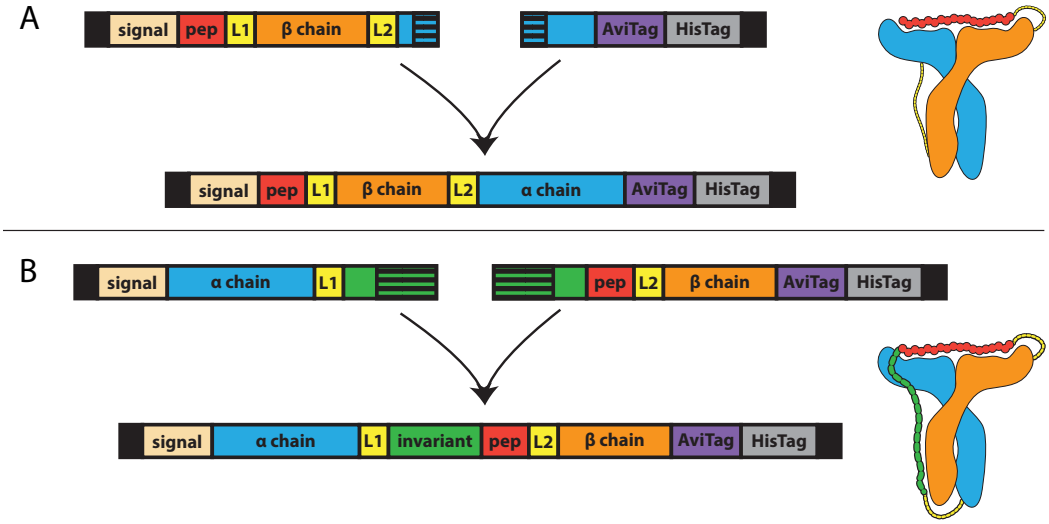


Figure 2.17: HLA Modularity of Class II SCT Designs
A. Linear map of Class II SCT design as reported by Zhu *et al.* with color-matched cartoon of expressed SCT construct (right). The fragment is split for Gibson assembly within the α chain to enable modular assembly of any α/β-encoded fragment. B. Linear map of Class II SCT design as reported by Thayer *et al.* with color-matched cartoon of expressed SCT construct (right). The fragment is split for Gibson assembly within the invariant chain to enable modular assembly of any α/β-encoded fragment.

2.2.2.2 SCT peptide library production

Traditional PCR methods were implemented for substitution of peptides into SCT-Z constructs. Universal binding sites for reverse primer (peptide_sub.REV, 5'-AGCAAGAGCAAGAGGAG-3') and forward primer (peptide_sub.FOR, 5'-GGTG GAGGAGGTTCTC-3') are implemented into the regions upstream and downstream of the peptide, respectively. Reverse complement codons encoding the target antigen

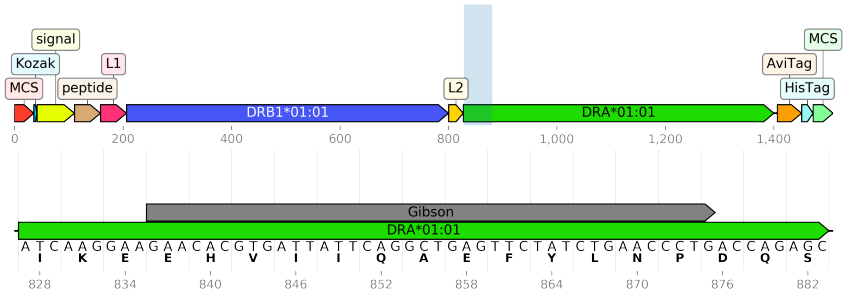


Figure 2.18: Detailed map of SCT-Z Gibson region
Top: Linear map of SCT-Z design. Bottom: Detailed view of the highlighted region from top, highlighting DNA sequence of Gibson overlap and invariant amino acids encoding α chain of DRA*01:01.

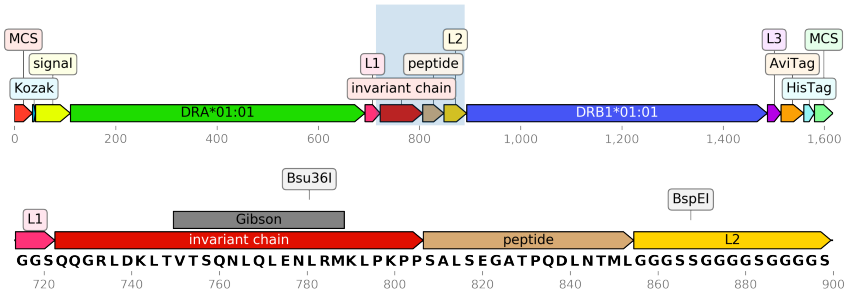


Figure 2.19: Detailed map of SCT-T Gibson region
Top: Linear map of SCT-T design. Bottom: Detailed view of the highlighted region from top, showing Gibson overlap and the restriction enzymes selected for peptide sequence ligation.

are appended onto the 5' end of the reverse primer. Inverse PCR of the Class II SCT with a peptide-encoded reverse primer (peptide_sub.REV appended with peptide) and universal forward primer (peptide_sub.FOR), followed by treatment of the PCR product with T4 DNA ligase, T4 polynucleotide kinase, and DpnI (New England Biolabs) allowed for re-construction of a plasmid with the replaced peptide. The plasmid was transformed into TOP10 chemically competent cells (Thermo Fisher Scientific) (**Fig. 2.20A**).

An alternative method to substitute peptides was utilized for SCT-T designs. For any given peptide, a pair of primers are designed in which the first primer encodes the carboxy-terminal region of the invariant chain and the former half of the peptide, while the second primer encodes the latter half of the peptide and the amino-terminal region of L2. These primers are designed to overlap and bind together at the region

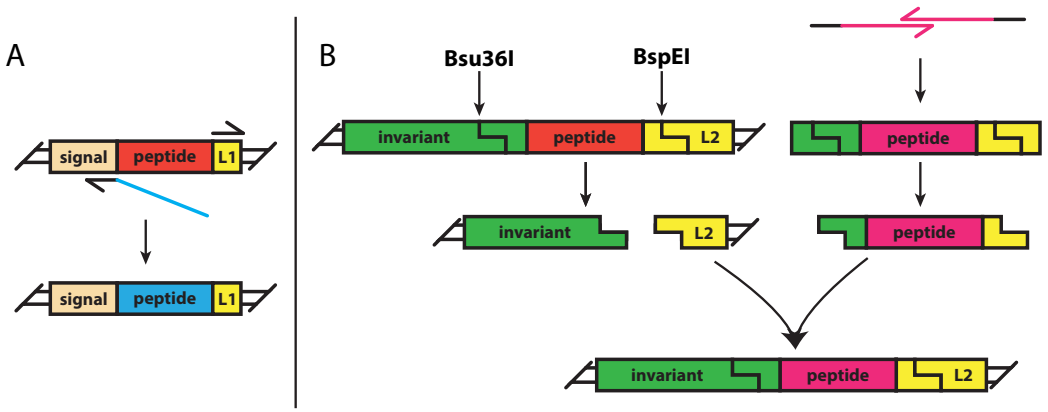


Figure 2.20: Peptide Modularity of Class II SCT Designs
A. Peptide substitution for SCT-Z plasmids using inverse PCR. A reverse primer encoding the reverse complement codons of a target peptide (blue) is used with a universal forward primer that binds to L1. B. Peptide substitution for SCT-T plasmids. Two ssDNA primers encoding a target peptide (pink) are used to assemble a dsDNA construct, which is digested by Bsu36I and BspEI to be ligated into the SCT template.

where the middle of the peptide is encoded, such that PCR amplification will result in a dsDNA product encoding the terminal region of the invariant chain, the entirety of the peptide, and the initial region of L2. The invariant chain and L2 are fixed sequences across all templates, and are encoded with Bsu36I and BspEI restriction enzyme cut sites, respectively. Double digest with these enzymes is conducted on both the template plasmid and the dsDNA PCR product (**Fig. 2.20B**). The desired products are gel-purified, ligated together, and transformed into TOP10 chemically competent cells.

2.2.2.3 SCT expression

Purified SCT plasmids were transfected into Expi293 cells within 24-well (2.5 ml capacity) plates. Briefly, 1.25 µg of plasmid was mixed with 75 µl Opti-MEM reduced serum media. 7.5 µl of ExpiFectamine Reagent was mixed with 70 µl Opti-MEM reduced serum media, incubated at room temperature for 5 minutes, and combined with the plasmid mixture. After a 15 minute room temperature incubation, the solution was added to 1.25 ml of Expi293 cells at 3 million cells/ml into a 24-well plate, which was then shaken at 225 RPM at 37 °C in 8% CO2 overnight. Twenty hours later, a solution containing 7.5 µl of ExpiFectamine Transfection Enhancer 1 and 75 µl of ExpiFectamine Transfection Enhancer 2 was added to each well. The

plate was kept on the shaker using aforementioned settings for a total of 4 days from start of transfection.

2.2.2.4 SCT biotinylation & purification

On day 4 of transfection, Expi293 cells were pelleted to enable collection of the supernatant (containing secreted SCTs). An aliquot of this supernatant was saved for SDS-PAGE gel analysis. If the SCTs were to be used for functional assays, they were re-suspended into 20 mM bicine PBS buffer to allow for biotinylation, purified by HisTag column-loaded pipet tips in an MEA 2 automated purification system (PhyNexus), and subsequently desalted into PBS buffer using Zeba 7k MWCO columns. SCTs that were used for downstream experiments involving tetramerization were stored at -20 °C in PBS buffer with 20% glycerol. SCTs that were to be used for thermal stability measurements were instead stored at 4 °C in PBS buffer.

2.2.2.5 Thermal stability characterization

SYPRO™ Orange Protein Gel Stain was purchased from ThermoFisher Scientific and diluted with H_2O to give a 100X working solution. To each 19 µl aliquot of Class II SCT protein solution (diluted to 10 µM, if possible), 1 µl of the 100X dye solution was added. A Bio-Rad thermal cycler equipped with a CFX96 real-time PCR detection system was used in combination with Precision Melt Analysis software to obtain melting curves of each SCT sample. Thermal ramp settings were 25 °C to 95 °C, 0.2 °C per 30 seconds.

2.2.2.6 De-glycosylation

De-glycosylation of SCTs was conducted with the PNGase F kit (New England Biolabs). Briefly, 20 µg of SCT was mixed with 1 µl of Glycoprotein Denaturing Buffer (10X) in a 10 µl H_2O solution. The mixture was denatured at 100 °C for 10 min, chilled on ice, and centrifuged for 10 seconds. To this solution, 2 µL GlycoBuffer 2 (10X), 2 µl 10% NP-40, and 6 µl H_2O was added. 1 µl PNGase F was mixed, and the entire solution was incubated at 37 °C for 1 hour. For analysis of non-denatured proteins, 1 µl PNGase F was added directly to the 10 µl $H2_2O$ solution without the addition of the mix containing GlycoBuffer 2 and NP-40.

2.2.2.7 Antigen-specific CD4$^+$ T cell isolation

Method 1: SCT-tetramer pool based. The monomer SCTs were individually tetramerized with PE or APC labeled streptavidin at a 4:1 molar ratio for 30 min at RT (or overnight at 4 °C). Biotin was added at an 8:1 molar ratio to streptavidin to block unoccupied biotin binding sites on streptavidin prior to mixing with the different tetramer samples. Each of the SCT-tetramers were pooled together and maintained at an individual tetramer concentration of 50 nM. The thawed PBMCs then were re-suspended in complete R10 media supplemented with IL-2 (50 IU/ml) and incubated for overnight recovery. On the next day, the PBMCs were washed and incubated in PBS added with tyrosine kinase inhibitor (Dasatinib, 50 nM) for 30 min. The PBMCs were then stained with Annexin V-BV421 (1µg/ml) and CD4-FITC antibody (1µg/ml) for 10 min at 4 °C followed by incubation with a pool of SCT-tetramers (each, 20 nM). Antigen-specific CD4$^+$ T cells captured by SCT-tetramer-PE were sorted into the tube using FACS sorter.

Method 2: Peptide stimulation based. A vial containing 1 million peripheral blood mononuclear cells (PBMCs) was thawed and incubated in complete R10 media (500 ml of RPMI 1640; 50 ml heat-inactivated fetal bovine serum (FBS); 5 ml of Pen/Strep (100 U/ml penicillin and 100 ug/ml streptomycin); 1x GlutaMAX) by adding 1µM of 15-mer peptide (or equivalent peptide pool, 15-mer with 11 aa overlap) and anti-CD40 antibody (1µg/ml) for 16 hrs. On the next day, the cells were washed and stained with Annexin V-BV421 (1µg/ml) and CD4-FITC antibody (1µg/ml) and CD154-PE antibody (1µg/ml) for 10 min at 4 °C. Activation-induced expression of CD154 by peptide stimulation permits the sorting of antigen-specific T-cells expressing these biomarkers.

2.2.2.8 Antigen-specific T cell functional assay

Method 1: SCT-Tetramer based. A vial of 0.5 M antigen-specific CD4$^+$ T cells suspended in serum-free cell culture media (CTL-Test™ Medium, Immunospot) was stimulated with SCT tetramers (0.1 µM final concentration) at 37 °C. After 16 hours of incubation, the supernatant of the cell solution was extracted for analysis by standard ELISA protocols for TNF-α (RD Systems, DY210-05), IFN-γ (RDSystems, DY285B-05), and IL-2 (BioLegend, 431804).

Method 2: Peptide stimulation based. 10^5-10^6 PBMCs or cloned CD4$^+$ T cells were incubated in 100 µL of serum-free cell culture media (CTL-Test™ Medium,

Immunospot) by adding 1 μM of 15-mer peptide (or equivalent peptide pool, 15-mer with 11 overlap) for 16 hrs. After 16 hours of incubation, the supernatant of the cell solution was extracted for analysis by standard ELISA protocols for TNF-α (RD Systems, DY210-05), IFN-γ (RDSystems, DY285B-05), and IL-2 (BioLegend, 431804).

2.2.2.9 $CD4^+$ T cell expansion

The FACS-sorted antigen-specific $CD4^+$ T cell T cells were directly transferred to the Rapid Expansion Protocol (REP) media [47]. REP media is composed of 2.5M irradiated PBMCs (4000 RAD), 0.5M TM-LCL (8000 RAD), IL-2 (50 IU/ml) and anti-CD3 antibody (30 ng/ml) per ml. On Day 3, half of the medium was removed without disturbing cells, and replaced with an equivalent volume of REP medium and cytokines. On Day 6-12, this media replacement step was repeated approximately every 2 days. On Day 14, cells were sorted with SCT tetramers to isolate the antigen-specific T cell populations. Multiple aliquots of cells were either frozen down or REP cycles were repeated to continue expansion, if needed for further analyses.

2.2.2.10 Sorting & TCR sequencing of expanded T cells

The expanded $CD4^+$ T cells were analyzed by flow cytometry using SCT-tetramers to enumerate individual antigen-specific T cell populations. 100k $CD4^+$ T cell were stained with Annexin V-BV421 (1 μg/ml) and CD4-FITC antibody (1 μg/ml) for 10 min at 4 °C followed by incubation with three tetramers with different dyes (PE, PE/Cy7, and APC) at 20 nM. The frequency of antigen-specific T cell population was measured by FlowJo software. The sorted cells were collected by antigen specificity in a 96-well plate, lysed, and RT-PCR was conducted to amplify the TCR α/β chains. The PCR product was extended with sequencing primers and the library was analyzed by Miseq software to extract antigen-paired TCR sequencing data. The sequencing data was further analyzed by customized R code and MIXCR.

2.2.3 Results & Discussion

To demonstrate viable SCT expression under these two designs, five initial SCTs were constructed (**Fig. 2.21A**). The first plasmid used an SCT-T template with

the Eα (52-68) peptide (ASFEAQGALANIAVDKA) and mouse MHC alleles I-Ab α/β as selected by Thayer *et al.* This sample was transfected as a biological triplicate to demonstrate reproducibility of the transfection method (**Fig. 2.21B**). Another four plasmids, designed under the SCT-Z template for the HLA allele pair DRA*01:01/DRB1*01:01, consist of a small library of peptides selected from citations identified through IEDB.org [48–50]. A positive control SCT consisting of a Class I SCT encoding HLA-A*02:01 with the WT1 (RMFPNAPYL) peptide was used to confirm transfection and to indicate the expected size for a Class I pMHC construct (approximately 50 kDa). The first Class II SCT plasmid produced a mass of approximately 60 kDa as previously reported, and all three biological replicates showed high and consistent yield. For the SCT-Z plasmids, we observed that the detected masses were also within the same range, but the last sample, interestingly, revealed a larger mass, which we attributed to the presence of an additional glycan group on this pMHC, given the presence of the NQT glycosylation motif within its peptide sequence (APIYNVLPTTSLVLGK**NQT**L). Furthermore, the expression yield of SCTs based on the Zhu *et al.* design appears to be peptide-dependent, implying that the degree of stabilization afforded by the selected peptide has a significant impact on expression level of the SCT construct.

Thermal shift assay measurements of the aforementioned five plasmids were performed to assess stability. As seen in **Figure 2.21C**, regardless of the template or peptide used, melting temperature (T_m) values based upon the absolute minimum of the derivative for fluorescence over temperature was approximately the same (70-75 °C). The high T_m values of these Class II SCTs compared to their Class I SCT counterparts (typically 50-65 °C) [51] suggests that these expressed Class II SCTs may be more stable than most Class I SCTs.

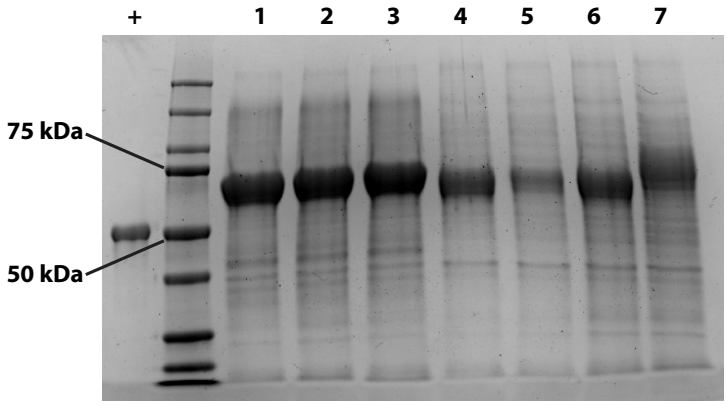
For SCT-T and SCT-Z designs, we expected masses of 55 kDa and 51 kDa, respectively. However, we observed SCT masses closer to the 75 kDa marker of the protein ladder (**Fig. 2.21B**). We suspected that this was a result of glycosylation. De-glycosylation of SCTs from both designs using PNGase treatment demonstrated a similar extent of decreased mass, closer to the calculated range of 51-55 kDa (**Fig. 2.22**).

Based upon characterization assays indicating higher yield and perhaps improved stability of SCT-T compared to SCT-Z, we chose SCT-T as the template to use for performing downstream functional validation assays. Toward this end, we rebuilt an SCT-T template to encode human MHC (HLA-DRA*01:01/DRB1*01:01),

A

well ID	SCT	peptide	MHC α chain	MHC β chain
1	T	ASFEAQGALANIAVDKA	I-Ab α	I-Ab β
2				
3				
4	Z	RFYKTLRAEQASQ	HLA-DRA*01:01	HLA-DRB1*01:01
5		SMRYQSLIPRLVEFF		
6		VGSDWRFLRGYHQYA		
7		APIYNVLPTTSLVLGKNQTL		

B



C

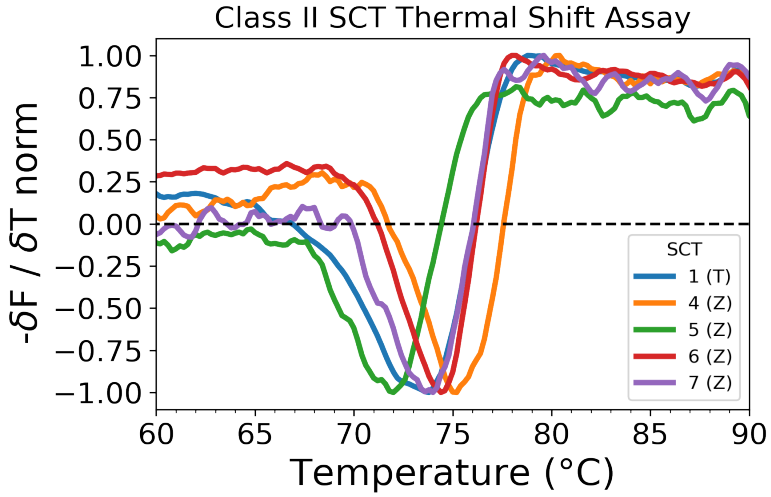


Figure 2.21: SCT Protein Expression & Thermal Stability
A. Table of the initial five SCT plasmids created using SCT-T and SCT-Z templates. B. SDS-PAGE of transfected SCTs. (+) represents positive control Class I SCT (A*02:01 with WT1 peptide, RMFPNAPYL). C. Thermal melting profiles of proteins expressed using SCT-T and SCT-Z templates. The negative of change in fluorescence over change in temperature is shown at each temperature. For (B) and (C), ID numbers correspond to the column "well ID" in (A).

and inserted a previously identified influenza virus peptide (PKYVKQNTLKLAT) as well as an irrelevant peptide (TRFQTLLALHRSYLT, from SARS-CoV-2 spike

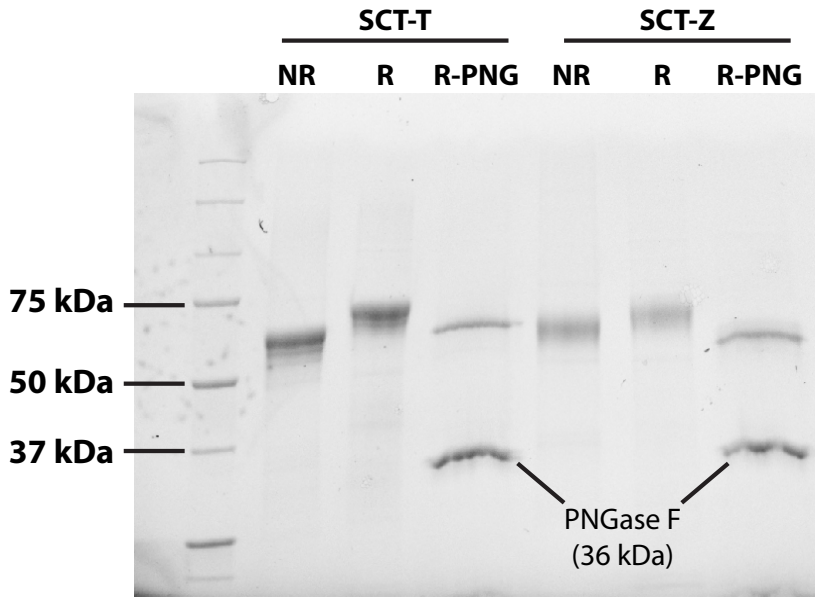


Figure 2.22: SCT De-glycosylation
SDS-PAGE of SCT-T and SCT-Z proteins de-glycosylated with PNGase F reveals that both class II SCTs display similar mass changes due to glycosylation. NR, non-reduced. R, reduced. R-PNG, reduced with PNGase F treatment.

protein). A $CD4^+$ T cell line specific to the influenza peptide and cognate pMHC-tetramers (produced by traditional methods requiring exogenous loading of peptide) were purchased from Benaroya Institute (BRI) to validate the SCT tetramers. Peptide specificity of the $CD4^+$ T cell line was confirmed via ELISA cytokine assay following overnight stimulation using influenza SCT tetramers (**Fig. 2.23**), which showed a significant increase in IFN-γ and TNF-α secretion only from the $CD4^+$ T cells. Flow cytometry assays indicated that the SCT tetramers performed similarly to the BRI tetramer variants in terms of binding sensitivity with the influenza-specific $CD4^+$ T cell line (**Fig. 2.24**). When both tetramer variants were incubated with a Jurkat cell line expressing an irrelevant TCR, however, the SCT tetramers showed significantly less binding, indicating reduced non-specific binding. Taken together, the class II SCT variants demonstrated similar capabilities as BRI variants in terms of binding to cognate TCRs, and are perhaps "cleaner" reagents as they appear to have less cross-reactivity. This difference may be due to the fact that SCTs undergo intracellular packaging and so can make use of native folding mechanisms, whereas for BRI variants, the class II α and β chains were co-expressed, extracted as an empty MHC construct, and then exogenously loaded with peptide. This additional *in vivo* step required for BRI tetramers may cause higher susceptibility to misfolding that

could explain the increased non-specificity.

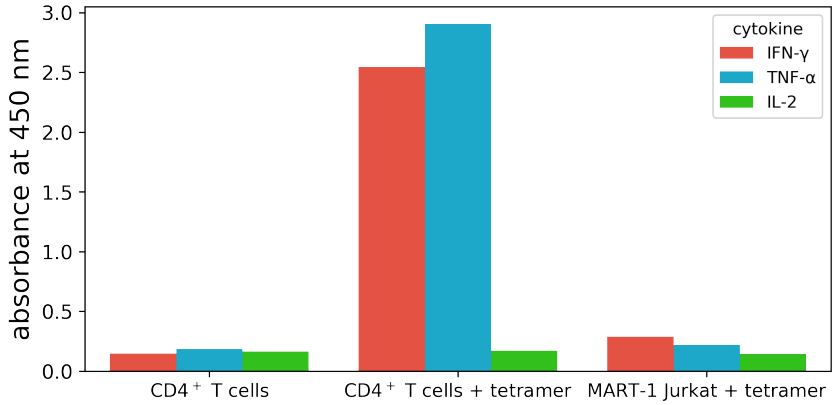


Figure 2.23: Tetramer Stimulation of Influenza-specific CD4$^+$ T Cells
ELISA cytokine assay measuring secretion of IFN-γ, TNF-α, and IL-2. "CD4$^+$ T cells" refers to influenza peptide-specific CD4$^+$ T cell line. "Tetramer" refers to SCT-T loaded with influenza peptide (PKYVKQNTLKLAT).

After validation of SCT-T design, we sought to utilize class II SCTs in a high-throughput application to demonstrate their implementation in a practical use case. Recently, CD4$^+$ T cells have been highlighted in numerous studies for their immuno-protective role against SARS-CoV-2. In order to better understand the functional role played by these T cells, we wanted to identify their antigen specificity and cognate TCR sequences. Toward this end, we identified 19 peptide candidates from SARS-CoV-2 structural proteins demonstrated by others to be immunogenic in patients with the paired HLA-DRA*01:01/DRB1*01:01 alleles (**Fig. 2.25A**) [52–55]. These peptides were substituted into the SCT-T template for transfection into Expi293 cells. Eighteen of the plasmids resulted in detectable SCT protein expression, although to varying degrees. Peptide #14, corresponding to an envelope protein epitope, resulted in no expression.

We plan to use these reagents to enumerate antigen-specific CD4$^+$ T cells from PBMCs of COVID-19 patients in future works. The SCT monomers will be tetramerized, pooled together, and then used to sort antigen-specific T cells from those patient samples. We will follow previously published expansion protocols to selectively expand these captured cells over two weeks [47] prior to individual tetramer sorting to isolate T cells by single antigen specificity. This will allow us to perform NGS bulk sequencing or 10X single-cell sequencing on the samples to obtain antigen-paired TCR sequences, which will then be cloned into CD4$^+$ T cell lines to demonstrate specificity and functionality. We will pulse APCs with the immunogenic peptides,

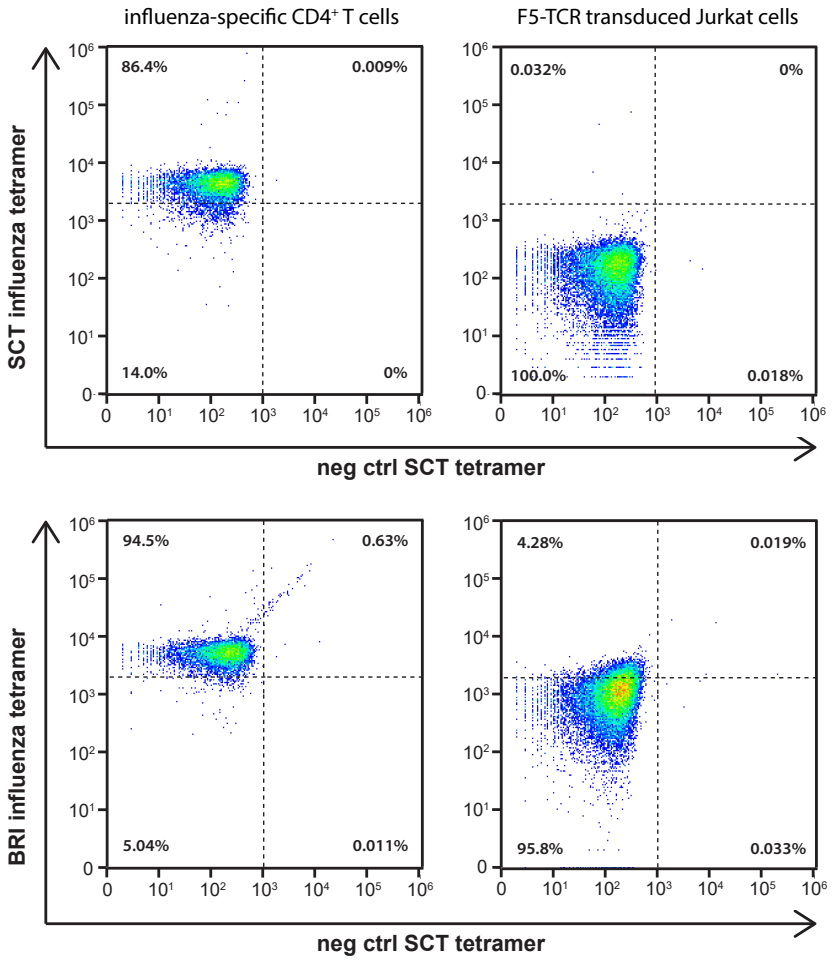


Figure 2.24: Class II SCT Flow Cytometry Validation
Flow cytometry assay of SCT-T tetramers against influenza-specific CD4$^+$ T cell line. SCT-T tetramers assembled with either an influenza peptide (PKYVKQNTLKLAT) or an irrelevant peptide were incubated with either influenza-specific CD4$^+$ T cells (top left) or Jurkat cells transduced with a TCR specific to ELAGIGILTV (MART-1) peptide (top right). The same experiment is performed (bottom row) with BRI tetramers loaded with the influenza peptide. Percentages in the corners of each plot represent the fraction of cells associated with each quadrant.

incubate them with the cell lines, and measure surface level and intracellular expression of activation/inflammation marker proteins. ELISA assays will be used to measure increases of activation/inflammatory protein secretion upon stimulation by the correct peptide. In summary, we plan to follow similar validation protocols as discussed in our COVID studies on HLA Class I epitopes (**Section 3.1.3**).

A

well ID	peptide	antigen	affinity (nM)	source
1	CTFEYVSQPFLMDLE	SP_166–180	22.5	◊
2	ITRFQTLLALHRSYL	SP_235-249	8.6	○
3	TRFQTLLALHRSYLT	SP_236-250	7.5	▪
4	FNFNGLTGTGVLTES	SP_541-555	9.3	▪
5	NLLLQYGSFCTQLNR	SP_751-765	64.6	◊▪
6	TQLNRALTGIAVEQD	SP_761-775	9.7	▪
7	CAQKFNGLTVLPPLL	SP_851-865	14	▪
8	TDEMIAQYTSALLAG	SP_866–880	17.8	◊
9	WTFGAGAALQIPFAM	SP_886-900	6.1	▪
10	IPFAMQMAYRFNGIG	SP_896-910	9.8	▪
11	TLVKQLSSNFGAISS	SP_961-975	8.3	▪
12	VQIDRLITGRLQSLQ	SP_991-1005	11.3	▪
13	LITGRLQSLQTYVTQ	SP_996-1010	6.1	▪
14	LAILTALRLCAYCCN	E_31-45	16.8	▪
15	SWFTALTQHGKEDLK	N_51–65	21.9	▫
16	KDGIIWVATEGALNT	N_127-141	10	○
17	AIVLQLPQGTTLPKG	N_156-170	25	○
18	GAVILRGHLRIAGHHLGR	M_141–158	26.1	◊
19	PKEITVATSRTLSYYKL	M_165–181	23.4	◊
IF	PKYVKQNTLKLAT	HA_306-318	116.7	

Sources: Peng *et al.* (◊); Nelde *et al.* (○); Mateus *et al.* (▪); Le Bert *et al.* (▫)

B

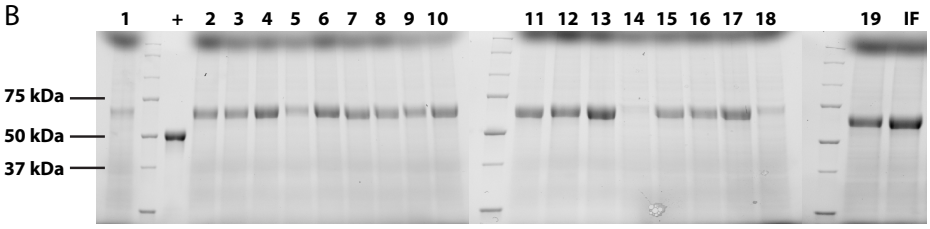


Figure 2.25: SCT Library for Structural Protein Epitopes of SARS-CoV-2
A. Table of peptides selected from the literature for assembly into SCT-T template plasmids. B. SDS-PAGE of SCT expression. Lane values correspond to the column "well ID" of table in (A). (+) represents positive control Class I SCT (A*02:01 with WT1 peptide, RMFPNAPYL).

2.3 Assembly of TCR Chains for Cloning

2.3.1 Introduction

The personalization aspect of antigen-based immunotherapies presents several major impediments not found in other therapeutic fields of development. Due to the peptide-based nature of personalized therapies, there have been numerous efforts to implement platforms that streamline production of peptide-bound major histocompatbility complex (pMHC) molecules to facilitate high-throughput enumeration of the identified cognate T-cell receptors. These technologies must be high-throughput by default to enable researchers to thoroughly assess the neoantigen landscape for any given patient, for which there may exist tens or hundreds of peptide candidates.

While the pMHC side of this workflow has recently seen successful improvements due to major contributions such as UV-facilitated peptide exchange or single-chain trimer technologies, TCR construction must also be concomitantly streamlined to accommodate the scale of therapeutic personalization. Any given pMHC, if found to be immunogenic and capable of capturing antigen-specific T cells, most likely will capture multiple TCR clonotypes, defined here as a pair of unique TCR α and β chains. Furthermore, unlike pMHCs which have 98% protein sequence consensus amongst those encoding for the same HLA allele, TCRs are extraordinarily diverse, consisting of an α and a β chain, each uniquely composed of a combination of variable (V) gene segment, a hyper-variable complementarity-determining region (CDR3), and a junctional (J) gene segment; this variability altogether accounts for at least 40% of the protein sequence. Therefore, there is an additional, daunting multiplicative factor of complexity to TCRs when considering methods to implement for their high-throughput generation. Additional constraints may arise depending on experimental requirements, such as the need for exclusion of restriction enzyme recognition sites and codon-optimization for human expression or for GC percent composition compatible with manufacturer recommendations.

Herein we introduce a standalone platform for the automated design and subsequent construction of TCR-encoded DNA constructs. One of the primary features of our platform is a user-friendly script that enables input of TCR sequences from various sources (e.g. bulk sequencing, 10X single-cell sequencing, or manually curated) toward generation of a plasmid library to clone into T cells for functional assessment and validation. The script is capable of using the TCR sequencing data to design optimized DNA fragments according the criteria previously outlined. Further, it can identify repeat alpha and beta chains within the library that may appear for certain cloning strategies (e.g. bulk and single-cell sequencing results often identify two alpha chains and two beta chains derived from the same cell). For such dual receptor results, there is no clear way to predict the correct α/β pairing that confers the desired antigen-specificity, and so one must assess all four potential α/β pairings. A system that can account for redundancy of unique chains will require the design and purchase of only four total chains versus eight, therefore significantly reducing reagent costs and production time. Accompanying this script is a generalized protocol outlining the experimental steps needed to assemble these fragments into DNA constructs compatible for cloning using traditional lentiviral plasmid systems or CRISPR/Cas9 methods.

Toward generation of these fragments, an explanation of the design scheme follows and may be visualized in **Figure 2.26A**. For the lentiviral strategy (blue path), four fragments are required per clonotype, two of which (f2_L and f4_L) are universal across all lentiviral TCR constructs because they encode only TCR constant gene segments. For the CRISPR strategy (red path), five fragments are required per clonotype, three of which (f0_C, f2_L, and f4_L) are universal across all CRISPR TCR constructs because they encode either TCR constant gene segments or universal CRISPR segments. Therefore, for each clonotype, the only variable fragments are f1_L, f1_C, and f3. Fragment f3 may be shared across both transduction strategies as it encodes the TCR alpha chain and contains sequences compatible with both transduction strategies. While both f1_L and f1_C fragments encode identical TCR beta chain sequences, f1_C must additionally encode CRISPR handle sequences upstream in order to overlap with f0_C. Therefore, f1_L and f1_C fragments must be uniquely created per clonotype. For each of the f2 fragments (f2_L and f2_C), there are two variants to account for TRBC1 and TRBC2 genes, which encode the two variations of C_β. In order to construct lentiviral chains for each clonotype, four fragments are assembled together, and the flanking regions of the product may be digested either with restriction enzymes (AscI/SalI) or (EcoRI/XhoI) to allow for insertion into pRRL or pcDNA3.1 vectors, respectively. In order to construct CRISPR chains for each clonotype, five fragments are assembled together, and the flanking regions of the product may be digested with restriction enzymes (EcoRI/XhoI) to allow for insertion into a compatible vector (e.g. pcDNA3.1 or pAAV). This vector serves to amplify the chain for sequencing verification and for the user to generate enough dsDNA for use in CRISPR transduction.

2.3.2 Methods

2.3.2.1 Data input

The design of TCR-encoded fragments requires the user to first generate a `csv` file containing a list of unique TCR clonotypes. As shown in **Figure 2.27A**, for each clonotype of interest, the user must designate a TCR identification code (TCR_ID), DNA consensus sequences encoding CDR3 regions along with partial sequences of the flanking variable and junctional regions (alpha_consensus and beta_consensus), and a string identifier (unique_tcr) which encapsulates all of the information regarding each gene segment constituting the clonotype. This file can readily be run in a Python terminal using the file name as an argument within the

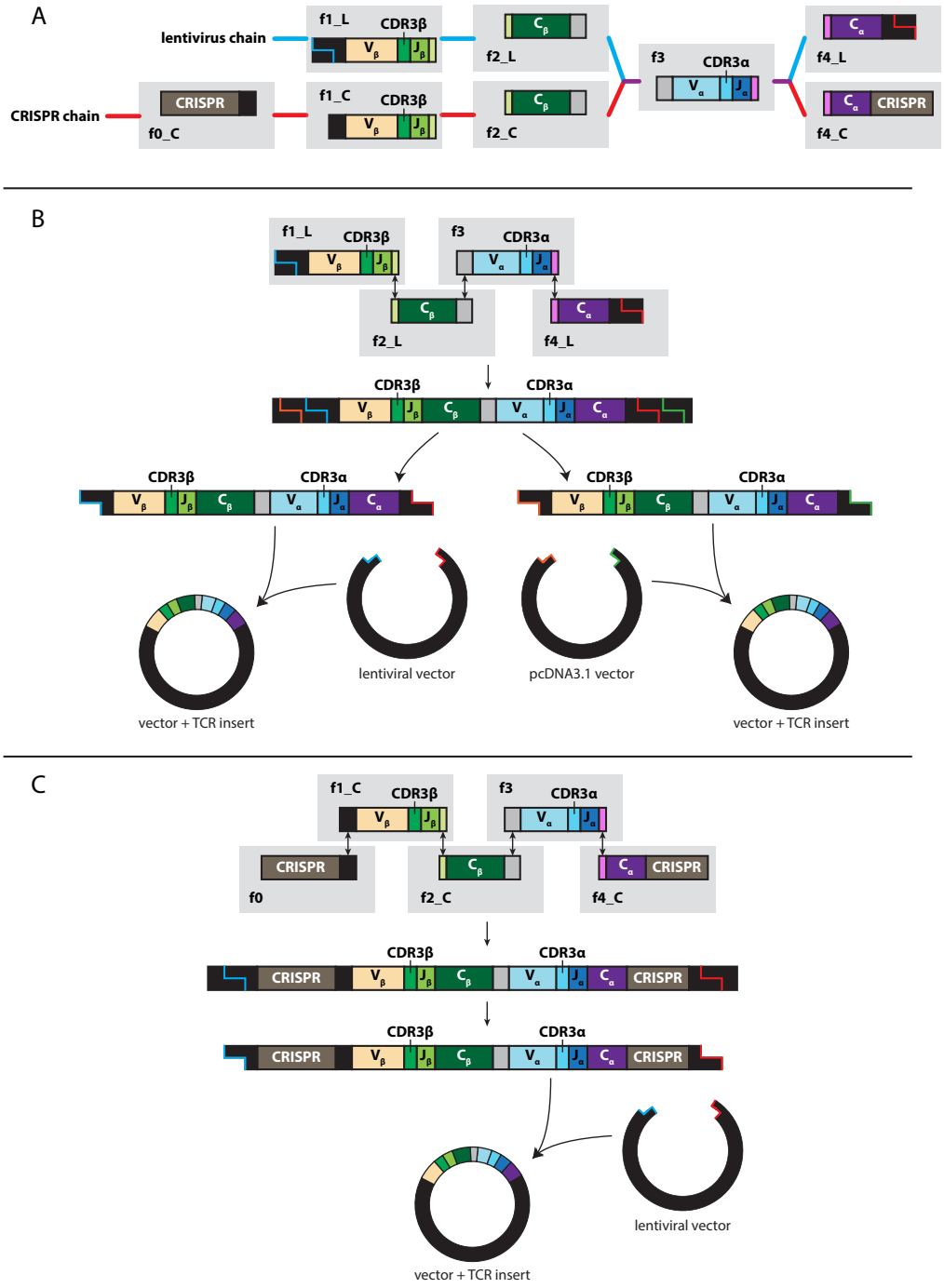


Figure 2.26: Schematic of TCR chain assembly
A. TCR clonotypes may be split into fragments for modular assembly into lentiviral or CRISPR transduction strategies. B. Diagram showing construction of lentiviral chains. C. Diagram showing construction of CRISPR chains.

function.

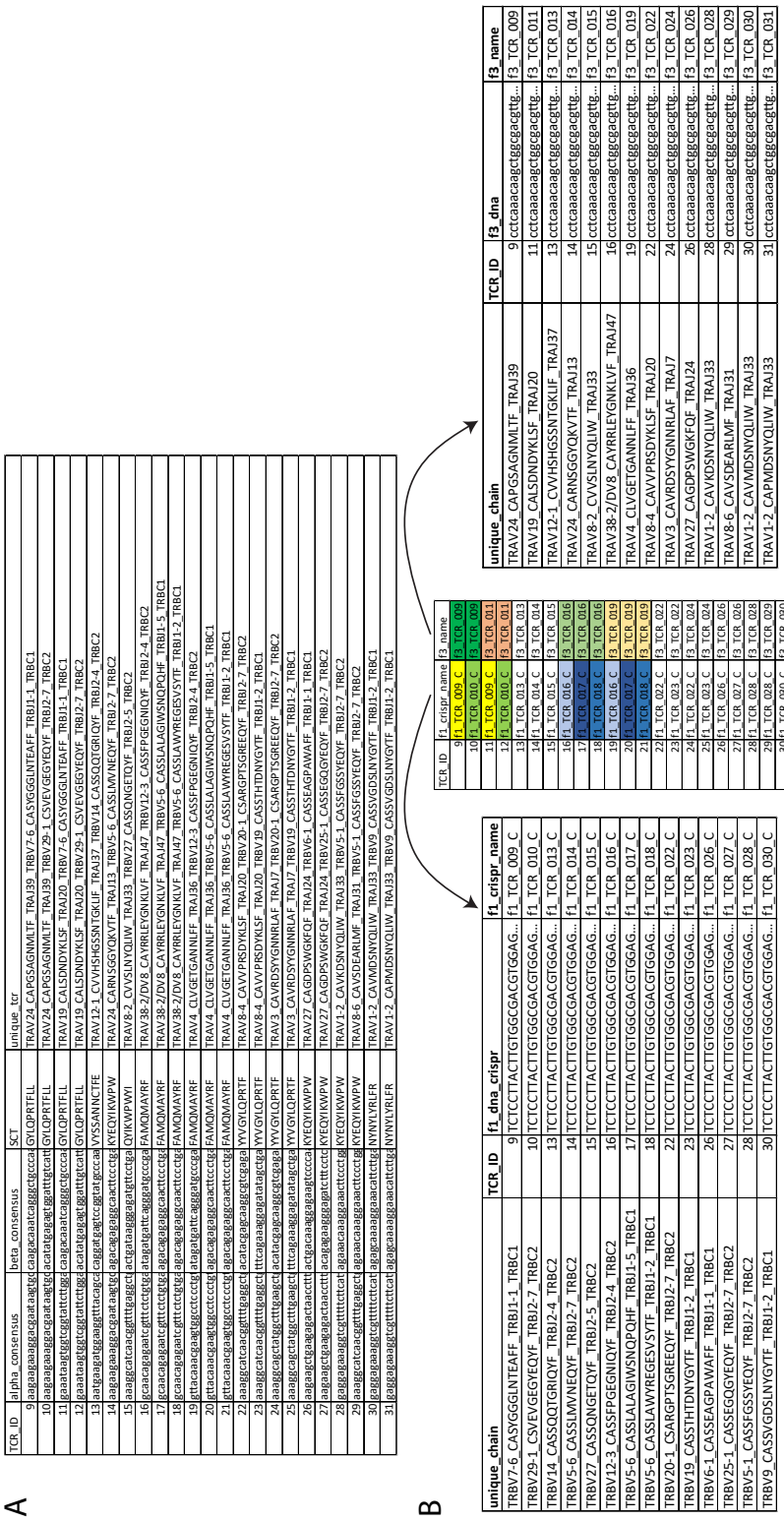


Figure 2.27: TCR chain generation script inputs and outputs
A. Example table of input required for TCR chain generation. B. Output table is generated (middle) that identifies the unique f1 and f3 chains (left and right, respectively) associated with each TCR clonotype.

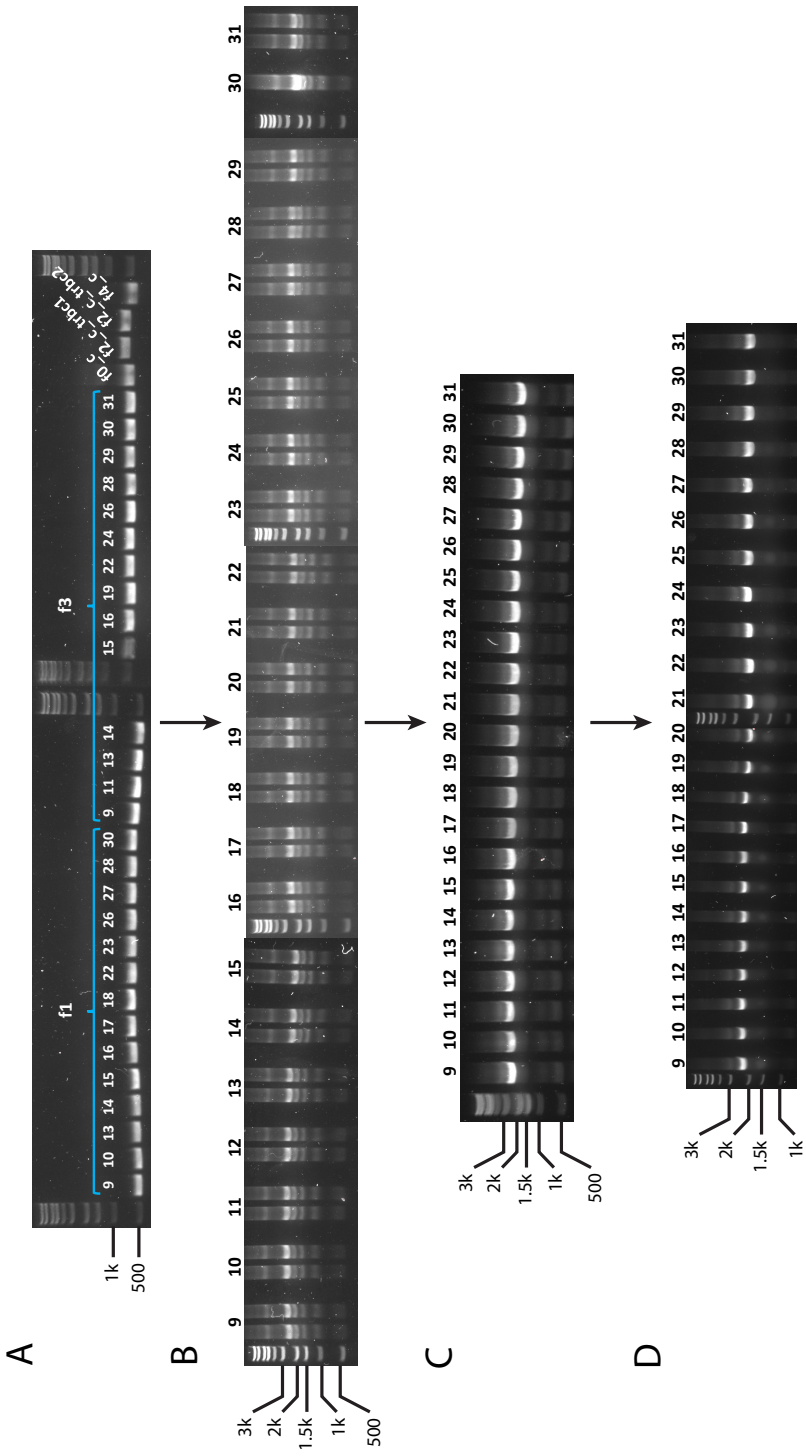


Figure 2.28: Gel electrophoresis of TCR chain assembly
A. Individual TCR chain fragments are amplified by PCR. B. Gibson assembly of each clonotype. C. Gel-purified and PCR-amplified Gibson assembly products from (B). D. Restriction enzyme digested TCR chains from (C). Numbers above lane correspond to the TCR_ID of input table in **Fig. 2.27A** and of **Appendix K**.

2.3.2.2 Data output

Three output `csv` files are generated (**Fig. 2.27B**). The first file serves as a reference file, where each unique clonotype is assigned one f1 fragment and one f3 fragment. This output accounts for redundant TCRα or TCRβ chains, so there may be cases where f1 or f3 chains are shared across clonotypes, as indicated by example entries which share the same color (middle table in **Fig. 2.27B**). Each clonotype also contains a column entry reporting the DNA sequence of the fully assembled TCR chain (with or without flanking restriction sites) to allow for the user to readily generate a Snapgene file. The second and third files report the DNA sequences of each designed f1 and f3 fragment (left and right tables of **Fig. 2.27B**, respectively). Two variants of the f1 fragments for each clonotype are generated to account for insertion into either CRISPR or lentiviral plasmid constructs. Sequences from fragments f0, f2, and f4 are constant, and so may be re-purchased and re-used for any desired clonotype. These sequences are reported in **Table G.1**.

2.3.2.3 Assembly of TCR chains

Sequences from DNA output files are ordered as dsDNA fragments (Twist Bioscience). They are amplified by PCR (KOD HiFi Polymerase) and then mixed together in appropriate combinations for Gibson assembly. The assembled products are then gel-purified and PCR-amplified. Subsequently, they are digested by appropriate restriction enzymes for incorporation into the designated vector. For fragments which are designed to be used with lentiviral transduction and therefore encode the entire TCR chain, they may be double digested by two different pairs of enzymes, to be inserted either into a lentiviral vector (AscI/SalI) or into a pcDNA3.1 vector for transfection (EcoRI/XhoI). For fragments which are designed to be used with CRISPR/Cas9 and therefore contain unique regions required for CRISPR transduction, the assembled fragments are digested by the enzyme pair (EcoRI/XhoI), to be inserted into pcDNA3.1 or pAAV vectors.

2.3.2.4 Functional validation of TCR chains

See **Section 3.1.3.4** for results from an example application of this TCR cloning strategy, where constructed TCR fragments are cloned into $CD8^+$ T cells and validated for antigen specificity and functionality.

References

[1] Nadine van Montfoort, Evelyn van der Aa, and Andrea M. Woltman. Understanding MHC Class I Presentation of Viral Antigens by Human Dendritic Cells as a Basis for Rational Design of Therapeutic Vaccines. *Frontiers in Immunology*, 5, April 2014. ISSN 1664-3224. doi: 10.3389/fimmu.2014.00182. URL http://journal.frontiersin.org/article/10.3389/fimmu.2014.00182/abstract.

[2] Xian-Hui He, Li-Hui Xu, and Yi Liu. Procedure for preparing peptide-major histocompatibility complex tetramers for direct quantification of antigen-specific cytotoxic T lymphocytes. *World Journal of Gastroenterology : WJG*, 11(27):4180–4187, July 2005. ISSN 1007-9327. doi: 10.3748/wjg.v11.i27.4180. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4615439/.

[3] Boris Rodenko, Mireille Toebes, Sine Reker Hadrup, Wim J E van Esch, Annemieke M Molenaar, Ton N M Schumacher, and Huib Ovaa. Generation of peptide–MHC class I complexes through UV-mediated ligand exchange. *Nature Protocols*, 1(3):1120–1132, August 2006. ISSN 1754-2189, 1750-2799. doi: 10.1038/nprot.2006.121. URL http://www.nature.com/articles/nprot.2006.121.

[4] Mireille Toebes, Miriam Coccoris, Adriaan Bins, Boris Rodenko, Raquel Gomez, Nella J Nieuwkoop, Willeke van de Kasteele, Guus F Rimmelzwaan, John B A G Haanen, Huib Ovaa, and Ton N M Schumacher. Design and use of conditional MHC class I ligands. *Nature Medicine*, 12(2):246–251, February 2006. ISSN 1078-8956, 1546-170X. doi: 10.1038/nm1360. URL http://www.nature.com/articles/nm1360.

[5] A. H. Bakker, R. Hoppes, C. Linnemann, M. Toebes, B. Rodenko, C. R. Berkers, S. R. Hadrup, W. J. E. van Esch, M. H. M. Heemskerk, H. Ovaa, and T. N. M. Schumacher. Conditional MHC class I ligands and peptide exchange technology for the human MHC gene products HLA-A1, -A3, -A11, and -B7. *Proceedings of the National Academy of Sciences*, 105(10):3825–3830, March 2008. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0709717105. URL http://www.pnas.org/cgi/doi/10.1073/pnas.0709717105.

[6] H. Kozono, J. White, J. Clements, P. Marrack, and J. Kappler. Production of soluble MHC class II proteins with covalently bound single peptides. *Nature*, 369(6476):151–154, May 1994. ISSN 0028-0836. doi: 10.1038/369151a0.

[7] P R Rhode, M Burkhardt, J Jiao, A H Siddiqui, G P Huang, and H C Wong. Single-chain MHC class II molecules induce T cell activation and apoptosis. page 8, June 1996.

[8] Y. Y. L. Yu, N. Netuschil, L. Lybarger, J. M. Connolly, and T. H. Hansen. Cutting Edge: Single-Chain Trimers of MHC Class I Molecules Form Stable Structures That Potently Stimulate Antigen-Specific T Cells and B Cells. *The Journal of Immunology*, 168(7):3145–3149, April 2002. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.168.7.3145. URL http://www.jimmunol.org/cgi/doi/10.4049/jimmunol.168.7.3145.

[9] Lonnie Lybarger, Y. Y. Lawrence Yu, Michael J. Miley, Daved H. Fremont, Nancy Myers, Tina Primeau, Steven M. Truscott, Janet M. Connolly, and Ted H. Hansen. Enhanced Immune Presentation of a Single-chain Major Histocompatibility Complex Class I Molecule Engineered to Optimize Linkage of a C-terminally Extended Peptide. *Journal of Biological Chemistry*, 278(29): 27105–27111, July 2003. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M303716200. URL http://www.jbc.org/lookup/doi/10.1074/jbc.M303716200.

[10] S. M. Truscott, L. Lybarger, J. M. Martinko, V. E. Mitaksov, D. M. Kranz, J. M. Connolly, D. H. Fremont, and T. H. Hansen. Disulfide Bond Engineering to Trap Peptides in the MHC Class I Binding Groove. *The Journal of Immunology*, 178(10):6280–6289, May 2007. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.178.10.6280. URL http://www.jimmunol.org/cgi/doi/10.4049/jimmunol.178.10.6280.

[11] Ted Hansen, Y.Y. Lawrence Yu, and Daved H. Fremont. Preparation of Stable Single-Chain Trimers Engineered with Peptide, 2 Microglobulin, and MHC Heavy Chain. In John E. Coligan, Barbara E. Bierer, David H. Margulies, Ethan M. Shevach, and Warren Strober, editors, *Current Protocols in Immunology*, page im1705s87. John Wiley & Sons, Inc., Hoboken, NJ, USA, November 2009. ISBN 978-0-471-14273-7. doi: 10.1002/0471142735.im1705s87. URL http://doi.wiley.com/10.1002/0471142735.im1705s87.

[12] Masanori Matsui, Masaaki Kawano, Sho Matsushita, and Toshitaka Akatsuka. Introduction of a point mutation into an HLA class I single-chain trimer induces enhancement of CTL priming and antitumor immunity. *Molecular Therapy - Methods & Clinical Development*, 1:14027, 2014. ISSN 23290501. doi: 10.1038/mtm.2014.27. URL https://linkinghub.elsevier.com/retrieve/pii/S2329050116300936.

[13] Fulian Lyu, Tatsuhiko Ozawa, Hiroshi Hamana, Eiji Kobayashi, Atsushi Muraguchi, and Hiroyuki Kishi. A novel and simple method to produce large amounts of recombinant soluble peptide/major histocompatibility complex monomers for analysis of antigen-specific human T cell receptors. *New Biotechnology*, November 2018. ISSN 18716784. doi: 10.1016/j.nbt.2018.11.005. URL https://linkinghub.elsevier.com/retrieve/pii/S1871678418316753.

[14] Songming Peng, Jesse M. Zaretsky, Alphonsus H.C. Ng, William Chour, Michael T. Bethune, Jongchan Choi, Alice Hsu, Elizabeth Holman, Xiaozhe Ding, Katherine Guo, Jungwoo Kim, Alexander M. Xu, John E. Heath, Won Jun Noh, Jing Zhou, Yapeng Su, Yue Lu, Jami McLaughlin, Donghui Cheng, Owen N. Witte, David Baltimore, Antoni Ribas, and James R. Heath. Sensitive Detection and Analysis of Neoantigen-Specific T Cell Populations from Tumors and Blood. *Cell Reports*, 28(10):2728–2738, September 2019. ISSN 22111247. doi: 10.1016/j.celrep.2019.07.106. URL https://linkinghub.elsevier.com/retrieve/pii/S2211124719310228.

[15] Z. Hein, H. Uchtenhagen, E. T. Abualrous, S. K. Saini, L. Janssen, A. Van Hateren, C. Wiek, H. Hanenberg, F. Momburg, A. Achour, T. Elliott, S. Springer, and D. Boulanger. Peptide-independent stabilization of MHC class I molecules breaches cellular quality control. *Journal of Cell Science*, 127(13):2885–2897, July 2014. ISSN 0021-9533, 1477-9137. doi: 10.1242/jcs.145334. URL http://jcs.biologists.org/cgi/doi/10.1242/jcs.145334.

[16] Andreas Moritz, Raghavendra Anjanappa, Claudia Wagner, Sebastian Bunk, Martin Hofmann, Gabriele Pszolla, Ankur Saikia, Maria Garcia-Alai, Rob Meijers, Hans-Georg Rammensee, Sebastian Springer, and Dominik Maurer. High-throughput peptide-MHC complex generation and kinetic screenings of TCRs with peptide-receptive HLA-A*02:01 molecules. *Science Immunology*, 4(37):eaav0860, July 2019. ISSN 2470-9468. doi: 10.1126/sciimmunol.aav0860. URL http://immunology.sciencemag.org/lookup/doi/10.1126/sciimmunol.aav0860.

[17] Sunil Kumar Saini, Tripti Tamhane, Raghavendra Anjanappa, Ankur Saikia, Sofie Ramskov, Marco Donia, Inge Marie Svane, Søren Nyboe Jakobsen, Maria Garcia-Alai, Martin Zacharias, Rob Meijers, Sebastian Springer, and Sine Reker Hadrup. Empty peptide-receptive MHC class I molecules for efficient detection of antigen-specific T cells. *Science Immunology*, 4(37), July 2019. ISSN 2470-9468. doi: 10.1126/sciimmunol.aau9039. URL http://immunology.sciencemag.org/lookup/doi/10.1126/sciimmunol.aau9039.

[18] Tripti Tamhane, Sunil Kumar Saini, Raghavendra Anjanappa, Ankur Saikia, Sofie Ramskov, Marco Donia, Inge Marie Swane, Søren Nyboe Jakobsen, Martin Zacharia, Rob Meijers, Sebastian Springer, and Sine Hadrup. Empty-loadable MHC class I tetramers for efficient detection of antigen-specific T cells. *The Journal of Immunology*, 204(1 Supplement):86.1–86.1, May 2020. ISSN 0022-1767, 1550-6606. URL https://www.jimmunol.org/content/204/1_Supplement/86.1. Publisher: American Association of Immunologists Section: Technological Innovations in Immunology 1.

[19] Filippo Mancia, Saurabh D Patel, Michael W Rajala, Philipp E Scherer, Adriana Nemes, Ira Schieren, Wayne A Hendrickson, and Lawrence Shapiro. Op-

timization of Protein Production in Mammalian Cells with a Coexpressed Fluorescent Marker. *Structure*, 12(8):1355–1360, August 2004. ISSN 09692126. doi: 10.1016/j.str.2004.06.012. URL https://linkinghub.elsevier.com/retrieve/pii/S0969212604002400.

[20] Aude G. Chapuis, Daniel N. Egan, Merav Bar, Thomas M. Schmitt, Megan S. McAfee, Kelly G. Paulson, Valentin Voillet, Raphael Gottardo, Gunnar B. Ragnarsson, Marie Bleakley, Cecilia C. Yeung, Petri Muhlhauser, Hieu N. Nguyen, Lara A. Kropp, Luca Castelli, Felecia Wagener, Daniel Hunter, Marcus Lindberg, Kristen Cohen, Aaron Seese, M. Juliana McElrath, Natalie Duerkopp, Ted A. Gooley, and Philip D. Greenberg. T cell receptor gene therapy targeting WT1 prevents acute myeloid leukemia relapse post-transplant. *Nature Medicine*, 25(7):1064–1072, July 2019. ISSN 1078-8956, 1546-170X. doi: 10.1038/s41591-019-0472-9. URL http://www.nature.com/articles/s41591-019-0472-9.

[21] Mikhail Shugay, Dmitriy V. Bagaev, Ivan V. Zvyagin, Renske M. Vroomans, Jeremy Chase Crawford, Garry Dolton, Ekaterina A. Komech, Anastasiya L. Sycheva, Anna E. Koneva, Evgeniy S. Egorov, Alexey V. Eliseev, Ewald Van Dyk, Pradyot Dash, Meriem Attaf, Cristina Rius, Kristin Ladell, James E. McLaren, Katherine K. Matthews, E. Bridie Clemens, Daniel C. Douek, Fabio Luciani, Debbie van Baarle, Katherine Kedzierska, Can Kesmir, Paul G. Thomas, David A. Price, Andrew K. Sewell, and Dmitriy M. Chudakov. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Research*, 46(D1):D419–D427, 2018. ISSN 1362-4962. doi: 10.1093/nar/gkx760.

[22] Pradyot Dash, Andrew J. Fiore-Gartland, Tomer Hertz, George C. Wang, Shalini Sharma, Aisha Souquette, Jeremy Chase Crawford, E. Bridie Clemens, Thi H. O. Nguyen, Katherine Kedzierska, Nicole L. La Gruta, Philip Bradley, and Paul G. Thomas. Quantifiable predictive features define epitope specific T cell receptor repertoires. *Nature*, 547(7661):89–93, July 2017. ISSN 0028-0836. doi: 10.1038/nature22383. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5616171/.

[23] Stéphanie Gras, Xavier Saulquin, Jean-Baptiste Reiser, Emilie Debeaupuis, Klara Echasserieau, Adrien Kissenpfennig, François Legoux, Anne Chouquet, Madalen Le Gorrec, Paul Machillot, Bérangère Neveu, Nicole Thielens, Bernard Malissen, Marc Bonneville, and Dominique Housset. Structural Bases for the Affinity-Driven Selection of a Public TCR against a Dominant Human Cytomegalovirus Epitope. *The Journal of Immunology*, 183(1):430–437, July 2009. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.0900556. URL http://www.jimmunol.org/lookup/doi/10.4049/jimmunol.0900556.

[24] Lydie Trautmann, Marie Rimbert, Klara Echasserieau, Xavier Saulquin, Bérangère Neveu, Julie Dechanet, Vincenzo Cerundolo, and Marc Bonneville.

Selection of T Cell Clones Expressing High-Affinity Public TCRs within Human Cytomegalovirus-Specific CD8 T Cell Responses. *The Journal of Immunology*, 175(9):6123–6132, November 2005. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.175.9.6123. URL http://www.jimmunol.org/lookup/doi/10.4049/jimmunol.175.9.6123.

[25] Eleanor Clancy-Thompson, Christine A. Devlin, Paul M. Tyler, Mariah M. Servos, Lestat R. Ali, Katherine S. Ventre, M. Aladdin Bhuiyan, Patrick T. Bruck, Michael E. Birnbaum, and Stephanie K. Dougan. Altered Binding of Tumor Antigenic Peptides to MHC Class I Affects CD8+ T Cell–Effector Responses. *Cancer Immunology Research*, 6(12):1524–1536, December 2018. ISSN 2326-6066, 2326-6074. doi: 10.1158/2326-6066.CIR-18-0348. URL https://cancerimmunolres.aacrjournals.org/content/6/12/1524.

[26] Marie Bodinier, Marie-Alix Peyrat, Claire Tournay, François Davodeau, François Romagne, Marc Bonneville, and François Lang. Efficient detection and immunomagnetic sorting of specific T cells using multimers of MHC class I and peptide with reduced CD8 binding. *Nature Medicine*, 6(6):707–710, June 2000. ISSN 1078-8956, 1546-170X. doi: 10.1038/76292. URL http://www.nature.com/articles/nm0600_707.

[27] Berangere Neveu, Klara Echasserieau, Timothy Hill, Kristine Kuus-Reichel, Elisabeth Houssaint, Marc Bonneville, and Xavier Saulquin. Impact of CD8–MHC class I interaction in detection and sorting efficiencies of antigen-specific T cells using MHC class I/peptide multimers: contribution of pMHC valency. *International Immunology*, 18(7):1139–1145, July 2006. ISSN 1460-2377, 0953-8178. doi: 10.1093/intimm/dxl048. URL https://academic.oup.com/intimm/article-lookup/doi/10.1093/intimm/dxl048.

[28] Immanuel F Luescher, Eric Vivier, Andreas Layer, and Jerome Mahiout. CDS modulation of T-cell antigen receptor-ligand interactions on living cytotoxic T lymphocytes. 373:4, 1995.

[29] Xiao-Ning Xu, Marco A Purbhoo, Nan Chen, Juthathip Mongkolsapaya, Jennifer H Cox, Ute-Christiane Meier, Sabrina Tafuro, P Rod Dunbar, Andy K Sewell, Christopher S Hourigan, Victor Appay, Vincenzo Cerundolo, Scott R Burrows, Andrew J McMichael, and Gavin R Screaton. A Novel Approach to Antigen-Specific Deletion of CTL with Minimal Cellular Activation Using ␣3 Domain Mutants of MHC Class I/Peptide Complex. page 12.

[30] Mikaël J. Pittet, Verena Rubio-Godoy, Gilles Bioley, Philippe Guillaume, Pascal Batard, Daniel Speiser, Immanuel Luescher, Jean-Charles Cerottini, Pedro Romero, and Alfred Zippelius. α3 Domain Mutants of Peptide/MHC Class I Multimers Allow the Selective Isolation of High Avidity Tumor-Reactive CD8 T Cells. *The Journal of Immunology*, 171(4):1844–1849, August 2003. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.171.4.1844. URL http://www.jimmunol.org/lookup/doi/10.4049/jimmunol.171.4.1844.

[31] Raghavendra Anjanappa, Maria Garcia-Alai, Janine-Denise Kopicki, Julia Lockhauserbäumer, Mohamed Aboelmagd, Janina Hinrichs, Ioana Maria Nemtanu, Charlotte Uetrecht, Martin Zacharias, Sebastian Springer, and Rob Meijers. Structures of peptide-free and partially loaded MHC class I molecules reveal mechanisms of peptide selection. *Nature Communications*, 11(1), December 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-14862-4. URL http://www.nature.com/articles/s41467-020-14862-4.

[32] Morten Nielsen, Claus Lundegaard, and Ole Lund. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC bioinformatics*, 8:238, July 2007. ISSN 1471-2105. doi: 10.1186/1471-2105-8-238.

[33] Morten Nielsen, Ole Lund, Søren Buus, and Claus Lundegaard. MHC Class II epitope predictive algorithms: MHC class II epitope predictive algorithms. *Immunology*, 130(3):319–328, July 2010. ISSN 00192805. doi: 10.1111/j.1365-2567.2010.03268.x. URL http://doi.wiley.com/10.1111/j.1365-2567.2010.03268.x.

[34] Peng Wang, John Sidney, Courtney Dow, Bianca Mothé, Alessandro Sette, and Bjoern Peters. A Systematic Assessment of MHC Class II Peptide Binding Predictions and Evaluation of a Consensus Approach. *PLoS Computational Biology*, 4(4), April 2008. ISSN 1553-734X. doi: 10.1371/journal.pcbi.1000048. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2267221/.

[35] Hiroyoshi Nishikawa and Shimon Sakaguchi. Regulatory T cells in cancer immunotherapy. *Current Opinion in Immunology*, 27:1–7, April 2014. ISSN 1879-0372. doi: 10.1016/j.coi.2013.12.005.

[36] Robert A. Seder, Patricia A. Darrah, and Mario Roederer. T-cell quality in memory and protection: implications for vaccine design. *Nature Reviews. Immunology*, 8(4):247–258, April 2008. ISSN 1474-1741. doi: 10.1038/nri2274.

[37] J. D. Altman, P. A. Reay, and M. M. Davis. Formation of functional peptide complexes of class II major histocompatibility complex proteins from subunits produced in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 90(21):10330–10334, November 1993. ISSN 0027-8424. doi: 10.1073/pnas.90.21.10330.

[38] J. Stöckel, E. Meinl, C. Hahnel, J. Malotka, R. Seitz, K. Drexler, H. Wekerle, and K. Dornmair. Refolding of human class II major histocompatibility complex molecules isolated from Escherichia coli. Assembly of peptide-free heterodimers and increased refolding-yield in the presence of antigenic peptide. *The Journal of Biological Chemistry*, 269(47):29571–29578, November 1994. ISSN 0021-9258.

[39] H Reijonen and Kwok, W.W. Use of HLA class II tetramers in tracking antigen-specific T cells and mapping T-cell epitopes. *Methods*, 29(3):282–288, March 2003. ISSN 10462023. doi: 10.1016/S1046-2023(02)00350-X. URL https://linkinghub.elsevier.com/retrieve/pii/S104620230200350X.

[40] Erik J. Novak, Andrew W. Liu, Gerald T. Nepom, and William W. Kwok. MHC class II tetramers identify peptide-specific human CD4+ T cells proliferating in response to influenza A antigen. *Journal of Clinical Investigation*, 104(12): R63–R67, December 1999. ISSN 0021-9738. doi: 10.1172/JCI8476. URL http://www.jci.org/articles/view/8476.

[41] Erik J. Novak, Andrew W. Liu, John A. Gebe, Ben A. Falk, Gerald T. Nepom, David M. Koelle, and William W. Kwok. Tetramer-Guided Epitope Mapping: Rapid Identification and Characterization of Immunodominant CD4+ T Cell Epitopes from Complex Antigens. *The Journal of Immunology*, 166(11):6665–6670, June 2001. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.166.11.6665. URL https://www.jimmunol.org/content/166/11/6665.

[42] Jane H. Buckner, Ursula Holzer, Erik J. Novak, Helena Reijonen, William W. Kwok, and Gerald T. Nepom. Defining antigen-specific responses with human MHC class II tetramers. *Journal of Allergy and Clinical Immunology*, 110(2):199–208, August 2002. ISSN 00916749. doi: 10.1067/mai.2002.125976. URL https://linkinghub.elsevier.com/retrieve/pii/S0091674902000805.

[43] Peter Braendstrup, Sune Justesen, Thomas Østerbye, Lise Lotte Bruun Nielsen, Roberto Mallone, Lars Vindeløv, Anette Stryhn, and Søren Buus. MHC Class II Tetramers Made from Isolated Recombinant α and β Chains Refolded with Affinity-Tagged Peptides. *PLoS ONE*, 8(9):e73648, September 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0073648. URL https://dx.plos.org/10.1371/journal.pone.0073648.

[44] Xiaojie Zhu, Sina Bavari, Robert Ulrich, Scheherazade Sadegh-Nasseri, Soldano Ferrone, Louise McHugh, and Michael Mage. 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 immunogenicityin vitro and reduced affinity for bacterial superantigens. *European Journal of Immunology*, 27(8):1933–1941, August 1997. ISSN 00142980, 15214141. doi: 10.1002/eji.1830270817. URL http://doi.wiley.com/10.1002/eji.1830270817.

[45] Wesley P Thayer, Chinh T Dao, Leszek Ignatowicz, and Peter E Jensen. A novel single chain I-Ab molecule can stimulate and stain antigen-specific T cells. *Molecular Immunology*, 39(14):861–870, May 2003. ISSN 01615890. doi: 10.1016/S0161-5890(03)00010-5. URL https://linkinghub.elsevier.com/retrieve/pii/S0161589003000105.

[46] Pau Serra, Nahir Garabatos, Santiswarup Singha, César Fandos, Josep Garnica, Patricia Solé, Daniel Parras, Jun Yamanouchi, Jesús Blanco, Meritxell Tort, Mireia Ortega, Yang Yang, Kristofor K. Ellestad, and Pere Santamaria. Increased yields and biological potency of knob-into-hole-based soluble MHC class II molecules. *Nature Communications*, 10(1), December 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-12902-2. URL http://www.nature.com/articles/s41467-019-12902-2.

[47] William Y. Ho, Hieu N. Nguyen, Matthias Wolfl, Juergen Kuball, and Philip D. Greenberg. In vitro methods for generating CD8+ T-cell clones for immunotherapy from the naïve repertoire. *Journal of Immunological Methods*, 310(1):40–52, March 2006. ISSN 0022-1759. doi: 10.1016/j.jim.2005.11.023. URL https://www.sciencedirect.com/science/article/pii/S0022175905004291.

[48] Iwona Strug, J. Mauricio Calvo-Calle, Karin M. Green, John Cruz, Francis A . Ennis, James E. Evans, and Lawrence J. Stern. Vaccinia peptides eluted from HLA-DR1 isolated from virus-infected cells are recognized by CD4+ T cells from a vaccinated donor. *Journal of proteome research*, 7(7):2703–2711, July 2008. ISSN 1535-3893. doi: 10.1021/pr700780x. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2515592/.

[49] William W. Kwok, Venus Tan, Laurie Gillette, Christopher T. Littell, Michele A. Soltis, Rebecca B. LaFond, Junbao Yang, Eddie A. James, and Jonathan H. DeLong. Frequency of epitope-specific naive CD4(+) T cells correlates with immunodominance in the human memory repertoire. *Journal of Immunology (Baltimore, Md.: 1950)*, 188(6):2537–2544, March 2012. ISSN 1550-6606. doi: 10.4049/jimmunol.1102190.

[50] Moran Galperin, Carine Farenc, Madhura Mukhopadhyay, Dhilshan Jayasinghe, Amandine Decroos, Daniela Benati, Li Lynn Tan, Lisa Ciacchi, Hugh H. Reid, Jamie Rossjohn, Lisa A. Chakrabarti, and Stephanie Gras. CD4+ T cell-mediated HLA class II cross-restriction in HIV controllers. *Science Immunology*, 3(24), June 2018. ISSN 2470-9468. doi: 10.1126/sciimmunol.aat0687.

[51] Dylan T. Blaha, Scott D. Anderson, Daniel M. Yoakum, Marlies V. Hager, Yuanyuan Zha, Thomas F. Gajewski, and David M. Kranz. High-Throughput Stability Screening of Neoantigen/HLA Complexes Improves Immunogenicity Predictions. *Cancer Immunology Research*, 7(1):50–61, January 2019. ISSN 2326-6066, 2326-6074. doi: 10.1158/2326-6066.CIR-18-0395. URL http://cancerimmunolres.aacrjournals.org/lookup/doi/10.1158/2326-6066.CIR-18-0395.

[52] Yanchun Peng, Alexander J. Mentzer, Guihai Liu, Xuan Yao, Zixi Yin, Danning Dong, Wanwisa Dejnirattisai, Timothy Rostron, Piyada Supasa, Chang Liu, César López-Camacho, Jose Slon-Campos, Yuguang Zhao, David I. Stuart, Guido C. Paesen, Jonathan M. Grimes, Alfred A. Antson, Oliver W. Bay-

field, Dorothy E. D. P. Hawkins, De-Sheng Ker, Beibei Wang, Lance Turtle, Krishanthi Subramaniam, Paul Thomson, Ping Zhang, Christina Dold, Jeremy Ratcliff, Peter Simmonds, Thushan de Silva, Paul Sopp, Dannielle Wellington, Ushani Rajapaksa, Yi-Ling Chen, Mariolina Salio, Giorgio Napolitani, Wayne Paes, Persephone Borrow, Benedikt M. Kessler, Jeremy W. Fry, Nikolai F. Schwabe, Malcolm G. Semple, J. Kenneth Baillie, Shona C. Moore, Peter J. M. Openshaw, M. Azim Ansari, Susanna Dunachie, Eleanor Barnes, John Frater, Georgina Kerr, Philip Goulder, Teresa Lockett, Robert Levin, Yonghong Zhang, Ronghua Jing, Ling-Pei Ho, Richard J. Cornall, Christopher P. Conlon, Paul Klenerman, Gavin R. Screaton, Juthathip Mongkolsapaya, Andrew McMichael, Julian C. Knight, Graham Ogg, and Tao Dong. Broad and strong memory CD4 + and CD8 + T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. *Nature Immunology*, 21(11):1336–1345, November 2020. ISSN 1529-2916. doi: 10.1038/s41590-020-0782-6. URL https://www.nature.com/articles/s41590-020-0782-6. Number: 11 Publisher: Nature Publishing Group.

[53] Annika Nelde, Tatjana Bilich, Jonas S. Heitmann, Yacine Maringer, Helmut R. Salih, Malte Roerden, Maren Lübke, Jens Bauer, Jonas Rieth, Marcel Wacker, Andreas Peter, Sebastian Hörber, Bjoern Traenkle, Philipp D. Kaiser, Ulrich Rothbauer, Matthias Becker, Daniel Junker, Gérard Krause, Monika Strengert, Nicole Schneiderhan-Marra, Markus F. Templin, Thomas O. Joos, Daniel J. Kowalewski, Vlatka Stos-Zweifel, Michael Fehr, Armin Rabsteyn, Valbona Mirakaj, Julia Karbach, Elke Jäger, Michael Graf, Lena-Christin Gruber, David Rachfalski, Beate Preuß, Ilona Hagelstein, Melanie Märklin, Tamam Bakchoul, Cécile Gouttefangeas, Oliver Kohlbacher, Reinhild Klein, Stefan Stevanović, Hans-Georg Rammensee, and Juliane S. Walz. SARS-CoV-2-derived peptides define heterologous and COVID-19-induced T cell recognition. *Nature Immunology*, 22(1):74–85, January 2021. ISSN 1529-2916. doi: 10.1038/s41590-020-00808-x. URL https://www.nature.com/articles/s41590-020-00808-x. Number: 1 Publisher: Nature Publishing Group.

[54] Jose Mateus, Alba Grifoni, Alison Tarke, John Sidney, Sydney I. Ramirez, Jennifer M. Dan, Zoe C. Burger, Stephen A. Rawlings, Davey M. Smith, Elizabeth Phillips, Simon Mallal, Marshall Lammers, Paul Rubiro, Lorenzo Quiambao, Aaron Sutherland, Esther Dawen Yu, Ricardo da Silva Antunes, Jason Greenbaum, April Frazier, Alena J. Markmann, Lakshmanane Premkumar, Aravinda de Silva, Bjoern Peters, Shane Crotty, Alessandro Sette, and Daniela Weiskopf. Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. *Science*, 370(6512):89–94, October 2020. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abd3871. URL https://science.sciencemag.org/content/370/6512/89. Publisher: American Association for the Advancement of Science Section: Research Article.

[55] Nina Le Bert, Anthony T. Tan, Kamini Kunasegaran, Christine Y. L. Tham, Morteza Hafezi, Adeline Chia, Melissa Hui Yen Chng, Meiyin Lin, Nicole Tan, Martin Linster, Wan Ni Chia, Mark I.-Cheng Chen, Lin-Fa Wang, Eng Eong Ooi, Shirin Kalimuddin, Paul Anantharajah Tambyah, Jenny Guek-Hong Low, Yee-Joo Tan, and Antonio Bertoletti. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature*, 584(7821): 457–462, August 2020. ISSN 1476-4687. doi: 10.1038/s41586-020-2550-z. URL https://www.nature.com/articles/s41586-020-2550-z. Number: 7821 Publisher: Nature Publishing Group.

Chapter 3

ENUMERATION OF ANTIGEN-SPECIFIC T CELLS AGAINST SARS-COV-2

3.1 HLA Haplotype Influence in Antigen-specific Cytotoxic T cell Responses against SARS-COV-2 Spike & PLpro Proteins

3.1.1 Introduction

In December 2019, an outbreak of unexplained pneumonia was recognized to be associated with an open market in Wuhan, China and reported to the World Health Organization (WHO) by mid-January. The causative agent of the disease outbreak, termed COVID-19, was noted to be a novel coronavirus [1]. Clinical features include fever, cough, and dyspnea with some individuals progressing to viral pneumonia and acute respiratory distress syndrome [2, 3]. The overall case fatality rate is estimated to be around 2.3%, but has also been reported to vary with age, sex [4], ethnicity [5], and comorbidities. These factors can influence immune responses against SARS-CoV-2 in both clear and subtle fashions. For example, certain autoimmune diseases, such as diabetes, can lead to altered immune responses against COVID-19 infections and more severe patient outcomes [6]. More subtle is the influence of HLA haplotype, which varies among different ethnic groups, and can dictate which T cell antigens elicit immune responses. For example, in North America, the HLA-A*02:01 allele is commonly expressed in Caucasian, Hispanic, African American, and Native American populations, while the A*24:02 allele is more commonly expressed within specific Native American and Asian populations. It has been documented that, for certain viral infections [7], including SARS 2002 [8], infection severity could be correlated with HLA haplotype. Here we report on the development and application of a method that allows for the rapid analysis and tracking of SARS-CoV-2 antigen-specific $CD8^+$ T cell populations in COVID-19 patients, for three different Class I HLA alleles.

We first describe the prediction of putative antigens derived from SARS-CoV-2 recognized by $CD8^+$ T cells, followed by the production of the single-chain trimer (SCT) libraries. We then describe results from the analysis of peripheral blood mononuclear cells (PBMCs) from COVID-19 participants using various assays to enumerate antigen-specific T cell populations. Finally, we validate the captured T

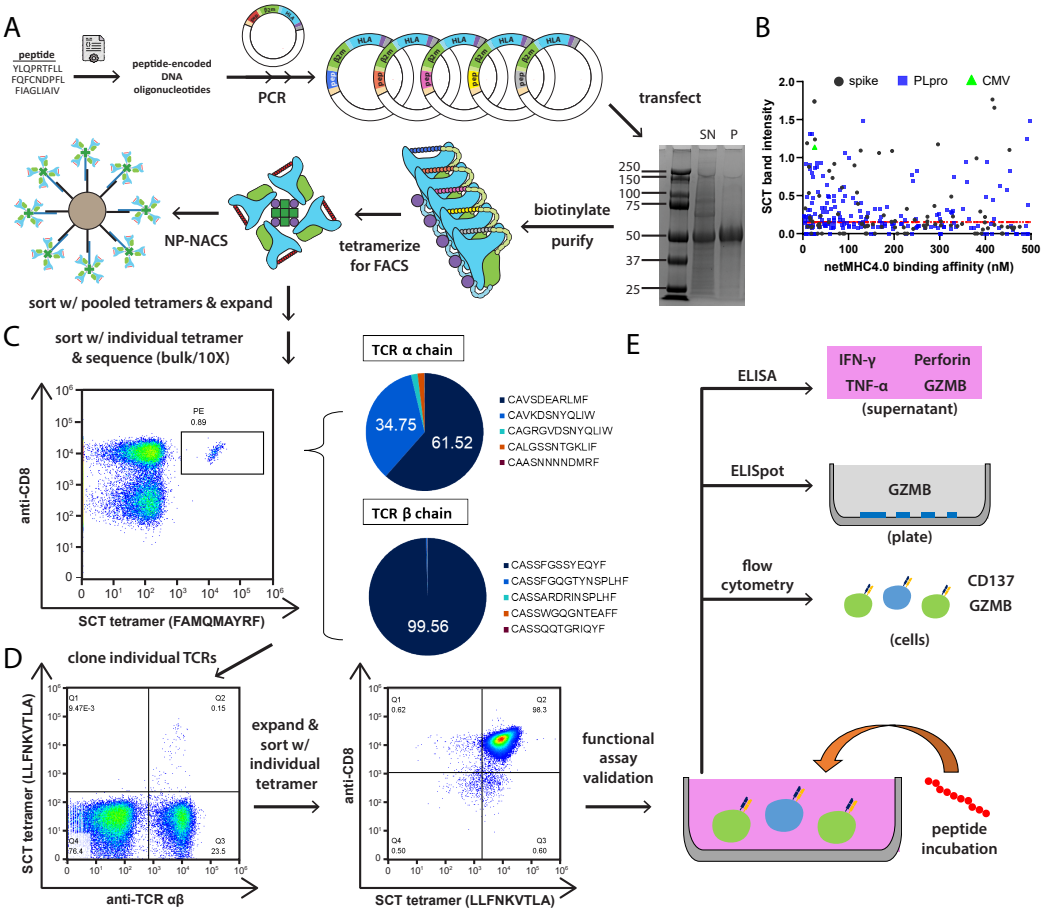


Figure 3.1: SCT-facilitated identification of antigen-specific TCRs
A. Schematic of SCT capture reagent production. The SCT platform allows for the production of pMHC libraries that may be multimerized for various downstream functional applications. The SDS-PAGE image (bottom right) depicts standard SCT proteins either directly from transfection supernatant (SN) or after HisTag purification (P). B. Scatterplot of SCT expression yield (band intensity) versus predicted binding affinity for each of the A*02:01 SARS-CoV-2 epitopes from spike protein or PLpro (Nsp3). "CMV" refers to the positive control A*02:01 SCT loaded with the peptide NLVPMVATV. C. SCT multimers are used in two rounds of PBMC sorting and expansion to identify antigen-specific T cells, whose TCRs are then sequenced. D. TCRs are cloned into $CD8^+$ T cells, purified by tetramer sorting, and expanded. E. Pure, antigen-specific T cell lines are functionally validated through various assays measuring cytokine and surface cell markers.

cell populations by cloning sequenced TCRs and demonstrating functional activity against specific SARS-CoV-2 epitopes.

3.1.2 Methods

3.1.2.1 Study design

The objective of this study was to explore the role of the antigen-specific adaptive immune response against SARS-CoV-2. To this end, pMHCs were designed in the form of a plasmid-encoded single-chain trimer comprising a candidate SARS-CoV-2-derived spike protein or Nsp3 epitope, β-2 microglobulin subunit of the MHC, and the human leukocyte antigen (HLA) subunit of the MHC. The optimized platform was utilized to express approximately 118 viable SCT constructs against the spike protein, and 75 against Nsp3. We then incorporated 88 of the spike SCTs and 75 of the Nsp3 SCTs as tetramers into our nanoparticle nucleic acid cell sorting (NP-NACS) system to generate high-avidity TCR capture agents. Last, NP-NACS was applied toward the identification and analysis of antigen-specific T cells derived from blood draws of eight COVID-19 participants covering three HLA alleles of interest, as well as from four HLA-matched healthy donor PBMC samples.

3.1.2.2 Study approval & sample collection

All human samples (blood) were obtained after approval from the Providence St. Joseph's Health IRB and participant-written informed consent, in accordance with 45 CFR 46, as part of the Swedish Institute's INCOV trial to study COVID-19 participants. Peripheral blood mononuclear cells (PBMCs) were isolated and cryopreserved by Bloodworks Northwest (North Seattle Donor Center, Seattle). They were collected from participants at up to three timepoints: T1 (diagnosis), T2 (4-5 days after diagnosis), and T3 (convalescence). 186 unique participant samples were submitted for HLA haplotyping (Cisco Genetics) (**Fig. H.1 & H.2**). Among all samples, we identified A*02:01, A*24:02, and B*07:02 alleles as the most prevalent, and therefore filtered for participants with these alleles for further analysis using SCT constructs.

3.1.2.3 SCT plasmid construction & protein expression

Single-chain trimer pMHC reagents consist of a peptide covalently linked to the β-2 microglobulin subunit, which is covalently linked to the HLA subunit (**Fig. 3.1A**). In order to build SARS-CoV-2 SCT libraries, identified peptides were encoded into primers for insertion into template SCT plasmids (**Sections 2.1.3.1, 2.1.3.2,**

Appendix B). The peptide-substituted SCT plasmid libraries were subsequently transfected into Expi293 cells for approximately four days. Secreted SCT proteins were collected from the supernatant, biotinylated, and purified by HisTag column.

3.1.2.4 SCT multimer assays

SCT monomer libraries can be biotinylated and incorporated into standard tetramer scaffolds for various downstream assays. The SCT tetramers can then be assembled onto the surface of magnetic nanoparticles to form pMHC-nanoparticle (pNP) libraries for hemocytometry fluorescence microscopy assays (**Fig. 3.1A**), similar to our previous reports [9, 10], or used in standard flow cytometry assays (**Fig. 3.1C**). Furthermore, these SCTs can be used with Immudex Klickmer reagents to form dextramers for use in 10X single-cell sequencing experiments. Each of the three platforms for antigen-specific T cell analysis has its own advantages and disadvantages. pNP libraries are advantageous in that all analysis is done in solution, thus avoiding risks from aerosolized COVID-19 patient biospecimens. Our prior work using the NP-NACS system highlights the enhanced sensitivity of this platform, which allows for its use with non-expanded $CD8^+$ T cells directly extracted from PBMCs. However, enumeration of TCR sequences from captured cells is difficult, and requires further microfluidic adaptations to enable single-cell sequencing [10, 11]. Compared to NP-NACS, flow cytometry assays making use of SCT tetramers are higher throughput and can be combined with bulk sequencing assays to identify antigen-specific TCR sequences, but the degree of specific binding by tetramers is more difficult to resolve as one cannot visualize tetramer staining at the microscopic level. Dextramer/10X assays are utilized in a similar manner to tetramers for flow cytometry and allow for antigen-pairing of TCR sequences, but compared to the other approaches is relatively more expensive and lower throughput, enabling analysis of only up to 10,000 cells per run. In order to maximize confidence that sequenced TCRs are derived from antigen-specific T cells, the latter two assays worked with $CD8^+$ T cells which had been expanded after either SCT capture or peptide stimulation.

3.1.2.5 Production of cysteine-modified streptavidin-DNA (SAC-DNA) conjugates

The SAC-DNA conjugate was produced following a previously published protocol [12]. Briefly, SAC was first expressed from the pTSA-C plasmid containing the SAC gene (Addgene). Before conjugation to DNA, SAC (1 mg/ml) was buffer exchanged to PBS containing Tris(2-Carboxyethyl) phosphine hydrochloride (TCEP, 5 mM) using Zeba desalting columns (Pierce). Then 3-N-Maleimido-6-hydraziniumpyridine hydrochloride (MHPH, 100 mM, Solulink) in DMF was added to SAC at a molar excess of 300:1. In the meantime, succinimidyl 4-formylbenzoate (SFB, 100mM, Solulink) in DMF was added to 5'-amine modified ssDNA (500 μM) in a 40:1 molar ratio. After reacting at rt for 4 hours, MHPH-labeled SAC and SFB-labeled DNA were buffer exchanged to citrate buffer (50 mM sodium citrate, 150 mM NaCl, pH 6.0), and then mixed at a 20:1 ratio of DNA to SAC to react at RT overnight. SAC-DNA conjugate was purified using the Superdex 200 gel filtration column (GE health) and concentrated with 10K MWCO ultra-centrifuge filters (Millipore).

3.1.2.6 COVID SCT pNP library construction

Streptavidin-coated NPs (500 nm radius, Invitrogen Dynabeads MyOne T1) were prepared according to the manufacturer's recommended protocol for biotinylated nucleic acid attachment. These NPs were mixed with barcoded biotin-ssDNA (100 μM) at 1:20 volume ratio to obtain NP-DNA. Excess DNA was removed by washing the NPs three times. In parallel, the SCT monomer library was added to SAC-DNA at a 4:1 ratio to form the SCT tetramer-DNA. To generate fluorescent pNPs, equimolar amounts (in terms of DNA ratio) of NP-DNA and pMHC tetramer-DNA were hybridized at 37°C for 20 min, along with 0.25 μl of 100 μM ssDNA bound to AlexaFluor 750, AlexaFluor 488, or Cy5 (IDT-DNA), and washed once with buffer (0.1% BSA, 2 mM MgCl2 PBS). The use of three dyes allows for multiplexing of up to three unique antigen pNPs per analysis. Typically, each NP-barcoded NACS analysis of <100,000 cells uses 2.5 μL of stock NPs (28.2 million particles total) per library element.

3.1.2.7 Preparation and isolation of CD8$^+$ T cells from PBMC suspensions

PBMCs were thawed and incubated in RPMI 1640 media supplemented with 10% FBS and IL2 (100 U/mL) for overnight recovery at 37 °C, 5% CO_2. Recovered cell viability was measured at >95% for all samples. CD8$^+$ T cell population was negatively selected using the CD8$^+$ T Cell Isolation Kit (Miltenyi Biotec, Bergisch Gladbach, Germany). Briefly, recovered cells were incubated with a biotinylated antibody cocktail that captures CD8$^-$ cells in PBMCs followed by streptavidin-coated microbeads. The untouched CD8$^+$ T cells were separated in a 15 mL Falcon tube using an LS column. The tube containing CD8$^+$ T cells was then centrifuged at 500 g for 5 minutes and the pellet was re-suspended in PBS buffer.

For the multiplex cell labeling, CD8$^+$ T cells were individually stained with Calcein Blue, AM (Thermo Fisher Scientific) or CellTracker™ Orange CMRA Dye (Thermo Fisher Scientific) at the concentration of 4 μM and 400 nM, respectively. After incubation for 10 minutes at 37°C under 5% CO_2, cells were washed twice with PBS and re-suspended in a cell suspension buffer (0.1% BSA, 2 mM $MgCl_2$ in PBS).

3.1.2.8 Identification of antigen-specific CD8$^+$ T cells by NP-NACS

The pNP library was combined into groups of 3 pNPs, with each pNP element in the group stained with one of three barcode dyes. From each pNP group, 7.5 μl was incubated with each aliquot of stained CD8$^+$ T cells at RT for 30 minutes. Antigen-specific cells were enriched by magnetic pulldown and re-suspended into 6 μl of 0.1% BSA 2 mM $MgCl_2$ PBS buffer. Captured cells were then loaded into a 4-chip disposable hemocytometer (Bulldog-Bio). The entire area in the hemocytometer chip was imaged to obtain the total pulldown cell number. Identification of antigen-specific T cells, including the detection and exclusion of non-specific binding events, was conducted with cellSens Olympus software and R programming language.

3.1.2.9 Tetramer binding flow assay

For use of SCTs in tetramer format for flow assays, see prior protocols (**Section 2.2.2.7**). Use of SCTs in dextramer format for 10X also followed similar protocols, where streptavidin was replaced with Immudex dextramer/Klickmer reagents, and downstream protocols for staining and washing were identical. For 10X single-cell

sample submission, we adhered to manufacturer's recommendations and protocols for best results.

3.1.3 Results & Discussion

3.1.3.1 SCT libraries enable rapid discovery of potential SARS-CoV-2 epitopes

To broadly survey antigen-specific $CD8^+$ T cell response against SARS-CoV-2, PBMC samples of hospitalized COVID participants were collected from blood draws across three timepoints, starting from diagnosis (T1) to 4-5 days post-diagnosis (T2) to convalescence (T3). ELISpot assays based on stimulation with peptide pools of SARS-CoV-2 structural proteins showed significantly increased IFN-γ production from two COVID participant PBMC samples versus health donor controls (**Fig. 3.2**). Among the INCOV participants, the increased IFN-γ signature primarily was detected at T2, indicating that an epitope-specific response against SARS-CoV-2 developed over time after infection.

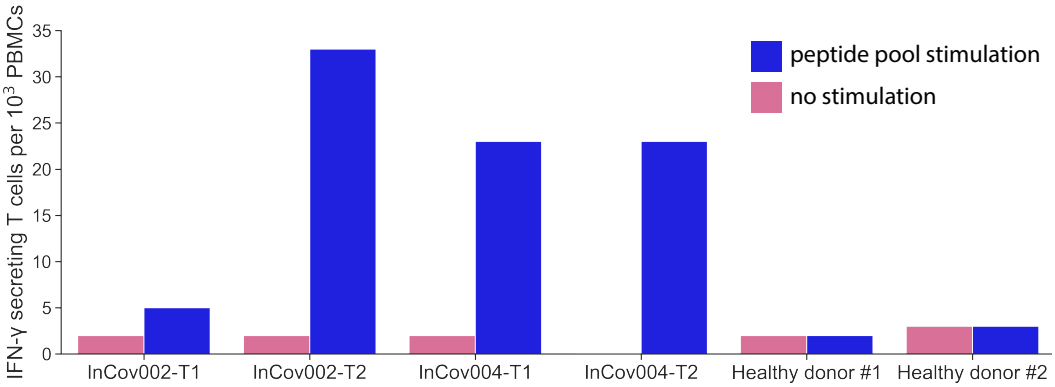


Figure 3.2: Peptide pool stimulation of PBMCs elicits cytokine response ELISpot assay of IFN-γ secreting $CD8^+$ T cells from PBMCs of COVID-19 participants and healthy donors stimulated with peptide pools derived from SARS-CoV-2 structural proteins.

Recent reports have indicated that the SARS-CoV-2-specific T cell repertoire of hospitalized COVID patients consists of a large proportion of the exhausted phenotype and overall low $CD8^+$ T cell counts correlated with disease severity. To enumerate the epitope landscape of SARS-CoV-2-specific $CD8^+$ T cells, including those which are potentially rare or exhausted, we opted to initiate our search by probing the PBMCs directly instead of relying on a stimulation/expansion-based

method. This approach prevents any potential bias against antigen-specific T cells with non-expandable phenotypes, which could skew the distribution of detected epitopes. To account for the absence of an expansion step, capture sensitivity was maximized using our NP-NACS platform, which affixes thousands of tetramers onto magnetic particles, enabling highly sensitive magnetic isolation and detection of clonal $CD8^+$ T cells at frequencies as low as 0.001% [9]. To capture as many antigen specificities as possible among unexpanded $CD8^+$ T cells, capture breadth was broadened using our SCT platform to generate hundreds of pMHCs. 9- to 11-mer peptide sequences from a protein of interest were entered into the NetMHC4.0 binding prediction algorithm [13]. For the spike protein, we identified 96, 33, and 51 peptides for HLA-A*02:01, B*07:02, and A*24:02 alleles, respectively, with 500 nM or stronger binding affinity (**Tables I.1, I.2, & I.3**). These matched to varying degrees with published lists of putative antigens [14–16].

This filtered peptide list was used to develop pMHC-encoded plasmids using our SCT platform. The distribution of SCT protein expression for epitopes along the spike protein domain map were unique for each haplotype. A*02:01 SCTs showed relatively heterogeneous levels of expression for epitopes throughout all domains except TM (weak expression) (**Fig. 3.3**). B*07:02 SCT expression showed preference for NTD, S1/S2 cleavage site, and parts of the S2 subunit, while highly expressed A*24:02 SCTs appeared to be concentrated around NTD, RBD, and TM regions (**Fig. I.1**). These distributions are partially skewed by artificial selection bias due to our use of NetMHC4.0 as a filtering step prior to SCT production. Therefore, the expression of these SCTs to some degree are a reflection of the prediction strength of the algorithm. Additionally, the results may be seen as an interpretation of the biological differences that exist across the HLA alleles. Differences in hydrophilic/hydrophobic preference within each HLA's binding will bias the stability of each pMHC construct for certain peptide motifs found in the spike domains.

3.1.3.2 SCT multimers can identify antigen-specific T cells from healthy and COVID-19 donors

We first asked whether immunogenic epitopes derived from the spike protein were shared among HLA-matched COVID participants. The highest expressing SCTs from each of the three libraries were utilized as NP-NACS reagents to identify antigen-specific T cells among COVID PBMCs from two participants and at least one healthy control per haplotype (**Fig. 3.4**). For each HLA haplotype, the NP-

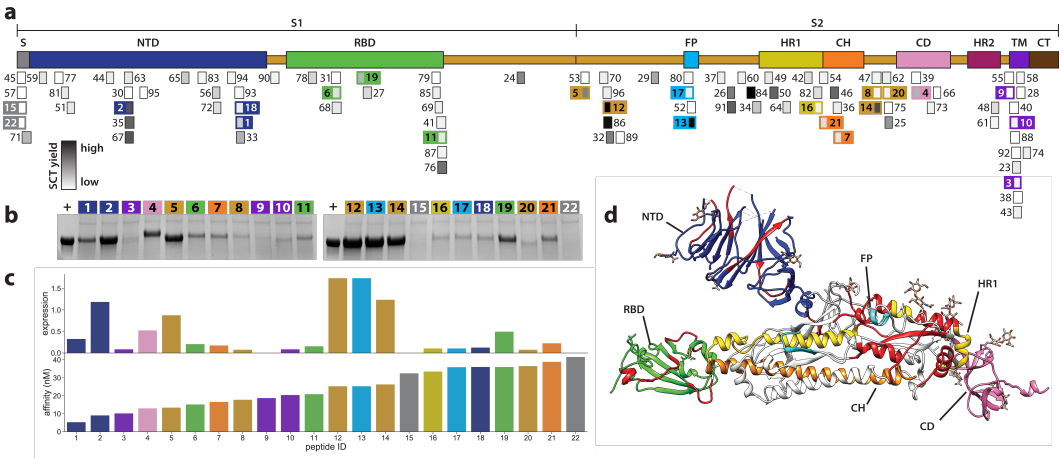


Figure 3.3: Expression of SCTs for A*02:01 SARS-CoV-2 spike protein epitopes A. Schematic of the spike protein domains. S, signal sequence; NTD, N-terminal domain; RBD, receptor binding domain; FP, fusion peptide; HR1, heptad repeat 1; CH, central helix; CD, connector domain; HR2, 5 heptad repeat 2; TM, transmembrane domain; CT, cytoplasmic tail; subunits denoted by S1 and S2. Shaded boxes denote relative position and expression yields of SCT proteins. Peptide ID numbers are indexed in descending order of predicted binding affinity. B. Reduced SDS-PAGE of a subset of spike epitope SCTs from (A). Lane number indicates peptide ID, with domain-matched background color. +, purified WT1 SCT. C. Barplots comparing relative SCT yield (quantified against WT1 SCT lanes) and predicted affinity for each peptide from the subset in (B). D. Crystal structure of spike monomer (24). Domain colors match those of the regions in schematic (A); S1 and S2 subunit backbones in white. Amino acids containing the 30 A*02:01 tested epitopes of Fig. 3A in red.

NACS assay was able to identify antigen-specific T cells against a shared subset of epitopes per library, regardless of disease state of the samples. However, COVID participants contained significantly higher frequencies of antigen-specific T cells against each of the top epitopes relative to their healthy controls. These shared immunodominant epitopes were detected at both time points of sample collection for the COVID participants with variations in relative frequency for each, indicating perhaps fluctuations in clonotype expansion against each epitope throughout the immune response. Here, our findings suggest that immunodominant epitopes are present among individuals of the same HLA haplotype, even among healthy controls, and that the degree of detection evolves throughout the course of disease state.

Although antigen-specific T cells were detected by NP-NACS, the degree to which these cells can be induced by those epitopes to produce an actual immune response remained in question. Five peptides comprising epitopes detected from either

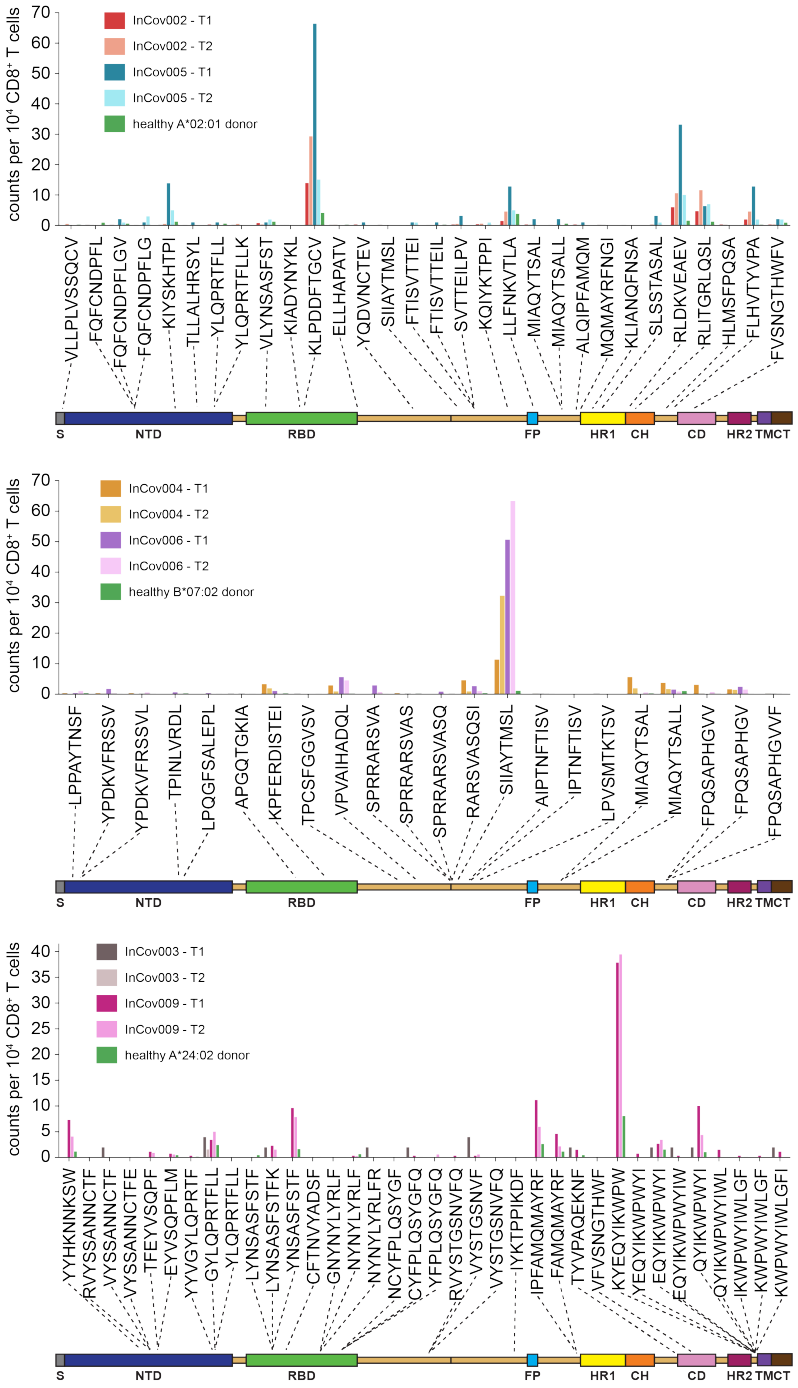


Figure 3.4: Spike protein-specific T cell populations from COVID-19 participants via NP-NACS

Peptides are plotted according to position on the spike protein, with dashed lines pointing to position along the domain map. In each plot, counts are from two COVID-19 participants and one HLA-matched donor sample (top: A*02:01, middle: B*07:02, bottom: A*24:02).

the A*02:01 or B*07:02 assay were synthesized and used in an ELISpot assay to stimulate HLA-matched PBMCs. In the A*02:01 assay (**Fig. 3.5A**), IFN-γ secretion was upregulated upon exposure to the peptides in both disease and healthy PBMCs. However, there was variation in the degree of IFN-γ upregulation per peptide. RLDKVEAEV induced the strongest response in the INCOV PBMCs, whereas for healthy PBMCs, KLPDDFTGCV elicited the strongest response, and to a greater extent when compared to other peptide responses seen in the INCOV samples. The fact that the KLPDDFTGCV SCT captured the highest frequency of cells in NP-NACS but gave a significantly reduced IFN-γ response in INCOV samples, while the healthy donor produced opposite results, indicates that this peptide perhaps is immunogenic but might cause T cell exhaustion in a disease state. We performed a similar assay for B*07:02 PBMCs using another set of peptides (**Fig. 3.5B**). Here we saw that the healthy B*07:02 donor PBMCs had no response to stimulation by any peptide, while the INCOV PBMCs secreted IFN-γ only with peptide stimulation. However, we did not expect KLPDDFTGCV to induce IFN-γ secretion for these PBMCs, as it was a predicted binder only to A*02:01 HLA alleles. A deeper HLA analysis of the INCOV-004 sample revealed that this participant also was positive for A*02:01, so activation by this peptide was expected. INCOV-006, however, did not possess the A*02:01 haplotype. One speculation to reconcile results with HLA haplotypes may be that the KLPDDFTGCV peptide can presented by this participant's other HLA alleles.

We next wanted to observe whether other regions of the SARS-CoV-2 proteome could elicit strong immune responses from patients in the form of higher antigen-specific T cell counts. As reported in other virus studies, non-structural proteins tend to be preferential for $CD8^+$ T cell activation. This finding, if applicable to the context of SARS-CoV-2, would be highly informative towards targeted vaccine developments. One such domain of interest, Nsp3, encodes for papain-like protease (PLpro), which has been identified in other coronavirus strains to play a significant role in the early stages of the infection cycle, processing other non-structural elements that are responsible for infection and assembly of structural virus elements. As such, Nsp3 is expressed much earlier than structural elements such as the spike protein. Therefore, we hypothesized that Nsp3 epitopes might be also be surveyed by the immune system earlier than epitopes derived from structural proteins. 191 Nsp3 peptide-encoded HLA-A*02:01 SCT plasmids were produced, approximately 100 of them expressed to a sufficient degree for biotinylation and tetramerization, and the top 75 expressed SCTs were utilized in NP-NACS to identify antigen-specific

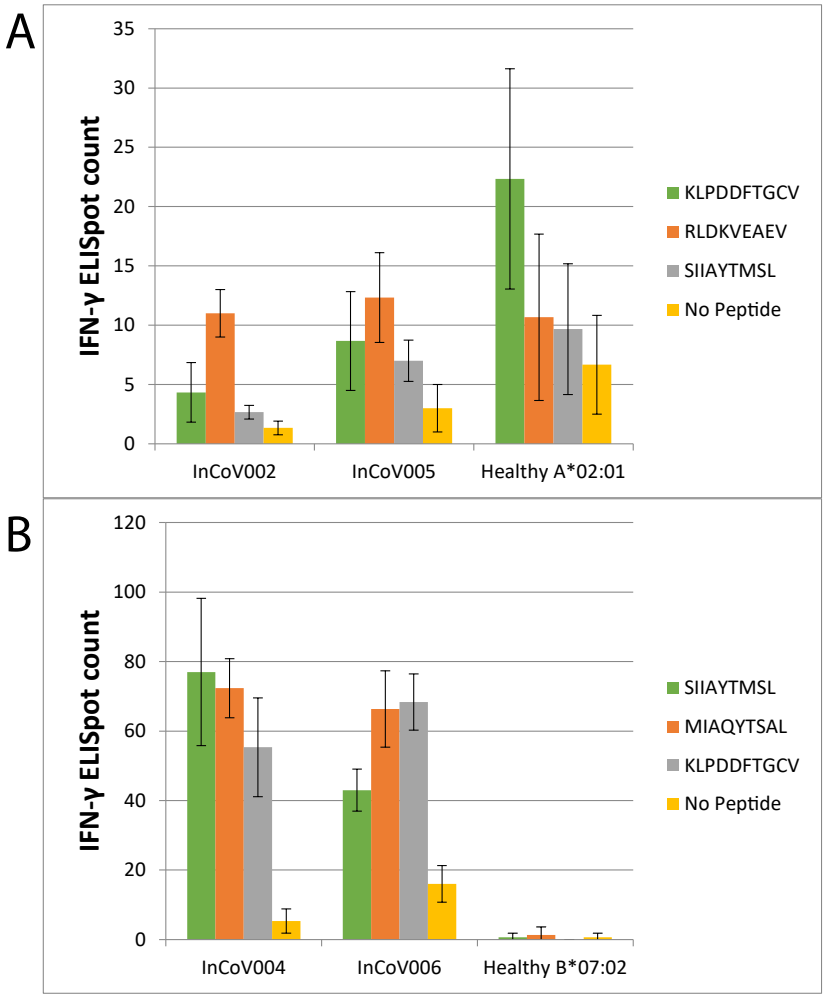


Figure 3.5: SARS-CoV-2 spike epitopes induce cytokine secretion in HLA-matched PBMCs

Peptides identified to be immunogenic from the NP-NACS assay were synthesized and used to stimulate HLA-matched PBMCs from InCoV participants and healthy donors for HLA-A*02:01 (A) and HLA-B*07:02 (B).

T cells in two COVID participants and two healthy controls (**Fig. 3.6 & Table I.4**). Again, we observed that both healthy and COVID PBMCs showed reactivity to the same epitopes. However, the relative counts for PLpro epitopes are much higher than for spike epitopes. Surprisingly, for some epitopes, healthy PBMCs gave just as high of a response. This finding implies perhaps prior exposure to coronavirus strains harboring similar epitopes.

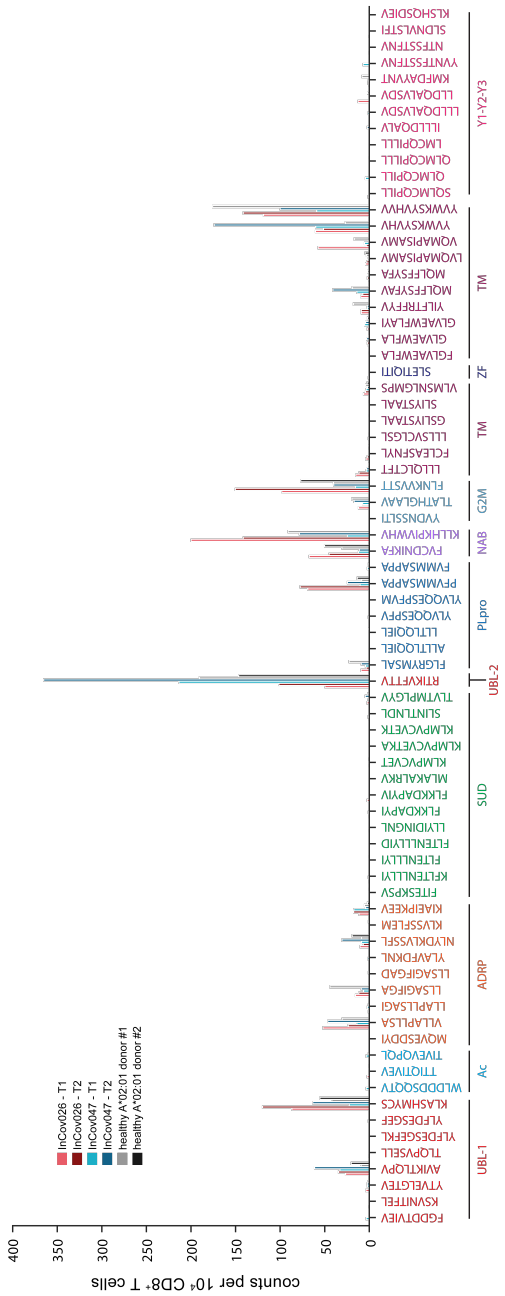


Figure 3.6: PLpro-specific T cell populations from A*02:01 COVID-19 participants via NP-NACS

Peptides are plotted along x-axis according to relative position on nsp3 protein and color-coded by nsp3 subunit (UBL: ubiquitin-like domain, Ac: Glu-rich acidic-domain, ADRP: ADP-ribose-1'-phosphatase domain, SUD: SARS unique domain, PLpro: papain-like protease, NAB: nucleic acid binding domain, G2M: marker domain, TM: transmembrane domain, ZF: zinc finger domain, Y1-Y2-Y3: Y domains preceding PLpro cleavage site).

3.1.3.3 SCTs enable high-throughput discovery of SARS-CoV-2-specific TCR sequences

While the NP-NACS platform allowed us to rapidly identify immunogenic antigens from primary $CD8^+$ T cells, we needed to enumerate their TCR sequences by flow cytometry to provide additional functional validation. Without the additional avidity conferred by the NP-NACS nanoparticle scaffold, we expected tetramer/dextramer binding assays to have some inherent degree of non-specific binding. This would render identification of antigen specificity difficult when working with primary $CD8^+$ T cells due to their low frequency and generally lower cell quality. Therefore, we first sorted for primary $CD8^+$ T cells using SCT tetramer pools for each patient (tetramer pools consisted of all SARS-CoV-2 SCTs synthesized matching the participant's HLA haplotype) (**Fig. 3.1C**). Each of the sorted populations were then expanded for approximately two weeks to improve quantity and viability. The cells were subsequently sorted by individual SCT tetramers within their respective libraries, such that we could associate each sorted population of TCR clonotypes with a targeted antigen. NGS bulk sequencing of the samples revealed antigen-specific populations against a subset of spike and PLpro antigens across most patients (**Fig. 3.7**). Of the 21 unique peptides which had detectable T cell populations, eight of them were found across multiple patients. Two of the patient samples had no cells captured by bulk sequencing after the expansion process, perhaps due to poor viability or biased expansion of non-specific T cells after low counts of tetramer-binding cells were collected during the pooled tetramer sorting step.

To complement the bulk sequencing approach, we wanted to perform single-cell sequencing on the expanded T cell populations to identify any prevalent clonotypes that may have been missed. The expanded cells were stained with a DNA hashtag to encode patient identity. Then, they were stained with a designated set of SCT dextramers, each containing an antigen-encoded DNA barcode. After excess dextramers were washed, the stained T cells from multiple patients were combined together and submitted for 10X single-cell sequencing. To assess the quality of SCT capture, we quantified dextramer binding frequency and heterogeneity. We first sorted the 10X data to identify the top 20 clonotypes which had the highest frequency of homogeneous dextramer binding (signal only from one unique dextramer barcode per cell), encompassing a frequency range of 24 to 959 antigen-specific cells detected against the dominant dextramer per clonotype (**Fig. 3.8**). The dextramer IDs of these 20 clonotypes were traced back to their associated SCT identities to reveal

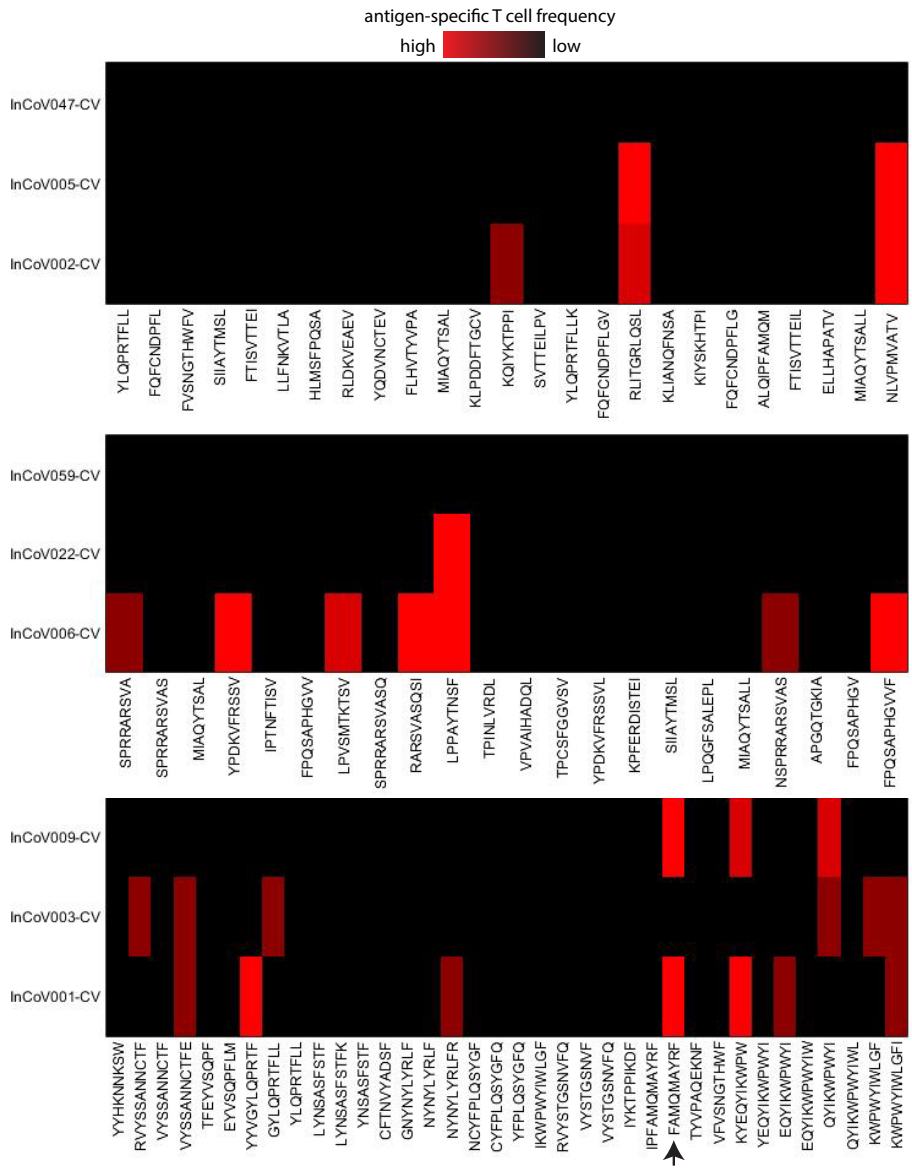


Figure 3.7: Enumeration of expanded antigen-specific T cell populations from COVID-19 participants via bulk sequencing

Frequencies of antigen-specific T cell populations identified by individual tetramer sorting from expanded T cells for COVID-19 participants (y-axis) of three HLA alleles (top, A*02:01; middle, B*07:02; bottom, A*24:02). The example flow and TCR data shown in **Fig. 3.1C** is obtained from InCoV001-CV sample for the peptide FAMQMAYRF (indicated by arrow).

specificity to six unique epitopes across A*02:01 and B*07:02. Five of the six epitopes are derived from spike protein, and one from PLpro. For each clonotype, cells with heterogeneously bound dextramers (non-specific) displayed a dominant dextramer signal derived from the same SCT as that of the homogeneously bound

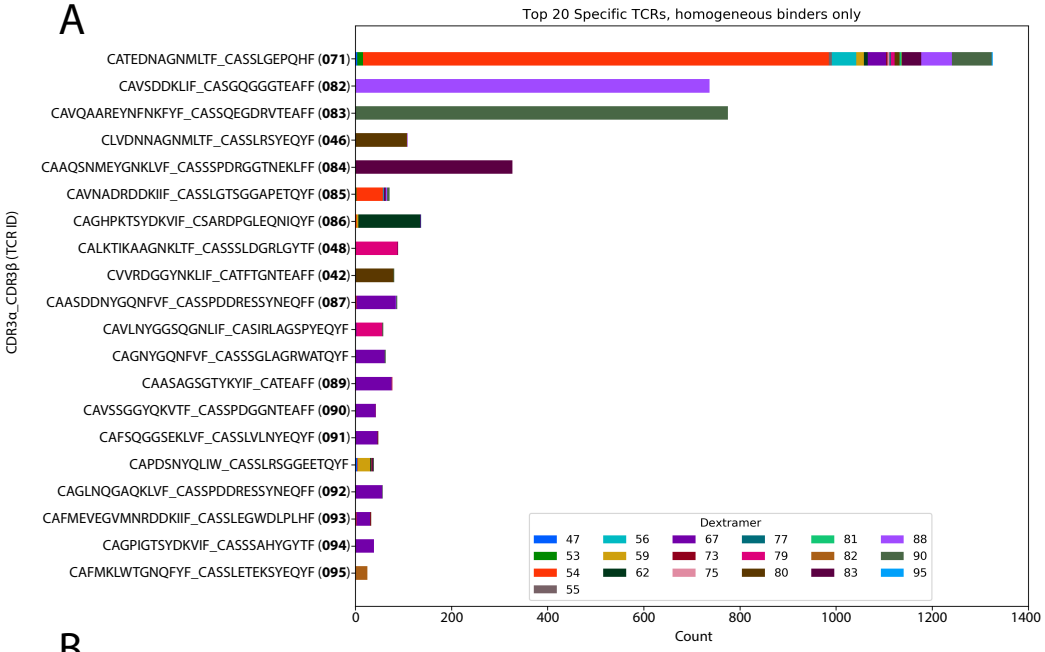
cells (**Fig. J.1**), but this signal comprised a significantly smaller fraction of the total dextramer signal. This indicates that our expansion step, followed by dextramer signal filtering, allows for successful reduction of background noise caused by non-specific dextramer binding. A comparison of the captured TCR data from single-cell against bulk sequencing revealed an overlap of six TCR clonotypes. The identified SCT specificity from five of these six clonotypes, whether in tetramer format (bulk sequencing) or dextramer format (10X single cell), were in agreement.

3.1.3.4 Identified TCRs are functionally responsive against SARS-CoV-2 peptides

In order to functionally validate our TCRs, sequencing results from bulk and 10X single-cell methods were sorted by prevalence, and 86 unique SARS-CoV-2-specific TCRs were selected for cloning into primary $CD8^+$ T cells by CRISPR/Cas9 transduction (**Figs. 2.28 & 3.1D; Tables K.1, K.2**). In order to thoroughly scan the most prevalent clonotypes for peptide specificity, several of the selected TCR clonotypes consist of different combinations of α/β pairs for cells in which dual TCR receptors were detected (e.g. TCR 087 & 092 share the same TCR β chain). The transduced T cells were sorted with SCT tetramers of corresponding antigen-specificity, and expanded for at least two weeks to generate cell lines. Of the 86 TCR sequences, we were able identify at least 13 which could specifically bind to SCT tetramers after expansion (**Fig. 3.9**). The lack of strong tetramer binding by the other T cell lines could be explained by the following causes: 1) non-productive TCR pairs derived from cells with dual TCRs; 2) collection of background cells from initial sorting of T cells from PBMCs via 10X or bulk method; 3) biased expansion of non-productive T cells. We found that a larger proportion of 10X-derived TCR sequences were productive versus bulk-derived TCR sequences, which is not surprising due to enhanced precision of the single-cell sequencing approach.

Initial functional validation of TCRs 001 and 002, which were obtained from healthy donors, demonstrated that peptide stimulation could induce CD137 expression (**Fig. 3.10**). This indicates that the TCRs identified are indeed capable of binding to biological pMHCs and inducing downstream activation signals. Furthermore, ELISA, ELISpot, and flow cytometry assays demonstrated that peptide-stimulated T cells could be induced to release cytokines (specifically, TNF-α was observed but not IFN-γ) and proteases (granzyme B), characteristic of a cytotoxic response from $CD8^+$ T cells upon activation (**Fig. 3.10**). We anticipate performing similar functional

A



B

Patient	Status	HLA	Antigen	Dextramer
InCoV003-CV	T3	A*24:02	KWPWYIWLGF	47
InCoV047-CV	T3	A*02:01	FCLEASFNYL	53
InCoV047-CV	T3	A*02:01	MLAKALRKV	54
InCoV047-CV	T3	A*02:01	YLQPRTFLLK	55
InCoV047-CV	T3	A*02:01	KQIYKTPPI	56
InCoV005-CV	T3	A*02:01	MLAKALRKV	59
InCoV005-CV	T3	A*02:01	RLITGRLQSL	62
InCoV002-CV	T3	A*02:01	MLAKALRKV	67
InCoV002-CV	T3	A*02:01	RLITGRLQSL	73
InCoV002-CV	T3	A*02:01	KQIYKTPPI	75
InCoV006-CV	T3	B*07:02	FPQSAPHGVVF	77
InCoV006-CV	T3	B*07:02	LPPAYTNSF	79
InCoV006-CV	T3	B*07:02	RARSVASQSI	80
InCoV006-CV	T3	B*07:02	YPDKVFRSSV	81
InCoV006-CV	T3	B*07:02	SPRRARSVA	82
GB17457	healthy	A*02:01	RLITGRLQSL	83
GB17457	healthy	A*02:01	LLFNKVTLA	88
GB18622	healthy	A*02:01	RLITGRLQSL	90
GB18622	healthy	A*02:01	LLFNKVTLA	95

Figure 3.8: Enumeration of expanded antigen-specific T cell populations from COVID-19 participants via 10X single-cell sequencing, filtered for homogeneously binding cells

A. Frequencies of antigen-specific T cell populations among the top 20 most common detected clonotypes, identified by multiplexed dextramer sorting from expanded T cells for COVID-19 participants. B. Table of dextramers used per patient sample, along with associated participant sample, antigen, and HLA allele of SCT. "Dextramer" column refers to the ID of the dextramer used for the legend in (A).

assays for the entire library of TCR-transduced cells in order to confirm peptide specificity. Further functional assays, such as cell-killing assays using APCs trans-

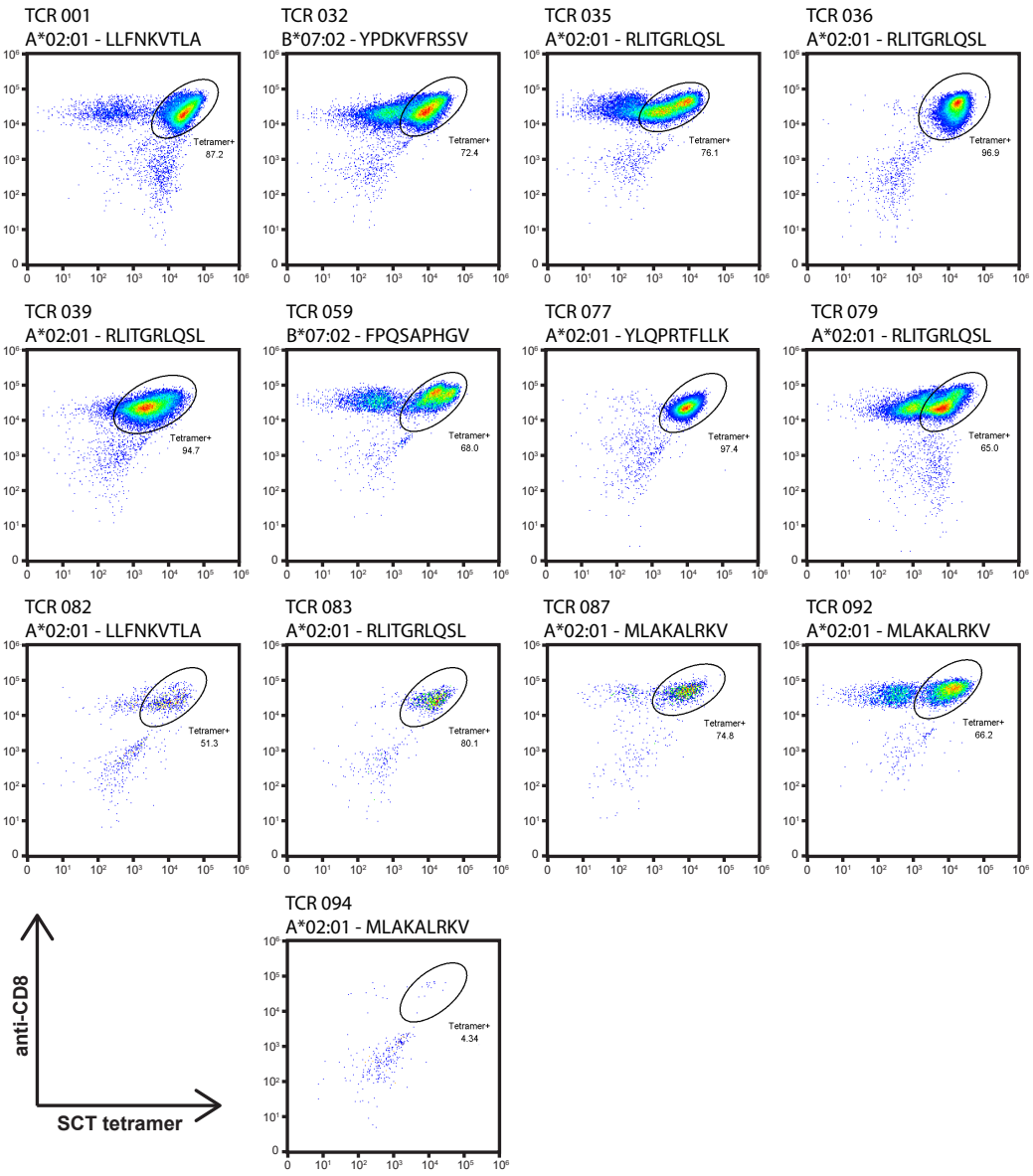
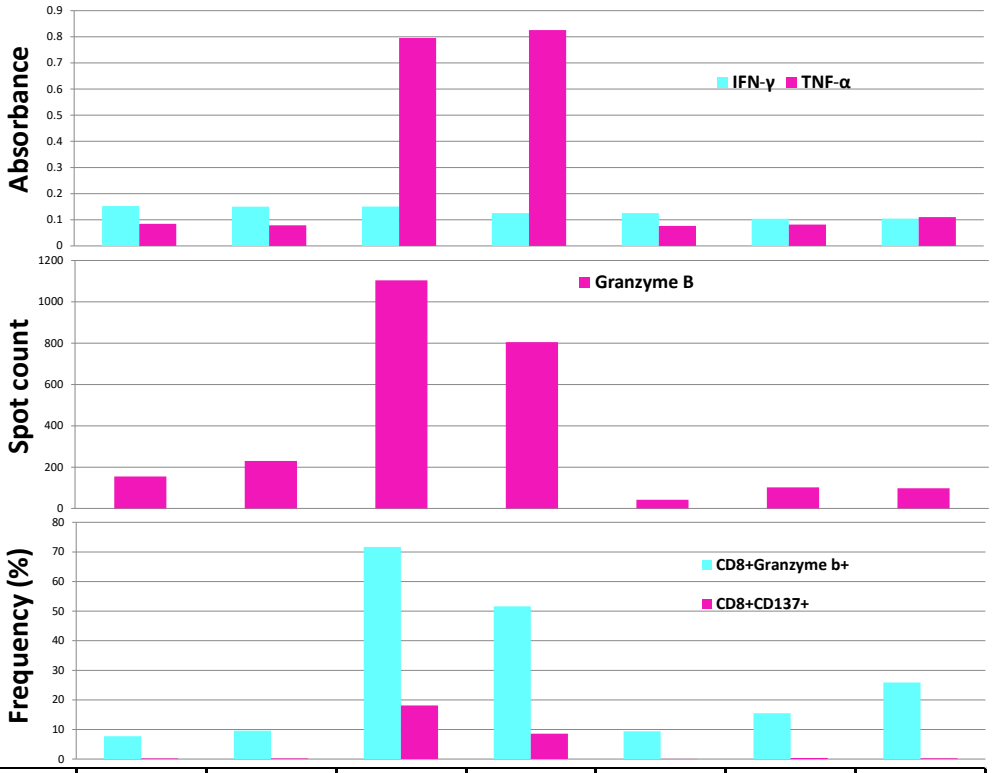


Figure 3.9: Transduced TCRs are specific to SARS-CoV-2 antigens
TCRs obtained by 10X or bulk sequencing methods from healthy donor or COVID-19 participant-derived T cells were transduced into HLA-matched $CD8^+$ T cells and selectively expanded after SCT tetramer binding to generate cell lines. Shown here are the tetramer binding results of the expanded cells, demonstrating SCT specificity and purity of the cell lines. Each cell line's TCR number corresponds to the TCR ID found in **Tables K.1-K.2**. Each cell line's HLA-restriction and antigen specificity is also labeled.

duced with spike protein or PLpro, will also be performed to assess cytotoxicity of these TCRs. Because these TCRs have been isolated using large SCT libraries against PBMCs from participants of different HLA haplotypes, we anticipate seeing

a spectrum of T cell responses that may highlight peptide-dependent, SARS-CoV-2 protein domain-dependent, or HLA-dependent immunogenicity.



TCR	Untransduced	Untransduced	TCR-001	TCR-002	Untransduced	TCR-001	TCR-002
Peptide 1: LLFNKVTLA	+	-	+	-	-	-	-
Peptide 2: RLITGRLQSL	-	+	-	+	-	-	-

Figure 3.10: Functional validation of TCR-transduced T cells
T cells transduced with TCRs 001 & 002 corresponding to peptides 1 & 2 (table), respectively, are functionally assessed after 16-hour overnight peptide stimulation. Top: ELISA assay measuring cytokine release. Middle: ELISpot assay counting cells with granzyme B expression. Bottom: Flow cytometry assay measuring percentage of cells activated ($CD137^+$) and cytotoxic (granzyme B^+).

3.2 Future Directions

3.2.1 SARS-CoV-2 Proteomic Analysis

In the work above, we have demonstrated the capacity for the SCT platform to be utilized in a timely manner to identify immunogenic antigens from SARS-CoV-2 proteins. However, some of the primary limitations in this study are that the libraries assessed were restricted to a few hundred total SCTs encompassing only two pro-

teins of SARS-CoV-2 (**Tables I.1, I.2, I.3, I.4**) and that the number of participants studied was small. It would be prudent, especially in the context of a pandemic, to develop a platform capable of assessing all potential epitopes from the SARS-CoV-2 proteome and to have the capability to screen more than just a handful of participants for T cell response with peptide-level resolution. Toward the former objective, this required the implementation of higher throughput approaches to each step of SCT production in our workflow to accommodate generation of 1,000+ proteins (**Appendix B-D**). Primarily, reduction of error in plasmid production, automation of plasmid sequence verification, and adaptation of 96-sample instruments across bottlenecking steps (PCR purification, transfection, SCT purification, protein desalting, etc.) are some of the major changes we adopted. Toward the latter objective, significant changes to our 10X single-cell sequencing methodology were required. Given the cost limitations of ordering a large multiplexed set of DNA-barcoded dextramers (approximately $50,000 for a set of 64 unique dextramers), multiple rounds of 10X sequencing are required to enable assessment of hundreds of SCTs covering the SARS-CoV-2 proteome. Furthermore, we needed to combinatorially assign participant-associated DNA hashtags to our samples, such that one could accommodate on the order of 10-20 participant T cell samples per round of 10X analysis. In this manner, the sample round, dextramer ID detected, and participant DNA hashtag detected would delineate the peptide-participant information associated with a TCR sequence. To our knowledge, this scale of pMHC tetramer-guided peptide/TCR mapping has not been implemented in the literature, primarily due to scaling and cost issues that exist for traditional technologies.

We anticipate that this platform will shed light on the question of immunodominance in SARS-CoV-2 antigens. Specifically, while recent literature and our study have focused on identifying immunogenic peptides from the spike protein, the question of which peptides throughout the proteome actually are primarily responsible for induction of an immunogenic response still remains to be answered. The complexity of this question is compounded by the biological origins of these peptides dictating mechanisms and efficiency of peptide presentation to $CD8^+$ T cells (for instance, do they derive from non-structural or structural proteins, are they involved in early or later stages of infection?), the analysis of participant samples at various checkpoints in their individual disease chronology, and the potential bias on peptide immunogenicity caused by HLA variation among the entire INCOV cohort, among numerous other factors. To narrow down this complexity, therefore, we assessed only A*02:01 participants and their respective peptides (covering approximately

40% of all participants due to the high prevalence of this allele, **Fig. H.1**). Additionally, we opted to study only PBMCs taken from samples at convalescent stage. Our rationale here was that SARS-CoV-2-specific T cells from samples at this timepoint should primarily consist of the memory subtype, whose antigen specificity arguably is of greater immunoprotective significance than that of exhausted or formerly active subtypes. In other words, examining convalescent stage PBMCs should filter only for T cells that have played an important role in establishing recovery and maintenance of protection.

We are currently developing such a platform to thoroughly assess all predicted peptides from the SARS-CoV-2 proteome in HLA-A*02:01 participants. The SCT platform has been advanced recently to improve production capacity by roughly an order of magnitude, such that approximately 1,000 SCTs were produced for this purpose (**Fig. B.3**). 9-mer and 10-mer epitopes were identified from the SARS-CoV-2 proteome using the same strategies previously discussed, and, in approximately two to three months, starting from a list of 951 peptides (**Tables L.1-L.4**), we produced SCT plasmids for every element and carried out transfection to identify those which expressed and could be used for antigen/TCR pairing assays. As seen in **Figure 3.11**, expressed SCTs included peptide elements from nearly all domains/proteins of SARS-CoV-2 without signficant bias for either 9-mer or 10-mer lengths. Altogether, 553 of the 951 SCTs were expressed to sufficient yields for downstream use. For TCR analysis, we have selected approximately 40 HLA-matched INCOV participants and 10 HLA-matched healthy donor samples as controls. Currently, we are optimizing barcoding configurations of dextramers and sample hashtags, and expect to report on antigen/TCR pairing results soon.

On a final note, based on the large size of the SCT proteome library, we wondered if the protein expression results might have correlation with NetMHC4.0 prediction. An initial plot of the data showed SCT expression to appear almost randomized for binding predictions made within this 0-500 nM window (**Fig. 3.12A**). A moving average plot to smoothen yield fluctuations when traversing across binding affinity demonstrated an extremely weak correlation (**Fig. 3.12B**). These results point to a need for further optimization of prediction software, and serve as a starting point for further expansion of these yield vs. prediction experiments. Given the ease of high-throughput SCT production as established herein, we anticipate future SCT expression datasets to be massive enough for sufficient training and testing of new and existing machine learning algorithms.

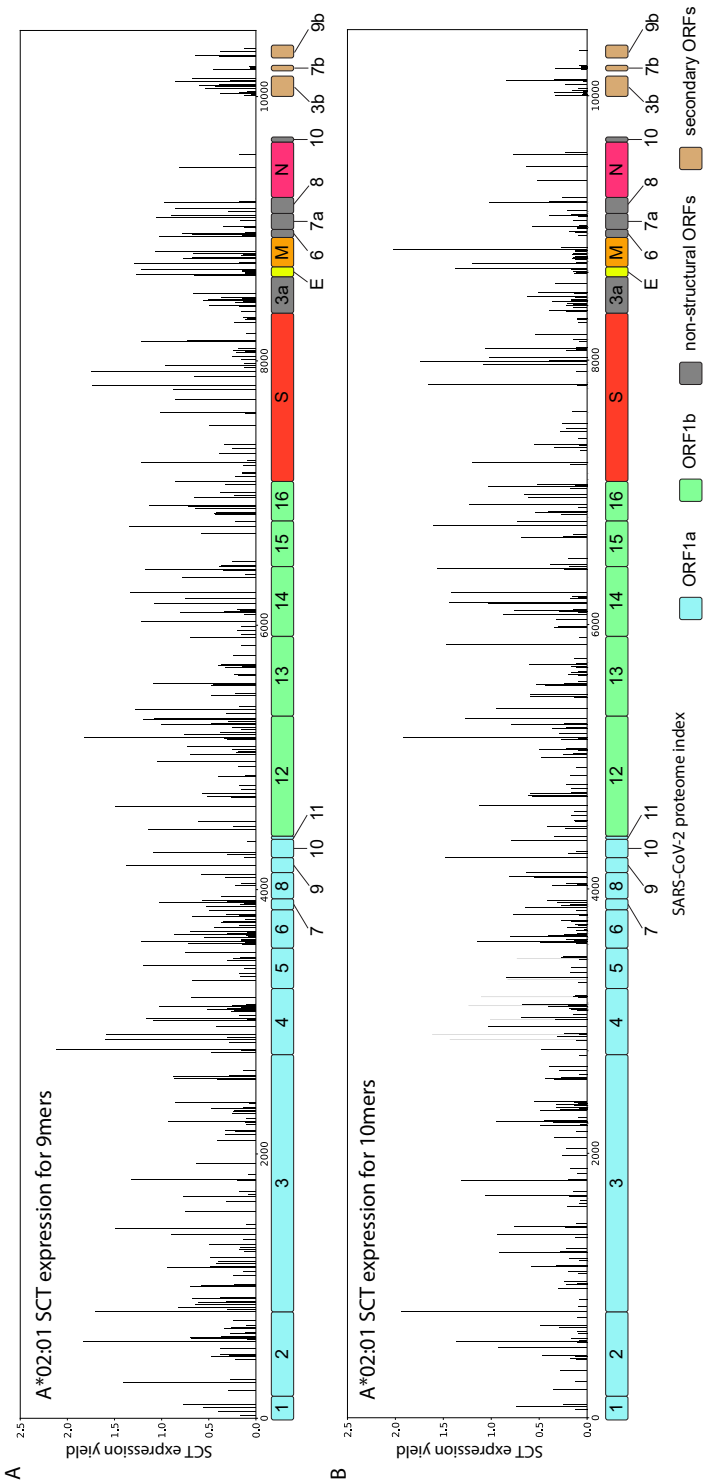


Figure 3.11: Expression of SCTs for A*02:01 SARS-CoV-2 proteomic epitopes Barplots of SCT protein band intensity (yield) for 9-mer (A) or 10-mer (B) epitopes predicted by NetMHC4.0 to have < 500 nM binding affinity.

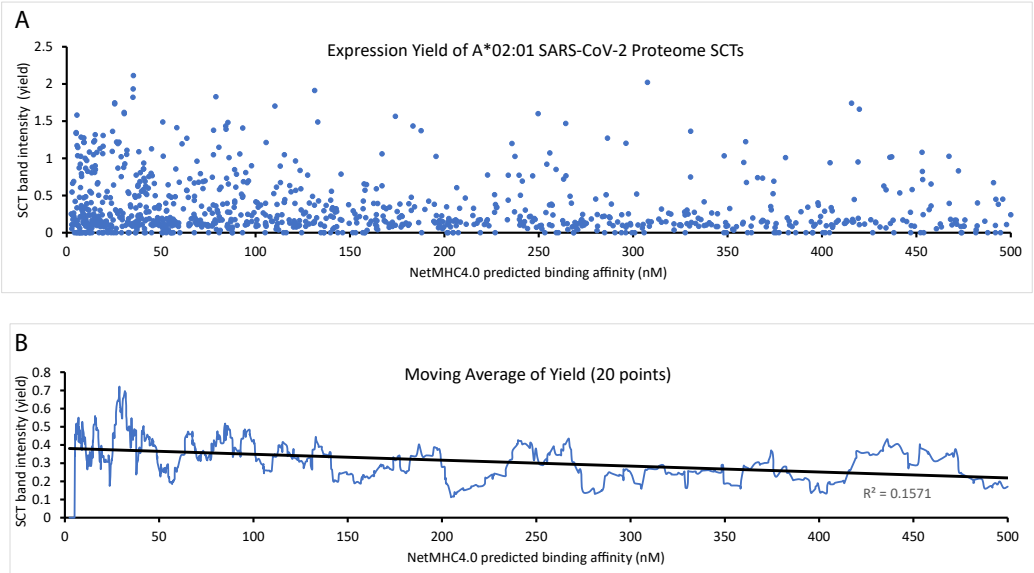
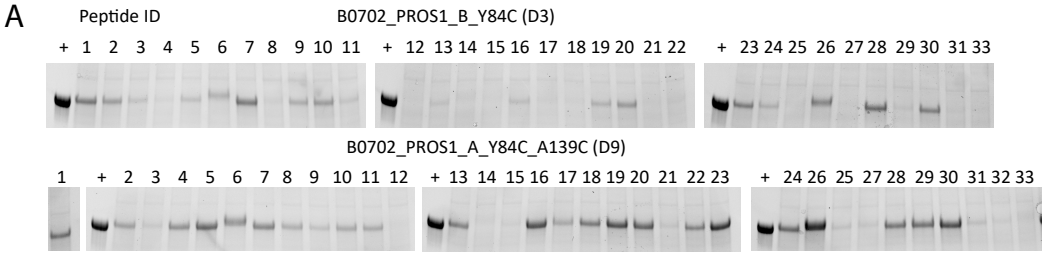


Figure 3.12: A*02:01 SCTs predicted for the SARS-CoV-2 proteome display poor correlation with yield
A. Scatterplot of SCT expression versus NetMHC4.0 predicted binding affinity. B. Moving average of expression yield (20 points) plotted against binding affinity.

3.2.2 Exploring HLA-B SCTs in the context of SARS-CoV-2

Another question we sought to explore during the onset of SCT development was whether our developed templates could be readily used in HLA alleles other than HLA-A. Given that the vast majority of existing Class I SCT literature focused on HLA-A, much of our initial proof-of-concept experiments implemented SCTs in the same context. However, the weight of HLA-A's role in disease protection relative to other Class I alleles is largely unknown, and it would be of great interest and importance to apply the sample principles of SCT design toward HLA-B and HLA-C. We had the opportunity to perform such an analysis for HLA-B with the study of B*07:02 SCTs against the spike protein (**Fig. I.1 & Table I.2**). While the presented data shows SCT yield only for the DS-SCT template of B*07:02, we actually settled on using this template only after exploring one of the traditional templates (**Fig. 3.13**), discovering that D3 generated relatively poor yield for this peptide library. In a deeper comparison of various B*07:02 templates (**Fig. 4.2**), the yields of this peptide library for all templates aside from DS-SCT were significantly worse, moreso than those seen in similar template comparison experiments conducted for HLA-A alleles (**Fig. 2.2B**), such that we contemplated whether this failure by non-DS-SCT templates to give good SCT yields was due to template mutation incompatibilities with the binding groove of B*07:02 or with all HLA-B alleles in general. The greater

sensitivity of HLA-B alleles to these modifications is corroborated by other studies concluding that minor mutations to binding grooves of HLA-B alleles tend to more drastically impact the peptide-binding repertoire compared to those of other Class I HLA alleles [17]. Our decision to use DS-SCT B*07:02 SCTs for further analysis of participant samples was made on the assumption that higher protein yields with this template indicated better stability of these constructs during protein translation and export. Better stability of such constructs would therefore be the preferred candidates for use in antigen-pairing assays to avoid non-binding or non-specific binding events.



B

ID	peptide	D3	D9
1	SPRRARSVA		
2	SPRRARSVAS		
3	NSPRRARSVA		
4	MIAQYTSAL		
5	YPDKVFRSSV		
6	IPTNFTISV		
7	FPQSAPHGVV		
8	LPVSMTKTSV		
9	TNSPRRARSVA		
10	SPRRARSVASQ		
11	RARSVASQSI		
12	QPYRVVVLSF		
13	LPPAYTNSF		
14	FPNITNLCPF		
15	WPWYIWLGF		
16	TPINLVRDL		
17	VPVAIHADQL		

ID	peptide	D3	D9
18	TPCSFGGVSV		
19	YPDKVFRSSVL		
20	KPFERDISTEI		
21	RAAEIRASA		
22	SIIAYTMSL		
23	LPQGFSALEPL		
24	MIAQYTSALL		
25	NSPRRARSVAS		
26	AIPTNFTISV		
27	SEPVLKGVKL		
28	APGQTGKIA		
29	FPQSAPHGV		
30	FPQSAPHGVVF		
31	ITRFQTLLAL		
32	VVNQNAQAL		
33	SLSSTASAL		

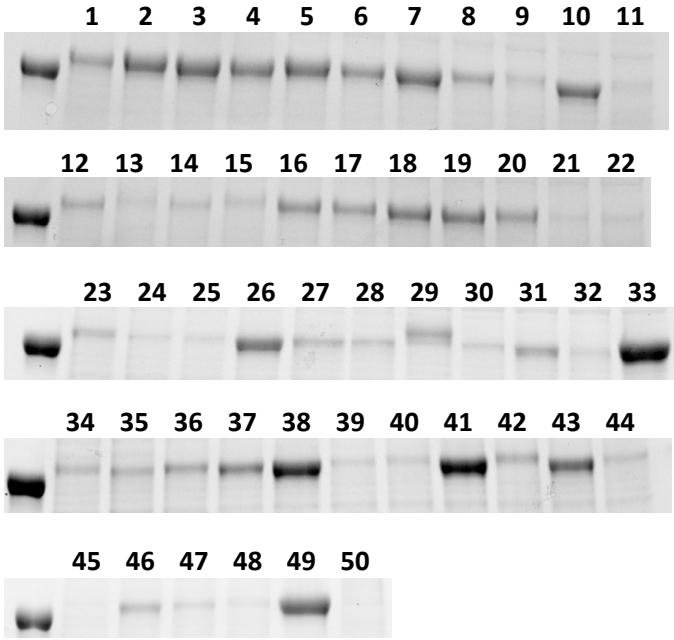
Figure 3.13: Comparison of B*07:02 spike SCT expression for two templates A. Reduced SDS-PAGE of B*07:02 spike epitope SCTs using two templates (B_Y84C corresponds to template D3 and A_Y84C_A139C corresponds to template D9 (DS-SCT) of table in **Figure 2.2B**). +, purified WT1 SCT. B. SCT yield heatmap comparing relative expression results for the two templates.

In a recent collaboration with Dr. Owen Witte's lab at UCLA, we sought to explore T cell immunity against the RNA-dependent RNA polymerase (RdRp) protein domain (Nsp12) of SARS-CoV-2. This non-structural protein plays a significant role in the early stages of infection, primarily dictating RNA translation. Our collaborators were able to identify A*02:01-restricted TCRs against RdRp epitopes through traditional peptide stimulation and T cell activation screening assays, and

plan to use SCTs from our proteomic library matching those epitopes to further verify functional binding (**Tables L.1-L.4**).

The Witte lab is also interested in exploring RdRp-specific T cells from B*35:01 participants. Taking the lessons learned from B*07:02 library development, we implemented the DS-SCT template toward production of the top 50 B*35:01 epitopes as predicted by NetMHCpan 4.1 algorithms, which uses EL_rank as a measure of peptide binding affinity (lower value is higher binding strength). Approximately 70% of the prepared elements resulted in high SCT expression (**Fig. 3.14**), and the magnitude of the protein yields was in agreement with DS-SCT B*07:02 results, supporting the notion that DS-SCT is a template design cross-compatible with HLA-B alleles in general, to a better extent than other traditional templates.

A



B

ID	peptide	EL_rank	len	yield
1	VPFVVSTGY	0.0234	9	0.38
2	YPLTKHPNQEY	0.0389	11	0.78
3	FVVEVVDKY	0.0552	9	0.90
4	FVSLAIDAY	0.1066	9	0.65
5	HPNQEYADVF	0.123	10	0.78
6	NPHLMGWDY	0.1539	9	0.44
7	NVIPTITQM	0.2217	9	0.84
8	SGVPVVDSY	0.2315	9	0.30
9	NAAISDYDY	0.2401	9	0.13
10	FVVSTGYHF	0.2907	9	0.68
11	EPEFYEAMY	0.3196	9	0.00
12	DVFHLYLQY	0.3204	9	0.17
13	SSGDATTAY	0.3515	9	0.00
14	AALTNNVAF	0.3537	9	0.10
15	AAISDYDYY	0.354	9	0.10
16	CPAVAKHDF	0.3581	9	0.34
17	LSFKELLVY	0.3635	9	0.27
18	LPYPDPSRI	0.4073	9	0.46
19	MVMCGGSLY	0.422	9	0.45
20	LVYAADPAM	0.4237	9	0.26
21	LTNDNTSRY	0.4243	9	0.00
22	QALLKTVQF	0.4809	9	0.00
23	TGTSTDVVY	0.483	9	0.13
24	EYADVFHLY	0.5551	9	0.00
25	YIKWDLLKY	0.5567	9	0.00

ID	peptide	EL_rank	len	yield
26	MPILTLTRAL	0.5711	10	0.55
27	TITQMNLKY	0.5809	9	0.15
28	VVIGTSKFY	0.6073	9	0.09
29	DAVVCFNSTY	0.6205	10	0.29
30	RLYYDSMSY	0.6342	9	0.07
31	HAASGNLLL	0.6439	9	0.21
32	TVVIGTSKF	0.6869	9	0.06
33	MSYEDQDAL	0.6977	9	1.29
34	TSSGDATTAY	0.7888	10	0.09
35	TTAYANSVF	0.798	9	0.10
36	LSDDAVVCF	0.8008	9	0.14
37	YADVFHLYL	0.8417	9	0.23
38	MPILTLTRA	0.853	9	0.57
39	YFKYWDQTY	0.8779	9	0.00
40	YWEPEFYEAMY	0.9286	11	0.00
41	MPNMLRIMA	0.9344	9	0.69
42	NAAISDYDYY	0.9574	10	0.11
43	YPDPSRILG	0.966	9	0.33
44	ECAQVLSEM	0.9717	9	0.00
45	EAMYTPHTV	0.9985	9	0.00
46	LANECAQVL	1.0253	9	0.22
47	DFVNEFYAY	1.0524	9	0.08
48	HVDTDLTKPY	1.0802	10	0.00
49	MSYEDQDALF	1.0927	10	1.10
50	AVVCFNSTY	1.1125	9	0.00

Figure 3.14: Expression of SCTs for B*35:01 SARS-CoV-2 RdRp protein epitopes A. Reduced SDS-PAGE of a RdRp epitope SCTs. Lane number indicates peptide ID (refer to (B) for peptide sequence). +, purified WT1 SCT. B. Quantified yield heatmap for SCTs from (A).

BIBLIOGRAPHY

[1] Peng Zhou, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, Wei Zhang, Hao-Rui Si, Yan Zhu, Bei Li, Chao-Lin Huang, Hui-Dong Chen, Jing Chen, Yun Luo, Hua Guo, Ren-Di Jiang, Mei-Qin Liu, Ying Chen, Xu-Rui Shen, Xi Wang, Xiao-Shuang Zheng, Kai Zhao, Quan-Jiao Chen, Fei Deng, Lin-Lin Liu, Bing Yan, Fa-Xian Zhan, Yan-Yi Wang, Geng-Fu Xiao, and Zheng-Li Shi. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798):270–273, March 2020. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-020-2012-7. URL http://www.nature.com/articles/s41586-020-2012-7.

[2] Huang Huang, Chunlin Wang, Florian Rubelt, Thomas J. Scriba, and Mark M. Davis. Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nature Biotechnology*, 38(10):1194–1202, October 2020. ISSN 1546-1696. doi: 10.1038/s41587-020-0505-4.

[3] Guang Chen, Di Wu, Wei Guo, Yong Cao, Da Huang, Hongwu Wang, Tao Wang, Xiaoyun Zhang, Huilong Chen, Haijing Yu, Xiaoping Zhang, Minxia Zhang, Shiji Wu, Jianxin Song, Tao Chen, Meifang Han, Shusheng Li, Xiaoping Luo, Jianping Zhao, and Qin Ning. Clinical and immunological features of severe and moderate coronavirus disease 2019. *The Journal of Clinical Investigation*, 130(5), April 2020. ISSN 0021-9738. doi: 10.1172/JCI137244. URL https://www.jci.org/articles/view/137244. Publisher: American Society for Clinical Investigation.

[4] Clare Wenham, Julia Smith, and Rosemary Morgan. COVID-19: the gendered impacts of the outbreak. *The Lancet*, 395(10227):846–848, March 2020. ISSN 0140-6736. doi: 10.1016/S0140-6736(20)30526-2. URL http://www.sciencedirect.com/science/article/pii/S0140673620305262.

[5] Manish Pareek, Mansoor N Bangash, Nilesh Pareek, Daniel Pan, Shirley Sze, Jatinder S Minhas, Wasim Hanif, and Kamlesh Khunti. Ethnicity and COVID-19: an urgent public health research priority. *The Lancet*, 395(10234): 1421–1422, May 2020. ISSN 0140-6736. doi: 10.1016/S0140-6736(20)30922-3. URL http://www.sciencedirect.com/science/article/pii/S0140673620309223.

[6] Ranganath Muniyappa and Sriram Gubbi. COVID-19 pandemic, coronaviruses, and diabetes mellitus. *American Journal of Physiology-Endocrinology and Metabolism*, 318(5):E736–E741, May 2020. ISSN 0193-1849, 1522-1555. doi: 10.1152/ajpendo.00124.2020. URL https://journals.physiology.org/doi/10.1152/ajpendo.00124.2020.

[7] Nguyen Thi Phuong Lan, Mihoko Kikuchi, Vu Thi Que Huong, Do Quang Ha, Tran Thi Thuy, Vo Dinh Tham, Ha Manh Tuan, Vo Van Tuong, Cao Thi Phi Nga, Tran Van Dat, Toshifumi Oyama, Kouichi Morita, Michio Yasunami, and Kenji Hirayama. Protective and Enhancing HLA Alleles, HLA-DRB1*0901 and HLA-A*24, for Severe Forms of Dengue Virus Infection, Dengue Hemorrhagic Fever and Dengue Shock Syndrome. *PLoS Neglected Tropical Diseases*, 2(10):e304, October 2008. ISSN 1935-2735. doi: 10.1371/journal.pntd.0000304. URL https://dx.plos.org/10.1371/journal.pntd.0000304.

[8] Marie Lin, Hsiang-Kuang Tseng, Jean A Trejaut, Hui-Lin Lee, Jun-Hun Loo, Chen-Chung Chu, Pei-Jan Chen, Ying-Wen Su, Ken Hong Lim, Zen-Uong Tsai, Ruey-Yi Lin, Ruey-Shiung Lin, and Chun-Hsiung Huang. Association of HLA class I with severe acute respiratory syndrome coronavirus infection. *BMC Medical Genetics*, 4(1):9, December 2003. ISSN 1471-2350. doi: 10.1186/1471-2350-4-9. URL http://bmcmedgenet.biomedcentral.com/articles/10.1186/1471-2350-4-9.

[9] Songming Peng, Jesse M. Zaretsky, Alphonsus H.C. Ng, William Chour, Michael T. Bethune, Jongchan Choi, Alice Hsu, Elizabeth Holman, Xiaozhe Ding, Katherine Guo, Jungwoo Kim, Alexander M. Xu, John E. Heath, Won Jun Noh, Jing Zhou, Yapeng Su, Yue Lu, Jami McLaughlin, Donghui Cheng, Owen N. Witte, David Baltimore, Antoni Ribas, and James R. Heath. Sensitive Detection and Analysis of Neoantigen-Specific T Cell Populations from Tumors and Blood. *Cell Reports*, 28(10):2728–2738, September 2019. ISSN 22111247. doi: 10.1016/j.celrep.2019.07.106. URL https://linkinghub.elsevier.com/retrieve/pii/S2211124719310228.

[10] Alphonsus H. C. Ng, Songming Peng, Alexander M. Xu, Won Jun Noh, Katherine Guo, Michael T. Bethune, William Chour, Jongchan Choi, Sung Yang, David Baltimore, and James R. Heath. MATE-Seq: microfluidic antigen-TCR engagement sequencing. *Lab on a Chip*, 19(18):3011–3021, 2019. ISSN 1473-0197, 1473-0189. doi: 10.1039/C9LC00538B. URL http://xlink.rsc.org/?DOI=C9LC00538B.

[11] Evan Z. Macosko, Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, Allison R. Bialas, Nolan Kamitaki, Emily M. Martersteck, John J. Trombetta, David A. Weitz, Joshua R. Sanes, Alex K. Shalek, Aviv Regev, and Steven A. McCarroll. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*, 161(5):1202–1214, May 2015. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2015.05.002. URL http://www.cell.com/cell/abstract/S0092-8674(15)00549-8.

[12] T. Sano and C. R. Cantor. Expression of a cloned streptavidin gene in Escherichia coli. *Proceedings of the National Academy of Sciences*, 87(1):142–

146, January 1990. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.87.1.142. URL http://www.pnas.org/cgi/doi/10.1073/pnas.87.1.142.

[13] Massimo Andreatta and Morten Nielsen. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics*, 32(4):511–517, February 2016. ISSN 1460-2059, 1367-4803. doi: 10.1093/bioinformatics/btv639. URL https://academic.oup.com/bioinformatics/article/32/4/511/1744469.

[14] Katie M Campbell, Gabriela Steiner, Daniel K Wells, Antoni Ribas, and Anusha Kalbasi. Prediction of SARS-CoV-2 epitopes across 9360 HLA class I alleles. preprint, Immunology, April 2020. URL http://biorxiv.org/lookup/doi/10.1101/2020.03.30.016931.

[15] Syed Faraz Ahmed, Ahmed A. Quadeer, and Matthew R. McKay. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. preprint, Bioinformatics, February 2020. URL http://biorxiv.org/lookup/doi/10.1101/2020.02.03.933226.

[16] Ethan Fast, Russ B. Altman, and Binbin Chen. Potential T-cell and B-cell Epitopes of 2019-nCoV. preprint, Microbiology, February 2020. URL http://biorxiv.org/lookup/doi/10.1101/2020.02.19.955484.

[17] Hanneke W. M. van Deutekom and Can Keşmir. Zooming into the binding groove of HLA molecules: which positions and which substitutions change peptide binding most? *Immunogenetics*, 67(8):425–436, August 2015. ISSN 0093-7711, 1432-1211. doi: 10.1007/s00251-015-0849-y. URL http://link.springer.com/10.1007/s00251-015-0849-y.

Chapter 4

AN EXPLORATION OF TUMOR ANTIGENS FOR SCT PRODUCTION

4.1 Introduction

A long-sought goal in the field of cancer immunotherapy is to develop technologies which may enable the rapid isolation and enumeration of tumor-specific T cells from patient tissue. Given the numerous side effects associated with radiotherapy and chemotherapy (weakened immunity, damage to healthy tissue, neuropathy, etc.), it would be advantageous to make use of immunotherapeutic approaches such as adoptive cell transfer (ACT) therapy, which can harness the antigen-specific nature of $CD8^+$ T cell cytotoxicity to eliminate tumor burden without off-target consequences. Toward this goal, advancement of ACT requires the identification of TCRs with tumor specificity, which may be defined as reactivities targeted against various types of antigens, either tumor-associated (TAAs) or tumor-specific (TSAs). TAAs are generally peptides over-expressed in tumor cells due to dys-regulated proteins in upstream pathways. In terms of amino acid sequence, they are no different from those of proteins found in healthy cells. Prominent examples include NY-ESO-1 and MART-1 peptides [1]. On the other hand, TSAs are, by definition, peptides found uniquely in tumor cells, and therefore contain mutated amino sequences. TSAs are further stratified into two groups, public and private. Public TSAs are found in high prevalence in cancer populations, and tend to contain so called "driver" mutations in certain proteins that play key roles in regulation of growth and metabolism. Such proteins have the capacity to become oncogenic if their normal functionality is either over- or under-regulated, resulting in uncontrolled cellular growth and expansion into healthy tissue. These protein "hotspots", or amino acid sequences that are particularly susceptible to oncogenicity if altered, are therefore the most prominent locations where one can identify public TSAs from the protein sequence. Private TSAs, in contrast, consist of "passenger" mutations, which are typically amino acid mutations that sporadically occur secondary to an initial driver mutation that triggers oncogenicity. Passenger mutations, as implied by the name, are typically non-oncogenic alterations to a protein sequence that nevertheless result in a sequential change and enable one to utilize epitopes containing such sequences as a specific

biomarker for tumor tissue. Because they tend to play a non-oncogenic role, they do not significantly contribute to selective pressure for tumor escape, and therefore are highly diverse, such that their presence across various tumor types and patients is essentially individualized.

From a commercial standpoint, the primary criterion to avoid off-target side effects favors an ACT approach that makes use of TSAs rather than TAAs. In terms of cost reduction and efficiency, the use of TCRs targeting public TSAs over private TSAs is strongly preferred, as it may enable off-the-shelf approaches to ACT administration with minimal patient customization required. There are a number of reviews which highlight the drawbacks of TAAs, implying that TSAs are perhaps the more favorable approach [2]. As for determining whether public or private TSAs are the best targets, however, there remains much room for insight, research, and development. The primary bottleneck to research output here is the limited capacity for high-throughput TCR enumeration, cloning, and validation of on-target cytotoxicity. The ACT workflow becomes vastly complex when one must take into account the broad TCR diversity which exists for any given epitope and the potentially large number of epitopes to assess for any patient. Here, the choice of public TSAs would address the peptide scope issue by focusing only on hotspot peptides rather than the individualized repertoire of passenger mutations. Public TSAs might also alleviate throughput concerns once viable TCRs have been selected, as their large prevalence across patient populations might only require the optimization of a few TCRs per peptide/HLA combination, which would circumvent the need to repeatedly identify and validate antigen-specific TCRs per individual patient. The SCT platform would also serve as a tool to streamline the aforementioned limitations behind TCR capture, as it enables high-throughput generation of capture reagents for large peptide libraries to such an extent where hundreds of SCTs may be generated within several weeks.

Using the SCT technologies outlined in the prior chapters, I discuss in this chapter an outgrowth of various projects emerging from our initial SCT library developments. While these projects are so far incomplete to various extents, the primary purpose is to inform the reader of the potential for SCT technology to impact ACT development and to lay the foundation for future SCT projects. The first section on TAAs demonstrates how the SCT platform may complement an existing method to identify prostate cancer epitopes (mass spectrometry) for additional validation. As for TSAs, I demonstrate the capacity of the high-throughput SCT platform to enable

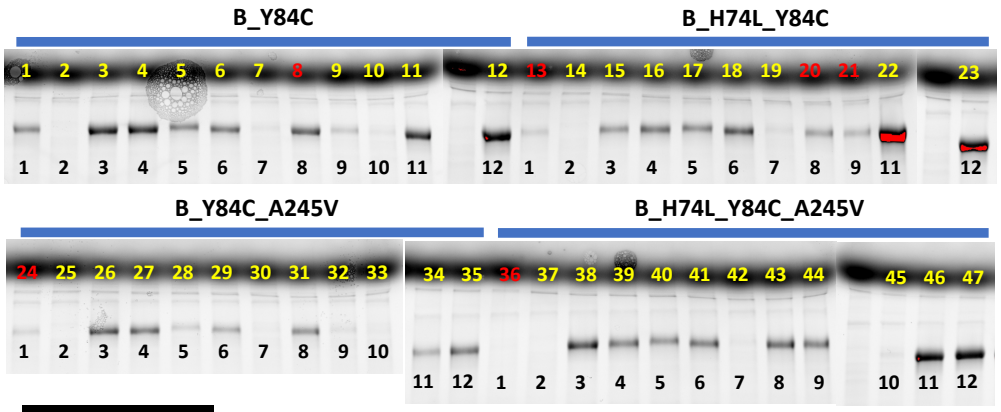
rapid discovery of immunogenic hotspot peptides for public TSAs on a per-HLA basis for the p53 and KRAS proteins. Next, the SCT platform can also be used in an individualized manner to identify potentially immunogenic private TSAs, as exemplified by its application for a melanoma patient. In all cases, the SCTs used can be further extended for use as tetramer reagents to identify cognate TCRs to these antigens, and some examples are provided for the above projects. Given the universality of SCT technologies for the study of any peptide/HLA pairing, these projects provide a small glimpse into the potential space of tumor antigen targets that can be explored, whether they originate from self-derived TAAs, public TSAs, or private TSAs. Furthermore, some of these projects will expand upon earlier research directions explored in the labs, such as glycosylation of peptides (**Fig. 2.22**), CD8-inhibiting mutations (**Fig. 2.8**), SCT mutation engineering (**Fig. 2.2**), or NP-NACS (**Fig. 3.4**). Their implementations, results, and potential future directions will also be discussed.

4.2 Tumor-Associated Antigens

We have so far assessed SCT expression for tumor-associated antigens (TAAs) in at least two different contexts. In the first case, we and our collaborators at the Fred Hutch (Dr. Roland Strong) have been able to express A*02:01 SCTs loaded with a Wilms Tumor 1 (WT1) antigen (RMFPNAPYL), which has been associated with certain leukemias and is of therapeutic interest at the level of clinical trials [3]. A functional analysis of these SCTs can be found in **Figures 2.4 & 2.8**. TAAs are also promising targets in the context of prostate cancer. We have worked with another Fred Hutch collaborator (Dr. John Lee) to assess prevalent A*02:01 and B*07:02 TAA candidates from various proteins (NKX3.1, prostatic acid phosphatase (PAP), etc) associated with prostate tumors. Our collaborators have identified these peptides to be potentially immunogenic in terms of activating $CD8^+$ T cell responses, and we have been able to confirm that these peptides can lead to pMHC expression in the SCT format. As seen in **Figure 4.1**, A*02:01 SCT libraries produced using various templates result in consistent expression patterns for a small library of twelve peptides. As for the B*07:02 library, we observed that SCT expression was overall much lower for this set of peptides (**Fig. 4.2**). We expected that this may be due to structural restrictions behind the SCT templates, rather than a true biological barrier (these peptides were detected by mass spectrometry assays by the Lee lab). This was confirmed by the last set of the B*07:02 library, in which the D9 design was

found to recover SCT yield for at least seven of the twelve SCTs.

A

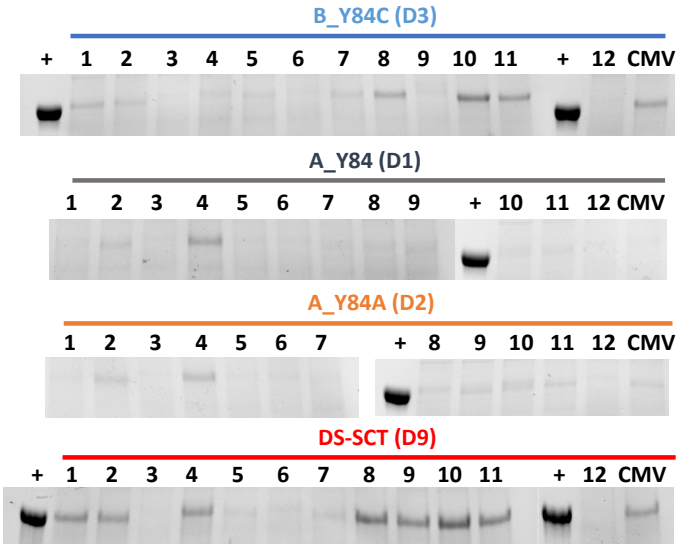


B

ID	antigen
1	PAP
2	PAP
3	PAP
4	PAP
5	PAP
6	SLC45a3
7	SLC45a3
8	SLC45a3
9	SLC45a3
10	SLC45a3
11	SLC45a3
12	NKX3.1

Figure 4.1: Expression of A*02:01 prostate cancer antigen SCT library (Lee). A. Reduced SDS-PAGE of SCT transfection results. Title above each bar represents the SCT template used for expression (B_Y84C is equivalent to D3 of **Fig. 2.2B**; B_H74L_Y84C is equivalent to D8 of **Fig. 2.2B**; A245V indicates a template with additional CD8-inhibiting mutation applied). Letters below each lane indicate the peptide ID, as referenced in (B). B. Table of indexed peptides used for SCT library assembly (protein origin is given, peptide sequences not disclosed due to pending publication from collaborators).

Our analysis of SCT expression in the context of prostate cancer TAAs extends to our collaboration with Dr. Owen Witte's lab at UCLA. In a parallel survey of PAP TAAs, they performed mass spectrometry of tumor samples to identify 27 peptides as potential candidate binders to A*02:01 MHC proteins. We utilized the SCT expression platform to assess whether these peptides could favorably bind with A*02:01's peptide groove to enable high SCT protein yield. SCT expression yields did not appear to correlate well with popular peptide prediction algorithms (NetMHC4.0 or NetMHCpan) (**Table 4.1**). The peptide library consisted of a broad range of predicted affinities, anywhere from within a probable binding range (5-1,000 nM) to an excessively high range indicative of poor binding prediction (> 15,000



200310			200331
B_Y84C	A_Y84	A_Y84A	DS-SCT
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12
CMV	CMV	CMV	CMV

Figure 4.2: Expression of B*07:02 prostate cancer antigen SCT library (Lee). Reduced SDS-PAGE SCT transfection results (left) are displayed for a library of B*07:02 PAP peptides (indexed 1-12) substituted into various B*07:02 templates as indicated above each bar (template design in parentheses refers to designs from **Fig. 2.2**). Heatmap (right) compares the relative SCT yield for each combination of peptide and template. +, positive control lane using A*02:01 WT1 (RMFPNAPYL) SCT.

nM). It was surprising then, that some SCTs containing peptides within probable predicted binding did not express (e.g. PAP_A2_27), while there were several SCTs containing peptides outside of predicted binding range (e.g. PAP_A2_01, 15, 25) that gave among the highest expression yield. These results indicate a potential limitation of prediction algorithms, that they may only be reliable predictors of conventional peptide lengths/motifs for a designated HLA allele, but may not have been trained with sufficient biological data to properly assess unconventional peptides that might still be functional binding candidates. Of note is the peptide pair PAP_A2_01 and PAP_A2_02, which differ in length by one amino acid, yet give extremely different expression yields. The large difference in SCT expression for this pair indicates that perhaps the C-terminal end of the peptide plays a significant role in determining stabilization of the binding groove, implying that the frames of these rather long peptides which interact with the binding groove might be biased toward the C-terminal region. In summary, the mass spectrometry-derived peptide list served as an experimental filter prior to tetramer production to highlight peptides which might be missed by simply adhering to prediction scores. Our SCT expression assay serves to complement the mass spectrometry data by demonstrating

that some of the poorly predicted peptides can indeed lead to successful production of a soluble pMHC construct. Beyond the scope of both of these assays is the ability to ascertain whether these peptides are truly functional from an immunogenic standpoint. Toward this end, future work with Lee and Witte labs will aim to utilize the SCTs to capture antigen-specific T cells using conventional methods and to functionally validate their TCR sequences for peptide specificity.

Name	Peptide subtypes	Length	Start position	End position	NetMHC A2(nM)	NetMHC A2(%)	NetMHCpan A2(%)	B_H74L_Y84C (D8) yield		B_H74L_Y84C _A245V yield	
PAP_A2_01	YSAHDTTVSGLQMA	14	286	299	19437.6	27	19.7418	0.75		0.27	
PAP_A2_02	YSAHDTTVSGLQM	13	286	298	27417.1	41	33.2593	0.14		0.00	
PAP_A2_03	LSELSLLSLYGIHK	14	238	251	13302.65	19	74.375	0.00		0.00	
PAP_A2_04	DFIATLGKLSG	11	186	196	28979.5	44	85	0.00		0.00	
PAP_A2_05	DFIATLGK	8	186	193	36583.3	70	65.5	0.00		0.00	
PAP_A2_06	IATLGKLSGLHGQDL	15	188	202	18809.84	#N/A	#N/A	0.00		0.00	
PAP_A2_07	IATLGKLSGLHGQD	14	188	201	32159.8	55	95	0.00		0.00	
PAP_A2_08	LSGLHGQDL	9	194	202	28193	42	42.8333	0.00		0.00	
PAP_A2_09	MEQHYELGEY	10	75	84	32300.7	55	34.8462	0.00		0.00	
PAP_A2_10	GEYFVEMYYR	10	323	332	21514.3	30	26.7288	0.03		0.00	
PAP_A2_11	SPIDTFPTDPIK	12	48	59	35161.7	65	41.3333	0.02		0.00	
PAP_A2_12	FQKRLHPYK	9	177	185	31660.2	55	32.2069	0.02		0.00	
PAP_A2_13	IRSTDVDRTL	10	104	113	25456.7	37	18.2108	0.00		0.00	
PAP_A2_14	ILLWQPIPV	9	135	143	5.2	0.04	0.1557	0.77		0.51	
PAP_A2_15	WQPIPVHTVPLS	12	138	149	22400.3	31	44	0.68		0.29	
PAP_A2_16	DVYNGLLPPYA	11	301	311	21620.5	30	8.6579	0.00		0.00	
PAP_A2_17	KELKFVTL	8	33	40	25341.8	36	24.7662	0.00		0.00	
PAP_A2_18	KELKFVTLVFRHGD	14	33	46	18618.7	25	85	0.00		0.00	
PAP_A2_19	IWSKVYDPLY	10	205	214	29598	46	35.1818	0.00		0.00	
PAP_A2_20	KVYDPLYCESV	11	208	218	615.5	3	0.401	0.62		0.39	
PAP_A2_21	LLLARAASLSL	11	6	16	977.9	4	5.1611	0.34		0.32	
PAP_A2_22	TLMSAMTNL	9	112	120	9.4	0.1	0.0794	0.69		0.72	
PAP_A2_23	LLFFWLDRSVLA	12	21	32	647.2	3.5	2.2948	0.08		0.00	
PAP_A2_24	SVHNFTLPSW	10	217	226	21204.2	29	24.2338	1.50		1.03	
PAP_A2_25	IMYSAHDTTV	10	284	293	82.3	0.9	0.8042	0.52		0.49	
PAP_A2_26	RHGDRSPIDTFPTD	14	43	56	37378.7	75	81	0.00		0.00	
PAP_A2_27	VLAKELKFV	9	30	38	51.2	0.7	0.1858	0.07		0.08	

Table 4.1: Expression of A*02:01 PAP SCT library (Witte).
Summary of SCT expression results for each of the PAP peptides as identified using mass spectrometry. SCT expression is quantified for each of the two template designs tested per peptide (rightmost two columns) to show relative yield.

We have preliminary assays confirming functionality of prostate cancer TAA-loaded SCTs. TCRs of various affinities specific to PAP_A2_14 were identified from both PBMCs and artificial thymic organoids by the Witte lab. They were sequenced and cloned into $CD8^+$ Jurkat cell lines, and subsequently used in flow cytometry staining assays against SCT tetramers to measure binding efficiency. Negative and positive control assays utilizing non-PAP pMHC tetramers (SCT and refolded variants) showed expected binding interactions with affiliated TCR targets (**Fig. 4.3**). When comparing the performance of our control tetramers in SCT and refolded format,

we noted that the refolded tetramer variants appeared to give stronger positive signal (**Fig. 4.4C vs. 4.4F**), but when they were incubated against T cells which were not transduced with the target TCR, these refolded tetramers tended to give higher background signal (**Fig. 4.4A&B vs. 4.4D&E**). The "cleaner" background signal of SCTs was again seen when we examined PAP SCTs, which gave minimal binding signal against off-target cell lines (**Fig. 4.5A-C**). When the PAP-specific TCR cell lines (PAP-TCR-135) were used instead, we observed gradually increased efficiencies in binding by the PAP SCTs (**Fig. 4.5D-F**), which positively correlated with the known level of reactivity of these TCRs against antigens based on IFNγ ELISA assays (data not shown). PAP pMHCs generated by refolding also displayed the same staining pattern when using the same peptide against these three TCRs (data not shown), supporting the notion that our TAA-loaded SCTs are capable of behaving in a manner similar to their biological counterparts. We further assessed the sensitivity of these SCT-tetramers by staining PAP-TCR-53 diluted into 1G4 (both on Jurkat-CD8 cells) across a range of dilutions (**Fig. 4.6**), showing that sensitivity was retained down to even 0.01% dilution. These results indicate that our SCT platform may be used as a discovery/validation assay for additional TAAs in terms of identifying ones which may stably bind onto MHC and be functionally useful for TCR identification.

Two of the PAP TAAs were also utilized in a separate direction of experiments to assess the relationship between SCT template and expression/functionality. As seen in **Figure 4.7A-B**, SCT templates incorporating several linker and HLA mutations (**Fig. 2.2B**) were used for the insertion of peptides ILLWQPIPV and SVHNFTLPSW. The SCT yields for both peptides varied in a template-dependent manner, but to differing capacities. For ILLWQPIPV, all templates showed some degree of expression, but appeared to give larger bands for templates containing cysteine linker modifications, implying that the additional dithiol stabilization may improve transfection yield (template D8 was not included in the gel in this figure, but was expressed and tabulated in (**Table 4.1**), providing a similar degree of expression to D3 and D4). For SVHNFTLPSW, this behavior was accentuated to a further extent, where the non-cysteine linker templates give much lower yields. When comparing this subset of templates for this peptide, it appears that yields for D1 and D2 are almost imperceptible, while a moderate yield still exists for D6 and D7. This suggests that perhaps, for this particular peptide, the H74L mutation serves to better stabilize the SCT construct, enabling yield recovery. The yield recovery obtained by the inclusion of cysteine linkers, however, are much higher than that of the H74L

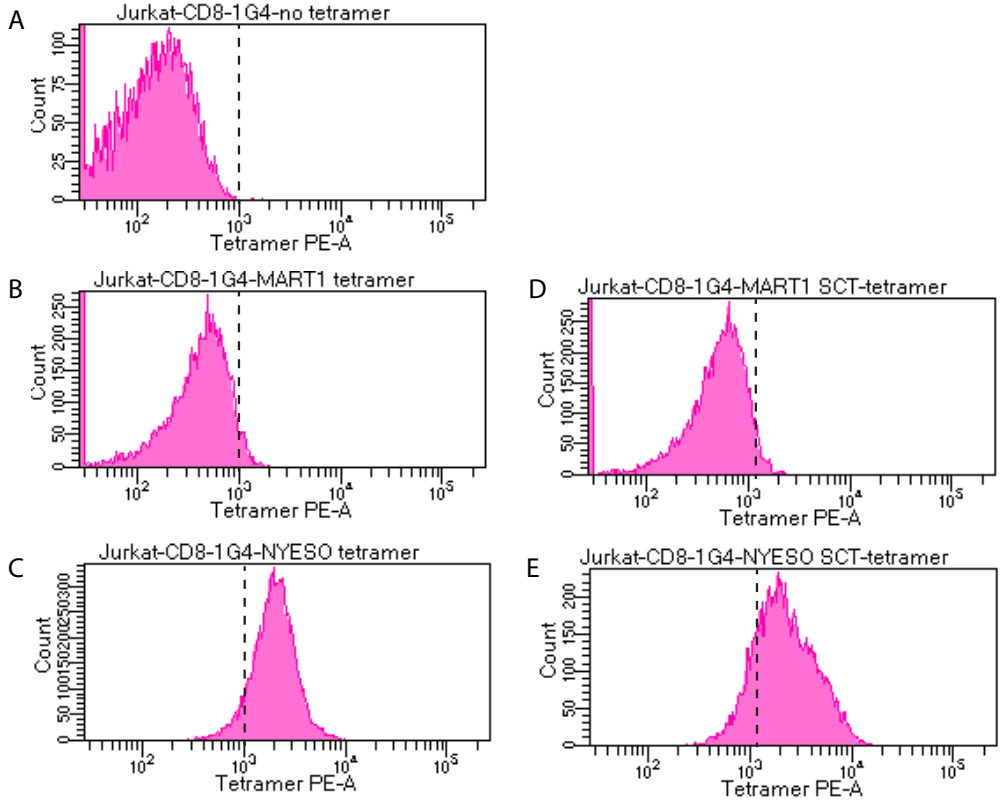


Figure 4.3: NY-ESO-1 tetramers (refolded and SCT variant) specifically stain 1G4 TCR-transduced Jurkat cells.
Jurkat cells transduced with 1G4 TCR (A), which is specific to the NY-ESO-1 antigen, were stained with either MART-1 (B & D) or NY-ESO-1 (C & E) tetramers. Tetramer variants utilized were either refolded (B & C) or SCT (D & E). Dashed line represents binding threshold of 10^3 MFI (mean fluorescence intensity) from tetramer staining.

mutation, as seen with the other designs (D3, D4, D5, D8, & D9) for this peptide. These expression correlations with SCT template appear to match our observations for a library previously explored (**Fig. 2.2B**).

To show that these SCTs are functional, we selected five of the ILLWQPIPV SCT variants of **Figure 4.7A** for tetramer binding assays against PAP TCR 53 (the D8 variant of this SCT was assessed already against this TCR in **Fig. 4.5F**). Despite the variations in SCT expression, we saw that all five of these SCT tetramers, in addition to the UV-exchanged tetramer variant, performed identically in terms of binding to the target TCR (**Fig. 4.8**). This particular assay's results stand in contrast to that of the WT1 SCT variants (**Fig. 2.4 & 2.8B**), where these same templates resulted

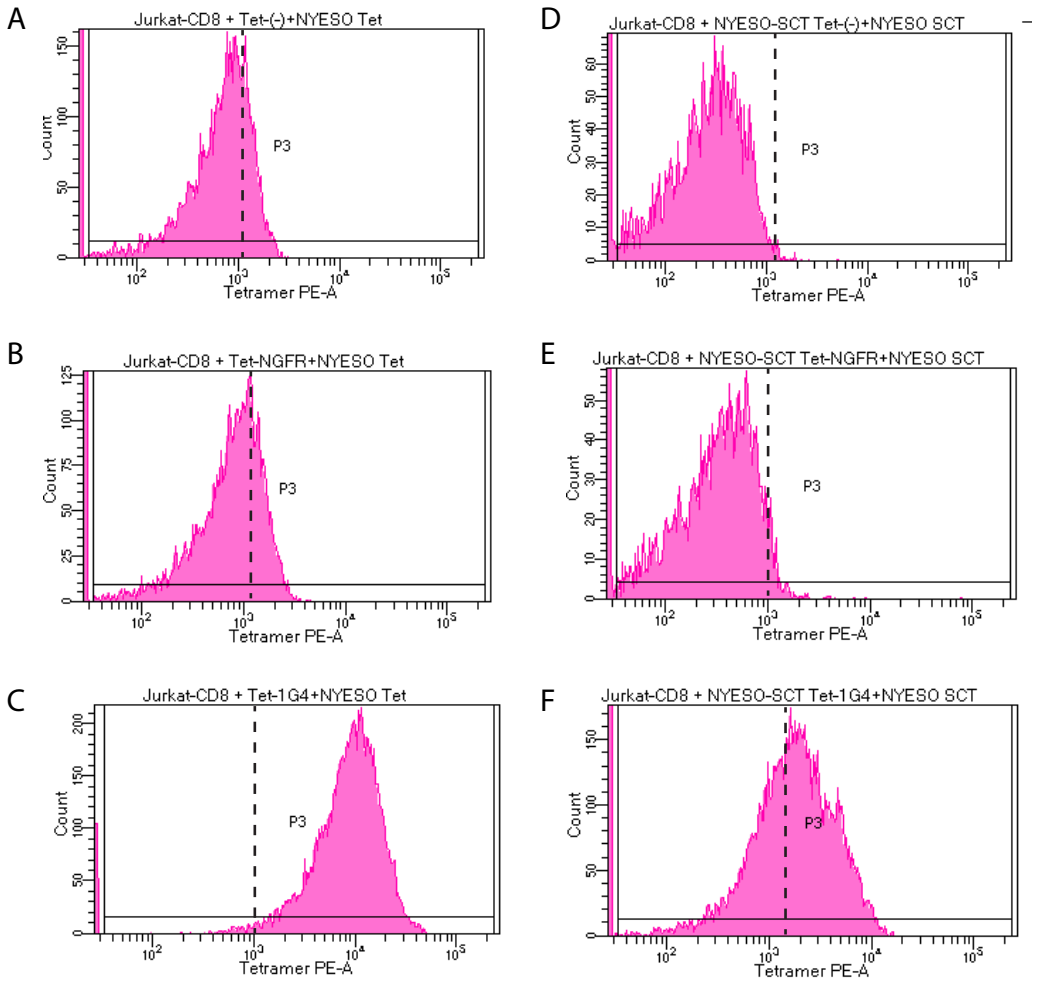


Figure 4.4: NY-ESO-1 SCTs have less background staining than refolded variants. Jurkat T cells without TCR transduction were stained with either NY-ESO-1 refolded tetramer or NY-ESO-1 SCT tetramer (A & D, respectively). The same procedure was performed with Jurkat T cells transduced with NGFR (B & E) and with Jurkat T cells transduced with the NY-ESO-1-specific 1G4 TCR (C & F). Dashed line represents binding threshold of 10^3 MFI (mean fluorescence intensity) from tetramer staining.

in significantly different tetramer binding efficiencies to the C4 TCR. This indicates that for the ILLWQPIPV-TCR interaction, SCT template modifications appear to have little impact on epitope presentation and interaction with the TCR. Whether this observation can be extended to other TCRs of the same antigen specificity is beyond the scope of our current work, but would be interesting to further explore.

Glycosylation of SCTs was another aspect that we were able to observe in **Figure 4.7B**. Glycosylation of the NXT motif in SVHNFTLPSW SCTs resulted in notice-

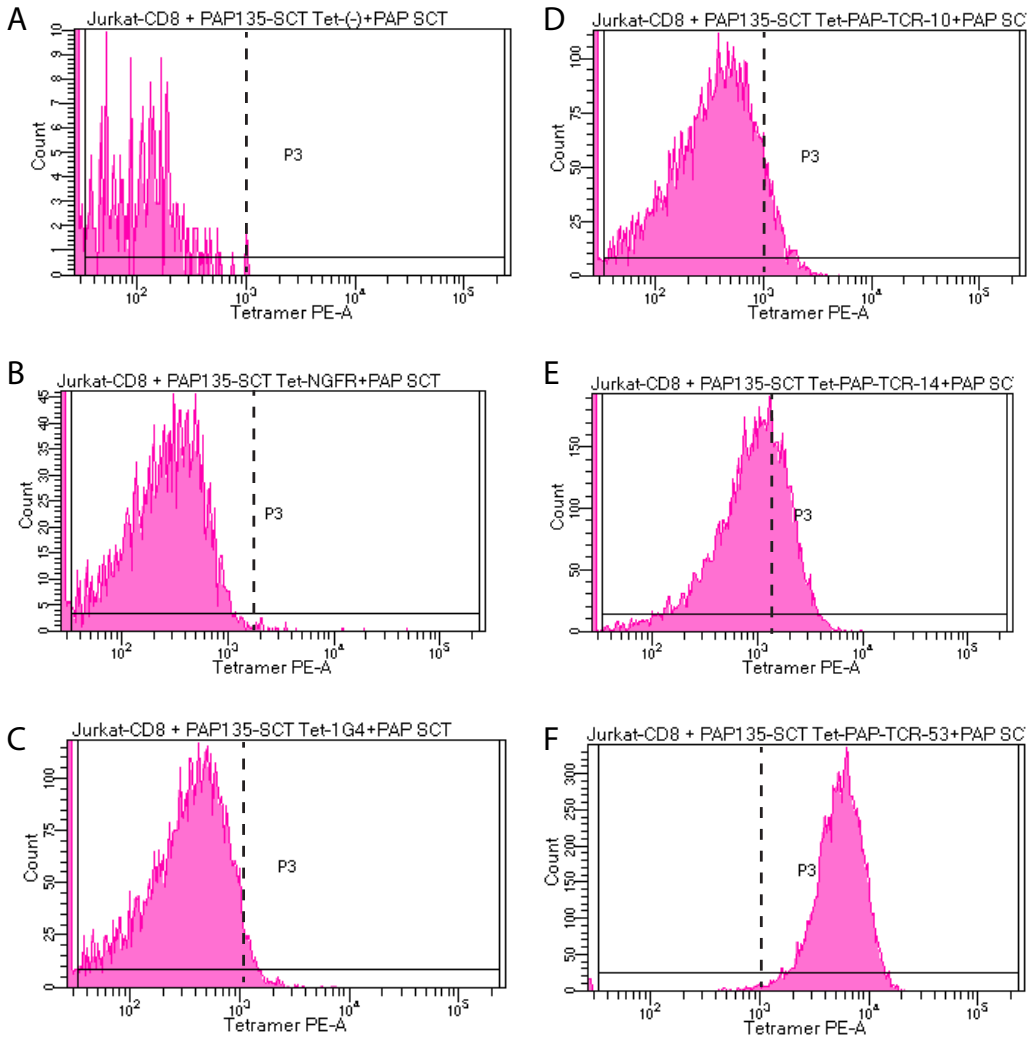


Figure 4.5: PAP SCTs stain PAP TCR-transduced Jurkat cells with variable degrees of efficiency.

SCT tetramers containing PAP_A2_14 peptide in the D8 template were used to stain Jurkat cells, which were either not transduced (A) or transduced with NGFR (B), 1G4 TCR (C), or a TCR specific to the PAP_A2_14 peptide (D-F). Dashed line represents binding threshold of 10^3 MFI (mean fluorescence intensity) from tetramer staining.

ably increased mass, as indicated by the upward shifted protein bands compared to WT1 control lane or to the MART-1 SCT. The increased mass of this SCT was attributed solely to glycosylation of the peptide's NXT motif in a de-glycosylation step **Figure 4.7C**, where SCTs containing peptides with and without the glycosylation motif (SVHNFTLPSW and IMYSAHDTTV, respectively) were treated with peptide-N-glycosidase (PNGase). Before PNGase treatment, reduced SDS-PAGE revealed that the former SCT was of a larger mass than the latter SCT, but after PN-

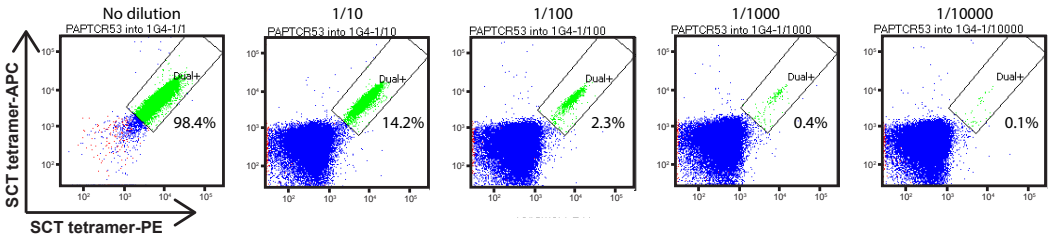


Figure 4.6: PAP SCTs capture PAP-TCR-53-transduced Jurkat cells with high sensitivity.

SCT tetramers (constructed with either streptavidin-PE or streptavidin-APC) containing PAP_A2_14 peptide in the D8 template were used to stain PAP-TCR-53-transduced Jurkat cells across a range of dilutions into 1G4-TCR-transduced Jurkat cells. Percentages indicate the proportion of cells double-stained by PAP SCT tetramer within each dilution.

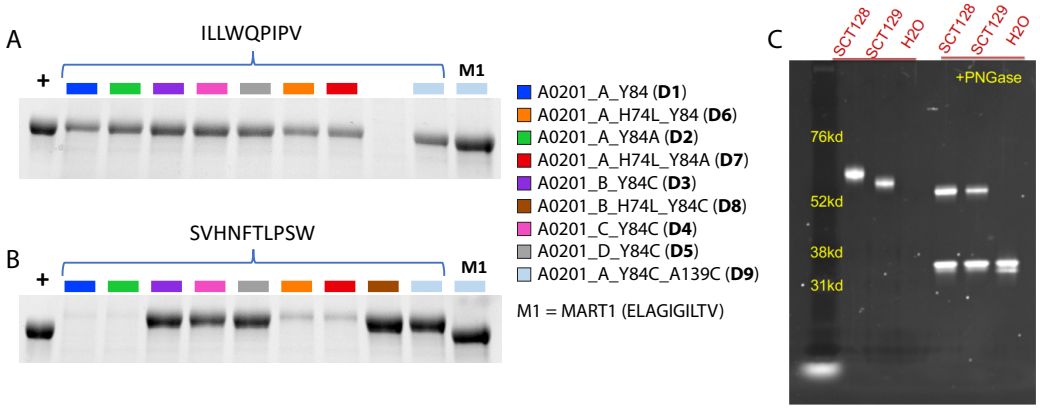


Figure 4.7: Template variants of PAP peptide SCTs.

Template variants of PAP peptide SCTs. A & B. Reduced SDS-PAGE gel showing expression of SCT template variants for ILLWQPIPV (A) and SVHNFTLPSW (B). Template designs (labeled in bold in the legend) refer to mutations from **Figure 2.2B**. +, positive control lane using A*02:01 WT1 (RMFPNAPYL) SCT. C. Reduced SDS-PAGE gel of SCTs designed under template D8 for SVHNFTLPSW (SCT128) and IMYSAHDTTV (SCT129). For lanes containing SCTs treated with PNGase, the protein mass between 31kd and 38kd markers is the PNGase protein.

Gase treatment, the masses of the two were approximately identical. This indicates that the additional mass of the former must be due to an alteration on the peptide, given that the rest of these A*02:01 SCTs encoded the same protein sequence. Further, all expressed template variants of the SVHNFTLPSW SCTs contained the same mass increase, as depicted by a uniform protein band per lane at approximately the same position (**Fig. 4.7B**), which suggests that the type and degree of glycosylation is identical regardless of other SCT mutations. An interesting future direction would

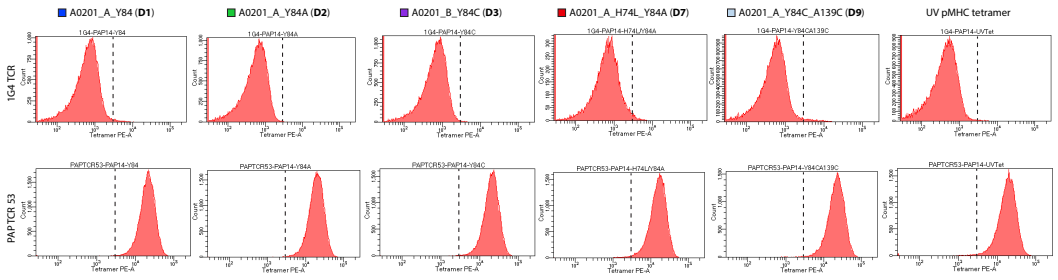


Figure 4.8: Template variants of PAP peptide SCTs.
SCT tetramers containing PAP_A2_14 peptide in various templates (**Fig. 4.7A**) were used to stain Jurkat cells, which were either transduced with a negative control 1G4 TCR (top row) or a TCR specific to the PAP_A2_14 peptide (bottom row). Dashed line represents binding threshold of 10^3 MFI (mean fluorescence intensity) from tetramer staining used to determine positive/negative signal.

be to utilize these SCTs for TCR discovery. Unlike traditional experimental methods to capture antigen-specific TCRs, which involve an initial peptide-facilitated stimulation/expansion step and subsequent tetramer binding, methods for the identification of glycosylated antigen-specific TCRs in a similar manner are not available due to the difficulty of synthesizing glycosylated peptides *in vitro* that mimic biological glycosylation. This bottleneck precludes the use of peptide stimulation and expansion to selectively enrich for cognate TCRs. Furthermore, this also renders the use of traditional refolding methods to synthesize pMHCs incompatible, as they would be presenting non-glycosylated epitopes in search of TCRs that may not be biologically relevant against glycosylated variants of the same target peptide. The use of SCTs, therefore, which enable eukaryotic glycosylating modifications to the peptides, is advantageous for this purpose. To our knowledge, there is no other existing alternative reagent containing glycosylated components that can be generated in such a facile manner.

4.3 Tumor-Specific Antigens

While TAAs have been shown to be promising antigenic targets for ACT, there have been cases reported in which TAA-specific ACT therapy induces cytotoxicity against healthy tissue, resulting in lifelong auto-immune symptoms that require maintenance by immunosuppression [2]. An alternative target for ACT therapies has been explored in the form of tumor-specific antigens (TSAs), which are only found in cancer tissue and in practice should reduce off-target T cell cytotoxicity and obviate the need for lifelong immune maintenance mediated by chemotherapeutic drugs. As

summarized previously, the primary attraction to employing TCRs targeting public versus private TSAs is the possibility of using a population-wide TSA-specific TCR repertoire to treat cancers containing these antigens. It would be advantageous then, to hone in on those public TSAs which occur frequently across populations of patients when accounting for HLA prevalence, as these might be the targets which, when functional cognate TCRs are identified, provide the best value in terms of broad applicability at the population rather than individual scale. Furthermore, because public TSAs encode driver mutations, they should have a high probability of being found homogeneously across tumor tissue, and therefore should be the most effective anti-tumor targets for ACT in terms of reducing immune escape.

4.3.1 Public TSAs - KRAS hotspot epitopes

Chief amongst these public TSAs for research interest are those encoding hotspot mutations from the KRAS protein. As a member of the GTPase family of proteins, KRAS relays cell proliferation signals to the nucleus when it is activated in a GTP-bound state. To reduce cell proliferation, KRAS is inactivated in a self-regulatory step where GTP is converted to GDP. Mutations to the KRAS protein, however, particularly at the G12 position, result in an inability to convert GTP to GDP, thus forcing KRAS to remain in an activated state, permanently inducing cell proliferation. It is estimated among all cancer cases which contain a KRAS mutation, 86% of them will contain a KRAS G12X mutation (X = D, V, C, R, S, A) [4]. If cognate TCRs against these KRAS G12X hotspot epitopes can be identified and shown to be cytotoxic against tumor cells expressing mutated KRAS, then these TCRs might serve as attractive off-the-shelf ACT candidates due to their potential broad applicability across multiple tumor types and large populations.

4.3.1.1 A*11:01-restricted KRAS G12D-specific TCRs may be captured by SCTs

The identification of such TCR candidates against A*11:01-restricted G12D and G12V peptides has been previously reported in the context of murine TCRs isolated from HLA-A*11:01 transgenic mice [5]. Recently, our collaborators at the Greenberg lab were able to identify A*11:01-restricted KRAS G12D- and G12V-specific TCRs, but from healthy human donors. A*11:01-positive healthy donors PBMCs were initially stimulated with various KRAS peptides, T cells reactive to particular G12X variants (**Fig. 4.9A**, peptides #1 & #2) were identified by measurement of

cell activation markers such as CD137, and their TCRs were sequenced and transduced into T cell lines for further functional validation via tests such as intracellular cytokin staining (IFN-γ), flow-based analysis, and cell-killing assays (**Fig. 4.10**). While the biological performance of these cell lines indicated that G12X peptides were properly loaded onto the MHC for activation of their TCRs, demonstration of the expected pMHC-TCR interaction *in vitro* by refolded tetramer binding assays often failed. In our initial proof-of-concept work with the Greenberg lab, we addressed this shortcoming of the refolded pMHCs by generating SCT variants against several of these peptides using our previously reported template D8 (**Fig. 2.2B**). We observed that SCTs containing the 10mer G12D peptide could specifically bind to T cells transduced with the cognate G12D-specific TCR4 with minimal off-target binding to G12V-specific TCRs (**Fig. 4.9C**, top row (D8 SCT variant); **Fig. 4.11A**, D9 SCT variant). This stands in contrast to the refolded pMHC tetramer variant of the G12D peptide (**Fig. 4.11B**), which binds to G12D-specific TCR4 to the same capacity as the refolded pMHC tetramer variant of the G12V peptide (**Fig. 4.11C**), indicating that the binding mechanism is essentially non-specific. Interestingly, in our comparison of SCT vs refolded pMHC performance for the G12V peptide (**Fig. 4.9C**, middle and bottom rows, respectively), both tetramers did not show expected on-target binding to G12V-specific T cell lines. However, the mechanism of failure between the two G12V constructs was different. Whereas the G12V SCT variant showed minimal binding against the G12V-specific TCRs, the refolded pMHCs exhibited a large degree of non-specific binding to all cell lines, including T cells transduced with A*02:01-restricted TCRs and T cells without any TCR. These results demonstrate two circumstances where SCTs may be favorably used as alternatives to the gold standard of refolded tetramers. In the first scenario (G12D), SCT tetramers give desired functionality where refolded tetramers show complete failure (**Fig. 4.11A & D**, respectively). In the second scenario (G12V), both reagent types fail, but refolded tetramers generate such a high false positive rate that their usage would obfuscate any capacity to identify truly G12V-specific TCRs from tetramer-bound populations.

4.3.1.2 A*11:01 and A*02:01 KRAS SCT libraries can be assessed across multiple variables to tune functionality

Due to the improved performance of G12D SCTs versus their refolded counterparts, we next wanted to investigate whether we could further tune the functionality of

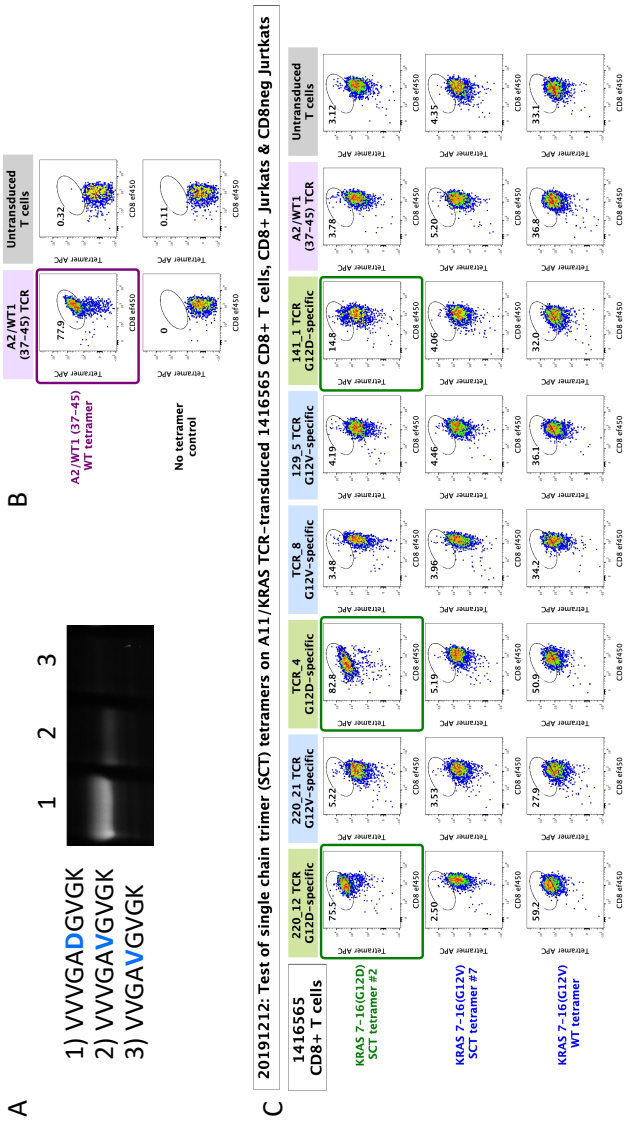


Figure 4.9: A*11:01 SCT with KRAS G12D peptide can bind specifically to cognate TCRs.

A. Reduced SDS-PAGE gel of A*11:01 SCTs expressed for three KRAS peptides using design D8 of **Figure 2.2B**. B. Tetramer binding positive and negative controls for flow cytometry. Tetramer used for WT1 is a refolded variant. C. Functional testing of various tetramers against TCR-transduced primary $CD8^+$ T cells from donor 1416565. Tetramers used include the G12D SCT tetramer as shown in (A, #1) (top row), G12V SCT tetramer as shown in (A, #2) (middle row), and a refolded variant of G12V tetramer (bottom row of flow plots). Primary $CD8^+$ T cells (column labels) are either transduced with G12D-specific TCRs (green boxes), G12V-specific TCRs (blue boxes), WT1-specific TCR (purple box), or no TCR (gray box). Number in each flow plot refers to the percentage of cells captured by the circled region to indicate tetramer staining.

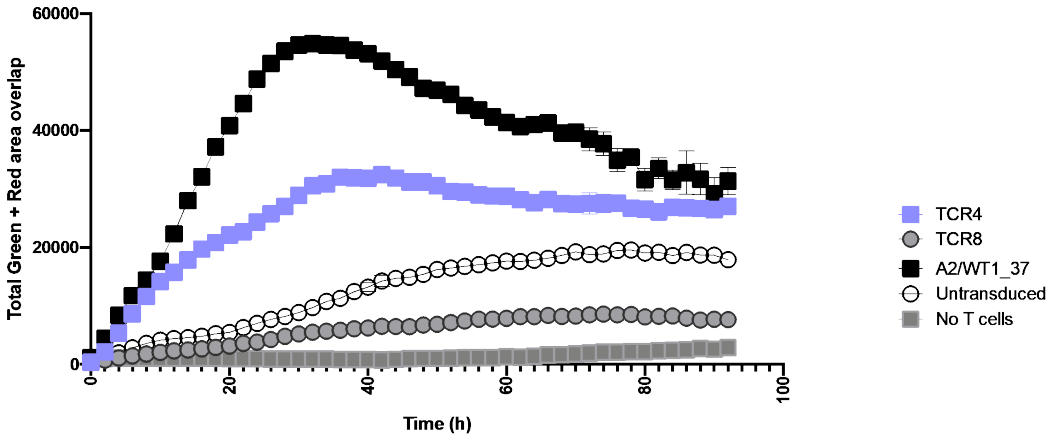


Figure 4.10: A*11:01-restricted G12D-specific TCR4-transduced T cells kill KRAS G12D-expressing tumor lines.
Panc-1 cells expressing the KRAS G12D mutation were measured for apoptosis (quantified by "Total Green + Red Cell area overlap") over time. The Panc-1 cells were incubated with $CD8^+$ T cells from healthy donors (3:1 T cell to target ratio), which were transduced with three different TCRs (TCR4: G12D-specific (VVVGADGVGK), A*11:01-restricted; TCR8: G12V-specific (VVVGAVGVGK), A*11:01-restricted; A2/WT1_37: WT1-specific (WT1_37-45 = VLDFAPPGA), A*02:01-restricted.

KRAS SCTs, given their facile capacity to accommodate numerous linker and HLA mutations. Heuristic analysis of the modified SCTs' binding performance against various G12X-specific cell lines might perhaps indicate an overall superior template design for epitope presentation. The most favorable SCT template design could subsequently be employed across a number of KRAS epitopes to generate highly effective tetramers for TCR enumeration in future patient samples. A library of nine A*11:01 SCT plasmids was first produced to encode several previously explored template designs (**Fig. 2.2B**). Next, based on NetMHC4.0 peptide binding affinity predictions, we identified fourteen favorable peptide KRAS G12X candidates (X = G, A, C, D, R, S, V) to HLA-A*11:01 (**Table 4.2A**), 'peptide' column). These peptide/HLA combinations resulted in 126 SCT-encoded plasmids that were synthesized and then transfected into Expi293 cells for SCT expression (**Fig. 4.12**). A heatmap of the relative SCT expression yield (**Table 4.2B**) revealed that SCT expression was dependent on numerous factors. Successful SCT expression beyond detectable thresholds in SDS-PAGE could only be achieved if the peptide length was ten amino acids (this is true for all templates except D6, which showed nearly uniform expression across all fourteen peptides). SCT expression was also dependent on the choice of template utilized, generally giving expression for 10mers for all templates

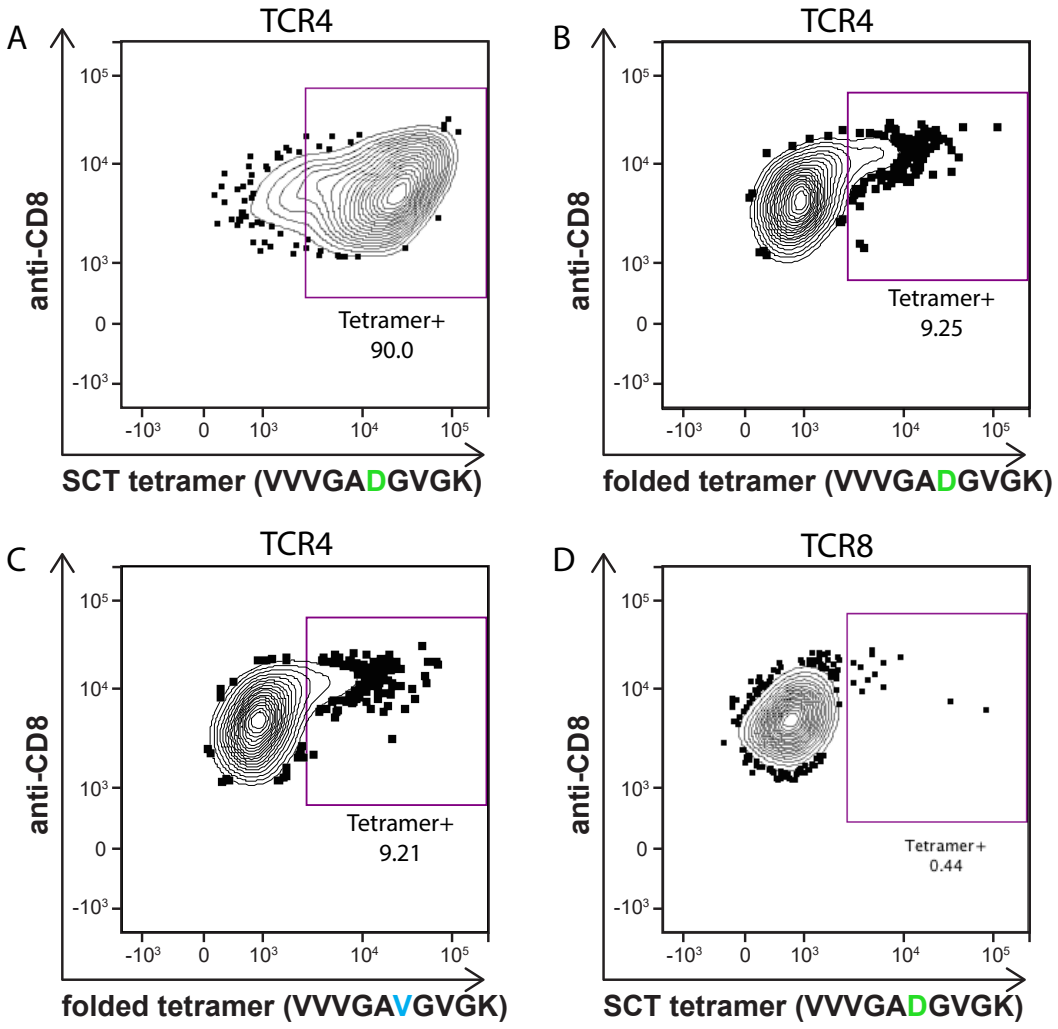


Figure 4.11: KRAS G12D A*11:01 SCTs specifically bind to cognate TCRs more efficiently than folded counterparts.
Tetramerized SCT #123 (refer to SCT ID in **Fig. 4.16 & Table M.1A**) was used to stain G12D-specific T cells (TCR4) (A). The TCR4 cell line was additionally stained with a folded tetramer variant of SCT #123 (B) and a folded G12V variant of the tetramer (C). Tetramerized SCT #123 was also used to stain a clonal population of G12V-specific T cells (TCR8) (D).

except D1 and D2. In addition, SCT yield was generally higher when a thiolated linker was utilized in the template (D3, D8, D4, D5, D9). These results mirror our prior observations with A*02:01 libraries, where D1 and D2 give poor expression, while thiolated templates give high expression (**Fig. 2.2B & Fig. 4.7B**). Another variable, the identity of G12X mutation, also seemed to play a role in affecting SCT expression. The expression data suggests that the G12 wildtype peptide may provide a baseline degree of stability given its minimal residue size. Substitution at

this position with a charged amino acid appears to give high SCT expression (G12D and G12R), while substitution with uncharged amino acids reduced SCT expression, implying a charge-dependent manner of stabilization between the G12X peptide and binding groove of HLA-A*11:01. Our claims should be further substantiated with additional experiments to assess peptide identity and SCT yield. The insertion of other amino acids into the G12 position, even those which have a low probability of occurring based on codon assignment of G12 in the human species, should serve to elucidate the phenomena discussed. Deeper insights might be revealed with thermal stability measurements and crystallography of select SCTs.

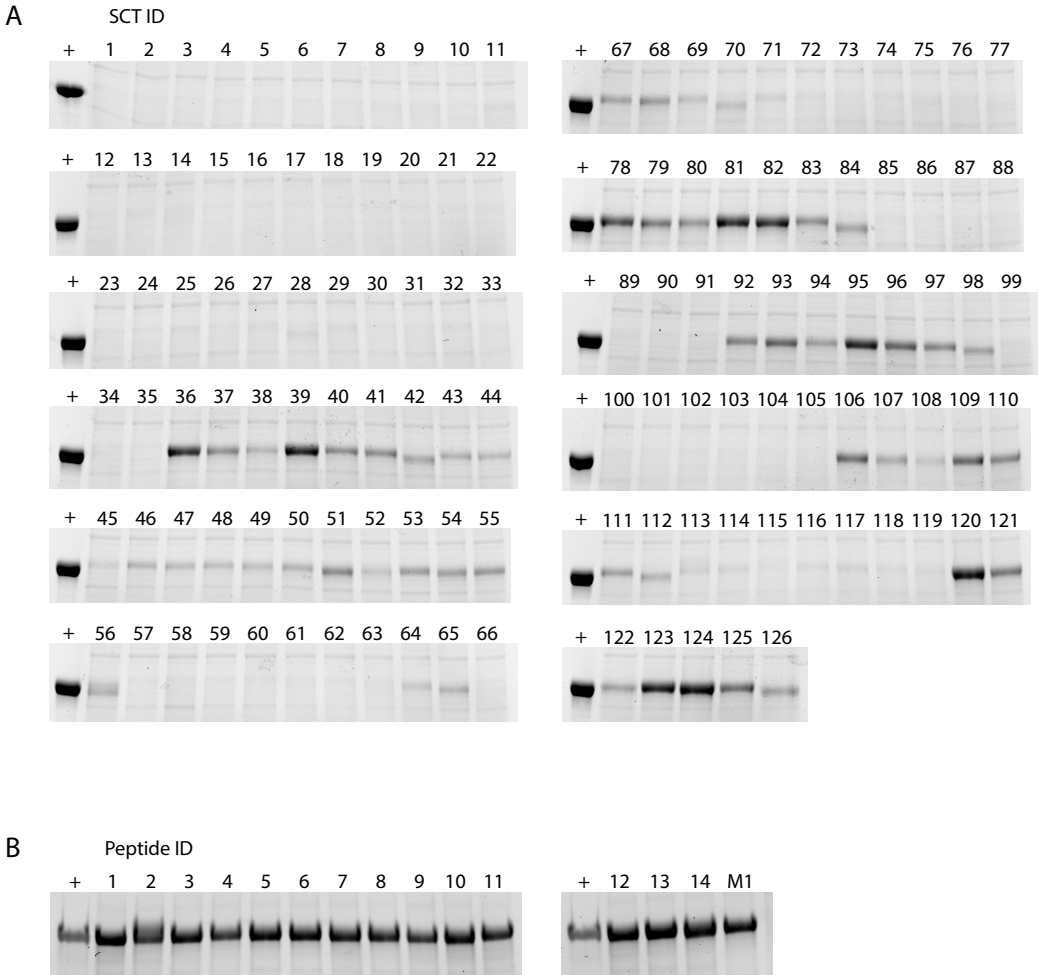


Figure 4.12: SDS-PAGE results for A*11:01 KRAS SCTs.
A. Reduced SDS-PAGE for initial A*11:01 KRAS SCT library. Number above each lane corresponds to SCT index as indicated by the matrix in **Table 4.3A**. B. Reduced SDS-PAGE for A*11:01 KRAS SCTs designed using a new template (D10), which is D9 with the additional D74L mutation. Number above each lane corresponds to the peptide ID number as indicated by ID column of **Table 4.3A**.

A

ID	peptide	antigen	netMHC affinity (nM)	A_Y84 (D1)	A_Y84A (D2)	B_Y84C (D3)	A_D74L_Y84 (D6)	A_D74L_Y84A (D7)	B_D74L_Y84C (D8)	C_Y84C (D4)	D_Y84C (D5)	A_Y84C_A139C (D9)
01	VVGAGGVGK	G12_8-16	154.38	1	15	29	43	57	71	85	99	113
02	VVGAAGVGK	G12A_8-16	147.74	2	16	30	44	58	72	86	100	114
03	VVGACGVGK	G12C_8-16	134.97	3	17	31	45	59	73	87	101	115
04	VVGADGVGK	G12D_8-16	368.19	4	18	32	46	60	74	88	102	116
05	VVGARGVGK	G12R_8-16	163.14	5	19	33	47	61	75	89	103	117
06	VVGASGVGK	G12S_8-16	114.36	6	20	34	48	62	76	90	104	118
07	VVGAVGVGK	G12V_8-16	65.47	7	21	35	49	63	77	91	105	119
08	VVVGAGGVGK	G12_7-16	299.73	8	22	36	50	64	78	92	106	120
09	VVVGAAGVGK	G12A_7-16	243.11	9	23	37	51	65	79	93	107	121
10	VVVGACGVGK	G12C_7-16	278.8	10	24	38	52	66	80	94	108	122
11	VVVGADGVGK	G12D_7-16	429.98	11	25	39	53	67	81	95	109	123
12	VVVGARGVGK	G12R_7-16	308.51	12	26	40	54	68	82	96	110	124
13	VVVGASGVGK	G12S_7-16	213.1	13	27	41	55	69	83	97	111	125
14	VVVGAVGVGK	G12V_7-16	137.28	14	28	42	56	70	84	98	112	126

B

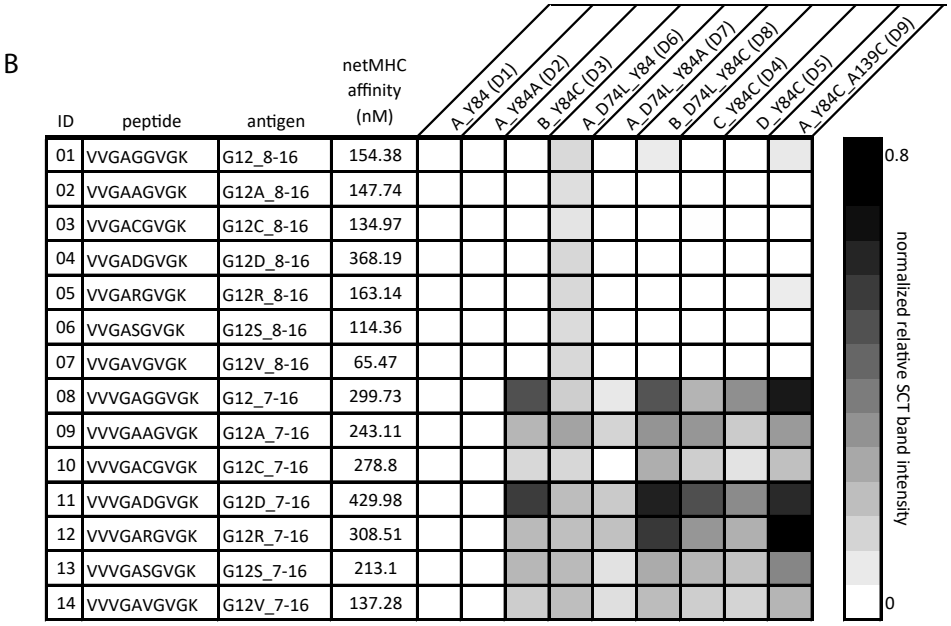


Table 4.2: A*11:01 SCT expression for KRAS peptides across various templates. A. SCT index table for 126 A*11:01 SCTs covering 14 KRAS peptides (peptide index in leftmost column) and 9 design templates (design ID refers to templates from **Fig. 2.2B**). B. Relative SCT expression heatmap as quantified by SDS-PAGE band intensity.

Examination of the reduced SDS-PAGE results showed that there were periodic SCT elements in the expression assay which display an irregular mass relative to the other expressed SCTs (**Fig. 4.12A**). In particular, SCTs #42, 65, 70, 84, 98, 112,

and 126 all showed a uniformly down-shifted protein mass. Intriguingly, this subset consists of all expressed SCTs encoding the 10mer G12V peptide. Regardless of the template selected for this peptide, expression resulted in a consistently smaller mass relative to other peptides. There are a number of hypotheses which might explain this finding. First, the G12V mutation version of the SCT may be curtailing some part of the protein translation or post-translational modification steps (e.g. a mass addition step like glycosylation) seen in the other SCT elements. Second, because the SCT signal peptide is appended directly before the peptide sequence (**Fig. 2.1A**), signal peptide cleavage might be occurring further downstream directly within the G12V peptide. This hypothesis might be validated by comparison of mass spectrometry analysis for an SCT from the G12V subset versus any other library element. Detection of the full sequence from a non-G12V peptide and detection of truncation from a G12V peptide would imply differences in protein processivity. Crystallography results from any element of this subset might display exotic epitope configurations on the peptide-binding groove of the SCT, to such an extent where it might substantially differ from the *in vivo* epitope configuration that leads to reactivity against the G12V-specific T cell lines.

KRAS SCTs which had sufficiently expressed to a moderate yield for purification were subsequently used for tetramer binding assays against the G12D-specific TCR4 cell line (**Fig. 4.3A**). All variants of the G12D SCTs were capable of binding to TCR4. Some other SCTs, primarily those containing the 10mer wildtype G12 peptide, were also found to bind to TCR4 beyond background levels. Some of these off-target binding results may have been due to human error or reagent quality issues over time, given the large binding differences for some SCTs when the assay was repeated (**Fig. 4.4**). In addition, while these results imply off-target binding of TCR4 against non-G12D peptides, functional testing of TCR4 against wildtype KRAS-expressing cells, for example, showed no upregulation of cytotoxic behavior (data not shown). Therefore, we interpreted the off-target results from the tetramer binding assay to be indicative of either human error, reagent quality issues, or epitope configurational artifacts introduced by particular peptide/template combinations in the SCT design that do not reflect true biological presentation.

When the SCTs were tested for binding against G12V-specific TCR8, no significant binding was observed. The negative binding results of all 10mer G12V variants expands upon the initial finding with the D8 variant (**Fig. 4.3B**). In the later course of this experiment, a new SCT design template was constructed, in which the D74L mu-

A G12D-specific TCR4

ID	peptide	antigen	netMHC affinity (nM)	B_Y84C (D3)	A_D74L_Y84 (D6)	A_D74L_Y84A (D7)	B_D74L_Y84C (D8)	C_Y84C (D4)	D_Y84C (D5)	A_Y84C_A139C (D9)
01	VVGAGGVGK	G12_8-16	154.38							
02	VVGAAGVGK	G12A_8-16	147.74							
03	VVGACGVGK	G12C_8-16	134.97							
04	VVGADGVGK	G12D_8-16	368.19		**1.08**					
05	VVGARGVGK	G12R_8-16	163.14							
06	VVGASGVGK	G12S_8-16	114.36							
07	VVGAVGVGK	G12V_8-16	65.47		**2.42**					
08	VVVGAGGVGK	G12_7-16	299.73	**50.7**	**1.79**			**21.5**	**46.6**	**30.3**
09	VVVGAAGVGK	G12A_7-16	243.11							**75.9**
10	VVVGACGVGK	G12C_7-16	278.8							**5.2**
11	VVVGADGVGK	G12D_7-16	429.98	**92**	**85.8**	**97.3**	**96.6**	**93.6**	**91.3**	**90**
12	VVVGARGVGK	G12R_7-16	308.51	**1.12**	**2.95**	**4.18**	**2.13**	**5.43**	**1.29**	**2.58**
13	VVVGASGVGK	G12S_7-16	213.1							
14	VVVGAVGVGK	G12V_7-16	137.28				**0.71**			

B G12V-specific TCR8

ID	peptide	antigen	netMHC affinity (nM)	B_Y84C (D3)	A_D74L_Y84 (D6)	A_D74L_Y84A (D7)	B_D74L_Y84C (D8)	C_Y84C (D4)	D_Y84C (D5)	A_Y84C_A139C (D9)
01	VVGAGGVGK	G12_8-16	154.38							
02	VVGAAGVGK	G12A_8-16	147.74							
03	VVGACGVGK	G12C_8-16	134.97							
04	VVGADGVGK	G12D_8-16	368.19		**0.9**					
05	VVGARGVGK	G12R_8-16	163.14							
06	VVGASGVGK	G12S_8-16	114.36							
07	VVGAVGVGK	G12V_8-16	65.47		**1.1**					
08	VVVGAGGVGK	G12_7-16	299.73	**0.27**	**0.57**			**0.07**	**0.2**	**0.57**
09	VVVGAAGVGK	G12A_7-16	243.11							**0.74**
10	VVVGACGVGK	G12C_7-16	278.8							**0.74**
11	VVVGADGVGK	G12D_7-16	429.98	**0.23**	**0.63**	**1.1**	**0.43**	**0.64**	**0.14**	**0.53**
12	VVVGARGVGK	G12R_7-16	308.51	**0.1**	**1.24**	**1.04**	**0.4**	**0.64**	**0.44**	**1.08**
13	VVVGASGVGK	G12S_7-16	213.1							
14	VVVGAVGVGK	G12V_7-16	137.28				**0.13**			

Table 4.3: KRAS A*11:01 SCT tetramer binding performance. A*11:01 SCTs containing various KRAS peptides were purified, tetramerized, and used to stain either a G12D-specific TCR-transduced cell line (A) or a G12V-specific TCR-transduced cell line (B). Values in each cell of the matrix represent the percentage of cells captured by tetramer staining. In (A), green numbers indicate expected positive binding, while red numbers indicate unexpected positive binding. Cells in which no values are present indicate SCTs which were not used for staining assay.

A

G12D-specific TCR4

ID	peptide	Antigen	B_Y84C (D3)	A_D74L_Y84 (D6)	A_D74L_Y84A (D7)	B_D74L_Y84C (D8)	C_Y84C (D4)	D_Y84C (D5)	A_Y84C_A139C (D9)	A1101_A_D74L_Y84C_A139C
01	VVGAGGVGK	G12_8-16		1.34						3.69
02	VVGAAGVGK	G12A_8-16								6.42
03	VVGACGVGK	G12C_8-16		0.59						10.7
04	VVGADGVGK	G12D_8-16		0.89						1.99
05	VVGARGVGK	G12R_8-16		0.75						4.18
06	VVGASGVGK	G12S_8-16		1.59						5.16
07	VVGAVGVGK	G12V_8-16		1.69						2.79
08	VVVGAGGVGK	G12_7-16	44	0.8			16.6	0	29.2	7.23
09	VVVGAAGVGK	G12A_7-16	64.2	1.94	1.94	2.14	68.4	49.2	60.6	4.42
10	VVVGACGVGK	G12C_7-16	4.98	0.7		2.54	6.67	1.19	3.67	8.24
11	VVVGADGVGK	G12D_7-16	85.6	62.8	91	91.4	87.9	83.4	87	92.2
12	VVVGARGVGK	G12R_7-16	2.23	2.53	2.49	5.58	2.34	1.74	15.1	8.4
13	VVVGASGVGK	G12S_7-16	79.8	1.49		4.49	86.5	77		7.04
14	VVVGAVGVGK	G12V_7-16	3.56	1.75		1.29		2.23	15.1	7.48

B

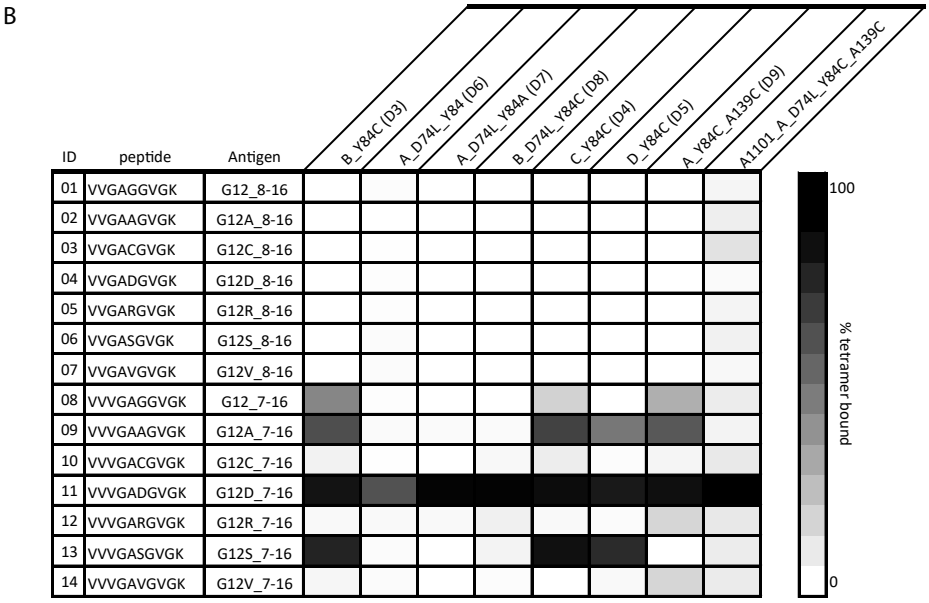


Table 4.4: KRAS A*11:01 SCT tetramer binding performance for TCR4 (repeated with inclusion of D10 SCT series).
A. A*11:01 SCTs containing various KRAS peptides were purified, tetramerized, and used to stain a G12D-specific TCR-transduced cell line. Values in each cell of the matrix represent the percentage of cells captured by tetramer staining. In (A), green numbers indicate expected positive binding, while red numbers indicate unexpected positive binding. B. Heatmap of tetramer binding values from (A).

tation was inserted into D9. This template, labeled A1101_A_D74L_Y84C_A139C, was employed in a similar manner to insert all G12X peptides for SCT expression

(**Fig. 4.12B**). The combination of the DS-SCT template with D74L appeared to have recovered SCT expression yield for all 14 variants, which had not been observed for any one of the prior design templates. Furthermore, the G12V SCT variants now displayed masses equivalent to SCTs for all other peptide variants. Binding assay results for this new series against TCR4 and TCR8 showed that these changes did not significantly change binding behavior versus previous designs ((**Table 4.4 & 4.5**). Only on-target binding to TCR4 by the 10mer G12D SCT was detected, and none of the SCTs could bind to the G12V-specific TCR8. Now that the assay was repeated with the inclusion of this new D74L template, another key feature could be seen from the binding assay. A heatmap of binding efficiency emphasizes the binding capacities of each peptide/template combination, showing a clear bifurcation of SCT binding behavior to TCR4 (**Table 4.4B**). All templates which lack the D74L mutation show off-target binding by G12, G12A, and G12S SCTs to TCR4. All templates which contain the D74L mutation, however, will only present the G12D peptide to TCR4 in a favorable manner. This data strongly denotes the importance of the D74L mutation, at least in the context of A*11:01 KRAS peptides, in terms of influencing epitope configuration on the binding groove for presentation to TCRs. In other words, SCTs with certain peptides and certain design templates may present the G12X epitopes in a manner to off-target TCRs that mimics on-target recognition or binding. D74L might serve to constrain these epitopes in such a way that enhances configurational weight of the G12 position to favor binding of TCRs only by the intended peptide target. One could then rationalize that in the search for a TCR specific to a particular G12X peptide, it might be best to utilize SCTs containing the D74L mutation, so as to enhance the SCT's capacity for identifying truly specific TCRs and to enrich the clonotype pool for productive TCR sequences.

Our collaboration with the Greenberg lab also extended to cover KRAS hotspot epitopes for A*02:01-restricted TCRs. G12V-specific, A*02:01-restricted TCRs with various degrees of cytotoxicity (**Fig. 4.13**) were identified and validated using methods similar to those described above for A*11:01-restricted TCRs. A small SCT library consisting of various KRAS hotspot peptides predicted to bind to A*02:01 or reported in the literature was prepared (**Fig. 4.14A & B**). These peptides are of various lengths (8-10 aa) and share the same initial KLVV motif. Based on a report that investigated A*02:01 compatibility with splice variants of the KRAS peptide, peptides #2, 6, and 7 of **Figure 4.14A** were included into the list to assess whether the identified TCRs might also respond to spliced G12X variants.

G12V-specific TCR8

ID	peptide	Antigen	B_Y84C (D3)	A_D74L_Y84 (D6)	A_D74L_Y84A (D7)	B_D74L_Y84C (D8)	C_Y84C (D4)	D_Y84C (D5)	A_Y84C_A139C (D9)	A1101_A_D74L_Y84C_A139C
01	VVGAGGVGK	G12_8-16		1.2						1.14
02	VVGAAGVGK	G12A_8-16								2.14
03	VVGACGVGK	G12C_8-16		0.54						3.77
04	VVGADGVGK	G12D_8-16		0.55						0.84
05	VVGARGVGK	G12R_8-16		0.69						1.19
06	VVGASGVGK	G12S_8-16		0.95						1.75
07	VVGAVGVGK	G12V_8-16		1.09						1.19
08	VVVGAGGVGK	G12_7-16	1	0.45			0.05	0	0.45	1.44
09	VVVGAAGVGK	G12A_7-16	0.45	0.8	0.54	0.35	0.35	0.25	0.25	1.14
10	VVVGACGVGK	G12C_7-16	0	0.45		0.15	0.2	0.05	0.3	2.93
11	VVVGADGVGK	G12D_7-16	0.35	0.45	0.3	0.7	0.5	0.25	0.4	5.13
12	VVVGARGVGK	G12R_7-16	0.6	1	0.2	1.68	0.45	0.55	5.51	2.39
13	VVVGASGVGK	G12S_7-16	0.5	0.9		0.55	0.15	0.35		1.64
14	VVVGAVGVGK	G12V_7-16	0.1	0.7		0.3		0.05	0.55	3.22

Table 4.5: KRAS A*11:01 SCT tetramer binding performance for TCR8 (repeated with inclusion of D10 SCT series).
A. A*11:01 SCTs containing various KRAS peptides were purified, tetramerized, and used to stain a G12V-specific TCR-transduced cell line. Values in each cell of the matrix represent the percentage of cells captured by tetramer staining.

The protein masses of the expressed SCTs in the reduced SDS-PAGE results bear striking resemblance to those of the A*11:01 KRAS library, as all G12V SCTs displayed a significantly down-shifted mass compared to their G12D counterparts (**Fig. 4.14B**, lanes 4 vs. 5, and 6 vs. 7). All eight SCTs were used in downstream tetramer binding assays against G12V-specific TCRs (**Fig. 4.14C**), demonstrating on-target binding by the desired SCT construct against all TCRs. There were three other SCTs which showed some appreciable degree of binding against some of the TCRs beyond background signal. SCTs #1 and #4 contain G12 and G12D variants of the target peptide. Because these TCRs were functionally shown to respond only to the G12V peptide, any binding of the TCRs with the G12 or G12D SCTs is indicative of off-target epitope configurational mimicry, an artifact of the SCT design. The third SCT to have significant levels of off-target binding was SCT #7, which is a splice variant of the G12V 10mer. Its binding efficiency against the TCR series reveals that the G12V-specific TCRs may all have some capacity to recognize

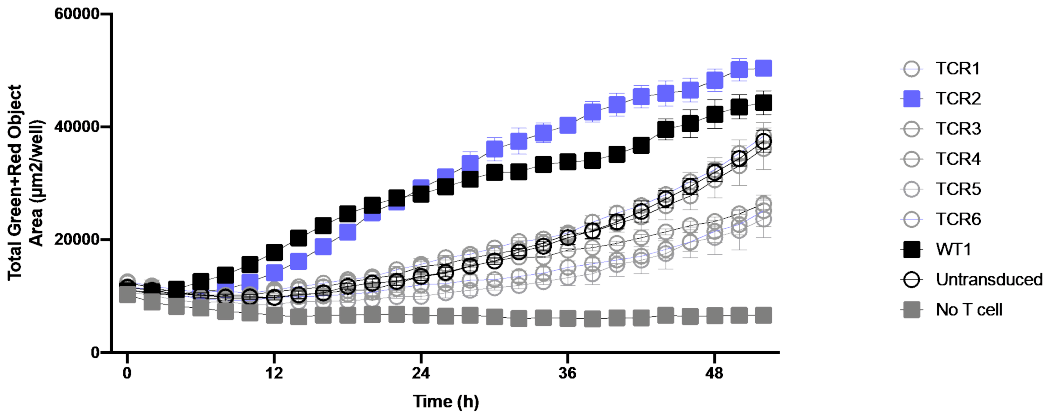


Figure 4.13: A*02:01-restricted G12V-specific TCR2 kills KRAS G12V-expressing CFPAC-1 tumor line.
CFPAC-1 cells expressing the KRAS G12V mutation were measured for apoptosis (quantified by "Total Green + Red Cell area overlap") over time. The CFPAC-1 cells were incubated with $CD8^+$ T cells from healthy donors (10:1 T cell to target ratio), which were transduced with various TCRs (see SCT binding performance in **Fig. 4.14**). WT1: WT1-specific (WT1_37-45 = VLDFAPPGA), A*02:01-restricted T cell line.

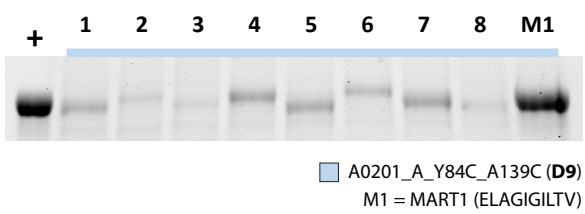
the splice variant, with TCR20 perhaps unable to distinguish between the splice and non-splice epitopes of G12V. When SCT and refolded variants of the tetramer were functionally validated by staining the TCR2 cell line, capture efficiency was virtually identical (**Fig. 4.15**). The excellent performance of this SCT reagent in spite of its low expression yield relative to G12D 10mer A*11:01 SCT (see **Fig. 4.16** for a direct comparison of yield for A*02:01 G12V 10mer SCT (#15) against A*11:01 G12V 10mer SCT (#129)) suggests no relationship between the degree of protein yield from SCT transfection and binding efficiency of the tetramerized construct.

In summary, we initially showed that SCTs may be utilized to capture A*11:01-restricted, G12D-specific TCRs. A subsequent expansion of the A*11:01 SCT library to encompass multiple KRAS G12X epitopes and various SCT template designs demonstrated that successful SCT expression could be achieved with careful selection of peptide length, peptide identity, and SCT template. A potential causative relationship between SCT mass and function was primarily hypothesized based upon the correlated mass and flow cytometry results for G12D and G12V SCTs in the A*11:01 assays. However, in the A*02:01 assay, we observed the same phenomenon of lower SCT mass for G12V SCTs whether the G12V peptide register was shifted to accommodate A*02:01 or A*11:01 binding grooves, and yet those A*02:01 SCTs

A

index	peptide	antigen
1	KLVVVGAGGV	wt-10
2	KLVVGAGGV	wt-9s
3	KLVVVGAG	wt-8
4	KLVVVGA**D**GV	G12D-10
5	KLVVVGA**V**GV	G12V-10
6	KLVVGA**D**GV	G12D-9s
7	KLVVGA**V**GV	G12V-9s
8	KLVVVGA**V**	G12V-8

B



C

index	peptide	antigen	TCR1	TCR2	TCR3	TCR4	TCR5	TCR6	TCR7	TCR8	TCR9	TCR15	TCR16	TCR17	TCR20	WT1	UNTR
1	KLVVVGAGGV	wt-10	1.64	35.7	94.8	97.4	8.16	11.1	2.49	70.2	45.5	1.16	13.7	40.3	93.9	0.3	0.34
2	KLVVGAGGV	wt-9s	0.046	0.095	0.17	0.07	0	0.023	0.046	0.21	0.14	0.093	0.069	0.023	0.16	0.2	0.22
3	KLVVVGAG	wt-8	0.046	0.092	0	0.051	0	0.046	0.099	0.026	0.099	0.025	0.025	0.075	0.29	0.08	0.06
4	KLVVVGADGV	G12D-10	0.67	1.12	5.8	97.6	2.67	97.3	15.2	5.19	14.2	1.13	21.5	89.4	1.85	0.3	0.23
5	KLVVVGAVGV	G12V-10	99.3	99.7	99.6	99.5	99.5	99.1	99.5	99.7	99.6	98.8	98.9	98.4	99.5	2.73	2.49
6	KLVVGADGV	G12D-9s	0.12	0.83	0.18	0.076	0.049	0.026	0.18	0.13	0.1	0.27	0.52	0.2	0.11	0.08	0.05
7	KLVVGAVGV	G12V-9s	57.4	34.8	33.2	15.1	23.7	8.9	22.2	13.6	5.89	9.21	8.39	9.59	97	2.22	3
8	KLVVVGAV	G12V-8	0.051	0.5	0.026	0	0.12	0.025	0.05	0.17	0.05	0.13	0.025	0.05	3.67	0.15	0.05

Figure 4.14: KRAS A*02:01 SCT expression and tetramer staining performance. A. Table of KRAS peptides selected for SCT expression into template D9 (**Fig. 2.2B**). Within peptide sequences, underlined amino acids represent a splice location, red letters represent G12 mutation. B. Reduced SDS-PAGE results of SCT expression. Numbers above lane correspond to indices of (A). +, positive control lane using A*02:01 WT1 (RMFPNAPYL) SCT. M1, MART-1 peptide (ELAGIGILTV). C. The expressed SCTs from (B) were purified, tetramerized, and used to stain multiple A*02:01-restricted, G12V-specific TCR-transduced cell lines (all specific to KLVVVGAVGV). Values in each cell of the matrix represent the percentage of cells captured by tetramer staining. WT1: WT1-specific TCR-transduced, A*02:01-restricted T cell line (WT1_37-45 = VLDFAPPGA). UNTR: non-transduced T cell line (no TCR). NOTE: TCR IDs in this figure may share the same indices as those of the A*11:01-restricted TCRs (**Fig. 4.9C**), but do not represent the same clonotypes.

proved to be highly functional. Coupled with the observation that our tenth template in the A*11:01 library (A_D74L_Y84C_A139C) recovers this mass for the 10mer G12V, it remains an open question as to whether these mass alterations in any way contribute to functional failure in an HLA-dependent manner. Overall, we show here that SCT platform can be used to design proteins according to a matrix of variables, enabling heuristic tuning for optimal yield or binding behavior. The insights we gained from downstream tetramer binding assays indicate that this library approach may allow one to stratify TCRs based on binding selectivity or behavior against particular peptide sequences, length, or splice variants.

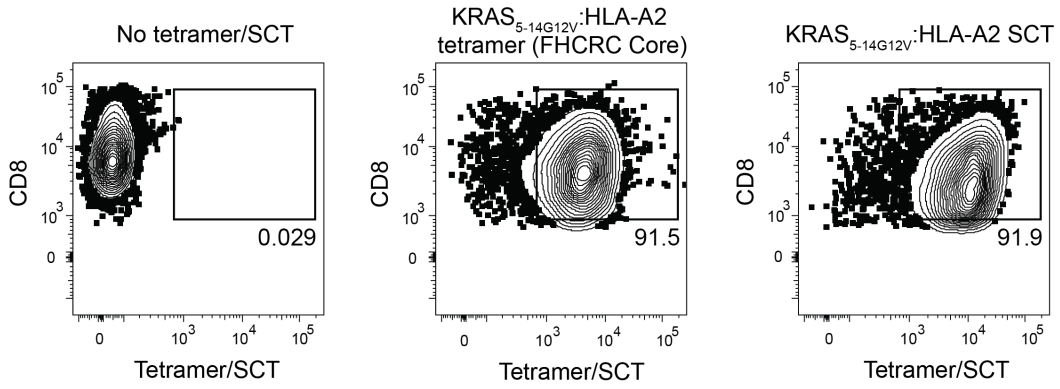


Figure 4.15: KRAS G12V A*02:01 folded and SCT tetramer variants display similar performance.
G12V-specific T cells (transduced with TCR2 of **Fig. 4.13 & 4.14**) were analyzed by flow cytometry using either no tetramer stain (left), a folded G12V (KLVVVGAVGV) tetramer variant (middle), or SCT G12V tetramer variant (same peptide, #5 of **Fig. 4.14A** or #15 of **Table M.1**) (right). Values in each flow plot represent the percentage of cells captured within bounding box, indicating tetramer binding efficiency.

4.3.1.3 A semi-biased approach to identify potentially immunogenic peptide/HLA pairs

The promising results from the KRAS SCTs covering peptides reported for A*02:01 and A*11:01 prompted us to subsequently investigate whether the library approach could be adapted for use in a broader manner to thoroughly assess more pairs of peptide and HLA haplotypes. Based on the demand for SCT template production from other projects, we generated SCT plasmids to cover a total of 24 Class I HLA haplotypes. This select list of haplotypes should be enough to overlap with at least one HLA from any given individual, according to HLA prevalence statistics from our COVID dataset (**Fig. H.1**). For all SCT templates, we chose to use the D9 design instead of its modified D74L counterpart (as studied in **Fig. 4.12B**) on the basis that we were still investigating the impact of the 74L mutation to a binding groove. Because this modification directly affects the interface between peptide and binding groove, unlike other linker or HLA mutations, it was not included due to our aim of demonstrating peptide/HLA interactions with the closest biological accuracy.

Given the high prevalence of KRAS G12 mutations among cancer cases worldwide, particularly G12D, G12R, and G12V, we wanted to focus on these particular peptides for SCT production to maximize clinical scope. Peptide fragments of 9 to 11 amino

acids in length and encompassing each of these mutations (and the wild-type variant) were identified, with the G12X position allowed to reside anywhere from the first to last position of the peptide. In total, this selection of peptides and HLAs would theoretically yield 2,880 potential SCT constructs for production (120 peptides x 24 HLAs). As we had learned from the expression of p53 hotspot encoded SCTs (**Fig. 4.19**), a completely unbiased approach to peptide selection (in other words, without assistance from peptide binding prediction algorithms) resulted in a very low expression rate (approximately 10%). To perform such an assay for 2,880 constructs (versus 240 constructs for the p53 library) within an academic setting most likely would also result in a similarly low expression rate but at an excessively high time cost due to lack of automation assistance. Thus, we aimed to reduce our final library size while simultaneously enriching for peptide/HLA pairs of high likelihood for stability (and therefore high expression) by conducting a semi-biased approach in which the peptide/HLA pairs were first passed through the NetMHCpan 4.1 peptide binding prediction algorithm. We ranked the results based on NetMHCpan's percentile EL_rank score, which served as a proxy for favorable binding affinity, for each peptide/HLA pair. Most predictions showed quite poor binding, with the best binder being an A*11:01-restricted, 9mer G12V variant, SCT #138 of **Table M.1** at EL_rank of 0.1277 (equivalent to NetMHC4.1 affinity of 65 nM), so we maximized the library size by selecting to express only peptide/HLA pairs which produced an EL_rank percentile of 20 or lower. This filter resulted in a final library size of 344 SCT elements, with the worst predicted one being SCT #290 (GARGVGKSA with B*44:02) (**Table M.3**), at EL_rank = 19.9, or equivalent to 29,927 nM binding affinity. By relaxing our filter parameter to encompass even pairs with such poor predictions, we hoped to find false negative pairs that could yield sufficient SCT expression.

The SCT library generated for the KRAS G12X library (**Fig. 4.16**) revealed some broad trends across HLAs. SCT expression was primarily observed for peptides paired with HLA-A alleles, with a few cases of measurable but low expression for those paired HLA-B alleles, and no instance of strong expression among those tested for the three HLA-C alleles. Cases of high SCT yield for each peptide appeared to be a function of peptide sequence and HLA. For example, HLA-A*02:01 and HLA-A*02:03, whose binding grooves are highly similar, generally were biased for better expression of peptides having the KLV motif in the initial region of their sequence, while A*03:01 and A*11:01's preferred peptides start with the VVV motif. A*02:11, whose binding groove is also similar to those of other A*02

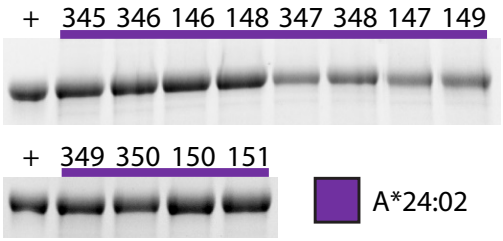
alleles, appeared to be significantly more compatible for these peptides, as almost all selected candidates resulted in expression. These findings serve to reinforce the notion that traditional mechanisms by which peptides stabilize each HLA's respective binding groove may play a role in the SCT system by influencing expression, and that therefore, the degree of SCT expression may serve as a proxy to understand and measure peptide/HLA interactions. While the expression of these constructs does appear to have a correlation with peptide/HLA compatibility, expression yield does not appear to play a significant role necessarily in determining the functional capacity of these reagents as tetramers. As discussed in earlier sections, A*02:01 and A*11:01 SCT candidates from this library (SCTs #15 & #129), which generate low and moderate yields, respectively, both have been shown to functionally behave in a desired manner, binding only to their cognate TCR-transduced cell lines, which have been shown to have peptide-specific cytotoxic functionality.

An interesting case of multiple register compatibility can be seen with A*23:01 and A*24:02 alleles, where we observe high SCT expression across multiple peptide registers (SCT #140-151) (**Fig. 4.16**). The similar degrees of expression across these two alleles for the same peptide subset is most likely due to consensus of their peptide-binding groove protein sequences (only 3 amino acids different). One exception, SCT #145 for A*23:01, failed to express to a similar manner as SCT #151 for A*24:02; this may be a case where the binding groove differences between the two HLAs are highly sensitive against this particular peptide (ARGVGKSAL). We note here two primary findings within these A*23:01 and A*24:02 SCTs. First, while some of these peptides have been identified as immunogenic for other HLA alleles, this report is, to our knowledge, the first instance where they have been identified as possibly immunogenic for A*23:01 and A*24:02 based on SCT expression and inference of potential peptide/HLA binding compatibility. The prediction scores of these identified peptides (e.g. A(V/R)GVGKSAL) for A*24:02 are extremely low (EL_rank from NetMHCpan: 19.5, binding affinity from NetMHC4.0: 40,000 nM), which may explain why they may not have been included for experimental assessment in prior works. Even in this experiment, in which we applied an extremely generous cut-off threshold of EL_rank = 20 (binding affinity = 40,000 nM) for all peptides containing the G12X mutations, the G12 and G12D versions of this peptide fell outside of the threshold (42,000 nM predicted binding) and were not initially included in this library. Given the high expression observed for G12V and G12R and assuming that high SCT expression implies stability and functionality, we believe that experimental analysis yields far more accurate insight into peptide-HLA

Figure 4.16: Multi-HLA KRAS SCT expression.
Reduced SDS-PAGE results of SCT expression for various KRAS peptide/HLA combinations using D9 template. Number above each lane corresponds to the SCT ID (**Table M.1-M.3**). Colored bars below each SCT correspond to HLA identity of the SCT, as depicted in legend (right). +, positive control lane using A*02:01 WT1 (RMFPNAPYL) SCT. SCT transfection yield is additionally quantified in **Table M.1-M.3**.

binding dynamics than current prediction algorithms. We therefore proceeded to generate G12 and G12D SCT variants for A*24:02 using these peptide registers to test these assumptions. As seen in the supplemental six SCTs of (**Fig. 4.16 & Table M.3**), these variants also led to high SCT expression. A rearrangement of these SCTs in reduced SDS-PAGE is depicted in (**Fig. 4.17**) to better highlight the similarities in SCT expression based on peptide register, regardless of G12X mutation.

A



B

ID	peptide	yield
345	GGVGKSALTI	1.26
346	DGVGKSALTI	1.37
146	VGVGKSALTI	1.36
148	RGVGKSALTI	1.35
347	EYKLVVVGAG	0.72
348	EYKLVVVGAD	0.87
147	EYKLVVVGAV	0.63
149	EYKLVVVGAR	0.60
349	AGGVGKSAL	1.36
350	ADGVGKSAL	1.30
150	AVGVGKSAL	1.48
151	ARGVGKSAL	1.32

Figure 4.17: A*24:02 KRAS peptides with poor prediction can express as SCTs. A. SDS-PAGE results for A*24:02 SCTs from **Figure 4.16**. Number above each lane corresponds to SCT ID (B). +, positive control lane using A*02:01 WT1 (RMFPNAPYL) SCT. B. Table of selected SCTs, showing peptide sequence associated with each SCT and SCT expression yield.

The second major finding from these results is that the registers of the peptides which stabilize A*23:01/A*24:02 is quite diverse. Of the six peptides tested, two contain G12X at the N-terminus, two at the C-terminus, and two contain mutations nestled within the peptide. Most reported immunogenic hotspot peptides contain the mutation in the middle of the peptide, but it would be interesting if one could identify cognate TCRs against hotspot peptides with terminally-located mutations. A worrisome possibility is that TCRs identified against C-terminus hotspot peptides may have higher non-specificity. We have seen numerous cases where the C-terminus of the peptide in an SCT can be extended beyond its natural assumed 9-11mer limit, and still provide high SCT expression or functionality. In the case of the YML series (**Section 2.1.6.2**), the Strong lab at Fred Hutch showed that A*02:01 SCTs loaded with the 14mer displayed the same conformation as the 12mer (data

not shown, manuscript in preparation). Most likely, a subset of their cognate TCRs would share similar contact points with the pMHC. For the case of hotspot peptides, if the mutation is located at the C-terminus, and the wild-type variant of peptide can also bind to the HLA groove in a similar manner, then the mutant SCT may possibly capture TCRs cross-specific to the wild-type peptide. Thus, the best peptide candidates containing C-terminus hotspot mutations are most likely those in which the terminus is presented in a configurationally distinct manner compared to the wildtype, resulting in capture of cognate TCRs distinct from those which might cross-react with the wild-type epitope. We intend to further assess this phenomenon by utilizing the crystallography approaches developed by the Strong lab to obtain crystal structures against these particular SCTs. The discovery of potentially favorable peptide/HLA interactions for this A*24:02 subset is intriguing given the high prevalence of this HLA allele (approximately 20% of world population). We plan to further validate the immunogenicity of these peptides by working together with the Greenberg lab to capture antigen-specific T cells from healthy donors and patient samples.

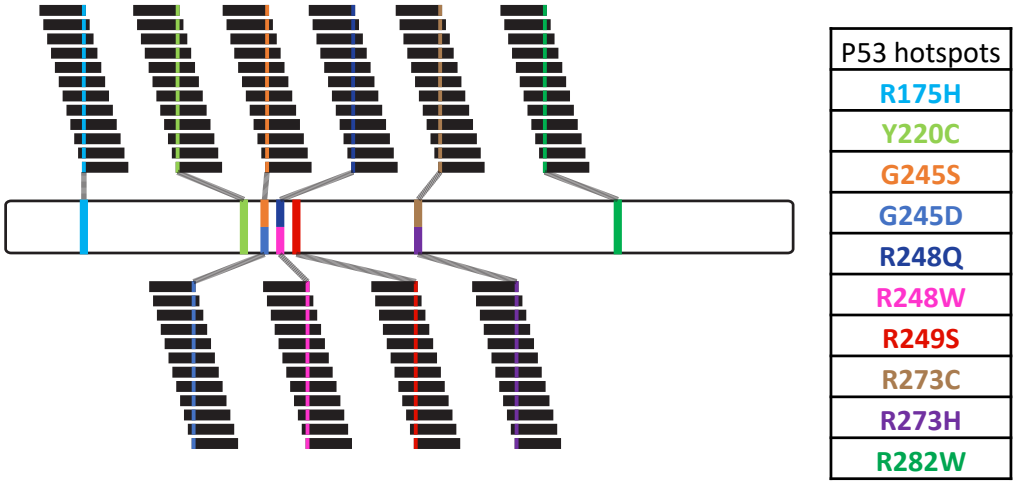
With regards to the lack of expression across HLA-B and HLA-C alleles, we provide two potential explanations. One possibility is that our SCT platform recapitulates findings in the literature stating higher sensitivities of these binding grooves to peptide repertoires, which might be exacerbated by linkers present in SCT design. A second possibility is that our A_Y84C_A139C (D9) template may have preferential expression compatibility for HLA-A alleles. This can be assessed by identifying commonly reported peptides compatible with these B and C alleles, and producing an SCT library of these peptide/HLA pairs using this template and a different template. Significant differences in SCT expression between the templates for each peptide would reveal stabilizing/destabilizing effects of the template mutations. If indeed higher SCT stability could be found with other mutations, it would require a more thorough, HLA-tailored approach to identify allele-dependent template-stabilizing mutations, similar to the experiments performed for A*02:01 (**Fig. 2.2**). Third, the observed failure of HLA-C alleles to express for KRAS peptides may simply be due to our decision to only assess three such alleles so far. A prudent follow-up experiment would be to generate a C*08:02 allele to test the KRAS G12V peptide as identified by Sim et al [6]. Expression of an SCT variant for this peptide/HLA pair would serve to validate our approach given that their report demonstrates biological functionality for the native construct.

In summary, the findings of the KRAS library demonstrate the power of a high-throughput SCT production platform for rapid identification of compatible peptide/HLA pairings against hotspot mutations from a highly prevalent oncogenic protein. We first performed a search for the best predicted G12X (X = G, D, R, V) peptide binders to 24 Class I HLA alleles. To avoid a low-yield result from a completely non-biased scan approach, we opted here for a semi-biased method in which we only tested peptide/HLA pairings that fell within the top 20th percentile of prediction scores. With the SCT platform, SCT-encoded plasmids against all 344 candidates were prepared in two weeks, and two additional weeks of transfection generated the primary results of this assay. The mismatch between predicted binding and actual SCT expression underscores the need for improved prediction algorithms or high-throughput experimental approaches similar to the SCT system for rapid testing. Our identification of significant SCT expression for peptide/HLA pairings at the prediction threshold (e.g. A*24:02 pairs, **Fig. 4.17**) indicates that this library is most likely not complete, and that there may be more false negative predictions not reported here that fell outside the range of our selection criteria. Much more work will be required, possibly requiring a full scan for some alleles (similar to p53, **Fig. 4.19**), especially those for which diverse frames were identified as leading to SCT expression. The methods outlined in this work, including the Python scripts covering peptide generation, primer generation, and SCT expression analysis, can be readily applied to any choice of protein, hotspot mutations, and HLA alleles as defined by the user.

4.3.2 Public TSAs - p53 hotspot epitopes

Whereas KRAS is considered a proto-oncogenic protein due to its role as a gatekeeper of cellular growth, the p53 protein is considered to be a tumor suppressor protein, acting as a supervisor of DNA damage to facilitate repair of faulty nucleotides. As such, the primary hotspot mutation locations within p53 are localized on the domains which make contact with DNA. Mutations within these p53 domains generally result in an increased propensity to acquire more mutations due to a lack of proofreading, thus giving way to immune escape and metastasis. Thus, there have been several efforts recently to examine the immunogenicity of epitopes arising from these hotspot mutations. Unlike KRAS, whose G12X mutations dominate the overall landscape of potential driver mutations, p53 hotspot mutations are quite diverse and tend to be spread out across the length of the entire protein, as depicted in

Figure 4.18, where the top 10 most prominent hotspot mutations of p53 are mapped out according to data from the COSMIC database of tumor patient genomes. These p53 hotspot mutations also tend to individually occur at a lower frequency, but taken altogether appear in approximately 50% of cancer cases.



R175H: SQHMTEVVRHCPHHERCSD

9mer	10mer	11mer
RHCPHHERC	RHCPHHERCS	RHCPHHERCSD
VRHCPHHER	VRHCPHHERC	VRHCPHHERCS
VVRHCPHHE	VVRHCPHHER	VVRHCPHHERC
EVVRHCPHH	EVVRHCPHHE	EVVRHCPHHER
TEVVRHCPH	TEVVRHCPHH	TEVVRHCPHHE
MTEVVRHCP	MTEVVRHCPH	MTEVVRHCPHH
HMTEVVRHC	HMTEVVRHCP	HMTEVVRHCPH
	QHMTEVVRHC	QHMTEVVRHCP
		SQHMTEVVRHC

Figure 4.18: p53 hotspot epitope selection
The top 10 most prominent hotspot mutations of p53 (top right table) are color-coded and displayed on a linear map of the protein (not to scale). For any given hotspot (R175H is depicted for example), peptides of length 9-11 aa and which contain the mutation within an internal position are selected for A*02:01 SCT production.

Recent efforts by members of the Rosenberg lab from the NIH have revealed that

antigen-specific T cells targeting p53 mutations can be isolated and demonstrated to have therapeutic potential. Deniger *et al.* first showed that p53-specific $CD4^+$ or $CD8^+$ T cells could be isolated from metastatic ovarian cancer patients [7]. The minimal epitope for T cell reactivity was elucidated using peptide stimulation assays to measure IFNγ upregulation. Subsequent HLA characterization of the samples revealed that p53 hotspot peptides were restricted to HLA-B, -C, -DP, -DQ, and -DR alleles. The authors concluded that the identified p53-specific TCRs might serve as potential candidates for ACT based on *in vitro* demonstrations of on-target cytotoxicity with no cross-reactivity to the wildtype p53 peptides. In a follow-up study, shared TCR recognition of an HLA-A*02:01-restricted p53 hotspot TSA harboring the R175H mutation was reported [8]. The minimal epitope was revealed to be HMTEVVRHC, a surprising find given that binding prediction algorithms deem this sequence to be a low affinity binder to A*02:01's binding groove. In other words, the authors here demonstrated that in the search for TCR recognition of key public TSAs such as p53 R175H-derived peptides, it is worthwhile to thoroughly comb the search space of potential peptides by synthesizing each one to perform *in vitro* assays. A complete reliance on prediction algorithms to guide peptide selection would have resulted in numerous false negatives which would have been dismissed from synthesis and downstream analysis. The drawback to such an approach, however, is the tedious synthesis of all potential peptide sequences and usage in peptide stimulation assays. Faced with this potential limitation, their report assessed only five R175H peptide candidates rather than all potential combinations if an unbiased approach were used.

4.3.2.1 An unbiased approach to rapidly identify potentially immunogenic peptide/HLA pairs

Herein, we demonstrate that the SCT platform may allow one to circumvent the experimental bottleneck of hotspot peptide synthesis, enabling a faster experimental approach to assess peptide/HLA binding interaction and identification of the best p53 hotspot epitope candidates for further testing. We systematically generated a list of 240 p53 epitopes which encompassed the top 10 hotspot mutations, ranging from 9 to 11 amino acids in length, and allowing for the mutated amino acid to reside within all amino acid positions of a peptide except the first and last position. These peptides were substituted into our A*02:01 SCT D9 template plasmids using molecular technologies previously discussed (**Fig. B.2**). The sequence-verified

plasmids were then transfected into Expi293 cells, and protein expression was analyzed by SDS-PAGE.

The SCT yield of this library (**Fig. 4.19**) shows that the majority of p53 epitopes do not successfully express for A*02:01 SCTs. This is an expected finding, due to the fact that we utilized all potential peptide sequences under the aforementioned criteria without implementing any prediction bias. As tabulated in **Tables N.1 & N.2**, virtually all of the 240 peptide candidates are predicted to be non-binders to A*02:01, with only one single peptide (#113) predicted to have less than 500 nM binding affinity, which is considered to be a weak binder by the NetMHC 4.0 algorithm. Therefore, it was interesting to see at least twenty peptides produce a moderate degree of SCT expression. Even more surprising was the observation that two of these peptides (#177 YMCNSSCMGSM and #186 YMCNSSCMGDM) resulted in a very high degree of SCT expression. For these two peptides, the reduced SDS-PAGE results showed a double band (**Table 4.6**), indicating potentially secondary products, which might be due to the dithiol present in both of these peptide sequences. Ultimately, their expression hints at potential biological compatibility between these peptides and the native A*02:01 MHC construct. The fact that both of these peptides, essentially identical in sequence but encoding either G245D or G245S mutations, give very similar SCT yields is a strong indication of the consistency behind our platform. The strong influence of these short 9-11 amino acid sequences on the degree of expression of a 50 kDa protein should warrant further investigation into the quality control mechanisms within eukaryotic cells that govern acceptance or rejection for further protein processing and secretion.

Taken altogether the unbiased approach to SCT library production performed here enabled us to thoroughly scan the broad p53 hotspot peptide repertoire for favorable binders to A*02:01. A ranked list of the highest expressing SCTs from the library reveals at least 26 potential candidates for subsequent functional analysis (**Table 4.6**). A strong positive indication of the reliability of our assay comes from the observation that among all of the possible R175H SCTs that were assembled, the best expressing one contained the same exact peptide (#007 HMTEVVRHC) as that reported by Lo *et al.* to be the most immunogenic [8]. In addition, we also observed that the other SCTs encoding R175H antigens which led to high expression shared the same register (**Table 4.6**, #157 & #077), which were also assessed by Lo *et al.* to elicit an elevated degree of IFNγ response from T cells relative to background. We anticipate SCT #007 (and perhaps #157 and #077) to be a functional reagent that can

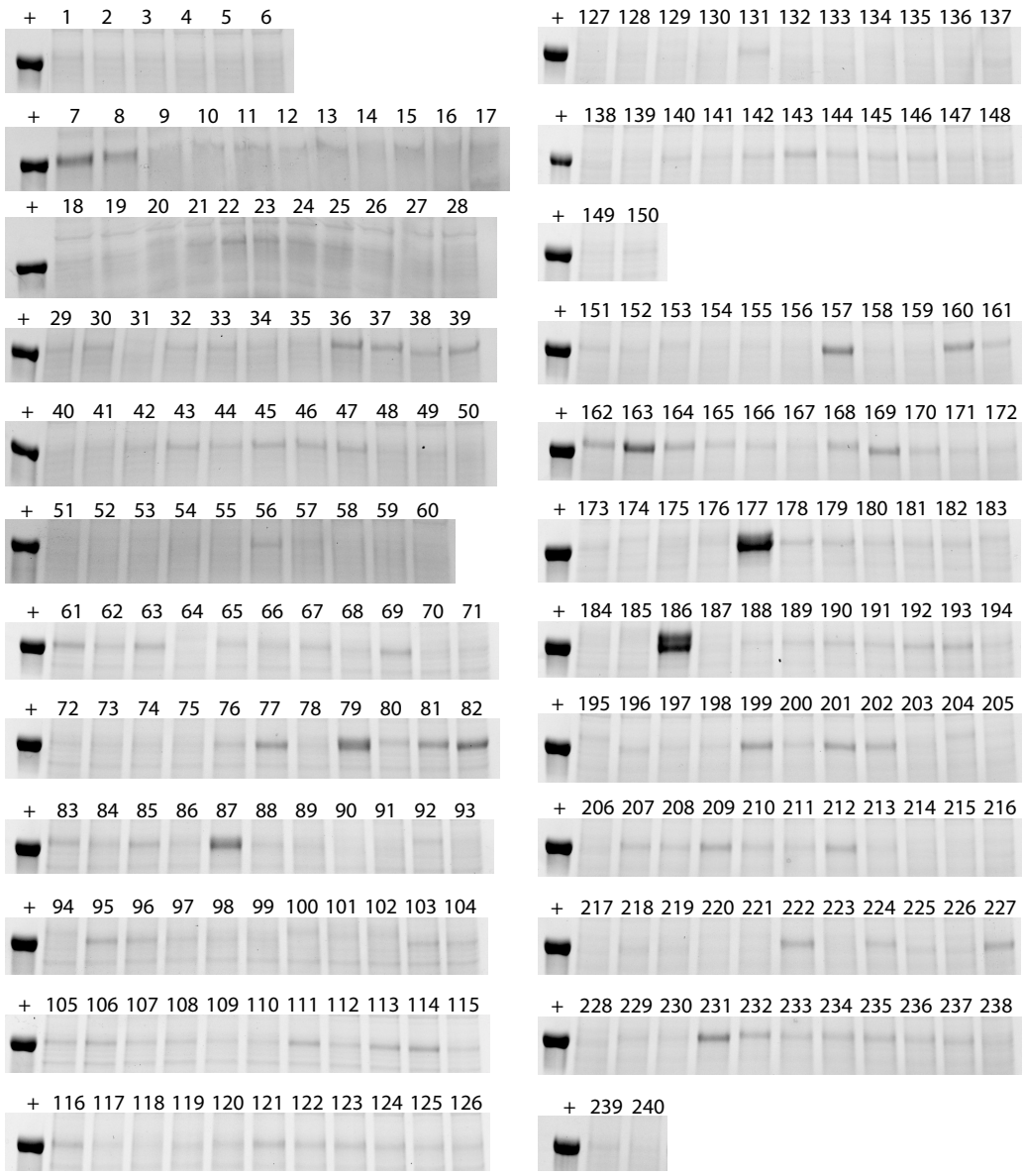


Figure 4.19: p53 hotspot A*02:01 SCT expression.
Reduced SDS-PAGE results for expression of the p53 hotspot A*02:01 SCT library using D9 template. +, positive control lane using A*02:01 WT1 (RMFPNAPYL) SCT. The SCT ID (number above each lane) and SCT transfection yield are tabulated in **Appendix N**.

be utilized in conjunction with peptide stimulation, expansion, and tetramer binding experiments to find more cognate TCRs than those previously reported. In the same vein, we hope to perform such TCR enumeration assays with our collaborators at the Fred Hutch from Dr. Greenberg's lab using the other peptides of **Table 4.6**. Most of these peptides have never been functionally reported for A*02:01 compatibility,

again most likely because they were dismissed from experimental follow-up on the basis of poor binding prediction and due to the tedium and cost associated with library-scale peptide synthesis.

ID	mutation	peptide	len	nM	yield
186	G245D	YMCNSSCMG**D**M	11	5350.3	0.98
177	G245S	YMCNSSCMG**S**M	11	3144.8	0.91
079	Y220C	P**C**EPPEVGSD	10	43910	0.38
163	Y220C	VVVP**C**EPPEVG	11	18368.6	0.37
007	R175H	HMTEVVR**H**C	9	7826.1	0.34
087	G245S	G**S**MNRRPILT	10	15779.7	0.32
082	Y220C	VVVP**C**EPPEV	10	1510.9	0.29
157	R175H	HMTEVVR**H**CPH	11	28604.7	0.27
231	R273H	LLGRNSFEV**H**V	11	1473.9	0.21
036	R248W	N**W**RPILTII	9	21658.4	0.20
081	Y220C	VVP**C**EPPEVG	10	12861.8	0.19
077	R175H	HMTEVVR**H**CP	10	16852.8	0.18
160	Y220C	P**C**EPPEVGSDC	11	42642.6	0.18
008	Y220C	P**C**EPPEVGS	9	42356.1	0.18
169	G245S	G**S**MNRRPILTI	11	9222	0.17
199	R248W	GGMN**W**RPILTI	11	1919	0.16
061	R273H	NSFEV**H**VCA	9	7518	0.16
039	R248W	GGMN**W**RPIL	9	22084.4	0.15
010	Y220C	VVP**C**EPPEV	9	681	0.15
015	G245S	G**S**MNRRPIL	9	21588.2	0.15
037	R248W	MN**W**RPILTI	9	7504	0.15
162	Y220C	VVP**C**EPPEVGS	11	30681.9	0.14
201	R248W	CMGGMN**W**RPIL	11	4245	0.14
164	Y220C	SVVVP**C**EPPEV	11	10286.8	0.14
069	R282W	ACPGRD**W**RT	9	32101.7	0.14
116	R248W	CMGGMN**W**RPI	10	178.5	0.13

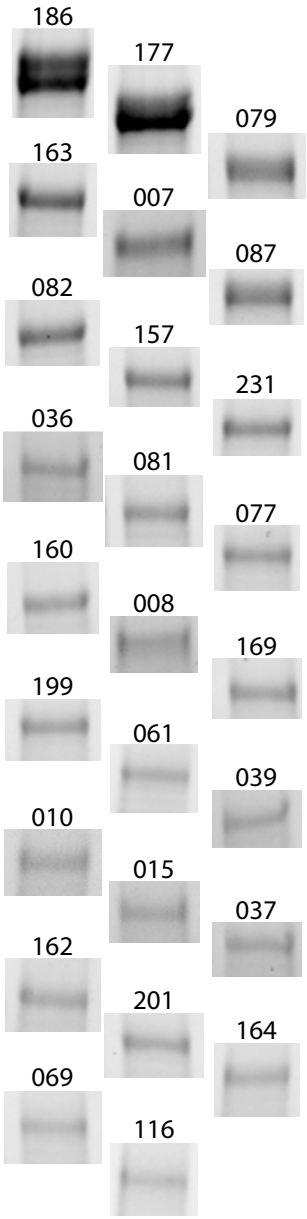


Table 4.6: Top p53 hotspot A*02:01 SCTs ranked by expression.
SCT expression is ranked by quantified SCT band and the top 26 are highlighted in table (left). Reduced SDS-PAGE results from **Figure 4.19** for each of these 26 SCTs are depicted (right).

4.3.3 Private TSAs

Wells, D. K. et al. Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction. Cell (2020) doi:10.1016/j.cell.2020.09.015.

Finally, we have also oriented our SCT technologies toward the study of private TSAs. In our collaborative efforts with the Parker Institute for Cancer Immunotherapy (PICI) and members of the Tumor Neoantigen Selection Alliance (TESLA) consortium, we experimentally validated putative neoantigens identified by prediction methods from various TESLA researchers. The main objective from this endeavor was to identify which prediction method was the most accurate in this regard, such that future prediction methods might be improved using their algorithms. The primary general findings of the combined effort by the TESLA consortium may be found in the published report cited above. Here, I summarize key contributions from the SCT technologies within this context.

TIL and PBMC samples from a melanoma cancer patient were obtained from our collaborators at the PICI. The samples were haplotyped and identified to be positive for HLA-A*01:01 and HLA-A*03:01. Lists of top private TSAs as predicted by members of the TESLA consortium for each of these two HLA alleles were also provided. Primers encoding predicted peptides for both of these HLAs were purchased and utilized in our SCT platform to generate two SCT libraries. The A*01:01 SCT library was subsequently utilized to identify antigen-specific T cells via hemocytometry analysis. This approach involves adaptation of the SCTs into our magnetic nanoparticle platform, termed NP-NACS, as previously discussed. For the A*03:01 SCT library, a library consisting of UV-exchanged pMHCs for the same peptides was prepared in parallel to allow for comparisons of SCT expression against UV-pMHC stability.

4.3.3.1 Enumeration of private TSA-specific T cells via SCT & NP-NACS technologies

Of the 27 A*01:01 peptides, 15 were able to be expressed as SCTs across at least one of the two templates (D3 and D8) tested. NP-NACS analysis of patient PBMC samples demonstrated that each of the peptides in SCT format, for either D3 or D8 designs, was capable of enumerating antigen-specific T cells (**Fig. 4.20A**). Due to the ease of cell visualization by optical microscopy for this assay, the measured

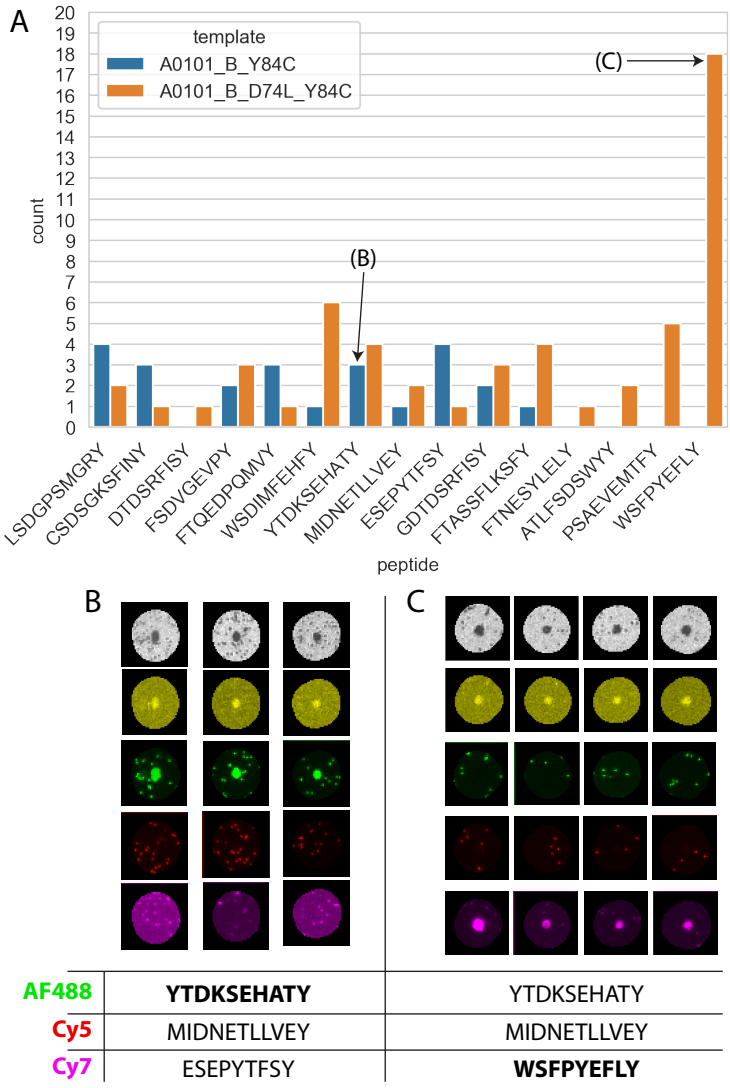


Figure 4.20: Enumeration of A*01:01-restricted, antigen-specific T cells from a melanoma patient via NP-NACS.

A. Counts of antigen-specific T cells per peptide as identified by NP-NACS assay using nanoparticles assembled with A*01:01 SCTs for each peptide. B. Example of hemocytometry analysis to identify T cells specific to the peptide YTDKSEHATY based on peptide/fluorophore associations. C. Example of hemocytometry analysis to identify T cells specific to the peptide WSFPYEFLY based on peptide/fluorophore associations. For B & C, each column represents one single cell isolated by NP-NACS. For B & C, from top to bottom row, hemocytometry images are displayed for cells imaged using brightfield, Cy3 filter (cell viability stain), AF488 filter, Cy5 filter, or Cy7 filter, to demonstrate coverage by nanoparticles, viability, and specificity to only one SCT reagent. In the table below B & C, peptide/fluorophore associations are provided for each of the two assays.

frequencies were declared to be antigen-specific with high confidence. Shown in **Figure 4.20B & C** are two examples of the NP-NACS hemocytometry counting results, whereby nanoparticle-coated objects within a chip may be individually isolated and confirmed to be healthy cells with a viability stain. The cells are then checked for specific binding to only one of three unique SCT elements per round of analysis due to each SCT element being associated with a particular fluorophore on the NP scaffold. While all SCTs could identify some T cells, the frequencies were quite low for all of them except WSFPYEFLY. This suggests possibly that T cells with reactivity to this peptide were prominently activated and proliferated in a biased manner compared to T cells of other specificities. Whether they play an immuno-protective role or might play a more active role during response to tumors within the time frame of sample extraction is beyond the scope of this work. However, it would be intriguing to further answer these questions by using these SCTs to further isolate TCRs for subsequent sequencing, characterization, and functional analysis.

4.3.3.2 Comparison of private TSA A*03:01 SCT expression and UV-pMHC stability

For the A*03:01 peptides, we initially generated an SCT library and saw various degrees of expression for each one under the same D3 and D8 designs (**Table 4.7**, col. 1 & 2). We questioned whether the expression variation might in some way be correlated with the stability of the peptide and binding groove interaction in a folded construct. Therefore, UV-labile pMHCs were expressed, and UV-facilitated peptide exchange was conducted to generate a UV-pMHC for each peptide element. The UV-pMHC library was then assessed by an ELISA assay to identify whether a strong positive absorbance signal could be detected from an anti-β2m antibody. Presence of the signal indicates retention of the β2m chain in the complex, whose HLA α chain is tethered on the ELISA plates. If β2m remains bound to the α chain, then this implies that the peptide is a compatible binder for the HLA's peptide groove, such that the three-component complex maintains stability and functionality. The absorbance data was normalized against background signal from a UV-labile pMHC exposed to UV light in the absence of peptide in solution (this will lead to complete pMHC instability), and the UV exchange efficiency of each construct was then assessed (**Table 4.7**, col. 3). The SCT expression data and UV exchange efficiencies were re-interpreted in a binary format. If SCTs expressed to such an extent where the quantified band intensity was above our cutoff score of 0.15, they were considered to

be robust and qualified for purification toward downstream analysis. If UV-pMHCs showed exchange efficiency beyond 30%, then its exchanged peptide was interpreted to be a favorable binder to the HLA construct so as to confer stability.

The results of this re-interpretation are shown in **Table 4.7** (col. 4, 5, & 6), and revealed a surprisingly high degree of agreement, where both SCT and UV-pMHC formats showed high expression and stability, respectively, or where both failed to generate a stable construct. Comparing D3 SCTs with UV-pMHCs (**Table 4.7**, col. 4 vs. 6), 29 of the 37 SCTs (78%) showed agreement (**Table 4.7**, col. 7). Comparing D8 SCTs with UV-pMHCs (**Table 4.7**, col. 5 vs 6), 31 of the 37 SCTs (84%) showed agreement (**Table 4.7**, col. 8). The difference between these two comparisons can be attributed to expression differences between the SCT designs for some peptides. In general, the D8 template appeared to increase SCT expression for most peptides (**Table 4.7**, col. 9), to such an extent where five of them were elevated from low or no expression status for D3 to moderate expression or better under D8 (peptides #12, 18, 20, 28, 36). However, peptide #26 was one instance where the D8 template appeared to have a negative impact, completely abolishing expression. The presence of D74L, which distinguishes the D8 template from the D3 template, most likely explains these expression differences. A further exploration of the impact this change has on various peptides that fall within the two subsets above (increased or decreased expression under D8 relative to D3) would highlight the motifs that causes such sensitivities to emerge within peptide sequences. As for the three UV-pMHCs (2, 3, 10) which were considered to be efficiently exchanged but could not be generated as an SCT under D3 or D8 formats (**Table 4.7**, compare col. 6 to col. 4 & 5), we caution over-interpretations of the biological nature of the UV-pMHCs based on these results. These three UV-pMHCs are close to the 30% threshold for efficient exchange, indicating perhaps that they may not be stable or functional reagents. Further interpretations into the peptide/HLA relationships of this data will require functional applications to identify antigen-specific TCRs for validation of pMHC stability.

4.3.4 Concluding Remarks on TSAs

In conclusion, I first presented in this section a proof-of-concept demonstration of our SCT platform in the context of two proteins harboring strong public TSA candidates for ACT therapy. For KRAS, we implemented a semi-biased approach in order to broadly account for numerous HLA alleles, and ranked the quality of pep-

tide/HLA pairings on the basis of SCT expression, some of which to our knowledge have not been previously reported. For p53, a completely unbiased approach was utilized, which, while more laborious per HLA, enabled us to definitively examine all potential p53 hotspot interactions with the prominent HLA-A*02:01 allele. The methodologies behind these two approaches can be readily implemented for the further study of other driver mutations in other proteins of major significance to the field (e.g. PIK3CA, BRAF, EGFR) [9]. The primary objective behind these protocols is to streamline the identification of key public TSA targets, which may accelerate TCR discovery and lead to translational applications of hotspot-oriented immunotherapeutics.

In addition, the SCT platform was used in the context of private TSAs across two HLAs for a melanoma patient. Here, we used the expressed SCT elements to augment our NP-NACS procedure for T cell enumeration. The SCTs were also characterized against UV-pMHC counterparts to demonstrate similarities between the two technologies.

The robustness of the SCT reagents across these projects was shown in various ways. KRAS reagent quality was demonstrated by our collaborators (Greenberg lab), who had hotspot-specific T cell lines for flow cytometry validation. As for p53, validation was assumed on the basis of observing high SCT expression for previously reported A*02:01 epitopes. The A*01:01 private TSA experiments used NP-NACS methods to showcase the technology's capacity for enabling visualization of clean specific binding between SCT and cognate TCRs. Finally, the A*03:01 private TSA experiments demonstrated that the degree of SCT expression was strongly correlated with UV-pMHC construct stability. A more complete approach to validation on these fronts will ensue upon resumption of cancer research after our efforts on COVID-19 projects.

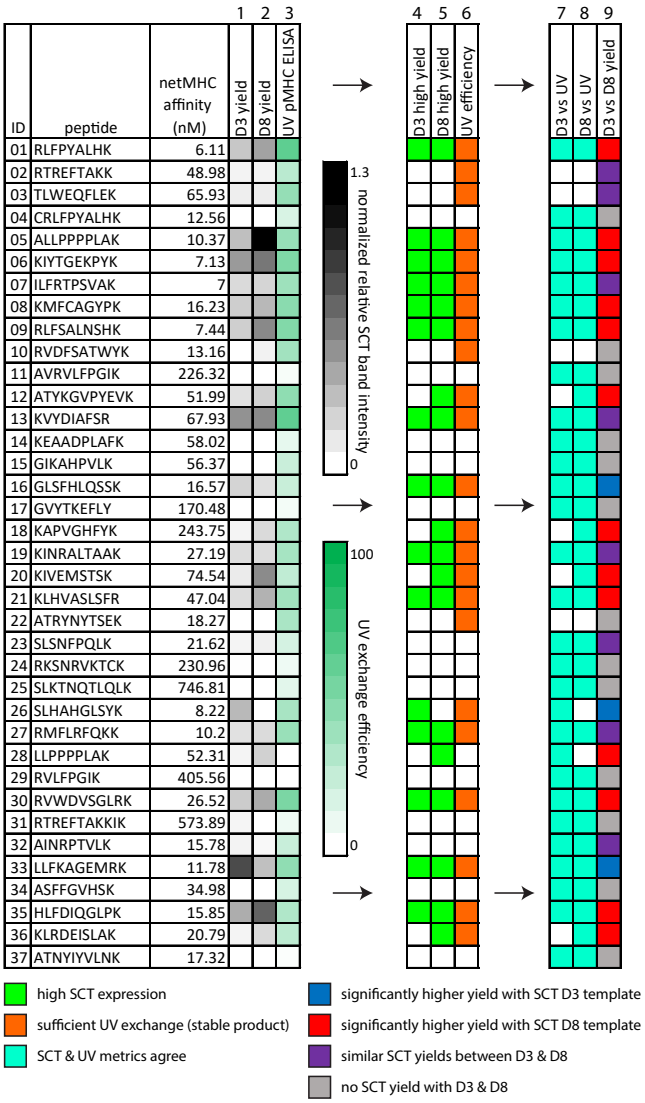


Table 4.7: A*03:01-restricted private neoantigens show agreement in stabilization of SCT or folded pMHC constructs.

Thirty-seven A*03:01-restricted private neoantigens predicted by collaborators of the TESLA consortium for a melanoma patient were used to generate SCTs (using templates D3 or D8 of **Fig. 2.2B**) or UV-exchanged pMHCs. Relative expression of D3 and D8 SCTs were quantified in a heatmap (**Table 4.7**, col. 1 & 2). Stabilization of the UV-exchanged pMHCs was correlated with an ELISA assay measuring peptide exchange efficiency (column 3). This initial data was converted into a binary format to identified stable pMHCs; for SCTs, high expression/stability was defined as relative expression above 0.15 ratio to WT1 positive control SCT (**Table 4.7**, col. 4 & 5); for UV pMHCs, stabilized pMHC product was defined as ELISA efficiency > 30%. Agreement of stabilization between SCT and UV-pMHC per peptide was assessed agreement (**Table 4.7**, col. 7 & 8). Significant differences in SCT yield between D3 and D8 templates per peptide was also assessed (**Table 4.7**, col. 9).

BIBLIOGRAPHY

[1] Rong-Fu Wang and Helen Y. Wang. Immune targets and neoantigens for cancer immunotherapy and precision medicine. *Cell Research*, 27(1):11–37, January 2017. ISSN 1001-0602. doi: 10.1038/cr.2016.155. URL http://www.nature.com/cr/journal/v27/n1/full/cr2016155a.html.

[2] Michael T Bethune and Alok V Joglekar. Personalized T cell-mediated cancer immunotherapy: progress and challenges. *Current Opinion in Biotechnology*, 48:142–152, December 2017. ISSN 09581669. doi: 10.1016/j.copbio.2017.03.024. URL https://linkinghub.elsevier.com/retrieve/pii/S095816691630283X.

[3] Aude G. Chapuis, Daniel N. Egan, Merav Bar, Thomas M. Schmitt, Megan S. McAfee, Kelly G. Paulson, Valentin Voillet, Raphael Gottardo, Gunnar B. Ragnarsson, Marie Bleakley, Cecilia C. Yeung, Petri Muhlhauser, Hieu N. Nguyen, Lara A. Kropp, Luca Castelli, Felecia Wagener, Daniel Hunter, Marcus Lindberg, Kristen Cohen, Aaron Seese, M. Juliana McElrath, Natalie Duerkopp, Ted A. Gooley, and Philip D. Greenberg. T cell receptor gene therapy targeting WT1 prevents acute myeloid leukemia relapse post-transplant. *Nature Medicine*, 25(7):1064–1072, July 2019. ISSN 1078-8956, 1546-170X. doi: 10.1038/s41591-019-0472-9. URL http://www.nature.com/articles/s41591-019-0472-9.

[4] Tatu Pantsar, Sami Rissanen, Daniel Dauch, Tuomo Laitinen, Ilpo Vattulainen, and Antti Poso. Assessment of mutation probabilities of KRAS G12 missense mutants and their long-timescale dynamics by atomistic molecular simulations and Markov state modeling. *PLOS Computational Biology*, 14(9):e1006458, September 2018. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1006458. URL http://dx.plos.org/10.1371/journal.pcbi.1006458.

[5] Q. J. Wang, Z. Yu, K. Griffith, K.-i. Hanada, N. P. Restifo, and J. C. Yang. Identification of T-cell Receptors Targeting KRAS-Mutated Human Tumors. *Cancer Immunology Research*, 4(3):204–214, March 2016. ISSN 2326-6066, 2326-6074. doi: 10.1158/2326-6066.CIR-15-0188. URL http://cancerimmunolres.aacrjournals.org/cgi/doi/10.1158/2326-6066.CIR-15-0188.

[6] Malcolm J. W. Sim, Jinghua Lu, Matthew Spencer, Francis Hopkins, Eric Tran, Steven A. Rosenberg, Eric O. Long, and Peter D. Sun. High-affinity oligoclonal TCRs define effective adoptive T cell therapy targeting mutant KRAS-G12D. *Proceedings of the National Academy of Sciences*, 117(23):12826–12835, June 2020. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1921964117. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1921964117.

[7] Drew C. Deniger, Anna Pasetto, Paul F. Robbins, Jared J. Gartner, Todd D. Prickett, Biman C. Paria, Parisa Malekzadeh, Li Jia, Rami Yossef, Michelle M. Langhan, John R. Wunderlich, David N. Danforth, Robert P. T. Somerville, and Steven A. Rosenberg. T-cell Responses to TP53 "Hotspot" Mutations and Unique Neoantigens Expressed by Human Ovarian Cancers. *Clinical Cancer Research*, 24(22):5562–5573, November 2018. ISSN 1078-0432, 1557-3265. doi: 10.1158/1078-0432.CCR-18-0573. URL https://clincancerres.aacrjournals.org/content/24/22/5562.

[8] Winifred Lo, Maria Parkhurst, Paul F. Robbins, Eric Tran, Yong-Chen Lu, Li Jia, Jared J. Gartner, Anna Pasetto, Drew Deniger, Parisa Malekzadeh, Thomas E. Shelton, Todd Prickett, Satyajit Ray, Scott Kivitz, Biman C. Paria, Isaac Kriley, David S. Schrump, and Steven A. Rosenberg. Immunologic Recognition of a Shared p53 Mutated Neoantigen in a Patient with Metastatic Colorectal Cancer. *Cancer immunology research*, 7(4):534–543, April 2019. ISSN 2326-6066. doi: 10.1158/2326-6066.CIR-18-0686. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6685528/.

[9] Fangjun Chen, Zhengyun Zou, Juan Du, Shu Su, Jie Shao, Fanyan Meng, Ju Yang, Qiuping Xu, Naiqing Ding, Yang Yang, Qin Liu, Qin Wang, Zhichen Sun, Shujuan Zhou, Shiyao Du, Jia Wei, and Baorui Liu. Neoantigen identification strategies enable personalized immunotherapy in refractory solid tumors. *The Journal of Clinical Investigation*, 129(5):2056–2070. ISSN 0021-9738. doi: 10.1172/JCI99538. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6486339/.

Chapter 5

SENSITIVE DETECTION AND ANALYSIS OF NEOANTIGEN-SPECIFIC T-CELL POPULATIONS FROM TUMORS AND BLOOD

Songming Peng[1,†], Jesse M. Zaretsky[2,†], Alphonsus H.C. Ng[1,3,†], **William Chour[3,4,†]**, Michael T. Bethune[4], Jongchan Choi[3], Alice Hsu[4], Elizabeth Holman[1], Xiaozhe Ding[1,4], Katherine Guo[1], Jungwoo Kim[1], Alexander M. Xu[1,3], John E. Heath[1], Won Jun Noh[4], Jing Zhou[1], Yapeng Su[1,3], Yue Lu[1,3], Jami McLaughlin[5], Donghui Cheng[5], Owen N. Witte[5,6,7], David Baltimore[4], Antoni Ribas[2], and James R. Heath[1,3,6,*]

[1] Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 East California Blvd, Pasadena, CA 91125, USA

[2] Department of Medicine, University of California Los Angeles and Jonsson Comprehensive Cancer Center 10833 Le Conte Avenue, Los Angeles, CA 90095, USA

[3] Institute for Systems Biology, Seattle, WA 98109, USA.

[4] Division of Biology and Biological Engineering, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

[5] Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, [6] Department of Microbiology, Immunology and Molecular Genetics, and [7] Howard Hughes Medical Institute, University of California, Los Angeles, CA 90095, USA

[†] These authors contributed equally to this work.

[*] Correspondence: jheath@systemsbiology.org

Peng, S. et al. Sensitive Detection and Analysis of Neoantigen-Specific T Cell Populations from Tumors and Blood. Cell Reports 28, 2728–2738 (2019).

5.1 Summary

Neoantigen-specific T cells are increasingly viewed as important immunotherapy effectors, but physically isolating these rare cell populations is challenging. Here, we describe a sensitive method for the enumeration and isolation of neoantigen-specific

$CD8^+$ T cells from small samples of patient tumor or blood. The method relies on magnetic nanoparticles that present neoantigen/MHC tetramers at high avidity via barcoded DNA linkers. The magnetic particles provide a convenient handle to isolate the desired cell populations, and the barcoded DNA enables multiplexed analysis. The method exhibits superior recovery of antigen-specific T cell populations relative to literature approaches. We applied the method to profile neoantigen-specific T-cell populations in the tumor and blood of patients with metastatic melanoma over the course of anti-PD1 checkpoint inhibitor therapy. We show that the method has value for monitoring clinical responses to cancer immunotherapy, and may help guide the development of personalized mutational neoantigen-specific T-cell therapies and cancer vaccines.

5.2 Introduction

Tumor neoantigens have been implicated in T-cell recognition of tumors, and are useful in the design of personalized cancer vaccines [1–3] and T-cell receptor (TCR)-engineered adoptive cell therapies [4, 5]. Neoantigens are mutation-containing peptide fragments of tumor-associated mutant proteins that can be presented by major histocompatibility complex (MHC) class I protein complexes for $CD8^+$ T-cell surveillance. These neoantigens are potentially recognized by highly specific TCRs, thus avoiding off-target interactions. The tumor-specificity of neoantigens, coupled with the ability of neoantigen-specific T cells to selectively kill cancer cells [6–8], have made them increasingly important for cancer immunotherapy.

Putative neoantigen peptides can be predicted by analyzing the tumor exome for mutated genes that may result in the presentation of a mutational peptide to T cells [7–11]. Candidates are typically ranked according to level of expression and the predicted peptide (p)-MHC binding affinity [12, 13]. Experimental testing of which candidate neoantigens are actually generating an anti-tumor T cell response is challenging. For example, considering only somatic mutations, a given tumor might yield 50 or more putative neoantigens with 500 nM or lower calculated binding constant (kd) to a given HLA allele, and each patient will have 6 or so such alleles. Second, any given neoantigen-specific T-cell clone is likely to exist in low abundance. Yet, harnessing neoantigen-specific T-cells for therapy has yielded promising clinical results, highlighting the value of meeting these challenges. One approach involves directly expressing putative neoantigens within antigen-presenting target cells that are HLA-genotype matched with the patient, and then incubating those cells

with tumor infiltrating lymphocytes (TILs) or T cells from peripheral blood mononuclear cells (PBMCs) to identify neoantigen reactive T-cell populations [8, 14]. This approach can identify such populations, but not quantitatively enumerate them. A second approach involves the use of multi-color labeled MHC tetramers for multiplex flow cytometry [15]. pMHC tetramers labeled for mass cytometry analysis [16, 17], or DNA-labeled tetramers designed for sequencing analysis [18, 19], have also been reported. These flow cytometry methods typically require reasonably large cell populations for analysis, and are often used to analyze *in vitro* expanded T cells [20]. Such expansion can significantly alter T-cell population profiles. Nevertheless, these methods have been used to identify neoantigen-specific $CD8^+$ T-cell populations in high mutation burden tumors [21].

A further measurement challenge is to match the neoantigen-specificity of a T cell with the TCR α and β chains. The (single cell) pairSEQ technique [22] provides an elegant approach for assembling the full TCR gene sequence, but does not establish the antigen specificity of that gene [23, 24]. In general, the dual challenge of identifying neoantigen-specific T cells and matching them with their cognate TCR genes increases in difficulty as abundance of the individual T-cell populations drops.

Here, we report on the method of nanoparticle (NP)-barcoded nucleic acid cell sorting (NACS), for the sensitive enumeration and isolation of antigen-specific $CD8^+$ T cells from small samples of patient tumor cells or blood. This method relies on peptide (p) MHC tetramers coupled, via DNA linkers, to magnetic nanoparticles (NPs). The NPs provide a convenient handle to isolate the desired cell populations and permit a high loading of pMHC tetramers on the NP surface to increase tetramer avidity. The DNA linkers can be further used as barcodes for multiplexed antigen-specific T-cell population profiling. We first demonstrate that the approach exhibits significantly superior recovery of antigen-specific T cells spiked in to a background of healthy human PBMCs, relative to the multi-color tetramer-based flow cytometry gold standard method [15]. We then apply the method to profile neoantigen-specific T-cell populations in the tumor and blood from patients with melanoma over the course of their treatment with anti-PD1 immunotherapy. We also demonstrate TCR gene sequencing of captured antigen- and neoantigen-specific T cells. These results demonstrate that NP-barcoded NACS is a sensitive tool for antigen specific $CD8^+$ T-cell capture and analysis. They also indicate that such analysis can provide a valuable and highly specific immune profiling tool for monitoring cancer immunotherapy responses in patients.

5.3 Results

5.3.1 Construction of NP-barcoded NACS reagents and sensitivity analysis for antigen-specific T-cell capture

The basic components of a NP-barcoded NACS reagent are shown in **Figure 5.1A**. The reagent is comprised of pMHC tetramers linked to magnetic NPs via ssDNA oligomers, and each component is designed for one or more specific purposes. The biotin-labeled pMHC components were prepared as described in the literature [5, 25, 26]. For the tetramer scaffold, we used cysteine-modified streptavidin (SAC) [27], which is engineered for ssDNA-labeling in a site-specific manner [28] that does not interfere with biotin binding. We explored a size range of iron oxide magnetic NPs to find that a radius (r) = 500 nm was optimal. Smaller particles (r = 50 nm) require strong magnetic fields and extensive processing for cell enrichment, which reduces cell recovery. Particles of r > 1 micrometer rapidly precipitate, making them difficult to manipulate. Each NP can present up to 300,000 ssDNA oligomers, which provide handles for hybridization with DNA-functionalized SAC molecules. This modularity of components, coupled with DNA-directed assembly, makes the preparation of a pNP library straightforward; most components can be prepared ahead of time and cryogenically stored until ready for use. In addition, as elaborated below, the sequence of each DNA oligomer can serve as a barcode for the identity of a specific peptide antigen, thus enabling multiplexed analysis. An assembled NP-barcoded NACS reagent is called a pNP. Each pNP presents >20,000 pMHC tetramers (**Fig. 5.5A**). To capture and enumerate a single antigen-specific T-cell population using NP-barcoded NACS, a small cocktail of pNPs is prepared, each with a different fluorophore. One pNP is specific to the T cell population of interest, and the others serve as controls. That cocktail is mixed with viability-stained $CD8^+$ T cells (**Fig. 5.1B**). The pNP-bound cells and free particles are then isolated with a magnet, and the captured cells are imaged and counted on a hemocytometer chip (**Fig. 5.1B**, top right). A cell is considered specific to the neoantigen of interest if it is decorated with a multiplicity (typically 5-20) NPs of a single fluorescent color. To enumerate mixed populations of antigen-specific T cells, one can either repeat this type of analysis in a serial manner (**Fig. 5.1B**, bottom right), or implement a parallel NP-barcoded NACS protocol using the oligomer barcode, as described below.

We first tested the sensitivity of a pNP reagent for antigen-specific $CD8^+$ T-cell capture by comparing its capture efficiency against the gold standard multi-color flow cytometry method [15] (**Fig. 5.1C**). For this comparison, we used the realistic

background of healthy donor $CD8^+$ PBMCs. Into this background we spiked a known number of anti-MART-1 (F5) TCR engineered T cells [29]. To engineer those cells, we used recently reported high efficiency CRISPR gene editing methods to knock out the endogenous TCR and knock in the F5 TCR [30]. The efficiency of that preparation, as assessed by flow cytometry analysis of $CD3^+$ cells, was around 34% (**Figure 5.5B**). Between 8 and 128 of these engineered cells were spiked into 10,000 $CD8^+$ donor PBMCs. The analysis preparation included viability staining as well as MART-1 and control pMHC to help differentiate spiked-in and background cells, and exclude non-specific cells (**Fig. 5.5C-5.5F**). We observed that non-specific binding of pNPs to cells is easily distinguished from specific cells (**Fig. 5.5D-5.5F**). Both NP-barcoded NACS and flow cytometry yielded linear performance for detecting the MART-1 specific T cells spiked into healthy donor PBMCs (**Fig. 5.1C**). However, the NP-barcoded NACS method captured 94% of the expected MART-1 specific T cells, while the flow method detected only 52%. The increased capture efficiency of the pNPs likely arises from the increased avidity enabled by the NP scaffold. A few individual cells identified as expressing MART-1-specific TCRs by either the NP-barcoded NACS method (n=3) or by the multicolor flow method (n=5) were analyzed and found, in fact, to express the F5 TCR. This analysis shows that while both methods capture the targeted T cell populations, the NP-barcoded NACS method exhibits an almost 2-fold increased capture efficiency.

5.3.2 Multiplexed analysis of TILs from biopsies of patients with metastatic melanoma using NP-barcoded NACS

We applied the NP-barcoded NACS approach to enumerate multiple antigen-specific T-cell populations from biopsies of four patients with metastatic melanoma responding to anti-PD1 (pembrolizumab) cancer immunotherapy within a phase I trial [31]. To this end, we constructed libraries of putative neoantigens by analyzing the exome and transcriptome of pre-therapy resected tumor materials from three patients expressing HLA-A*02:01 allele (**Table 5.1**). The fourth patient expressed the HLA-A*03:01 allele. The analysis provided input for *in silico* prediction of putative neoantigens, rank-ordered by p-MHC binding affinity, and filtered for gene expression. The putative neoantigen lists for the patients, along with gene expression and wild type peptide information, are found in **Table 5.2**. For the fourth patient, we did a more limited search to demonstrate the generality of the approach, as described below.

We developed a parallel NP-barcoded NACS protocol by modifying the process flow of **Figure 5.1B** in two ways, as illustrated in **Figure 5.2**. First, we created a barcoded pNP library with each library element defined by a 3-position, 3-color barcode, to yield 27 distinct DNA barcodes to be paired to a unique putative neoantigen (**Fig. 5.2A**). Second, we developed a microfluidic chip to trap individual pNP-labeled (barcoded) cells (**Fig. 5.2B & 5.6A**) for decoding. The chip design contained a series of 60 microchambers, each with 10 cell traps, and connected via microfluidic channels. This chip design permitted about 10% capture efficiency of barcoded cells, which is suitable to sample population trends of expanded T-cell specimens. A near term goal is the design of a cell-trap microchip that permits more efficient cell capture. By comparison, the hemocytometer chip yielded near 100% cell capture.

We carried out the parallel NP-barcoded NACS protocol on patient #1 $CD8^+$ TILs by mixing a 27-element barcoded pNP library with the specimen (**Fig. 5.2A**). The barcoded cells were isolated from unbound cells using a magnet, purified of free NPs using a transwell membrane, and immobilized in individual traps on the microchip. To decode the barcodes, a set of dye-labeled ssDNA was hybridized to the first position and read out using fluorescence microscopy. The dye-labeled ssDNAs were removed using displacement ssDNAs, and a second set of dye-labeled ssDNAs was added to read position two and so forth, until all 3 positions were decoded for each cell (**Fig. 5.2A & 5.2C**). Each of the 27 antigen-specificities was associated with a unique barcode sequence (**Fig. 5.2D**). The chemistry associated with barcode readout, along with validation data, is provided in **Figures 5.6B-5.6E**. The sample sorting scheme prior to analysis is provided in **Figure 5.6F**. For the full list of the ssDNA reagents and their corresponding barcode usage, refer to **Table 5.3**.

Figures 5.2B-5.2D are representative data from the analysis of TILs from patient #1. The fluorescent micrographs (**Fig. 5.2C**) are sequential reads of the 9 trapped cells seen in **Figure 5.2B**. The DNA barcode key (**Fig. 5.2D**) pairs the 27 color sequences with the pNP antigen identity. For example, the cell at position iii reads 'YRG', which assigns it as a neoantigen #12-specific $CD8^+$ T cell. Typically, we obtained either a high fidelity read (position viii) or a nonsense read to be discarded (position ix). In some instances, a trap may capture two cells (see position iv), which can be separately barcoded by microscopic inspection. The combination of fluorescent barcoding plus cell imaging thus yields a high fidelity to the approach.

We tested the reproducibility of parallel NP-barcoded NACS approach by separately

analyzing 2 separate vials of TILs from the same tumor biopsy, using freshly prepared reagents each time, and carrying out each analysis more than 1 week apart. From these analyses, we determined that 4% of the $CD8^+$ TILs exhibited specificity to library elements 1-27 (**Fig. 5.7A**). We also recorded a 0.5% non-selective capture rate, based upon testing the library against CD4+ T cells from the same sample. Importantly, both analyses yielded the same seven neoantigen-specific T-cell populations (**Fig. 5.7B**). Here, we define neoantigen-specific T-cell populations as detected if identified by 3 or more unambiguous reads, or identified in more than one analysis of patient #1 TILs.

The above TIL analyses revealed a larger number of neoantigen-specific T-cell populations than have been reported using flow cytometry [1, 2, 21, 32]. This prompted us to analyze expanded TILs from patient #1 using the multiplexed flow cytometry method [15]. For this analysis, we prepared a 14-element tetramer library presenting a subset of 13 putative neoantigens (9 of which were detected using parallel NP-barcoded NACS) and MART-1. With a conservative gating scheme reflective of the multiplexed flow method [15], we only identified neoantigen #12 in patient #1 TILs (**Fig. 5.7C-5.7E**). With a non-conservative gating scheme (**Fig. 5.7F & 5.7G**), we identified neoantigen #12, plus signals (>7 cells) for 3 populations (5, 15, and 27). The 7 cell cutoff, for the non-conservative gating scheme, was selected based upon a background level of 7 cells identified as specific for the conditional (J) pMHC tetramer.

We also used the serial hemocytometer NP-barcoded NACS method (**Fig. 5.1B**) to analyze patient TILs. We first established the selectivity of this method using the full pNP library (**Table 5.2**) associated with patient #1 to analyze TILs from a different patient with metastatic melanoma on the same clinical trial, as well as PBMCs from a healthy donor. The list of neoantigens is unique to patient #1 and should not capture T-cell populations from the other specimens. The results for both controls were similar. The patient #1 pNP library captured on average 2 cells per library element from a total of 10,000 $CD8^+$ TILs, with a standard deviation of 1.4 (**Fig. 5.7H & 5.7I**). Thus, we set a detection threshold of 5 cells (2 standard deviations above the mean). For either control sample, no library element captured more than 5 cells, and there was no correlation between the controls, indicating that none of the patient #1 library elements exhibited intrinsically low selectivity. This finding was consistent with the analysis of spiked specimens. We used serial NP-barcoded NACS to analyze patient #2 TILs and observed a similar number of

neoantigen-specific T cell populations as were found for patient #1 (**Fig. 5.8A & 5.8B**). In contrast, serial analysis of patient #3 TILs yielded fewer populations compared to patients #1 and #2 (**Fig. 5.8C-5.8E**).

5.3.3 Kinetics of neoantigen-specific T-cell populations in a patient with metastatic melanoma responding to anti-PD1 immunotherapy

Since therapeutic responses were mediated by tumor infiltrating $CD8^+$ cells [33], we reasoned that the kinetics of neoantigen-specific T-cell populations might provide insights into cancer patient responses to anti-PD1 checkpoint inhibitor treatments. We thus carried out a comparative analysis of neoantigen-specific TILs and PBMCs from patient #1, and explored how those detected neoantigen-specific T cell populations evolved over time, and how they related to independent measures of tumor volume in that patient. The response of patient #1 to therapy is documented in the computed tomography (CT) scans of **Figure 5.3A**, and in the lesion size timeline of **Fig. 5.3B**.

The above described data sets from the patient #1 TILs analyses are combined in the top histogram of Figure 3C, along with a third parallel NP-barcoded NACS analysis that employed a pNP library built from 23 additional putative neoantigens predicted to be more weakly binding (**Fig. 5.7B**). Each bin on the histogram is associated with a specific putative neoantigen. The numerical identifier of each neoantigen indicates its rank order of calculated antigen-MHC binding strength. A lower numerical index indicating a stronger binder. In the histograms, the neoantigens are arranged, from left to right, according to the measured expression level of the associated transcript (bottom histogram of **Fig. 5.3C**). The transcriptome analysis was from a tumor biopsy collected 28 days prior to the start of therapy (start of **Fig. 5.3B** timeline). For the ten rightmost putative neoantigens (nos. 3, 5, 11, etc.), zero transcripts were detected.

In addition to TILs analyzed at day 187 following the start of anti-PD1 therapy, we performed serial NP-barcoded NACS to analyze non-expanded PBMCs from patient #1 at days 41, 187, 208, and 439 (middle three histograms of **Fig. 5.3C**). There are several things to note about **Figure 5.3C**. First, all but one of the neoantigen-specific T cell populations observed in the TILs were also observed in at least one PBMC analysis. Second, for the neoantigen-specific T-cell populations seen in more than one analysis, the detectable numbers of T cells were strongly correlated with the relevant transcript expression level (R2=0.76) (**Fig. 5.7J**), and only loosely

correlated with the predicted pMHC binding affinity (R2=0.20) (**Fig. 5.7K**). Of the 10 putative neoantigens predicted from rare transcripts below the RNA-seq detection limit, only T cells specific to the strongly binding neoantigen #5 were detected. Third, the detected numbers of neoantigen-specific $CD8^+$ TILs strongly correlated with measured lesion size (**Fig. 5.3B**). By day 439, no T-cell populations were detected at all (**Fig. 5.7L**). By contrast, for PBMCs collected at day 41, when the tumors appeared to be growing even after the start of therapy (an effect known as pseudo progression [34], neoantigen-specific T cells were present in their greatest relative abundance.

We also performed serial NP-barcoded NACS analysis of PBMCs from patient #3 (**Fig. 5.8C-5.8E**). For this patient, two neoantigen-specific T-cell populations (nos. 13 and 20) were detected in TILs. The most dominant population (13) was also detected in PBMCs at day 25 following start of therapy, along with four additional neoantigen-specific populations. Only T-cell populations specific to neoantigens 13 and 14 were detected at the time point (Day 87) closest to maximum therapeutic response. Although patient #3 was participating in the sample trial as patient #1 and #2, patient #3 showed a very different response profile to anti-PD1 therapy, and had a lower abundance of neoantigen-specific T cells.

5.3.4 Functional and genetic validation of identified neoantigen-specific T cells

To validate the neoantigen-specific populations identified by NP-barcoded NACS, we analyzed functional activity by an ELISpot cell secretion assay, and, for one population, carried out single cell TCR gene sequencing. Exposure of antigen-specific $CD8^+$ T cells to their cognate antigen-MHC leads to functional activation, which is commonly detected by interferon gamma (IFN-γ) release. To confirm functional activation of the neoantigen-specific $CD8^+$ populations from patient #1 PBMCs collected at day 41, we prepared a 14-element tetramer library, stimulated the PBMCs with the individual tetramers, and measured IFN-γ secretion with an ELISpot single cell assay (**Fig. 5.3D**). The baseline (dashed line) was established as the average background level using an identical number of $CD8^+$ T cells from a healthy donor. Of the 12 neoantigen-specific elements in the 14-element library, 11 of them exhibited ELISpot counts above background, while no ELISpots were detected in the negative controls (neoantigen nos. 1 and 10).

We also used NP-barcoded NACS to isolate a single neoantigen-specific cell from patient #1 PBMCs, and determined both its antigen specificity and TCRα/β genes

(**Fig. 5.4**). For this experiment, we designed a cell capture microchip with a low-density of cell traps so that the cell could be barcoded to identify the antigen specificity (**Fig. 5.4A**), and then punched out of the device for TCR sequencing. For this case, the identity of the cell was neoantigen #12. The sequenced TCR genes were cloned into a retroviral vector. Jurkat cells (a $CD3^+$ T cell immortalized line) that were transduced to express the TCR were found to bind to the neoantigen #12 tetramer (**Fig. 5.4B**), demonstrating that matched TCR-antigen pairs can be identified from single cells isolated by NP-barcoded NACS. Importantly, these two neoantigen-specific T-cell characterization assays confirm that the populations detected by NP-barcoded NACS are, in fact, present in the patient-derived specimens and are functionally active.

5.3.5 Extension to HLA-A*03:01

We carried out a limited library analysis (patient #4, 5 neoantigens + MART-1, **Table 5.2**) of non-expanded PBMCs from the patient expressing the HLA-A*03:01 allele, and, using the hemocytometry approach, identified a single neoantigen-specific T cell population (**Fig. 5.8F & 5.8G**). The neoantigen identified (SLHAHGLSYK, gene F5) had a high predicted p-MHC binding strength (8.22 nM), although cell populations for putative neoantigens with similar binding strength (7.44, 16.23 nM) were not found. We subsequently expanded the neoantigen-specific cells and used tetramer staining to sort single cells of that same neoantigen-specific T cell population into individual microwells for TCR sequencing [24]. Of 24 wells where single cells were sorted, TCR β chains were sequenced from 10 wells (**Fig. 5.8H**). One β chain was identified from 4 different wells and another β chain was identified in 3 different wells, demonstrating a polyclonal response to this neoantigen, but also demonstrating generality of this method beyond HLA-A*02:01.

5.4 Discussion

Individual populations of neoantigen-specific T cells are typically scarce in patient tumor tissues or blood, and this can make them extremely challenging to study. Nevertheless, knowledge about those populations can provide guidance for various vaccination [2] or cell-based cancer immunotherapies [35]. The NP-barcoded NACS approaches described here are simple to implement, and yet outperform state-of-the-art multi-color flow cytometry for the detection of these low-abundance cells. The high capture efficiency of the NP-barcoded NACS approach likely arises from the

increased avidity of the NP-presentation format for the pMHC tetramers, which is especially useful for small sample sizes. The parallel and serial NP-barcoded NACS approaches used similar reagents for cell capture, but with distinct advantages. The serial approach has limited multiplexing, but features near unity cell capture efficiency (**Fig. 5.1C**, while the parallel approach has excellent multiplexing (**Fig. 5.2**), but lower capture efficiency. The captured cells remain viable for further functional or genetic investigations (**Fig. 5.4**).

In the analysis of TILs and/or PBMCs collected from patients responding to anti-PD1 therapy, we made three major observations. First, we observed that 6-32% of the top predicted neoantigens for HLA-A*02:01 can be paired with T-cell populations (**Fig. 5.3, 5.7B, 5.8B, & 5.8D**). Including additional HLA alleles for each patient (**Table 5.1**), as well as including weaker binding neoantigens would almost certainly yield additional T-cell populations. The implication is that a few tens of % of the $CD8^+$ TILs within these patient tumors may be neoantigen-specific. This number is high relative to previous reports from patient sample analyses [1, 21], but is consistent with a prior result in mice [2]. The neoantigen-specific T-cell population numbers detected here are consistent with immune responses to tumor antigens that have been observed in healthy donors [36]. In that work, PBMCs from the healthy donor were expanded in the presence of a library of neoantigens, so as to amplify those rare populations that exhibit specificity against library elements. With the additional sensitivity gained by the NP-barcoded NACS method, cell expansion and the associated distortion of T-cell populations may no longer be necessary.

A second observation is that the measured spectrum of neoantigen-specific T-cell populations only correlates loosely with the calculated antigen-MHC binding affinity, but there is a strong correlation with mRNA expression levels (**Fig. 5.7J & 5.7K**), which is consistent with recent work on antigen profiling [37]. These data should provide guidance for the refinement of neoantigen prediction algorithms [38].

A third observation is that the same neoantigen-specific populations detected in the tumor are also found in the blood, albeit at a lower abundance relative to all $CD8^+$ PBMCs. This is consistent with a recent observation in patients with melanoma [39]. In patient #3, a neoantigen-specific T cell population was found in blood 25 days after the start of checkpoint inhibitor immunotherapy, which was also observed in TILs a month prior to start of therapy (**Fig. 5.8D**). In the analysis of patient #1, neoantigen-specific populations were detected in blood a full 2 months prior to observation of actual tumor shrinkage via CT scan (**Fig. 5.3B & 5.3C**). In this

patient, we observed a strong correlation between the kinetics of tumor shrinkage, and the abundance kinetics of neoantigen-specific T cells in PBMCs. The NP-barcoded NACS method for enumeration is currently challenging to integrate with flow-cytometry for single cell isolation and sequencing of TCRs due to the NP light scattering properties. It is likely that smaller diameter NP scaffolds may be useful for this purpose [18].

5.5 Acknowledgments

We acknowledge the following agencies and foundations for support: The Stand Up to Cancer Foundation and the Cancer Research Institute (A.R. and J.R.H.), the National Cancer Institute (1U54 CA199090 and R01-CA170689 to J.R.H. and R35 CA197633 to A.R.), the Jean Perkins Foundation (J.R.H. PI), and Caltech internal support through a CI2 grant. O.N.W. is supported by the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research. A.H.C.N. is supported by a Banting Postdoctoral Fellowship from the Government of Canada. We thank Eugene Barsov and Richard Morgan (NCI) for providing the retroviral MSGV vector.

5.6 Author Contributions

S.P., J.M.Z., A.H.C.N., W.C., M.T.B., O.N.W., D.B., A.R. and J.R.H. designed the experiments. A.R. ran the trials that provided patient samples in this study. S.P., J.M.Z., A.H.C.N., W.C., M.T.B., A.H., E.H., X.D., K.G., J.K., A.M.X., J.E.H., W.J.N., J.Z., Y.S., Y.L., J.M., and D.C. did the experiments. S.P., J.M.Z., A.H.C.N, M.T.B., A.R., and J.R.H. wrote the paper.

5.7 Declaration of Interests

D.B., A.R., and J.R.H. are scientific co-founders of PACT Pharma, a company that is seeking to commercialize certain aspects of the NP-barcoded NACS technology. S.P., A.H.C.N., W.C., J.C., and J.R.H. have at least one patent related to this work.

5.8 Figure Captions

Figure 5.1: Peptide-MHC (pMHC) tetramer functionalized magnetic nanoparticles (NPs) for antigen-specific T-cell enumeration. (A) Biotinylated pMHC, DNA-labeled cysteine-modified streptavidin (SAC-DNA), and an NP decorated with DNA (NP-DNA) are coupled, via biotin-streptavidin interaction and DNA hybridization. This produces a pMHC tetramer NP (pNP), with each particle presenting at least 20,000 pMHC tetramers. In addition to a SAC hybridization site, the DNA on the NP also has a fluorescent oligo docking site (barcode) for hybridizing dye-labeled ssDNA. (B) In a typical antigen-specific T-cell enumeration, a specific pNP reagent (e.g. peptide 1) is mixed with around 10,000 viability-stained $CD8^+$ T cells. For sensitivity comparison in panel (C), two irrelevant pNPs labeled with different oligo dyes are also added to characterize non-specific binding. After incubating the pNP reagents with the $CD8^+$ T cells, pNP-bound cells and free particles are isolated with a magnet, and spread across a hemocytometry chip for enumeration via microscopy. The optical/fluorescent micrograph series shows a typical antigen-specific T cell (top right). In the bright field image, the T cell appears black due to the pNPs that decorate its surface. In fluorescent images, the viability stain (green) confirms the presence of a live cell, and its co-localization with only peptide 1 pNP (red) confirms its specificity. Scale bars are 10 μm. Serial NP-barcoded NACS: To analyze more than one antigen-specific population using this method, the supernatant containing the free cells is sequentially analyzed with different pNPs (e.g, peptide 2, 3, n). (C) Plots comparing sensitivity of pNP capture (left, n=4) and flow cytometry (right, n=3) for the analysis of MART-1 TCR-engineered T cells spiked into 10,000 $CD8^+$ donor PBMCs. Each plot includes a straight-line fit, the slope, percent recovery, and the R2 fitting metric. See also **Figure 5.5**.

Figure 5.2: Parallel NP-barcoded NACS and sample TILs analysis. (A) A barcoded pNP library is formed, with each library element containing a unique peptide and 3-position oligo docking site (barcode). Each position can be one of three unique sequences for hybridizing green, yellow, or red-labeled ssDNA, creating 27 (3×3×3) unique DNA barcodes. All elements of the barcoded pNP library are combined and mixed with $CD8^+$ cells, and the particle-bound cells (barcoded cells) are isolated and physically immobilized in an array of microfluidic cell traps. To decode the barcode, sets of dye-labeled ssDNA are sequentially hybridized, imaged, and displaced for three rounds (one for each position). Thus, each of the 27 unique peptides is associated with a unique sequence of three colors from the fluorescent

readout. (B) Optical micrograph shows a microfluidic chamber equipped with 10 cell traps, 9 of which contain single barcoded T cells. This microchip has 60 such cell capture microchambers. Scale bar is 50 μm. (C) The fluorescent micrograph series are the sequential reads of the three barcode positions, with the fluorescent readouts for the 9 trapped cells provided to the right (R = red, G = green; Y = yellow). The cell at position viii does not provide a clean read, while the cell at position ix was lost during the process. Some traps contain 2 cells (for example, iv), and so reads are only done on those cells that are clearly delineated in the images. Scale bar is 5 μm. (D) The DNA barcode key shows each of the 27 color sequences and their corresponding neoantigen identity or, at position #8, the MART-1 tumor antigen. For example, the cell at position #3 reads 'YRG' and corresponds to neoantigen 12. The Y(D) notation (red font) implies a Y for D mutation in the neoantigen sequence. See also **Figure 5.6**.

Figure 5.3: Analysis of neoantigen-specific CD8$^+$ T-cell populations from patient #1 TILs and PBMCs over the course of treatment with anti-PD1 therapy. (A) Representative CT scan of patient #1, who had metastatic melanoma to the chest wall, pleura and lung progressing after three prior lines of therapy with high dose interleukin-2, vemurafenib, and TIL adoptive cell therapy. Upon administration of pembrolizumab, the patient had a transient progression for three months, with increase in size of the chest wall masses and pleural effusion (Dec 2012), followed by a long-lasting tumor regression. (B) Timeline of the lesion size and neoantigen-specific T-cell counts in CD8$^+$ PBMCs. Day 0 is the start of treatment. A baseline tumor biopsy was collected for genomic and transcriptomic analysis at day -28 (star symbol). Black dots represent CT-scan measurement of lesion size (left y-axis), while black squares represent total neoantigen-specific T cells counted from 104 CD8$^+$ PBMCs (right y-axis). PBMCs analyzed at day 439 yielded no detectable neoantigen-specific populations. Earlier points of analyses are represented by arrows, color-coded for the bar graphs in panel (C). (C) Neoantigen-specific T-cell populations detected from Day 187 TILs (top graph) and PBMCs (middle graphs) over the course of the therapy, along with mutation-containing mRNA read counts for the mutant proteins (bottom graph) from the baseline RNA-seq. All plots are arranged by descending mRNA read counts, and putative neoantigens with undetectable transcripts (on the right) are controls. The horizontal dashed lines in the TIL and PBMC plots represent the signal threshold above which the identification of a T-cell population is statistically significant. (D) ELISpot assays of IFN-γ secretion from each of the neoantigen-specific T-cell populations detected from Day

41 PBMCs. The baseline (dashed line) was established as the average background level using an identical number of $CD8^+$ T cells from a healthy donor. The left y-axis represents the neoantigen-specific cells detected in panel (C) and the right y-axis represents, in an analysis of a different vial of Day 41 $CD8^+$ PBMCs, the numbers of ELISpots detected. Neoantigens nos. 1 and 10 (red font x-axis label) are controls. Micrographs of representative ELISpot assays are shown for a healthy donor and for a neoantigen-specific T-cell population. Scale bar is 500 μm. See also **Figures 5.6, 5.7, & 5.8**.

Figure 5.4: Determination of T-cell receptor (TCR) genes for a corresponding neoantigen-specific T cell using NP-barcoded NACS. (A) The optical micrograph show a captured barcoded T cell from patient #1, followed by the 3 sequential fluorescent readout steps to identify specificity against neoantigen 12. Scale bars are 20 μm. The captured single T cell was punched out for RT-PCR to obtain TCRα and TCRβ gene sequences (DNA ladder: 100bp). The TCR gene was then assembled and inserted into a retroviral vector, and transduced into Jurkat T cells for analysis by flow cytometry. (B) Flow cytometry results for untransduced Jurkats (left), Jurkats transduced with LNGFR expression reporter only (center), and Jurkats transduced with both LNGFR and the TCR specific for neoantigen 12 (right). See also **Figures 5.7 & 5.8**.

Figure 5.5: Characterization of NP and engineered T cells and sample analysis of NP-barcoded NACS and flow cytometry. Related to Figure 5.1. (A) Plot of bicinchoninic acid assay (BCA) calibration curve fitted using a cubic polynomial (R2=1.0), resulting in a limit of detection of 1.64 μg/mL. The standards were prepared and analyzed in triplicates. The number of pMHC tetramers per NP was measured to be 20270, assuming that the molecular weight of pMHC tetramer is 236.8 kDa. (B) Flow cytometry analysis of CRISPR-Cas9-based TCR gene-editing T cells. CD3 was used as an indirect measure of TCR gene editing efficiency since TCR-CD3 complexes are presented on the cell surface only when both CD3 and TCR subunits are co-expressed. Density plots of donor cells are shown for no electroporation conditions (control), knock out-knock in at endogenous TCR-alpha locus using F5 TCR homology-directed repair template, knock out at endogenous TCR-beta locus, and simultaneous alpha knock out and in and beta knock out (TCR engineered cells). The gene editing efficiency for the TCR engineered cells is about 34%. Only 2% of these cells expressed endogenous TCR (IP26). (C) Gating scheme for flow cytometry analysis of a spiked specimen. MART-1 specific T

cells are defined as Calcein green positive, NY-ESO tetramer negative, and double MART-1 tetramer positive. (D-F) NP-barcoded NACS analysis of a MART-1 spiked specimen. Cells imaged in the hemocytometer can either be (1) specific, (2) non-specific, or (3) unbound. (D) Representative fluorescent plots with each type of cell highlighted. (E) Representative x-y coordinate map of the hemocytometer chip. (F) Representative bright field and fluorescent images. Scale bars are 10 μm. Here, MART-1 pNP (stained red) is accompanied by two irrelevant pNP (stained yellow and purple). Thus, specific cells must be coated with black pNP in bright field microscopy, stain positive for red and Calcein green, and stain negative for yellow and purple.

Figure 5.6: Parallel NP-barcoded NACS device design, readout validation, and sample selection. Related to Figures 5.2 & 5.3. (A) Design of microchip for the capture and isolation of antigen-specific T cells. Cells are injected into the input, and flow through one of 5 channels, before exiting to waste. Each channel contains 12 microchambers, each of which has 10 cell traps. The micrograph at right shows one of those microchambers, with 9 of the 10 cell traps filled. The trapped cells appear black because of the pNPs that coat the cell surface. Scale bar is 50 μm. (B) Illustration and (C) test data of the DNA sequential barcode readout process, reflecting 3 fluorescent read-out steps, and 3 displacement steps. The fluorescence images illustrate this process for DNA barcodes appended to nanoparticles. Scale bar is 50 μm. (D) Barcode cross reactivity evaluation. To check for cross reactivity of displacement and read DNA during the decoding process, an equal mixture of three DNA-barcoded 10-μm beads is analyzed after isolating individual beads in the cell traps of a microfluidic chip. Scale bar is 40 μm. (E) Histogram of the three labeled barcode positions as read out from beads isolated within a single column of cell traps in the microchip of panel D. The counts for each of the three barcoded identities 8, 12, and 22 (inset) indicated that each population accounted for approximately 1/3 of the trapped beads, as expected. Negligible reads (<2) were detected from absent barcodes. (F) Sample density plots of $CD8^+$ T cells from patient #1 TILs sorted for parallel NP-barcoded NACS analysis.

Figure 5.7: Patient #1 analysis. Related to Figures 5.3 & 5.4. (A) Number of neoantigen-specific $CD8^+$ T cells captured for patient #1 expanded TILs from two independent analyses using pNP library 1-27. Non-selective capture is estimated to be around 0.5%, as gauged by the numbers of CD4+ T cells that were captured using the same method. The two independent captures were carried out more than

one week apart. (B) Neoantigen populations detected in expanded TILs from the tumor of patient #1. For each run, approximately 10,000 $CD8^+$ TILs were analyzed by NP-barcoded NACS, resulting in around 500 cells pulled down by pNPs. After free particles are removed using a 5 μm pore transwell membrane, the pNP-bound cells are loaded into the microfluidic chip (Figure S2A), and cells are isolated within the cell capture chambers of the microfluidic chip for fluorescence-based readouts of the attached pNPs. Capture efficiency within the microfluidic chip was about 10% (i.e. of the 500 cells that are loaded, 50 cells can be captured for analysis). Runs #1 and #2 used the same 27-element NP-barcoded NACS library, corresponding to the top-ranked 27 putative neoantigens. All populations shown for runs #1 and #2 were detected in both runs, except neoantigen #5, which was detected in 4 clean reads only in run #2. Run #3 utilized a NP-barcoded NACS library designed to capture $CD8^+$ T cells specific to neoantigens rank ordered 28-50. (C-G) Multiplexed flow cytometry analysis of patient #1 expanded TILs and detection statistics for two selections of $CD8^+$ T cells. (C) Color encoding scheme of neoantigen-MHC tetramers. For example, neoantigen #7 tetramer is stained with QD605 and PE. The CD8++ subpopulation is analyzed in panels D and F, while the $CD8^+$ T cells are analyzed in panels E and G. Each panel incorporated either a tight or loose gating condition for counting those cells that stain with exactly 2 colors. Each panel provides the numbers of cells analyzed (single stain positive for CD8), as well as the numbers of cells that stained positive for single stains and two stains. Note that neoantigen #12 specific $CD8^+$ T cells are almost completely localized within the CD8++ subpopulation. (H and I) Results of serial NP-barcoded NACS control experiments in which the pNP library designed for patient #1 was used to capture $CD8^+$ T cells from a different patient with melanoma (blue) and from a healthy donor (orange). (H) Average numbers of pulled-down T cells, per library element, from healthy donor PBMC and PBMC from an unrelated patient with melanoma, on the same clinical trial as patient #1. Data is presented as mean ± standard deviation. (I) Number of pulled-down T cells for each library element. The numerical labels on the x-axis corresponds to the rank-order of the putative neoantigens. The data has been sorted according to frequency of detection for the patient, so as to illustrate that there is no correlation between these two controls. A correlation would likely indicate capture by non-selective library elements. The noise threshold was set at the average plus twostandard deviations. (J and K) Correlation analysis between the number of neoantigen-specific T-cell populations and measured neoantigen-related mRNA levels (J), or calculated binding constant for neoantigen-MHC (K). The total

T cell numbers are a summation of each neoantigen-specific population counts from PBMCs (day 41, 187, and 207) and TILs at day 187 (**Figure 5.3C**). Only neoantigen populations that appeared in more than one analysis are included in this correlation (i.e. neoantigen #1, 9, 30, and 35 were excluded). (L) Analysis of patient #1 PBMCs from a blood draw collected 439 days following the start of anti-PD-1 therapy. No populations above two standard deviations of the mean (2.2±1.4 cells per library element) were recorded.

Figure 5.8: Patient #2, #3, and #4 analysis. Related to Figures 5.3 & 5.4. (A and B) TIL analysis of patient #2. (A) The timeline of the lesion size of subcarinal lymph node, supraclavicular lymph node, and neck in patient #2. Day 0 corresponds to the start of anti-PD-1 therapy. A baseline tumor biopsy (indicated by the purple arrow) was collected for genomic and transcriptomic analysis at day -11. Black markers represent CT-scan measurement dates, while the green arrow corresponds to the time point of analysis. (B) Serial NP-barcoded NACS analysis of patient #2 TILs using a pNP library formed from the top 28 putative neoantigens predicted for patient #2, plus the MART-1 tumor antigen. Only T-cell populations detected at more than 5 cells per 3000 TILs were considered statistically significant. The bottom graph shows the mRNA copies measured for the mutated proteins from which the neoantigens are derived. (C-E) TIL and PBMC analysis of Patient #3 over the course of response to anti-PD-1 therapy. (C) The timeline of the lesion size of chest, lung, liver and intraabdominal in Patient #3. Day 0 corresponds to the start of anti-PD-1 therapy. A baseline tumor biopsy (indicated by the purple arrow) was collected for genomic and transcriptomic analysis at day -29. Black markers represent CT-scan measurement dates, while the arrows correspond to the time points of analysis, and are color coded for the bar graphs in panel D. (D) Results of serial NP-barcoded NACS analysis using patient #3 specific 34-element pNP library. The plots are neoantigen specific T-cell populations detected from expanded TILs collected from baseline (top graph), PBMCs over the course of the therapy (middle two graphs), and mutation-containing mRNA read counts for the mutant proteins (bottom graph) from the baseline RNA-seq. The horizontal dashed lines in the TIL and PBMC analysis graphs represent the signal threshold above which the identification of a T cell population is statistically significant, which is determined in panel E. The vertical gray dashed lines indicate T cell populations detected across the different time points and patient materials, and their correlation with RNA transcripts reads. (E) Results of control experiments in which a patient #3 specific 34-element pNP library was used to capture $CD8^+$ PBMCs from a healthy donor. The average pulled-

down T cells numbers from healthy donor PBMCs was 0.9 ± 1.1 (mean ± standard deviation). The noise threshold was set at average plus two standard deviations. (F-H) Detection and single cell TCR sequencing results of a neoantigen-specific T cell population from a patient expressing the HLA-A*03:01 allele (patient #4). (F) Of 5 potential neoantigens, the second highest predicted binder was detected above the signal threshold. (G) Micrographs showing a representative neoantigen-specific cell in bright field and fluorescence channels. Scale bars are 10 μm. (H) Cells were tetramer-sorted using the neoantigen SLHAHGLSYK into 24 wells. TCR β chains were identified in 10 wells, and 7 wells shared one of two repeated TCR β sequences.

5.9 Figures & Tables

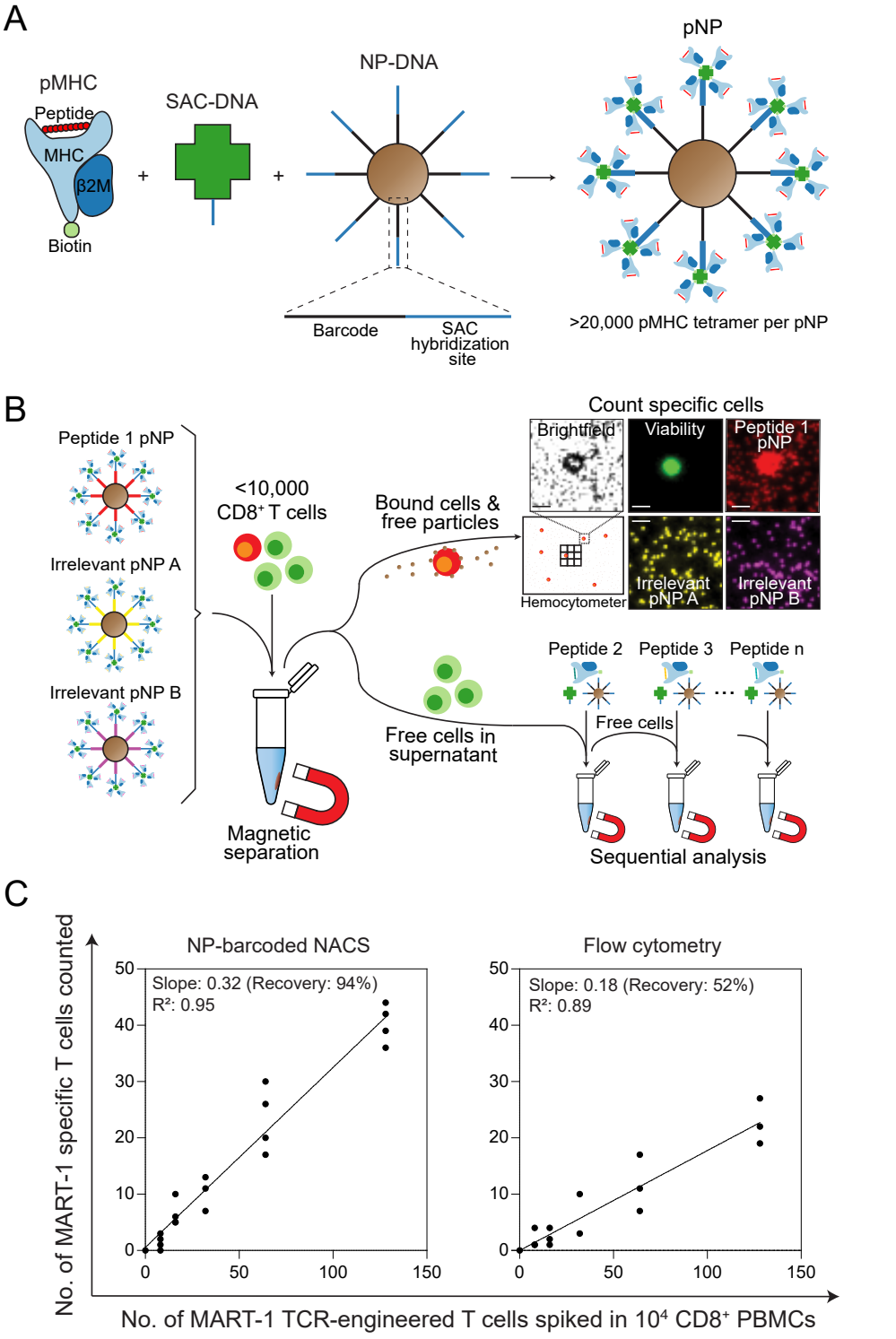


Figure 5.1: Peptide-MHC (pMHC) tetramer functionalized magnetic nanoparticles (NPs) for antigen-specific T-cell enumeration.

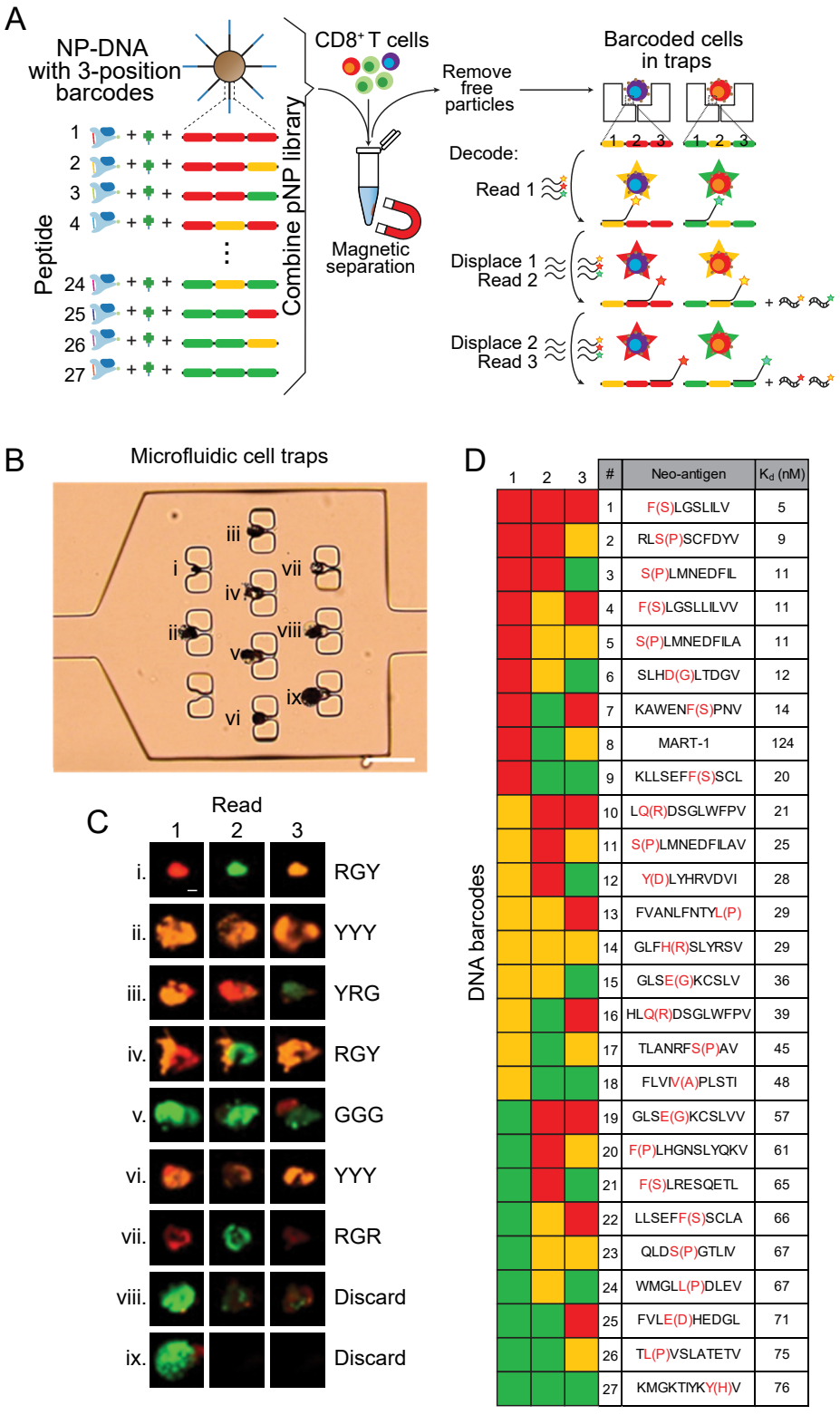


Figure 5.2: Parallel NP-barcoded NACS and sample TILs analysis.

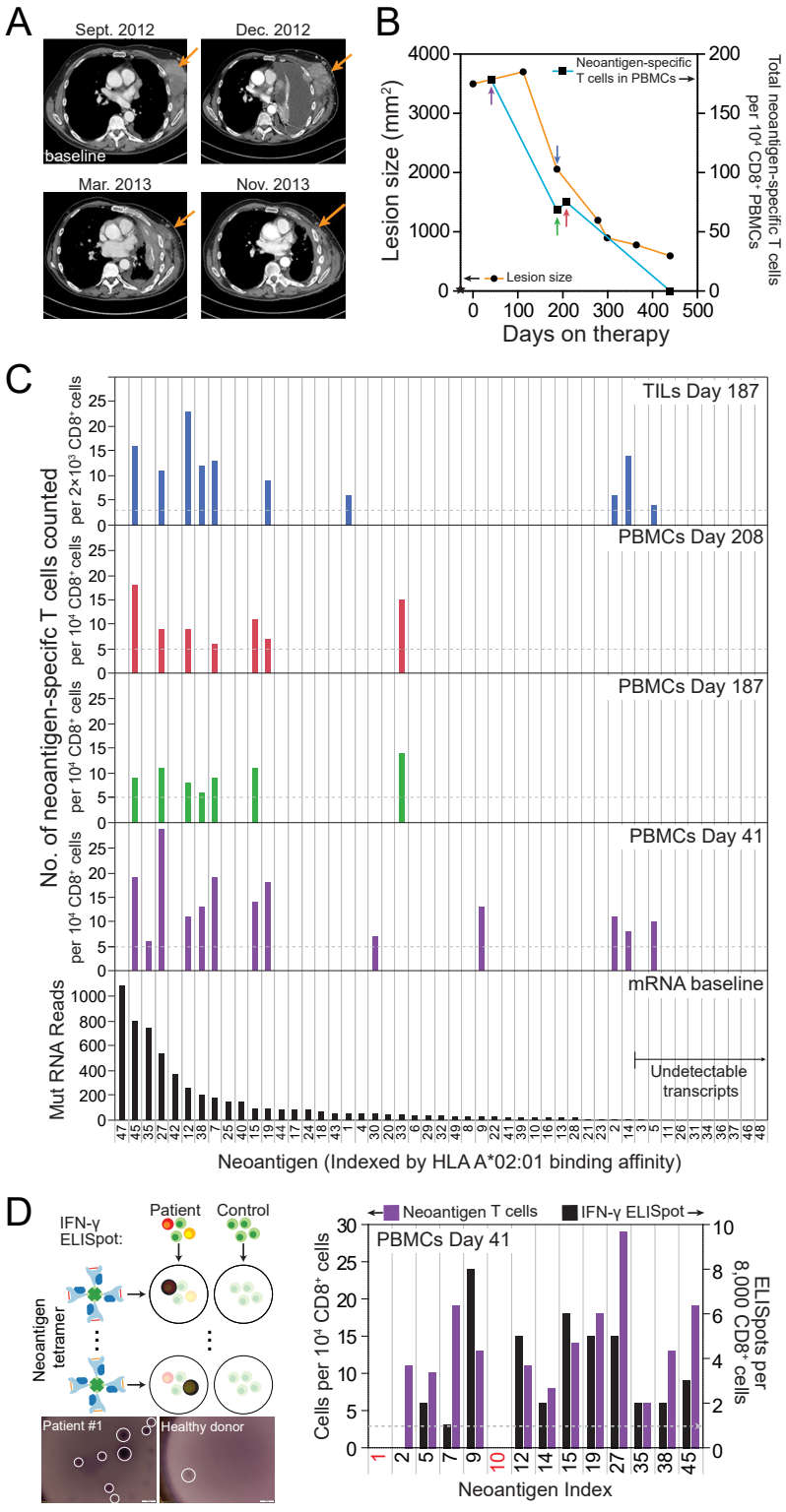


Figure 5.3: Analysis of neoantigen-specific CD8$^+$ T-cell populations from patient #1 TILs and PBMCs over the course of treatment with anti-PD1 therapy.

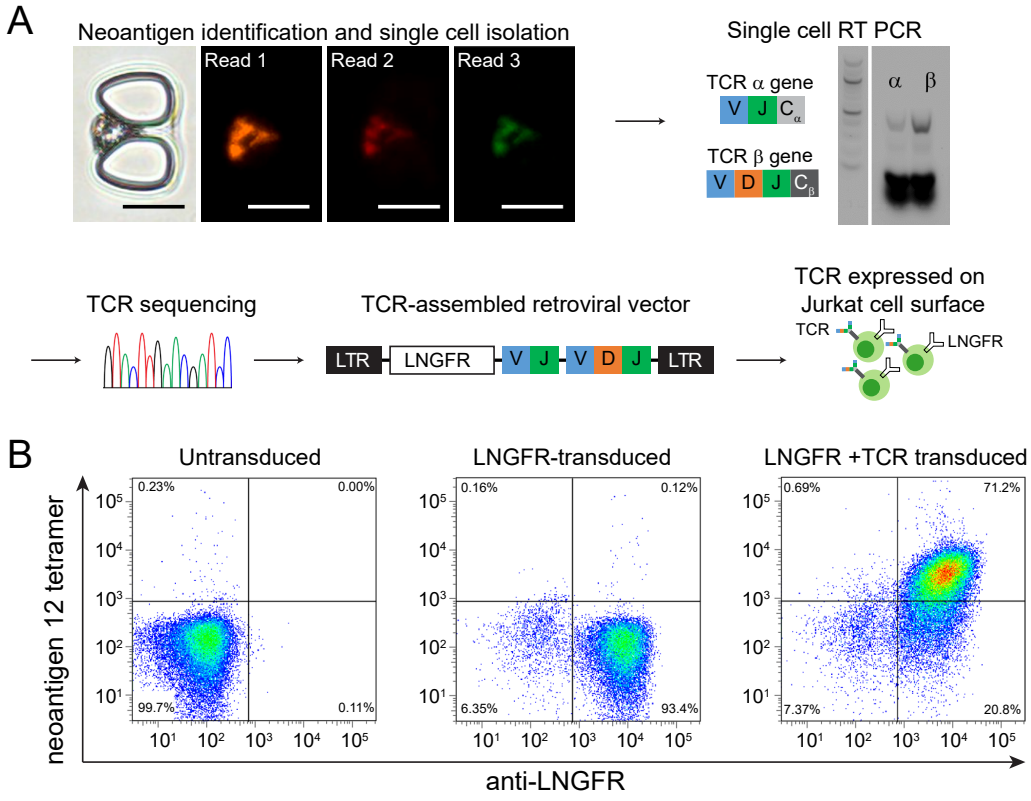


Figure 5.4: Determination of T-cell receptor (TCR) genes for a corresponding neoantigen-specific T cell using NP-barcoded NACS.

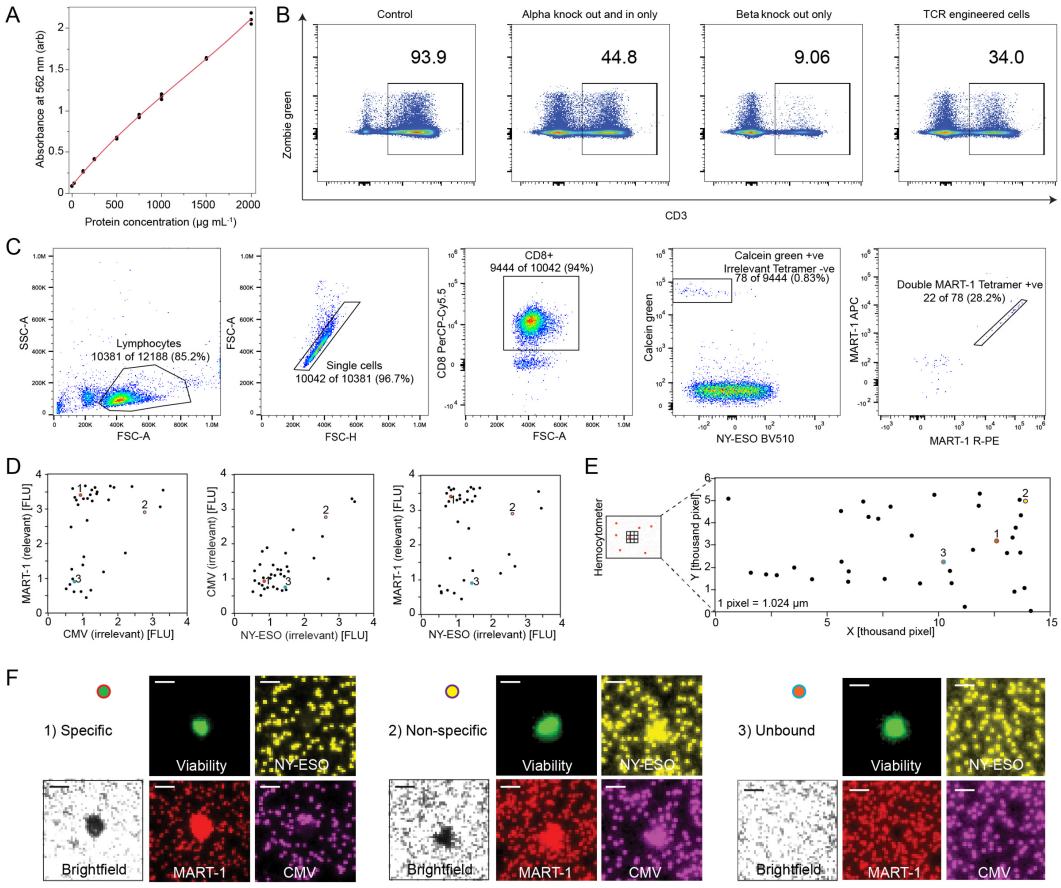


Figure 5.5: Characterization of NP and engineered T cells and sample analysis of NP-barcoded NACS and flow cytometry. Related to Figure 5.1.

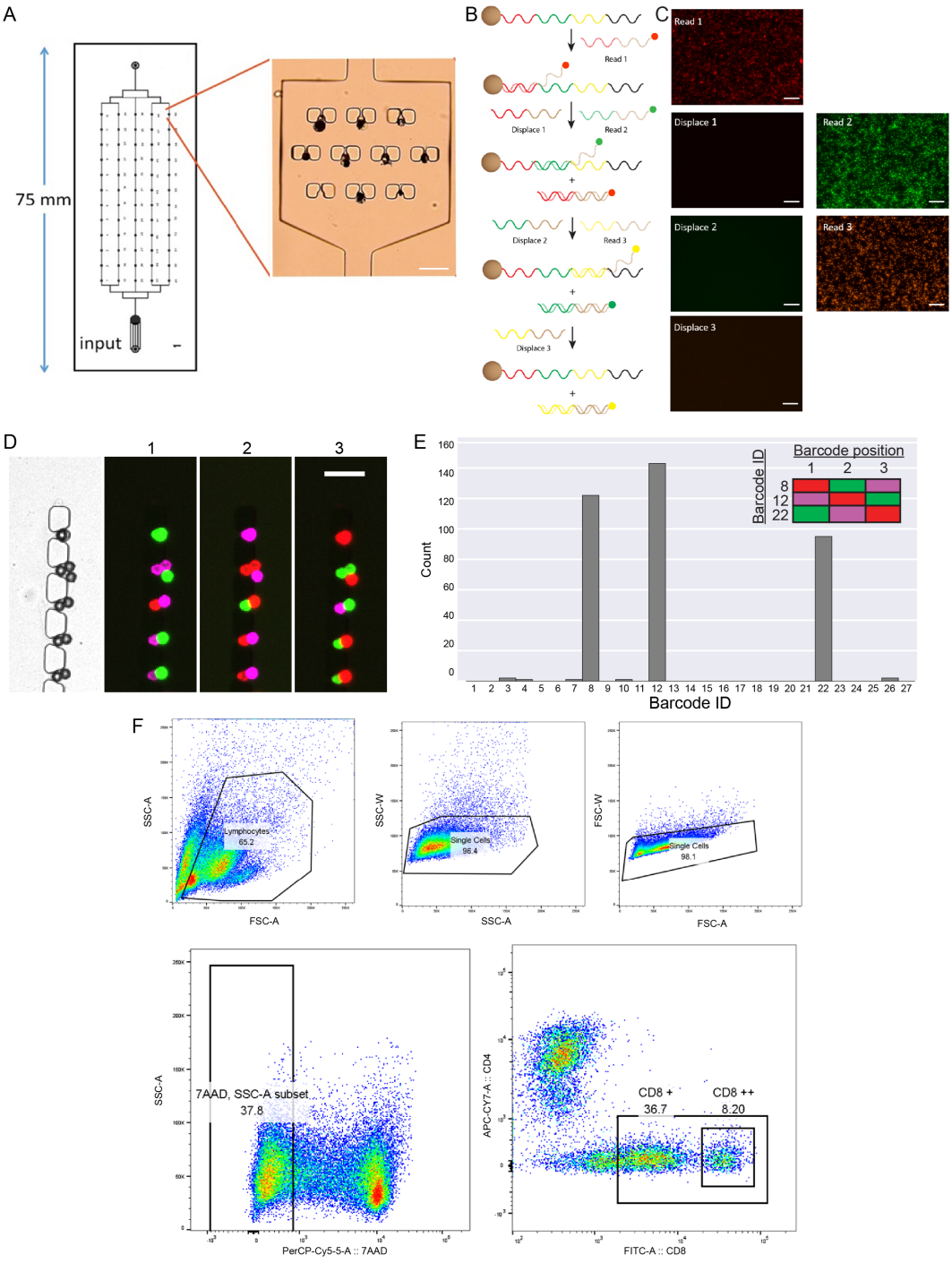


Figure 5.6: Parallel NP-barcoded NACS device design, readout validation, and sample selection. Related to Figures 5.2 & 5.3.

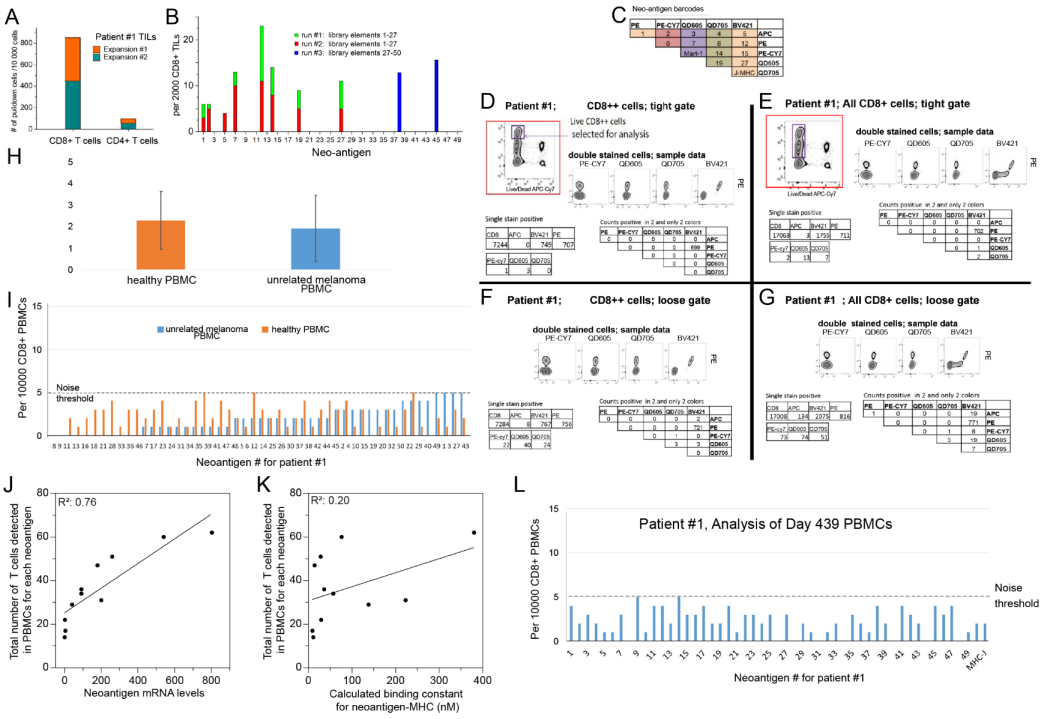


Figure 5.7: Patient #1 analysis. Related to Figures 5.3 & 5.4.

Table S1. Patient Information. Related to Figures 2, 3, and 4.

	Patient #1	Patient #2	Patient #3
irRECIST	Partial Response	Partial Response	Partial Response
Study	Merck MK-3475-001	Merck MK-3475-001	Merck MK-3475-001
Age at Tx start	61	70	70
Sex	M	M	M
ECOG Status at Baseline	0	0	0
Disease Status at Baseline	M1c	M1b	M1c
BRAF/NRAS	BRAF V600E	NRAS G13D	NRAS Q61K
Melanoma Sub-Type	Cutaneous	Cutaneous	Cutaneous
HLA Type	HLA-A*02:01, HLA-A*68:01 HLA-B*15:07, HLA-B*44:02 HLA-C*03:03, HLA-C*07:04	HLA-A*02:01, HLA-A*24:02 HLA-B*51:01, HLA-B*58:01 HLA-C*01:02, HLA-C*03:02	HLA-A*02:01, HLA-A*03:01 HLA-B*07:02, HLA-B*51:01 HLA-C*02:02, HLA-C*07:02
Prior Systemic Therapies	1. Vemurafenib 2. Ipilimumab 3. Chemotherapy+IL-2 4. TIL adoptive cell transfer	1. Adjuvant GM-CSF	1. Ipilimumab
Date of first evidence of tumor regression	Day 166	Day 134	Day 87
Date of first irRECIST Response (-50%)	Day 250	Day 280	Day 186
Site of Baseline Biopsy	Left Chestwall	Subclavicular Lymph Node	Right Chestwall
Date of Baseline Biopsy for Exome/RNA-seq	Day -28	Day -11	Day -29
# Somatic Nonsynonymous Mutations	350	745	660
Date of Biopsy for TIL collection	Day 187	Day 23	Day -29
Date of PBMCs Sampled	Day 41, Day 187, Day 208, Day 439	N/A	Day 25, Day 87

Table 5.1: Patient Information. Related to Figures 5.2, 5.3, & 5.4.

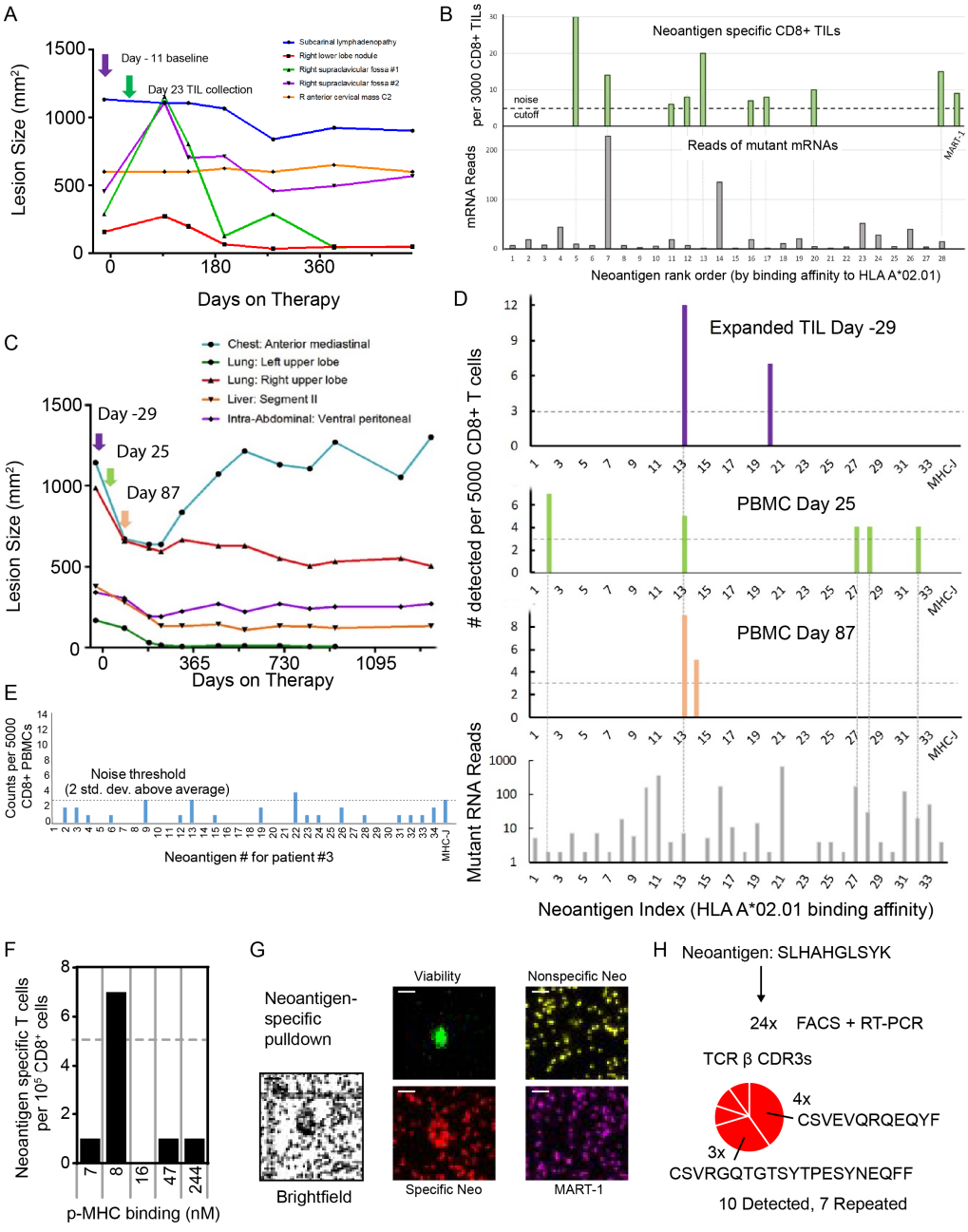


Figure 5.8: Patient #2, #3, and #4 analysis. Related to Figures 5.3 & 5.4.

Patient #	HLA	neoantigen	Gene	Protein_Change	FPKM	Mut_RNA_Reads_	pepLen	mutantpep Position	peptideTumor	K_d (nM)	peptideNormal	K_d (nM)	DeltaAbs (nM)
1	A2:01	1	DYM	p.S434F	20.5835	53	9	1	FLGSLLILV	5	SLGSLLILV	13	-8
1	A2:01	2	SLC8A2	p.P719S	0.025223	4	9	3	RLSSCFDYV	9	RLPSCFDYV	15	-6
1	A2:01	3	IFNA7	p.P133S	0	0	9	1	SLMNEDFIL	11	PLMNEDFIL	820	-809
1	A2:01	4	DYM	p.S434F	20.5835	53	10	1	FLGSLLILVV	11	SLGSLLILVV	33	-22
1	A2:01	5	IFNA7	p.P133S	0	0	10	1	SLMNEDFILA	11	PLMNEDFILA	550	-539
1	A2:01	6	RFTN1	p.G348D	5.27924	36	9	4	SLHDLTDGV	12	SLHGLTDGV	36	-24
1	A2:01	7	PRRC2C	p.S2260F	27.70395	179	9	6	KAWENFPNV	14	KAWENSPNV	22	-8
1	A2:01	8	FASTKD1	p.S159F	10.8846	26	9	6	LLSEFFSCL	18	LLSEFSSCL	26	-8
1	A2:01	9	FASTKD1	p.S159F	10.8846	26	10	7	KLLSEFFSCL	20	KLLSEFSSCL	32	-12
1	A2:01	10	LDHD	p.R148Q	4.41145	20	10	2	LQDSGLWFPV	21	LRDSGLWFPV	2629	-2608
1	A2:01	11	IFNA7	p.P133S	0	0	11	1	SLMNEDFILAV	25	PLMNEDFILA	94	-69
1	A2:01	12	USP7	p.D789Y	41.2267	259	9	1	YLYHRVDVI	28	DLYHRVDVI	10865	-10837
1	A2:01	13	PLS1	p.P376L	4.53658	19	10	10	FVANLFNTYL	29	FVANLFNTYP	3399	-3370
1	A2:01	14	CPVL	p.R25H	5.87024	2	10	4	GLFHSLYRSV	29	GLFRSLYRSV	49	-20
1	A2:01	15	PCDHB6	p.G327E	11.7725	91	9	4	GLSEKCSLV	36	GLSGKCSLV	132	-96
1	A2:01	16	LDHD	p.R148Q	4.41145	20	11	3	HLQDSGLWFPV	39	HLRDSGLWF	126	-87
1	A2:01	17	GOSR1	p.P213S	14.3655	84	9	7	TLANRFSAV	45	TLANRFPAV	13	32
1	A2:01	18	CHD8	p.A866V	10.04205	64	10	5	FLVIVPLSTI	48	FLVIAPLSTI	38	10
1	A2:01	19	PCDHB6	p.G327E	11.7725	91	10	4	GLSEKCSLVV	57	GLSGKCSLVV	107	-50
1	A2:01	20	ENTHD2	p.P110F	2.401075	47	11	1	FLHGNSLYQKV	61	PLHGNSLYQK	1934	-1873
1	A2:01	21	PPFIA4	p.S43F	0.305613	6	9	1	FLRESQETL	65	SLRESQETL	522	-457
1	A2:01	22	FASTKD1	p.S159F	10.8846	26	10	6	LLSEFFSCLA	66	LLSEFSSCLA	133	-67
1	A2:01	23	LGALS8	p.P29S	11.86285	6	9	4	QLDSGTLIV	67	QLDPGTLIV	70	-3
1	A2:01	24	OSGIN1	p.P233L	7.949295	84	9	5	WMGLLDLEV	67	WMGLPDLEV	202	-135
1	A2:01	25	NUMA1	p.D136E	29.5143	144	9	4	FVLEHEDGL	71	FVLDHEDGL	79	-8
1	A2:01	26	PCLO	p.P2787L	0.07751	0	10	2	TLVSLATETV	75	TPVSLATETV	20558	-20483
1	A2:01	27	SNRNP200	p.H1175Y	51.15825	540	10	7	KMGKTIYKYV	76	KMGKTIHKYV	117	-41
1	A2:01	28	PLS1	p.P376L	4.53658	19	9	7	NLFNTYLCL	77	NLFNTYPCL	42	35
1	A2:01	29	MCM6	p.A630V	11.7303	33	9	5	RLSEVMARM	82	RLSEAMARM	117	-35
1	A2:01	30	DYM	p.S434F	20.5835	53	10	6	VLTEIFLGSL	105	VLTEISLGSL	185	-80
1	A2:01	31	IGSF1	p.G718E	0.029302	0	11	6	ALYKEEEQEPV	133	ALYKEGEQEP	133	0
1	A2:01	32	TAOK1	p.D299V	9.740685	32	11	11	VLIDLIQRTKV	134	VLIDLIQRTKD	21537	-21403
1	A2:01	33	QSER1	p.R1185C	9.301975	41	10	3	MVCTFCPPPL	138	MVRTFCPPPL	1307	-1169
1	A2:01	34	FAM83B	p.P127S	0.003497	0	10	5	LLFHSPRAHL	139	LLFHPPRAHL	222	-83
1	A2:01	35	KIF1C	p.M81V	26.47415	744	11	1	VLLHAFEGYNV	147	MLLHAFEGYN	73	74
1	A2:01	36	PCLO	p.P2787L	0.07751	0	9	8	VTSSIVTLV	158	VTSSIVTPV	57	101
1	A2:01	37	C8orf86	p.P192L	0	0	10	2	SLAPPRTPEL	212	SPAPPRTPEL	26625	-26413
1	A2:01	38	RETSAT	p.S492F	55.40045	200	9	1	FFVEASMSV	223	SFVEASMSV	2721	-2498
1	A2:01	39	ADCK1	p.C379S	4.45008	22	11	1	SMLTARSWDSV	248	CMLTARSWD	729	-481
1	A2:01	40	NUMA1	p.D136E	29.5143	144	11	4	FVLEHEDGLNL	261	FVLDHEDGLN	358	-97
1	A2:01	41	RPL13	p.A112T	403.4245	23	9	4	SLQTNVQRL	273	SLQANVQRL	152	121
1	A2:01	42	COPS7A	p.Y113F	48.0616	372	9	7	KVKCIPFAV	313	KVKCIPYAV	465	-152
1	A2:01	43	LYST	p.R395C	36.5952	55	10	7	FVFSKYCHRA	366	FVFSKYRHRA	1426	-1060
1	A2:01	44	SLC39A10	p.N393S	20.23725	86	10	1	SLVPEDEANI	368	NLVPEDEANI	1530	-1162
1	A2:01	45	DUSP4	p.L203F	54.64275	803	10	5	ILPFEYLGSA	380	ILPELYLGSA	786	-406
1	A2:01	46	SNTG1	p.N127I	0	0	11	2	RIAGEEVTLTV	416	RNAGEEVTLT	7235	-6819
1	A2:01	47	TTC34	p.A88T	0.016071	1089	10	3	VLTRLALLQL	418	VLARLALLQL	146	272
1	A2:01	48	F13A1	p.S305I	0.51986	0	11	6	LLEYRISENPV	440	LLEYRSSENP	813	-373
1	A2:01	49	PIAS2	p.P138Q	7.271005	27	11	7	MQQPSPQIPPV	449	MQQPSPPIPP	367	82
1	A2:01	50	CPVL	p.R25H	5.87024	2	9	4	GLFHSLYRS	463	GLFRSLYRS	1048	-585
2	A2:01	1	LRBA	p.S1325L	1.881915	6	9	8	LLTDLLFLI	4	LLTDLLFSI	4	0
2	A2:01	2	TK2	p.H137Y	4.86209	18	10	8	RLMERSIYSA	7	RLMERSIHSA	19	-12
2	A2:01	3	PKD1	p.S3220L	6.81945	7	9	8	FLVNDWLLV	8	FLVNDWLSV	7	1
2	A2:01	4	ATMIN	p.S715C	8.58103	43	10	5	FLDSCPHLPL	8	FLDSSPHLPL	9	-1
2	A2:01	5	CASK	p.H512Y	2.38613	9	9	7	KMNELNYCI	10	KMNELNHCI	24	-14
2	A2:01	6	LRBA	p.S1325L	1.88192	6	10	9	RLLTDLLFLI	10	RLLTDLLFSI	8	2
2	A2:01	7	PIGT	p.R570W	161.945	228	9	7	RLANLIWRA	11	RLANLIRRA	763	-752
2	A2:01	8	LRBA	p.S1325L	1.881915	6	9	9	RLLTDLLFL	12	RLLTDLLFS	909	-897
2	A2:01	9	EPHB1	p.R90C	1.78296	2	10	7	RIYTEMCFTV	13	RIYTEMRFTV	41	-28
2	A2:01	10	TRAPPC9	p.P181L	5.145625	5	9	2	VLFEKKDFV	16	VPFEKKDFV	18017	-18001
2	A2:01	11	MROH2A	p.S1064F	0.257644	18	9	7	KIFCASFRI	17	KIFCASSRI	261	-244
2	A2:01	12	LRBA	p.S1325L	1.881915	6	9	2	FLIETDIQM	17	FSIETDIQM	4758	-4741
2	A2:01	13	ELOVL3	p.P17L	0.64389	1	10	5	QLFQLYNFEL	23	QLFQPYNFEL	35	-12
2	A2:01	14	POLE2	p.H664N	8.53243	135	10	5	VLYFNHYFPA	24	VLYFHHYFPA	22	2
2	A2:01	15	SLC5A8	p.S76F	0.112721	1	9	7	TVLGTPFEV	26	TVLGTPSEV	391	-365
2	A2:01	16	TK2	p.H137Y	4.86209	18	10	3	SIYSARYIFV	26	SIHSARYIFV	205	-179
2	A2:01	17	DYDC2	p.H39Y	0.328192	1	9	8	YLAHWLYYY	27	YLAHWLYHY	99	-72
2	A2:01	18	SASH1	p.P981S	3.694035	10	9	3	KISSQPPPV	28	KIPSQPPPV	51	-23
2	A2:01	19	TERF1	p.S252L	7.794875	20	9	2	KLSTFLMKA	29	KSSTFLMKA	9473	-9444
2	A2:01	20	TTC39B	p.S634F	1.13525	4	9	8	FTLFELAFL	31	FTLFELASL	55	-24
2	A2:01	21	CDK12	p.S433I	3.62697	1	10	8	FLPRKENISV	31	FLPRKENSSV	75	-44
2	A2:01	22	EPB41L4A	p.R127C	3.31517	3	10	5	VLQGCLPCPV	31	VLQGRLPCPV	75	-44
2	A2:01	23	CCDC61	p.S469F	10.9458	51	9	1	FLANSGGWV	33	SLANSGGWV	300	-267
2	A2:01	24	EPDR1	p.D100N	18.869	27	9	7	YILLYKNGV	37	YILLYKDGV	59	-22
2	A2:01	25	TTC39B	p.S634F	1.13525	4	9	1	FLYKSQGEI	40	SLYKSQGEI	286	-246
2	A2:01	26	TUBG1	p.P350S	30.40205	39	9	3	FISWGPASI	44	FIPWGPASI	74	-30
2	A2:01	27	PCDHA4	p.S608L	0.590446	3	9	2	LLYELQPGT	44	LSYELQPGT	11654	-11610
2	A2:01	28	B3GALNT	p.H304Y	5.16678	14	9	3	NLYEEDALL	45	NLHEEDALL	273	-228
3	A2:01	1	DDHD1	p.S873F	0.71471	5	9	1	FLDVALFLL	6	SLDVALFLL	19	-13
3	A2:01	2	TTC28	p.S1823F	2.30807	2	9	1	FLLGLPNPA	7	SLLGLPNPA	34	-27
3	A2:01	3	STAG3	p.P212S	0.88299	2	9	1	SMDDLISLL	7	PMDDLISLL	371	-364
3	A2:01	4	KLHL13	p.G441E	2.77748	7	9	4	ALKEYLYAV	7	ALKGYLYAV	12	-5
3	A2:01	5	TTC28	p.S1823F	2.30807	2	10	1	FLLGLPNPAL	8	SLLGLPNPAL	26	-18
3	A2:01	6	SUSD2	p.M155I	20.3220	7	9	3	SMIEKSELV	9	SMMEKSELV	5	4
3	A2:01	7	NAV2	p.V2374I	23.3417	2	10	6	YLLEAIREGL	9	YLLEAVREGL	10	-1
3	A2:01	8	ABCC1	p.R1046C	7.15493	18	9	5	ILASCCLHV	10	ILASRCLHV	26	-16
3	A2:01	9	KIAA0556	p.S707L	2.72127	6	9	2	TLMGDMPSA	10	TSMGDMPSA	4711	-4701
3	A2:01	10	MCHR1	p.P377S	65.2719	164	9	4	CLNSFVYIV	12	CLNPFVYIV	14	-2
3	A2:01	11	RAC1	p.P29S	158.091	363	9	2	FSGEYIPTV	14	FPGEYIPTV	294	-280
3	A2:01	12	NBPF1	p.V1130I	1.31404	4	10	10	TLMGTSLHLI	14	TLMGTSLHLV	6	8
3	A2:01	13	SLC22A23	p.E456K	3.92437	7	9	6	SMMGHKVKV	15	SMMGHEVKV	9	6
3	A2:01	14	UTRN	p.G1513R	1.29967	1	9	1	RMDEQLTSL	15	GMDEQLTSL	17	-2
3	A2:01	15	SLC7A4	p.A348V	1.366	5	10	2	MVADGLFFQV	15	MAADGLFFQV	34	-19
3	A2:01	16	SLC9A1	p.P598L	23.1068	176	9	3	KILSAVSTV	16	KIPSAVSTV	107	-91
3	A2:01	17	SULT1A2	p.P19L	0.06452	11	9	1	LLIKYFAEA	17	PLIKYFAEA	917	-900
3	A2:01	18	SNX14	p.P253S	4.52371	2	10	8	KLTELLFSYI	19	KLTELLFPYI	10	9
3	A2:01	19	IQGAP3	p.S1429F	9.07844	14	10	6	SLTAHFLLPL	21	SLTAHSLLPL	27	-6
3	A2:01	20	STAG3	p.P212S	0.88299	2	9	2	FSMDDLISL	22	FPMDDLISL	603	-581
3	A2:01	21	UQCC2	p.E15K	47.996	660	10	7	FLKLCEKWPV	26	FLKLCEEWPV	13	13
3	A2:01	22	NUBPL	p.L266F	1.84902	1	10	6	KLAQTLGLEV	27	KLAQTLGLEV	30	-3
3	A2:01	23	AKR1B15	p.R46C	0.03659	1	9	3	LLLCPYPASL	28	LLLRPYPASL	135	-107
3	A2:01	24	GRIK2	p.P6S	0.74939	4	10	8	IIFSILSNPV	31	IIFPILSNPV	31	0
3	A2:01	25	NBPF1	p.V1130I	1.31404	4	9	9	LMGTSLHLI	32	LMGTSLHLV	9	23
3	A2:01	26	ABCB4	p.G1054R	0.94703	2	9	4	VLQRLSLEV	32	VLQGLSLEV	19	13
3	A2:01	27	SLC9A1	p.P598L	23.1068	176	9	6	GMGKILSAV	41	GMGKIPSAV	125	-84
3	A2:01	28	RXRG	p.A425V	17.8726	29	9	8	KLLLRLPVL	43	KLLLRLPAL	19	24
3	A2:01	29	NBPF1	p.V1130I	1.31404	4	10	5	SLHLIFQMGV	45	SLHLVFQMGV	49	-4
3	A2:01	30	STAG3	p.P212S	0.88299	2	10	2	FSMDDLISLL	45	FPMDDLISLL	688	-643
3	A2:01	31	PPAN	p.P165S	35.152	123	10	8	TMFQNLFSSI	46	TMFQNLFPSI	20	26
3	A2:01	32	NBAS	p.R2365C	2.47855	20	9	3	ALCAAQHWV	49	ALRAAQHWV	301	-252
3	A2:01	33	TLN1	p.S714F	46.3297	50	9	1	FQLVACTKV	50	SQLVACTKV	783	-733
3	A2:01	34	ISLR	p.E163K	117.196	4	10	4	TLAKGTFTPL	50	TLAEGTFTPL	14	36
4	A3:01	1	PHF1	p.L356F	N/A	N/A	10	3	RLFSALNSHK	7	RLLSALNSHK	12	-5
4	A3:01	2	F5	p.E1728K	N/A	N/A	10	10	SLHAHGLSYK	8	SLHAHGLSYE	8527	-8519
4	A3:01	3	HPN	p.E393K	N/A	N/A	9	9	KMFCAGYPK	16	KMFCAGYPE	12755	-12739
4	A3:01	4	SLC4A3	p.P411L	N/A	N/A	10	2	KLHVASLSFR	47	KPHVASLSFR	9042	-8995
4	A3:01	5	A2M	p.E1248K	N/A	N/A	9	9	KAPVGHFYK	244	KAPVGHFYE	21524	-21280

Table 5.2: Analysis of binding affinity and gene expression of putative neoantigens in patients. Related to Figures 5.2, 5.3, & 5.4.

DNA for NP modification (5'-biotin-)					
Barcode	Name	DNA sequence	Fluorescence		
1	D1-D4-D7	AA AAA AAA A GTG ATG AGT TTC AA ATC AGT CAA GAG AA CTC GTT CAC TAT AA CTG AAT CCT CGG GAT GCC TA	D1	D4	D7
2	D1-D4-D8	AA AAA AAA A GTG ATG AGT TTC AA ATC AGT CAA GAG AA CTT ACG AGT GTA AA CTG AAT CCT CGG GAT GCC TA	D1	D4	D8
3	D1-D4-D9	AA AAA AAA A GTG ATG AGT TTC AA ATC AGT CAA GAG AA TGT CTC TAA GTG AA CTG AAT CCT CGG GAT GCC TA	D1	D4	D9
4	D1-D5-D7	AA AAA AAA A GTG ATG AGT TTC AA GTA TTC GTC ATC AA CTC GTT CAC TAT AA CTG AAT CCT CGG GAT GCC TA	D1	D5	D7
5	D1-D5-D8	AA AAA AAA A GTG ATG AGT TTC AA GTA TTC GTC ATC AA CTT ACG AGT GTA AA CTG AAT CCT CGG GAT GCC TA	D1	D5	D8
6	D1-D5-D9	AA AAA AAA A GTG ATG AGT TTC AA GTA TTC GTC ATC AA TGT CTC TAA GTG AA CTG AAT CCT CGG GAT GCC TA	D1	D5	D9
7	D1-D6-D7	AA AAA AAA A GTG ATG AGT TTC AA GTC AGA TAG TTC AA CTC GTT CAC TAT AA CTG AAT CCT CGG GAT GCC TA	D1	D6	D7
8	D1-D6-D8	AA AAA AAA A GTG ATG AGT TTC AA GTC AGA TAG TTC AA CTT ACG AGT GTA AA CTG AAT CCT CGG GAT GCC TA	D1	D6	D8
9	D1-D6-D9	AA AAA AAA A GTG ATG AGT TTC AA GTC AGA TAG TTC AA TGT CTC TAA GTG AA CTG AAT CCT CGG GAT GCC TA	D1	D6	D9
10	D2-D4-D7	AA AAA AAA A CTA TGT CGA TAC AA ATC AGT CAA GAG AA CTC GTT CAC TAT AA CTG AAT CCT CGG GAT GCC TA	D2	D4	D7
11	D2-D4-D8	AA AAA AAA A CTA TGT CGA TAC AA ATC AGT CAA GAG AA CTT ACG AGT GTA AA CTG AAT CCT CGG GAT GCC TA	D2	D4	D8
12	D2-D4-D9	AA AAA AAA A CTA TGT CGA TAC AA ATC AGT CAA GAG AA TGT CTC TAA GTG AA CTG AAT CCT CGG GAT GCC TA	D2	D4	D9
13	D2-D5-D7	AA AAA AAA A CTA TGT CGA TAC AA GTA TTC GTC ATC AA CTC GTT CAC TAT AA CTG AAT CCT CGG GAT GCC TA	D2	D5	D7
14	D2-D5-D8	AA AAA AAA A CTA TGT CGA TAC AA GTA TTC GTC ATC AA CTT ACG AGT GTA AA CTG AAT CCT CGG GAT GCC TA	D2	D5	D8
15	D2-D5-D9	AA AAA AAA A CTA TGT CGA TAC AA GTA TTC GTC ATC AA TGT CTC TAA GTG AA CTG AAT CCT CGG GAT GCC TA	D2	D5	D9
16	D2-D6-D7	AA AAA AAA A CTA TGT CGA TAC AA GTC AGA TAG TTC AA CTC GTT CAC TAT AA CTG AAT CCT CGG GAT GCC TA	D2	D6	D7
17	D2-D6-D8	AA AAA AAA A CTA TGT CGA TAC AA GTC AGA TAG TTC AA CTT ACG AGT GTA AA CTG AAT CCT CGG GAT GCC TA	D2	D6	D8
18	D2-D6-D9	AA AAA AAA A CTA TGT CGA TAC AA GTC AGA TAG TTC AA TGT CTC TAA GTG AA CTG AAT CCT CGG GAT GCC TA	D2	D6	D9
19	D3-D4-D7	AA AAA AAA A TAC ATC CAA GAC AA ATC AGT CAA GAG AA CTC GTT CAC TAT AA CTG AAT CCT CGG GAT GCC TA	D3	D4	D7
20	D3-D4-D8	AA AAA AAA A TAC ATC CAA GAC AA ATC AGT CAA GAG AA CTT ACG AGT GTA AA CTG AAT CCT CGG GAT GCC TA	D3	D4	D8
21	D3-D4-D9	AA AAA AAA A TAC ATC CAA GAC AA ATC AGT CAA GAG AA TGT CTC TAA GTG AA CTG AAT CCT CGG GAT GCC TA	D3	D4	D9
22	D3-D5-D7	AA AAA AAA A TAC ATC CAA GAC AA GTA TTC GTC ATC AA CTC GTT CAC TAT AA CTG AAT CCT CGG GAT GCC TA	D3	D5	D7
23	D3-D5-D8	AA AAA AAA A TAC ATC CAA GAC AA GTA TTC GTC ATC AA CTT ACG AGT GTA AA CTG AAT CCT CGG GAT GCC TA	D3	D5	D8
24	D3-D5-D9	AA AAA AAA A TAC ATC CAA GAC AA GTA TTC GTC ATC AA TGT CTC TAA GTG AA CTG AAT CCT CGG GAT GCC TA	D3	D5	D9
25	D3-D6-D7	AA AAA AAA A TAC ATC CAA GAC AA GTC AGA TAG TTC AA CTC GTT CAC TAT AA CTG AAT CCT CGG GAT GCC TA	D3	D6	D7
26	D3-D6-D8	AA AAA AAA A TAC ATC CAA GAC AA GTC AGA TAG TTC AA CTT ACG AGT GTA AA CTG AAT CCT CGG GAT GCC TA	D3	D6	D8
27	D3-D6-D9	AA AAA AAA A TAC ATC CAA GAC AA GTC AGA TAG TTC AA TGT CTC TAA GTG AA CTG AAT CCT CGG GAT GCC TA	D3	D6	D9
DNA for barcoding					
Name		DNA sequence			
M1		5-Cy5-AGC ACA GGG AAA CTC ATC AC			
M2		5-Cy3(or Alex Fluor 750)- GCA TCA TCG TAT CGA CAT AG			
M3		5-Alex488-ATG GTT CGG TCT TGG ATG TA			
M4		5-Cy5-CGC CAA TGC TCT TGA CTG AT			
M5		5-Cy3(or Alex Fluor 750)- AGG ACT TCG ATG ACG AAT AC			
M6		5-Alex488-ATC CTT GCG AAC TAT CTG AC			
M7		5-Cy5-GCC GTA TCA TAG TGA ACG AG			
M8		5-Cy3(or Alex Fluor 750)- CCA GCG ATT ACA CTC GTA AG			
M9		5-Alex488-CAG ACC TGC ACT TAG AGA CA			
DNA for displacement					
Name		DNA sequence			
M1 comp		GTG ATG AGT TTC CCT GTG CT			
M2 comp		CTA TGT CGA TAC GAT GAT GC			
M3 comp		TAC ATC CAA GAC CGA ACC AT			
M4 comp		ATC AGT CAA GAG CAT TGG CG			
M5 comp		GTA TTC GTC ATC GAA GTC CT			
M6 comp		GTC AGA TAG TTC GCA AGG AT			
M7 comp		CTC GTT CAC TAT GAT ACG GC			
M8 comp		CTT ACG AGT GTA ATC GCT GG			
M9 comp		TGT CTC TAA GTG CAG GTC TG			
DNA for streptavidin labeling					
Name		DNA sequence			
DNA-SAC		5-NH_2-AAA AAA AAA A TAG GCA TCC CGA GGA TTC AG			

Table 5.3: DNA sequence for library construction and barcoding. Related to Figure 5.2.

Vα-gene-specific primers for cloning TCRα genes		
TRAV gene	**Signal peptide sequence**	**TRAV gene-specific sequence**
TRAV1-1*01	5'-TACAGGAAGCCTCAGCA	GGACAAAGCCTTGAGCAGCCCTC-3'
TRAV1-2*01	5'-TACAGGAAGCCTCAGCA	GGACAAAACATTGACCAGCCCACTG-3'
TRAV2*01	5'-TACAGGAAGCCTCAGCA	AAGGACCAAGTGTTTCAGCCTTCCAC-3'
TRAV3*01	5'-TACAGGAAGCCTCAGCA	GCTCAGTCAGTGGCTCAGCCGGA-3'
TRAV4*01	5'-TACAGGAAGCCTCAGCA	CTTGCTAAGACCACCCAGCCCATC-3'
TRAV5*01	5'-TACAGGAAGCCTCAGCA	GGAGAGGATGTGGAGCAGAGTCTTTTCC-3'
TRAV6*01	5'-TACAGGAAGCCTCAGCA	AGCCAAAAGATAGAACAGAATTCCGAGGC-3'
TRAV6*03	5'-TACAGGAAGCCTCAGCA	GAGGCCCTGAACATTCAGGAGGG-3'
TRAV7*01	5'-TACAGGAAGCCTCAGCA	GAAAACCAGGTGGAGCACAGCCC-3'
TRAV8-1*01	5'-TACAGGAAGCCTCAGCA	GCCCAGTCTGTGAGCCAGCATAACC-3'
TRAV8-2*01	5'-TACAGGAAGCCTCAGCA	GCCCAGTCGGTGACCCAGCTTG-3'
TRAV8-2*02	5'-TACAGGAAGCCTCAGCA	GCCCAGTCGGTGACCCAGCTTAG-3'
TRAV8-3*01	5'-TACAGGAAGCCTCAGCA	GCCCAGTCAGTGACCCAGCCTG-3'
TRAV8-4*06	5'-TACAGGAAGCCTCAGCA	CTCTTCTGGTATGTGCAATACCCCAACC-3'
TRAV8-4*07	5'-TACAGGAAGCCTCAGCA	GTTGAACCATATCTCTTCTGGTATGTGCAATACC-3'
TRAV8-6*01	5'-TACAGGAAGCCTCAGCA	GCCCAGTCTGTGACCCAGCTTGAC-3'
TRAV8-7*01	5'-TACAGGAAGCCTCAGCA	ACCCAGTCGGTGACCCAGCTTG-3'
TRAV9-1*01	5'-TACAGGAAGCCTCAGCA	GGAGATTCAGTGGTCCAGACAGAAGGC-3'
TRAV9-2*01	5'-TACAGGAAGCCTCAGCA	GGAAATTCAGTGACCCAGATGGAAGG-3'
TRAV9-2*02	5'-TACAGGAAGCCTCAGCA	GGAGATTCAGTGACCCAGATGGAAGG-3'
TRAV10*01	5'-TACAGGAAGCCTCAGCA	AAAAACCAAGTGGAGCAGAGTCCTCAGTC-3'
TRAV11*01	5'-TACAGGAAGCCTCAGCA	CTACATACACTGGAGCAGAGTCCTTCATTCC-3'
TRAV12-1*01	5'-TACAGGAAGCCTCAGCA	CGGAAGGAGGTGGAGCAGGATCC-3'
TRAV12-2*01	5'-TACAGGAAGCCTCAGCA	CAGAAGGAGGTGGAGCAGAATTCTGG-3'
TRAV12-2*03	5'-TACAGGAAGCCTCAGCA	GGACCCCTCAGTGTTCCAGAGGG-3'
TRAV12-3*01	5'-TACAGGAAGCCTCAGCA	CAGAAGGAGGTGGAGCAGGATCCTG-3'
TRAV13-1*02	5'-TACAGGAAGCCTCAGCA	GGAGAGAATGTGGAGCAGCATCCTTC-3'
TRAV13-2*01	5'-TACAGGAAGCCTCAGCA	GGAGAGAGTGTGGGGCTGCATCTTC-3'
TRAV14/DV4*01	5'-TACAGGAAGCCTCAGCA	GCCCAGAAGATAACTCAAACCCAACCAG-3'
TRAV14/DV4*04	5'-TACAGGAAGCCTCAGCA	CAGAAGATAACTCAAACCCAACCAGGAATG-3'
TRAV16*01	5'-TACAGGAAGCCTCAGCA	GCCCAGAGAGTGACTCAGCCCGA-3'
TRAV17*01	5'-TACAGGAAGCCTCAGCA	AGTCAACAGGGAGAAGAGGATCCTCAGG-3'
TRAV18*01	5'-TACAGGAAGCCTCAGCA	GGAGACTCGGTTACCCAGACAGAAGG-3'
TRAV19*01	5'-TACAGGAAGCCTCAGCA	GCTCAGAAGGTAACTCAAGCGCAGACTG-3'
TRAV20*01	5'-TACAGGAAGCCTCAGCA	GAAGACCAGGTGACGCAGAGTCCC-3'
TRAV21*01	5'-TACAGGAAGCCTCAGCA	AAACAGGAGGTGACGCAGATTCCTGC-3'
TRAV22*01	5'-TACAGGAAGCCTCAGCA	GGAATACAAGTGGAGCAGAGTCCTCCAG-3'
TRAV23/DV6*01	5'-TACAGGAAGCCTCAGCA	CAGCAGCAGGTGAAACAAAGTCCTCA-3'
TRAV23/DV6*04	5'-TACAGGAAGCCTCAGCA	CAGCAGGTGAAACAAAGTCCTCAATCTTTG-3'
TRAV24*01	5'-TACAGGAAGCCTCAGCA	ATACTGAACGTGGAACAAAGTCCTCAGTCAC-3'
TRAV25*01	5'-TACAGGAAGCCTCAGCA	GGACAACAGGTAATGCAAATTCCTCAGTACC-3'
TRAV26-1*01	5'-TACAGGAAGCCTCAGCA	GATGCTAAGACCACCCAGCCCCC-3'
TRAV26-1*02	5'-TACAGGAAGCCTCAGCA	GATGCTAAGACCACCCAGCCCACC-3'
TRAV26-2*01	5'-TACAGGAAGCCTCAGCA	GATGCTAAGACCACACAGCCAAATTCAATG-3'
TRAV27*01	5'-TACAGGAAGCCTCAGCA	ACCCAGCTGCTGGAGCAGAGCC-3'
TRAV29/DV5*01	5'-TACAGGAAGCCTCAGCA	GACCAGCAAGTTAAGCAAAATTCACCATC-3'
TRAV30*01	5'-TACAGGAAGCCTCAGCA	CAACAACCAGTGCAGAGTCCTCAAGC-3'
TRAV34*01	5'-TACAGGAAGCCTCAGCA	AGCCAAGAACTGGAGCAGAGTCCTCAG-3'
TRAV35*01	5'-TACAGGAAGCCTCAGCA	GGTCAACAGCTGAATCAGAGTCCTCAATC-3'
TRAV36/DV7*01	5'-TACAGGAAGCCTCAGCA	GAAGACAAGGTGGTACAAAGCCCTCTATCTC-3'
TRAV36/DV7*02	5'-TACAGGAAGCCTCAGCA	GAAGACAAGGTGGTACAAAGCCCTCAATC-3'
TRAV38-1*01	5'-TACAGGAAGCCTCAGCA	GCCCAGACAGTCACTCAGTCTCAACCAG-3'
TRAV38-1*04	5'-TACAGGAAGCCTCAGCA	GCCCAGACAGTCACTCAGTCCCAGC-3'
TRAV38-2/DV8*01	5'-TACAGGAAGCCTCAGCA	GCTCAGACAGTCACTCAGTCTCAACCAGAG-3'
TRAV39*01	5'-TACAGGAAGCCTCAGCA	GAGCTGAAAGTGGAACAAAACCCTCTGTTC-3'
TRAV40*01	5'-TACAGGAAGCCTCAGCA	AGCAATTCAGTCAAGCAGACGGGC-3'
TRAV41*01	5'-TACAGGAAGCCTCAGCA	AAAAATGAAGTGGAGCAGAGTCCTCAGAAC-3'
Vβ-gene-specific primers for cloning TCRα genes		
TRBV gene	**Signal peptide sequence**	**TRBV gene-specific sequence**
TRBV1*01	5'-CAGGAGGGCTCGGCA	GATACTGGAATTACCCAGACACCAAAATACCTG-3'
TRBV2*01	5'-CAGGAGGGCTCGGCA	GAACCTGAAGTCACCCAGACTCCCAG-3'
TRBV3-1*01	5'-CAGGAGGGCTCGGCA	GACACAGCTGTTTCCCAGACTCCAAAATAC-3'
TRBV3-2*01	5'-CAGGAGGGCTCGGCA	GACACAGCCGTTTCCCAGACTCCA-3'
TRBV4-1*01	5'-CAGGAGGGCTCGGCA	GACACTGAAGTTACCCAGACACCAAAACAC-3'

Table 5.4: Single cell TCRα and TCRβ cloning primers.

TRBV4-1*02	5'-CAGGAGGGCTCGGCA	CACCTGGTCATGGGAATGACAAATAAGAAG-3'
TRBV4-2*01	5'-CAGGAGGGCTCGGCA	GAAACGGGAGTTACGCAGACACCAAG-3'
TRBV4-3*04	5'-CAGGAGGGCTCGGCA	AAGAAGTCTTTGAAATGTGAACAACATCTGGG-3'
TRBV5-1*01	5'-CAGGAGGGCTCGGCA	AAGGCTGGAGTCACTCAAACTCCAAGATATC-3'
TRBV5-1*02	5'-CAGGAGGGCTCGGCA	AGGGCTGGGGTCACTCAAACTCC-3'
TRBV5-3*01	5'-CAGGAGGGCTCGGCA	GAGGCTGGAGTCACCCAAAGTCCC-3'
TRBV5-4*01	5'-CAGGAGGGCTCGGCA	GAGACTGGAGTCACCCAAAGTCCCAC-3'
TRBV5-4*03	5'-CAGGAGGGCTCGGCA	CAGCAAGTGACACTGAGATGCTCTTCTCAG-3'
TRBV5-4*04	5'-CAGGAGGGCTCGGCA	ACTGTGTCCTGGTACCAACAGGCCCT-3'
TRBV5-5*01	5'-CAGGAGGGCTCGGCA	GACGCTGGAGTCACCCAAAGTCC-3'
TRBV5-8*01	5'-CAGGAGGGCTCGGCA	GAGGCTGGAGTCACACAAAGTCCCAC-3'
TRBV5-8*02	5'-CAGGAGGGCTCGGCA	AGGACAGCAAGCGACTCTGAGATGC-3'
TRBV6-1*01	5'-CAGGAGGGCTCGGCA	AATGCTGGTGTCACTCAGACCCCA-3'
TRBV6-4*01	5'-CAGGAGGGCTCGGCA	ATTGCTGGGATCACCCAGGCAC-3'
TRBV6-4*02	5'-CAGGAGGGCTCGGCA	ACTGCTGGGATCACCCAGGCAC-3'
TRBV7-1*01	5'-CAGGAGGGCTCGGCA	GGTGCTGGAGTCTCCCAGTCCCTG-3'
TRBV7-2*01	5'-CAGGAGGGCTCGGCA	GGAGCTGGAGTCTCCCAGTCCCC-3'
TRBV7-2*04	5'-CAGGAGGGCTCGGCA	GGAGCTGGAGTTTCCCAGTCCCC-3'
TRBV7-3*01	5'-CAGGAGGGCTCGGCA	GGTGCTGGAGTCTCCCAGACCC-3'
TRBV7-3*05	5'-CAGGAGGGCTCGGCA	TGGGAGCTCAGGTGTGATCCAATTTC-3'
TRBV7-4*01	5'-CAGGAGGGCTCGGCA	GGTGCTGGAGTCTCCCAGTCCC-3'
TRBV7-6*01	5'-CAGGAGGGCTCGGCA	GGTGCTGGAGTCTCCCAGTCTCCC-3'
TRBV7-9*01	5'-CAGGAGGGCTCGGCA	GATACTGGAGTCTCCCAGAACCCCAG-3'
TRBV7-9*03	5'-CAGGAGGGCTCGGCA	GATACTGGAGTCTCCCAGGACCCCAG-3'
TRBV7-9*04	5'-CAGGAGGGCTCGGCA	ATATCTGGAGTCTCCCACAACCCCAGAC-3'
TRBV7-9*07	5'-CAGGAGGGCTCGGCA	CACAACCGCCTTTATTGGTACCGACAG-3'
TRBV9*01	5'-CAGGAGGGCTCGGCA	GATTCTGGAGTCACACAAACCCCAAAGC-3'
TRBV10-1*01	5'-CAGGAGGGCTCGGCA	GATGCTGAAATCACCCAGAGCCCAAG-3'
TRBV10-2*01	5'-CAGGAGGGCTCGGCA	GATGCTGGAATCACCCAGAGCCCA-3'
TRBV10-2*02	5'-CAGGAGGGCTCGGCA	AAGGCAGGTGACCTTGATGTGTCACC-3'
TRBV11-1*01	5'-CAGGAGGGCTCGGCA	GAAGCTGAAGTTGCCCAGTCCCC-3'
TRBV11-2*01	5'-CAGGAGGGCTCGGCA	GAAGCTGGAGTTGCCCAGTCTCCCAG-3'
TRBV11-3*01	5'-CAGGAGGGCTCGGCA	GAAGCTGGAGTGGTTCAGTCTCCCAGA-3'
TRBV11-3*03	5'-CAGGAGGGCTCGGCA	GGTCTCCCAGATATAAGATTATAGAGAAGAAACAGC-3'
TRBV12-1*01	5'-CAGGAGGGCTCGGCA	GATGCTGGTGTTATCCAGTCACCCAGG-3'
TRBV12-2*01	5'-CAGGAGGGCTCGGCA	GATGCTGGCATTATCCAGTCACCCAAG-3'
TRBV12-3*01	5'-CAGGAGGGCTCGGCA	GATGCTGGAGTTATCCAGTCACCCC-3'
TRBV12-5*01	5'-CAGGAGGGCTCGGCA	GATGCTAGAGTCACCCAGACACCAAGG-3'
TRBV13*01	5'-CAGGAGGGCTCGGCA	GCTGCTGGAGTCATCCAGTCCCC-3'
TRBV14*01	5'-CAGGAGGGCTCGGCA	GAAGCTGGAGTTACTCAGTTCCCCAGC-3'
TRBV15*01	5'-CAGGAGGGCTCGGCA	GATGCCATGGTCATCCAGAACCCAAG-3'
TRBV16*01	5'-CAGGAGGGCTCGGCA	GGTGAAGAAGTCGCCCAGACTCCA-3'
TRBV17*01	5'-CAGGAGGGCTCGGCA	GAGCCTGGAGTCAGCCAGACCC-3'
TRBV18*01	5'-CAGGAGGGCTCGGCA	AATGCCGGCGTCATGCAGAAC-3'
TRBV19*01	5'-CAGGAGGGCTCGGCA	GATGGTGGAATCACTCAGTCCCCAAAG-3'
TRBV20-1*01	5'-CAGGAGGGCTCGGCA	GGTGCTGTCGTCTCTCAACATCCGAG-3'
TRBV20/OR9-2*01	5'-CAGGAGGGCTCGGCA	AGTGCTGTCGTCTCTCAACATCCGAG-3'
TRBV21-1*01	5'-CAGGAGGGCTCGGCA	GACACCAAGGTCACCCAGAGACCTAGAC-3'
TRBV21/OR9-2*01	5'-CAGGAGGGCTCGGCA	GACACCAAGGTCACCCAGAGACCTAGATTTC-3'
TRBV23-1*01	5'-CAGGAGGGCTCGGCA	CATGCCAAAGTCACACAGACTCCAGG-3'
TRBV24-1*01	5'-CAGGAGGGCTCGGCA	GATGCTGATGTTACCCAGACCCCAAG-3'
TRBV25-1*01	5'-CAGGAGGGCTCGGCA	GAAGCTGACATCTACCAGACCCCAAGATAC-3'
TRBV26*01	5'-CAGGAGGGCTCGGCA	GATGCTGTAGTTACACAATTCCCAAGACACAG-3'
TRBV26/OR9-2*01	5'-CAGGAGGGCTCGGCA	GATGCTGTAGTTACACAATTCTCAAGACACAGAATC-3'
TRBV27*01	5'-CAGGAGGGCTCGGCA	GAAGCCCAAGTGACCCAGAACCC-3'
TRBV28*01	5'-CAGGAGGGCTCGGCA	GATGTGAAAGTAACCCAGAGCTCGAGATATC-3'
TRBV29-1*01	5'-CAGGAGGGCTCGGCA	AGTGCTGTCATCTCTCAAAAGCCAAGC-3'
TRBV29-1*03	5'-CAGGAGGGCTCGGCA	ACGATCCAGTGTCAAGTCGATAGCCAAG-3'
TRBV30*01	5'-CAGGAGGGCTCGGCA	TCTCAGACTATTCATCAATGGCCAGCG-3'
TRBV30*04	5'-CAGGAGGGCTCGGCA	ACTATTCATCAATGGCCAGCGACCC-3'

Table 5.5: Single cell TCRα and TCRβ cloning primers (continued).

BIBLIOGRAPHY

[1] Beatriz M. Carreno, Vincent Magrini, Michelle Becker-Hapak, Saghar Kaabinejadian, Jasreet Hundal, Allegra A. Petti, Amy Ly, Wen-Rong Lie, William H. Hildebrand, Elaine R. Mardis, and Gerald P. Linette. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science*, 348(6236):803–808, May 2015. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aaa3828. URL https://science.sciencemag.org/content/348/6236/803. Publisher: American Association for the Advancement of Science Section: Report.

[2] Matthew M. Gubin, Xiuli Zhang, Heiko Schuster, Etienne Caron, Jeffrey P. Ward, Takuro Noguchi, Yulia Ivanova, Jasreet Hundal, Cora D. Arthur, Willem-Jan Krebber, Gwenn E. Mulder, Mireille Toebes, Matthew D. Vesely, Samuel S. K. Lam, Alan J. Korman, James P. Allison, Gordon J. Freeman, Arlene H. Sharpe, Erika L. Pearce, Ton N. Schumacher, Ruedi Aebersold, Hans-Georg Rammensee, Cornelis J. M. Melief, Elaine R. Mardis, William E. Gillanders, Maxim N. Artyomov, and Robert D. Schreiber. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature*, 515(7528):577–581, November 2014. ISSN 1476-4687. doi: 10.1038/nature13988. URL https://www.nature.com/articles/nature13988. Number: 7528 Publisher: Nature Publishing Group.

[3] Patrick A. Ott, Zhuting Hu, Derin B. Keskin, Sachet A. Shukla, Jing Sun, David J. Bozym, Wandi Zhang, Adrienne Luoma, Anita Giobbie-Hurder, Lauren Peter, Christina Chen, Oriol Olive, Todd A. Carter, Shuqiang Li, David J. Lieb, Thomas Eisenhaure, Evisa Gjini, Jonathan Stevens, William J. Lane, Indu Javeri, Kaliappanadar Nellaiappan, Andres M. Salazar, Heather Daley, Michael Seaman, Elizabeth I. Buchbinder, Charles H. Yoon, Maegan Harden, Niall Lennon, Stacey Gabriel, Scott J. Rodig, Dan H. Barouch, Jon C. Aster, Gad Getz, Kai Wucherpfennig, Donna Neuberg, Jerome Ritz, Eric S. Lander, Edward F. Fritsch, Nir Hacohen, and Catherine J. Wu. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature*, 547(7662):217–221, July 2017. ISSN 1476-4687. doi: 10.1038/nature22991. URL https://www.nature.com/articles/nature22991. Number: 7662 Publisher: Nature Publishing Group.

[4] David F Stroncek, Carolina Berger, Martin A Cheever, Richard W Childs, Mark E Dudley, Peter Flynn, Luca Gattinoni, James R Heath, Michael Kalos, Francesco M Marincola, Jeffrey S Miller, Gustavo Mostoslavsky, Daniel J Powell, Mahendra Rao, Nicholas P Restifo, Steven A Rosenberg, John O'Shea, and Cornelis JM Melief. New directions in cellular therapy of cancer: a summary of the summit on cellular therapy for cancer. *Journal of Translational Medicine*,

10:48, March 2012. ISSN 1479-5876. doi: 10.1186/1479-5876-10-48. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3362772/.

[5] Nikolaos Zacharakis, Harshini Chinnasamy, Mary Black, Hui Xu, Yong-Chen Lu, Zhili Zheng, Anna Pasetto, Michelle Langhan, Thomas Shelton, Todd Prickett, Jared Gartner, Li Jia, Katarzyna Trebska-McGowan, Robert P. Somerville, Paul F. Robbins, Steven A. Rosenberg, Stephanie L. Goff, and Steven A. Feldman. Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. *Nature Medicine*, 24 (6):724–730, June 2018. ISSN 1546-170X. doi: 10.1038/s41591-018-0040-8. URL https://www.nature.com/articles/s41591-018-0040-8. Number: 6 Publisher: Nature Publishing Group.

[6] Michael F. Berger and Elaine R. Mardis. The emerging clinical relevance of genomics in cancer medicine. *Nature Reviews Clinical Oncology*, 15(6):353–365, June 2018. ISSN 1759-4782. doi: 10.1038/s41571-018-0002-6. URL https://www.nature.com/articles/s41571-018-0002-6. Number: 6 Publisher: Nature Publishing Group.

[7] Yong-Chen Lu, Xin Yao, Jessica S. Crystal, Yong F. Li, Mona El-Gamil, Colin Gross, Lindy Davis, Mark E. Dudley, James C. Yang, Yardena Samuels, Steven A. Rosenberg, and Paul F. Robbins. Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 20(13):3401–3410, July 2014. ISSN 1557-3265. doi: 10.1158/1078-0432.CCR-14-0433.

[8] Paul F. Robbins, Yong-Chen Lu, Mona El-Gamil, Yong F. Li, Colin Gross, Jared Gartner, Jimmy C. Lin, Jamie K. Teer, Paul Cliften, Eric Tycksen, Yardena Samuels, and Steven A. Rosenberg. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nature Medicine*, 19(6):747–752, June 2013. ISSN 1546-170X. doi: 10.1038/nm.3161.

[9] Marvin H. Gee, Arnold Han, Shane M. Lofgren, John F. Beausang, Juan L. Mendoza, Michael E. Birnbaum, Michael T. Bethune, Suzanne Fischer, Xinbo Yang, Raquel Gomez-Eerland, David B. Bingham, Leah V. Sibener, Ricardo A. Fernandes, Andrew Velasco, David Baltimore, Ton N. Schumacher, Purvesh Khatri, Stephen R. Quake, Mark M. Davis, and K. Christopher Garcia. Antigen Identification for Orphan T Cell Receptors Expressed on Tumor-Infiltrating Lymphocytes. *Cell*, 172(3):549–563.e16, January 2018. ISSN 1097-4172. doi: 10.1016/j.cell.2017.11.043.

[10] Nienke van Rooij, Marit M. van Buuren, Daisy Philips, Arno Velds, Mireille Toebes, Bianca Heemskerk, Laura J. A. van Dijk, Sam Behjati, Henk Hilkmann, Dris El Atmioui, Marja Nieuwland, Michael R. Stratton, Ron M.

Kerkhoven, Can Kesmir, John B. Haanen, Pia Kvistborg, and Ton N. Schumacher. Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 31(32):e439–442, November 2013. ISSN 1527-7755. doi: 10.1200/JCO.2012.47.7521.

[11] Mahesh Yadav, Suchit Jhunjhunwala, Qui T. Phung, Patrick Lupardus, Joshua Tanguay, Stephanie Bumbaca, Christian Franci, Tommy K. Cheung, Jens Fritsche, Toni Weinschenk, Zora Modrusan, Ira Mellman, Jennie R. Lill, and Lélia Delamarre. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature*, 515(7528):572–576, November 2014. ISSN 1476-4687. doi: 10.1038/nature14001. URL https://www.nature.com/articles/nature14001. Number: 7528 Publisher: Nature Publishing Group.

[12] Edward F. Fritsch, Mohini Rajasagi, Patrick A. Ott, Vladimir Brusic, Nir Hacohen, and Catherine J. Wu. HLA-binding properties of tumor neoepitopes in humans. *Cancer Immunology Research*, 2(6):522–529, June 2014. ISSN 2326-6074. doi: 10.1158/2326-6066.CIR-13-0227.

[13] Morten Nielsen, Claus Lundegaard, Thomas Blicher, Kasper Lamberth, Mikkel Harndahl, Sune Justesen, Gustav Røder, Bjoern Peters, Alessandro Sette, Ole Lund, and Søren Buus. NetMHCpan, a Method for Quantitative Predictions of Peptide Binding to Any HLA-A and -B Locus Protein of Known Sequence. *PLoS ONE*, 2(8):e796, August 2007. ISSN 1932-6203. doi: 10.1371/journal.pone.0000796. URL https://dx.plos.org/10.1371/journal.pone.0000796.

[14] Carsten Linnemann, Marit M. van Buuren, Laura Bies, Els M. E. Verdegaal, Remko Schotte, Jorg J. A. Calis, Sam Behjati, Arno Velds, Henk Hilkmann, Dris el Atmioui, Marten Visser, Michael R. Stratton, John B. A. G. Haanen, Hergen Spits, Sjoerd H. van der Burg, and Ton N. M. Schumacher. High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4 + T cells in human melanoma. *Nature Medicine*, 21 (1):81–85, January 2015. ISSN 1546-170X. doi: 10.1038/nm.3773. URL https://www.nature.com/articles/nm.3773. Number: 1 Publisher: Nature Publishing Group.

[15] Rikke Sick Andersen, Pia Kvistborg, Thomas Mørch Frøsig, Natasja W. Pedersen, Rikke Lyngaa, Arnold H. Bakker, Chengyi Jenny Shu, Per thor Straten, Ton N. Schumacher, and Sine Reker Hadrup. Parallel detection of antigen-specific T cell responses by combinatorial encoding of MHC multimers. *Nature Protocols*, 7(5):891–902, April 2012. ISSN 1750-2799. doi: 10.1038/nprot.2012.037.

[16] M. Fehlings, Y. Simoni, H. L. Penny, E. Becht, C. Y. Loh, M. M. Gubin, J. P. Ward, S. C. Wong, R. D. Schreiber, and E. W. Newell. Checkpoint blockade

immunotherapy reshapes the high-dimensional phenotypic heterogeneity of murine intratumoural neoantigen-specific CD8 + T cells. *Nature Communications*, 8(1):562, September 2017. ISSN 2041-1723. doi: 10.1038/s41467-017-00627-z. URL https://www.nature.com/articles/s41467-017-00627-z. Number: 1 Publisher: Nature Publishing Group.

[17] Evan W. Newell, Natalia Sigal, Nitya Nair, Brian A. Kidd, Harry B. Greenberg, and Mark M. Davis. Combinatorial tetramer staining and mass cytometry analysis facilitate T-cell epitope mapping and characterization. *Nature Biotechnology*, 31(7):623–629, July 2013. ISSN 1546-1696. doi: 10.1038/nbt.2593. URL https://www.nature.com/articles/nbt.2593. Number: 7 Publisher: Nature Publishing Group.

[18] Amalie Kai Bentzen, Andrea Marion Marquard, Rikke Lyngaa, Sunil Kumar Saini, Sofie Ramskov, Marco Donia, Lina Such, Andrew J. S. Furness, Nicholas McGranahan, Rachel Rosenthal, Per thor Straten, Zoltan Szallasi, Inge Marie Svane, Charles Swanton, Sergio A. Quezada, Søren Nyboe Jakobsen, Aron Charles Eklund, and Sine Reker Hadrup. Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nature Biotechnology*, 34(10):1037–1045, October 2016. ISSN 1546-1696. doi: 10.1038/nbt.3662. URL https://www.nature.com/articles/nbt.3662. Number: 10 Publisher: Nature Publishing Group.

[19] Shu-Qi Zhang, Ke-Yue Ma, Alexandra A. Schonnesen, Mingliang Zhang, Chenfeng He, Eric Sun, Chad M. Williams, Weiping Jia, and Ning Jiang. High-throughput determination of the antigen specificities of T cell receptors in single cells. *Nature Biotechnology*, November 2018. ISSN 1546-1696. doi: 10.1038/nbt.4282.

[20] Naiyer A. Rizvi, Matthew D. Hellmann, Alexandra Snyder, Pia Kvistborg, Vladimir Makarov, Jonathan J. Havel, William Lee, Jianda Yuan, Phillip Wong, Teresa S. Ho, Martin L. Miller, Natasha Rekhtman, Andre L. Moreira, Fawzia Ibrahim, Cameron Bruggeman, Billel Gasmi, Roberta Zappasodi, Yuka Maeda, Chris Sander, Edward B. Garon, Taha Merghoub, Jedd D. Wolchok, Ton N. Schumacher, and Timothy A. Chan. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science (New York, N.Y.)*, 348(6230):124–128, April 2015. ISSN 1095-9203. doi: 10.1126/science.aaa1348.

[21] Nicholas McGranahan, Andrew J. S. Furness, Rachel Rosenthal, Sofie Ramskov, Rikke Lyngaa, Sunil Kumar Saini, Mariam Jamal-Hanjani, Gareth A. Wilson, Nicolai J. Birkbak, Crispin T. Hiley, Thomas B. K. Watkins, Seema Shafi, Nirupa Murugaesu, Richard Mitter, Ayse U. Akarca, Joseph Linares, Teresa Marafioti, Jake Y. Henry, Eliezer M. Van Allen, Diana Miao, Bastian Schilling, Dirk Schadendorf, Levi A. Garraway, Vladimir Makarov, Naiyer A. Rizvi, Alexandra Snyder, Matthew D. Hellmann, Taha Merghoub, Jedd D. Wolchok, Sachet A. Shukla, Catherine J. Wu, Karl S. Peggs, Timothy A. Chan,

Sine R. Hadrup, Sergio A. Quezada, and Charles Swanton. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science*, 351(6280):1463–1469, March 2016. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aaf1490. URL https://science.sciencemag.org/content/351/6280/1463. Publisher: American Association for the Advancement of Science Section: Report.

[22] Bryan Howie, Anna M. Sherwood, Ashley D. Berkebile, Jan Berka, Ryan O. Emerson, David W. Williamson, Ilan Kirsch, Marissa Vignali, Mark J. Rieder, Christopher S. Carlson, and Harlan S. Robins. High-throughput pairing of T cell receptor α and β sequences. *Science Translational Medicine*, 7(301):301ra131–301ra131, August 2015. ISSN 1946-6234, 1946-6242. doi: 10.1126/scitranslmed.aac5624. URL https://stm.sciencemag.org/content/7/301/301ra131. Publisher: American Association for the Advancement of Science Section: Research Article.

[23] Jacob Glanville, Huang Huang, Allison Nau, Olivia Hatton, Lisa E. Wagar, Florian Rubelt, Xuhuai Ji, Arnold Han, Sheri M. Krams, Christina Pettus, Nikhil Haas, Cecilia S. Lindestam Arlehamn, Alessandro Sette, Scott D. Boyd, Thomas J. Scriba, Olivia M. Martinez, and Mark M. Davis. Identifying specificity groups in the T cell receptor repertoire. *Nature*, 547(7661):94–98, July 2017. ISSN 1476-4687. doi: 10.1038/nature22976. URL https://www.nature.com/articles/nature22976. Number: 7661 Publisher: Nature Publishing Group.

[24] Arnold Han, Jacob Glanville, Leo Hansmann, and Mark M. Davis. Linking T-cell receptor sequence to functional phenotype at the single-cell level. *Nature Biotechnology*, 32(7):684–692, July 2014. ISSN 1546-1696. doi: 10.1038/nbt.2938. URL https://www.nature.com/articles/nbt.2938. Number: 7 Publisher: Nature Publishing Group.

[25] Patrick H. N. Celie, Mireille Toebes, Boris Rodenko, Huib Ovaa, Anastassis Perrakis, and Ton N. M. Schumacher. UV-Induced Ligand Exchange in MHC Class I Protein Crystals. *Journal of the American Chemical Society*, 131(34):12298–12304, September 2009. ISSN 0002-7863. doi: 10.1021/ja9037559. URL https://doi.org/10.1021/ja9037559. Publisher: American Chemical Society.

[26] Boris Rodenko, Mireille Toebes, Sine Reker Hadrup, Wim J. E. van Esch, Annemieke M. Molenaar, Ton N. M. Schumacher, and Huib Ovaa. Generation of peptide–MHC class I complexes through UV-mediated ligand exchange. *Nature Protocols*, 1(3):1120–1132, August 2006. ISSN 1750-2799. doi: 10.1038/nprot.2006.121. URL https://www.nature.com/articles/nprot.2006.121. Number: 3 Publisher: Nature Publishing Group.

[27] Vasanthi Ramachandiran, Vitalii Grigoriev, Lan Lan, Eugene Ravkov, Suzanne A. Mertens, and John D. Altman. A robust method for produc-

tion of MHC tetramers with small molecule fluorophores. *Journal of Immunological Methods*, 319(1-2):13–20, January 2007. ISSN 0022-1759. doi: 10.1016/j.jim.2006.08.014.

[28] Gabriel A. Kwong, Caius G. Radu, Kiwook Hwang, Chengyi J. Shu, Chao Ma, Richard C. Koya, Begonya Comin-Anduix, Sine Reker Hadrup, Ryan C. Bailey, Owen N. Witte, Ton N. Schumacher, Antoni Ribas, and James R. Heath. Modular Nucleic Acid Assembled p/MHC Microarrays for Multiplexed Sorting of Antigen-Specific T Cells. *Journal of the American Chemical Society*, 131 (28):9695–9703, July 2009. ISSN 0002-7863. doi: 10.1021/ja9006707. URL https://doi.org/10.1021/ja9006707. Publisher: American Chemical Society.

[29] Laura A. Johnson, Bianca Heemskerk, Daniel J. Powell, Cyrille J. Cohen, Richard A. Morgan, Mark E. Dudley, Paul F. Robbins, and Steven A. Rosenberg. Gene transfer of tumor-reactive TCR confers both high avidity and tumor reactivity to nonreactive peripheral blood mononuclear cells and tumor-infiltrating lymphocytes. *Journal of Immunology (Baltimore, Md.: 1950)*, 177(9):6548–6559, November 2006. ISSN 0022-1767. doi: 10.4049/jimmunol.177.9.6548.

[30] Theodore L. Roth, Cristina Puig-Saus, Ruby Yu, Eric Shifrut, Julia Carnevale, P. Jonathan Li, Joseph Hiatt, Justin Saco, Paige Krystofinski, Han Li, Victoria Tobin, David N. Nguyen, Michael R. Lee, Amy L. Putnam, Andrea L. Ferris, Jeff W. Chen, Jean-Nicolas Schickel, Laurence Pellerin, David Carmody, Gorka Alkorta-Aranburu, Daniela del Gaudio, Hiroyuki Matsumoto, Montse Morell, Ying Mao, Min Cho, Rolen M. Quadros, Channabasavaiah B. Gurumurthy, Baz Smith, Michael Haugwitz, Stephen H. Hughes, Jonathan S. Weissman, Kathrin Schumann, Jonathan H. Esensten, Andrew P. May, Alan Ashworth, Gary M. Kupfer, Siri Atma W. Greeley, Rosa Bacchetta, Eric Meffre, Maria Grazia Roncarolo, Neil Romberg, Kevan C. Herold, Antoni Ribas, Manuel D. Leonetti, and Alexander Marson. Reprogramming human T cell function and specificity with non-viral genome targeting. *Nature*, 559(7714): 405–409, July 2018. ISSN 1476-4687. doi: 10.1038/s41586-018-0326-5. URL https://www.nature.com/articles/s41586-018-0326-5. Number: 7714 Publisher: Nature Publishing Group.

[31] Antoni Ribas, Omid Hamid, Adil Daud, F. Stephen Hodi, Jedd D. Wolchok, Richard Kefford, Anthony M. Joshua, Amita Patnaik, Wen-Jen Hwu, Jeffrey S. Weber, Tara C. Gangadhar, Peter Hersey, Roxana Dronca, Richard W. Joseph, Hassane Zarour, Bartosz Chmielowski, Donald P. Lawrence, Alain Algazi, Naiyer A. Rizvi, Brianna Hoffner, Christine Mateus, Kevin Gergich, Jill A. Lindia, Maxine Giannotti, Xiaoyun Nicole Li, Scot Ebbinghaus, S. Peter Kang, and Caroline Robert. Association of Pembrolizumab With Tumor Response and Survival Among Patients With Advanced Melanoma. *JAMA*, 315(15): 1600–1609, April 2016. ISSN 1538-3598. doi: 10.1001/jama.2016.4059.

[32] Ton N. Schumacher and Robert D. Schreiber. Neoantigens in cancer immunotherapy. *Science*, 348(6230):69–74, April 2015. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aaa4971. URL https://science.sciencemag.org/content/348/6230/69. Publisher: American Association for the Advancement of Science Section: Review.

[33] Paul C. Tumeh, Christina L. Harview, Jennifer H. Yearley, I. Peter Shintaku, Emma J. M. Taylor, Lidia Robert, Bartosz Chmielowski, Marko Spasic, Gina Henry, Voicu Ciobanu, Alisha N. West, Manuel Carmona, Christine Kivork, Elizabeth Seja, Grace Cherry, Antonio J. Gutierrez, Tristan R. Grogan, Christine Mateus, Gorana Tomasic, John A. Glaspy, Ryan O. Emerson, Harlan Robins, Robert H. Pierce, David A. Elashoff, Caroline Robert, and Antoni Ribas. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature*, 515(7528):568–571, November 2014. ISSN 1476-4687. doi: 10.1038/nature13954.

[34] Dieta Brandsma, Lukas Stalpers, Walter Taal, Peter Sminia, and Martin J. van den Bent. Clinical features, mechanisms, and management of pseudoprogression in malignant gliomas. *The Lancet. Oncology*, 9(5):453–461, May 2008. ISSN 1474-5488. doi: 10.1016/S1470-2045(08)70125-6.

[35] Eric Tran, Paul F. Robbins, Yong-Chen Lu, Todd D. Prickett, Jared J. Gartner, Li Jia, Anna Pasetto, Zhili Zheng, Satyajit Ray, Eric M. Groh, Isaac R. Kriley, and Steven A. Rosenberg. T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. *New England Journal of Medicine*, 375(23):2255–2262, December 2016. ISSN 0028-4793. doi: 10.1056/NEJMoa1609279. URL https://doi.org/10.1056/NEJMoa1609279. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa1609279.

[36] Erlend Strønen, Mireille Toebes, Sander Kelderman, Marit M. van Buuren, Weiwen Yang, Nienke van Rooij, Marco Donia, Maxi-Lu Böschen, Fridtjof Lund-Johansen, Johanna Olweus, and Ton N. Schumacher. Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science (New York, N.Y.)*, 352(6291):1337–1341, June 2016. ISSN 1095-9203. doi: 10.1126/science.aaf2288.

[37] Jennifer G. Abelin, Derin B. Keskin, Siranush Sarkizova, Christina R. Hartigan, Wandi Zhang, John Sidney, Jonathan Stevens, William Lane, Guang Lan Zhang, Thomas M. Eisenhaure, Karl R. Clauser, Nir Hacohen, Michael S. Rooney, Steven A. Carr, and Catherine J. Wu. Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction. *Immunity*, 46(2):315–326, February 2017. ISSN 1097-4180. doi: 10.1016/j.immuni.2017.02.007.

[38] Vinod P. Balachandran, Marta Łuksza, Julia N. Zhao, Vladimir Makarov, John Alec Moral, Romain Remark, Brian Herbst, Gokce Askan, Umesh Bhanot, Yasin Senbabaoglu, Daniel K. Wells, Charles Ian Ormsby Cary,

Olivera Grbovic-Huezo, Marc Attiyeh, Benjamin Medina, Jennifer Zhang, Jennifer Loo, Joseph Saglimbeni, Mohsen Abu-Akeel, Roberta Zappasodi, Nadeem Riaz, Martin Smoragiewicz, Z. Larkin Kelley, Olca Basturk, Mithat Gönen, Arnold J. Levine, Peter J. Allen, Douglas T. Fearon, Miriam Merad, Sacha Gnjatic, Christine A. Iacobuzio-Donahue, Jedd D. Wolchok, Ronald P. DeMatteo, Timothy A. Chan, Benjamin D. Greenbaum, Taha Merghoub, and Steven D. Leach. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature*, 551(7681):512–516, November 2017. ISSN 1476-4687. doi: 10.1038/nature24462. URL `https://www.nature.com/articles/nature24462`. Number: 7681 Publisher: Nature Publishing Group.

[39] Alena Gros, Maria R. Parkhurst, Eric Tran, Anna Pasetto, Paul F. Robbins, Sadia Ilyas, Todd D. Prickett, Jared J. Gartner, Jessica S. Crystal, Ilana M. Roberts, Kasia Trebska-McGowan, John R. Wunderlich, James C. Yang, and Steven A. Rosenberg. Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nature Medicine*, 22 (4):433–438, April 2016. ISSN 1546-170X. doi: 10.1038/nm.4051.

Appendix A

DNA OPTIMIZATION METHODS

Codon optimization of proteins for various applications in this work was performed using a Python script. Here, I will describe how the script generally behaves, highlighting the method by which codons are selected, and how user-defined criteria are considered.

Codon selection: Starting from an initial peptide or protein sequence, DNA fragments are created using codons that adhere to three criteria:

1. occur at high frequencies in the human species
2. maintain a specified GC percentage along user-defined base intervals
3. avoid incorporation of any restriction digest cut sites that are found at flanking regions of the product for use in downstream steps

Criteria 1: Codon selection. Fixed codon assignment for every amino acid is non-optimal for primer design (as detailed in [refer to neosub section of Appendix]), and also for DNA fragment design due to incompatibilities with manufacturer guidelines. Therefore, the approach I used was to probabilistically select codons for every amino acid by referencing a table of codon frequencies. Prior to selection, low-frequency codons were removed from the codon table, and frequencies were recalculated. For example, arginine's cumulative frequency array is `[.08, .26, .37, .57, .78, 1]` for the codons `['cgt', 'cgc', 'cga', 'cgg', 'aga', 'agg']`. `'cgt'` and `'cga'` occur at a low frequency (0.08 and 0.11, respectively), so those are removed, generating the frequency array `[.225, .475, .7375, 1]` for the codons `['cgc', 'cgg', 'aga', 'agg']`. For each amino acid in a given protein sequence, a random float value between 0 and 1 is generated, and the associated codon is selected according to the frequency table. This selection process occurs in a segmented manner user-defined by the peptide length for a given protein sequence.

Criteria 2: GC content. After a DNA fragment is generated for a segment, the GC content is calculated and checked for adherence to a user-defined range. If it

falls outside of that range, the codon selection process is looped (up to 1,000 cycles) until criteria is met. The decision to optimize codon selection and GC content in a segmented manner is crucial to avoid imbalances in GC content throughout the length of a protein sequence. For example, it is theoretically possible to achieve 50% GC content in two extremes. In one scenario, without a segmented approach, the initial half of the protein could consist of 75% GC content while the latter half consists of 25% GC content, averaging out to 50% overall GC content. In the alternative scenario, with a segmented approach to codon selection and content optimization, the entire length of the protein for any given window of bases (e.g. 30-60 bp) would approximate 50% GC content. The second scenario is desired to overcome manufacturing hurdles and potential hairpin formation.

Criteria 3: Sequence avoidance. A table of most available restriction digest enzymes and their recognition sites serves as a reference. Enzymes which are crucial to downstream steps, and therefore whose recognition sites must not be present in the designed DNA fragment, may be listed within an array. The designed DNA sequence is checked for overlap with entries within this array using the reference table, and codon selection is repeated if any recognition sites are present.

Examples demonstrating codon optimization and experimental execution of assembly these fragments may be found in **Figures A.1A & A.2**, respectively. In **Figure A.1A**, a linear map of an example Class I SCT design as reported in subsection 2.1.2 is shown, and below it is a plot showing the calculated GC content of each 30 bp window for its DNA sequence. The results of codon optimization of the HLA allele for a GC content of 40-60% are shown by the region of the plot encapsulating the optimized segment (between the dashed vertical lines). The contrast of larger GC content variance in non-optimized regions of the SCT map demonstrate that the script is indeed capable of tuning GC content to user-defined ranges. Experimentally, this optimization translates to faster turnaround times for ordering and constructing dsDNA structures. For an example set of twelve HLA alleles (**Fig. A.2**), the f2 fragment (encoding the HLA) for each unique HLA was ordered, amplified by PCR, and Gibson assembled with a universal f1 fragment (encoding the upstream half of the SCT). The Gibson products were subsequently gel-purified, PCR-amplified, double-digested into pcDNA3.1 vector, and sequenced to validate integrity. These template plasmids were later used in constructing a KRAS library to show that they could lead to expressed SCT proteins when re-encoded with compatible peptides (**Fig. 4.16**).

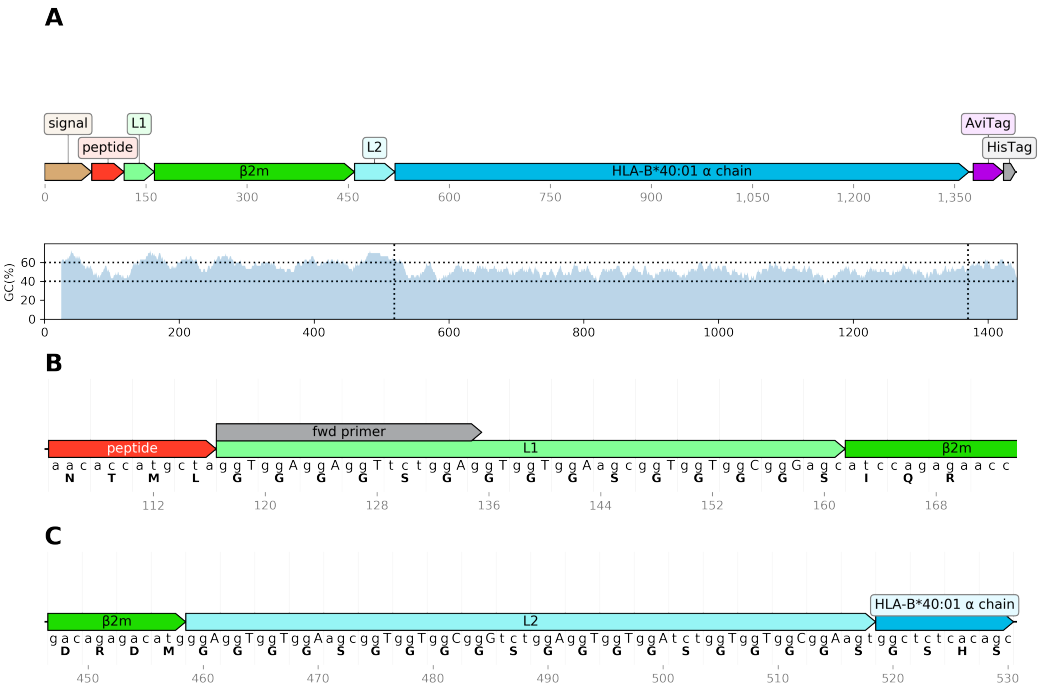


Figure A.1: Codon optimizations to Class I SCT design
A. Linear map of Class I SCT design (top) and GC content calculations (bottom) performed on a 30 bp window. The HLA region (darker blue in top diagram, and encapsulated by vertical dashed lines in bottom plot) was designed using codon optimization script. B. Detailed view of L1 segment, highlighting codon selection of GGGGS motifs. C. Detailed view of L2 segment, highlighting codon selection of GGGGS motifs. In (B) and (C), the third base of each glycine residue is emphasized in upper-case.

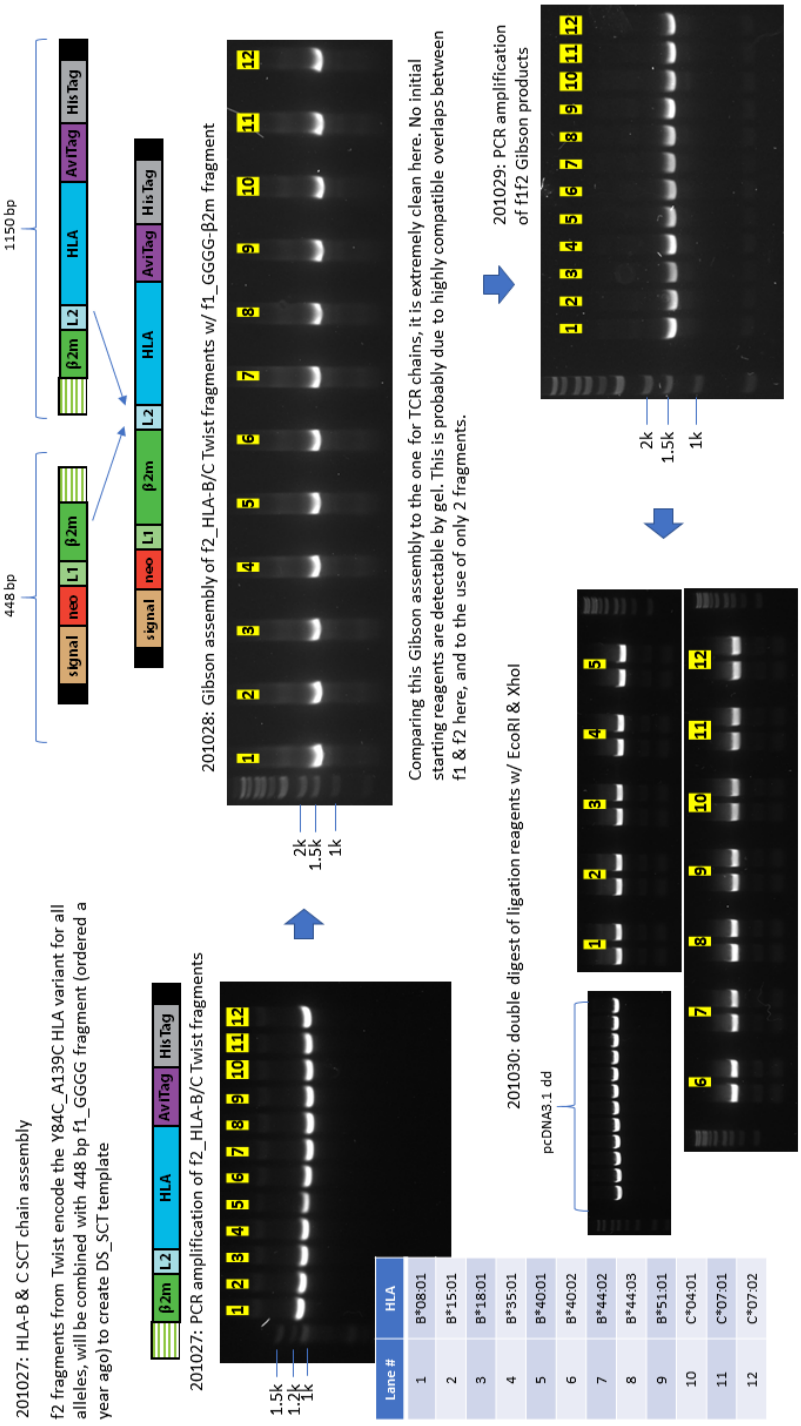


Figure A.2: Demonstration of Class I SCT Gibson assembly
DNA fragments encoding the latter half of Class I SCTs are purchased and PCR-amplified (top left gel), and assembled together with initial half of Class I SCTs (top right diagram and gel). These fragments are gel-purified, PCR-amplified (bottom right gel), and then double-digested for ligation into pcDNA3.1 vector (bottom left gel). Numbers above each lane correspond to the HLA identity of each fragment, as referenced by the table (bottom left).

Appendix B

PCR PRIMER OPTIMIZATION

B.1 single-codon assignment

Given the objective of designing SCT libraries encompassing large numbers of peptides, initial attempts to streamline the primer production process for extension PCR used Python scripts to automate generation of peptide-encoded primers. These primers consist of two regions (**Fig. B.1A**): 1) a constant binding site to the plasmid template (in the case of a peptide-encoded reverse primer, the constant binding site is found on the IFNα2 signal, upstream of the peptide region on the plasmid), and 2) a variable DNA region consisting of codons encoding the desired peptide. The constant region of the primer encodes the initial 19 bases of the L1 region and required careful codon selection so as to not significantly overlap with any of the GGGGS motifs found in the rest of L1 or in the L2 region. This was an important consideration, as all codons encoding glycine residues are highly similar (ggx), and L1 and L2 together contain up to seven such repeats of this motif. Off-target primer binding was avoided by ensuring that each GGGGS motif in L1 and L2 contained a different combination glycine codons (this is highlighted in **Figures A.1B & A.1C**, for L1 and L2 regions, respectively, by upper-case letters denoting the third base of each glycine codon). The variable region of the primer was designed according to automatic single codon assignment for any given amino acid, selected based on the highest frequency in the human species. Primer sequences generated for any given peptide sequence were therefore pre-determined, and PCR experiments for small library sizes (e.g. 10-100) generally produced electrophoresis product bands in approximately 70% of samples, reflecting the efficiency of appropriate primer binding and initiation of PCR.

Subsequent repeat PCR experiments to obtain bands across all elements of a library required at least two additional rounds of PCR, in which only primers which failed to produce an initial product band in electrophoresis were used, and eventually could result in a visible gel product. After re-ligating the dsDNA product into a plasmid and Sanger sequencing, errors were often detected in the peptide region, generating either multi-base deletions (possible hairpin within primer's peptide-encoded region) or failed peptide replacement (**Fig. B.1C**). As seen in **Figure B.2A-D**, for an especially

large library, such as the 240-element p53 A*02:01 SCT library, using primers designed in a such a manner proved to be extremely laborious, requiring up to four rounds of PCR until all reactions generated a band. Furthermore, repeated PCR rounds to obtain the correct sequence would not necessarily be successful even if PCR product bands were observed and plasmids were isolated after transformation due to the unique failure mechanism behind these primers (**Fig. B.1C**). The difficulty in arriving at a product arose from issues inherent to each primer's unique free energy structure. NUPACK software predicted hairpin structures for many of these primers at 60 °C (**Fig. B.1D**). Therefore, even though bands could be obtained for difficult primers after many PCR attempts, it does not necessarily mean that the sequence obtained was correct from the plasmid, and consequently required selection of a large number of colonies for culture and sequencing.

B.2 probabilistic codon assignment & mean free energy optimization

Based on this finding, a revised script was written to heuristically obtain primers that could avoid the structural trappings of hairpin structures which prevent proper PCR amplification. This Python script utilizes NUPACK modules to design ssDNA strands whose mean free energies (MFE) at 60 °C result in linearized conformations (**Fig. B.1D**). Briefly, the primer generation process was no longer fixed, in the sense that each amino acid was no longer assigned to a specific codon. Instead, codons were selected in a probabilistic manner according to a revised frequency table for the human species (rare codons were eliminated and probabilities re-calculated, in a manner similar to methods discussed in **Appendix A**). The variable region of the peptide was constructed from these codons, linked to the constant binding site, and then its MFE was calculated for the target structure of a linearized ssDNA strand. This calculation was conducted approximately 1,000 times per peptide through randomized codon selection per round, requiring about 45 seconds per peptide, and selects the result with the lowest MFE to be the final output. The decision to run 1,000 optimization cycles was made after showing that designed primers under this parameter were sufficiently capable of generating PCR products consistently (**Fig. B.3**). In theory, the ideal approach for finding the best primer for any peptide would be to scan the entire space of potential codon combinations, calculating MFE of each structure. However, this approach is impractical for certain peptides which contain highly degenerate amino acids. For example, the peptide FLTENLLLYI has 2 x 6 x 4 x 2 x 2 x 6 x 6 x 6 x 2 x 3 = 248,832 potential codon combinations

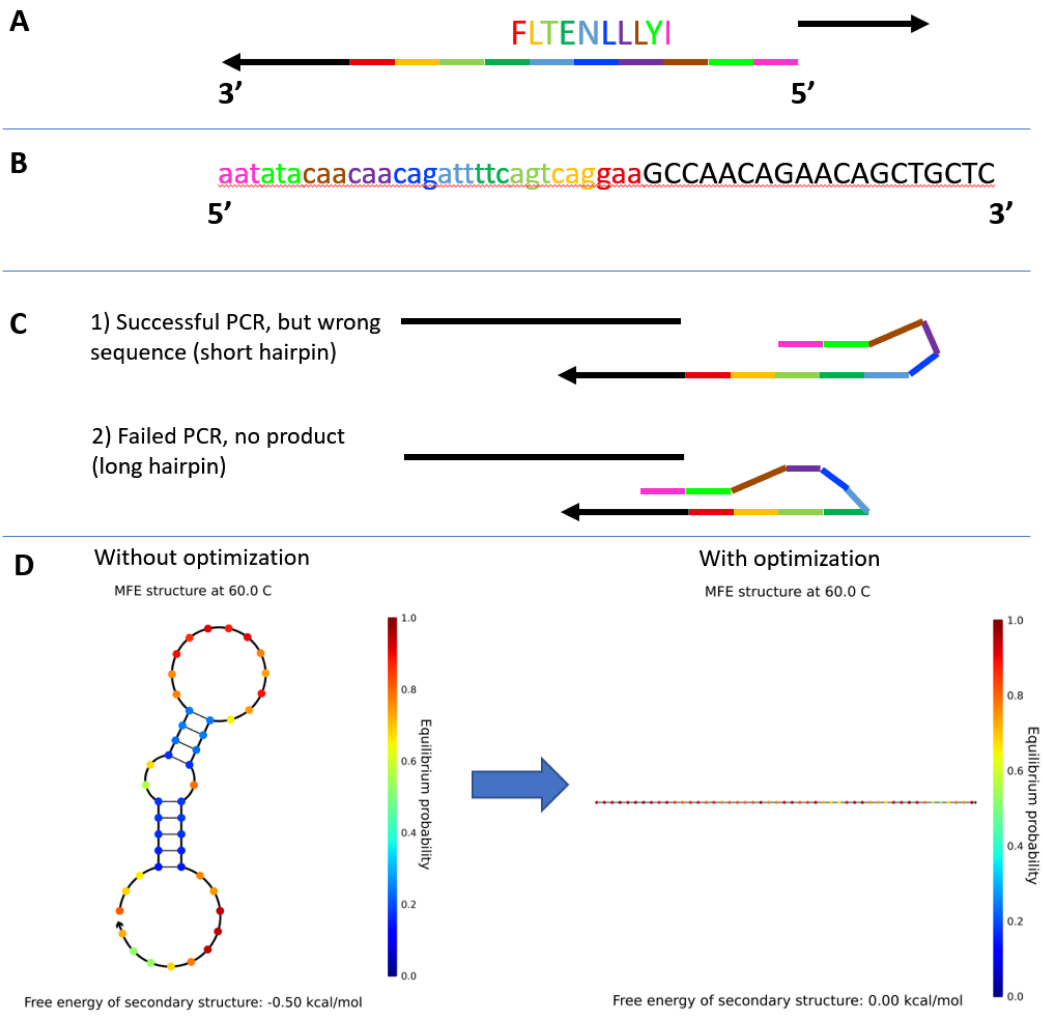


Figure B.1: Primer optimization
A. Setup of primers for PCR peptide substitution. A reverse primer encodes the entirety of the peptide sequence (FLTENLLLYI), while the universal forward primer binds to L1 region of Class I SCT. B. 5'-to-3' orientation of the reverse primer, with codons color-coded according to amino acids in (A). C. Examples of PCR failure mechanisms when primers are un-optimized. D. Comparison of MFE structures for primers that are not optimized (left, using single codon assignment method) versus those which are optimized (right, using randomized codon assignment and NUPACK MFE calculations).

(after eliminating low-frequency codons). This is a 250-fold increase in the number of MFE calculations required, and would require approximately three hours per peptide, versus the 45-second, 1,000 cycle method using randomized combinations. Primer optimization for a 100-element library would therefore require weeks versus an overnight computation session.

The finalized script was designed to accommodate any desired constant binding

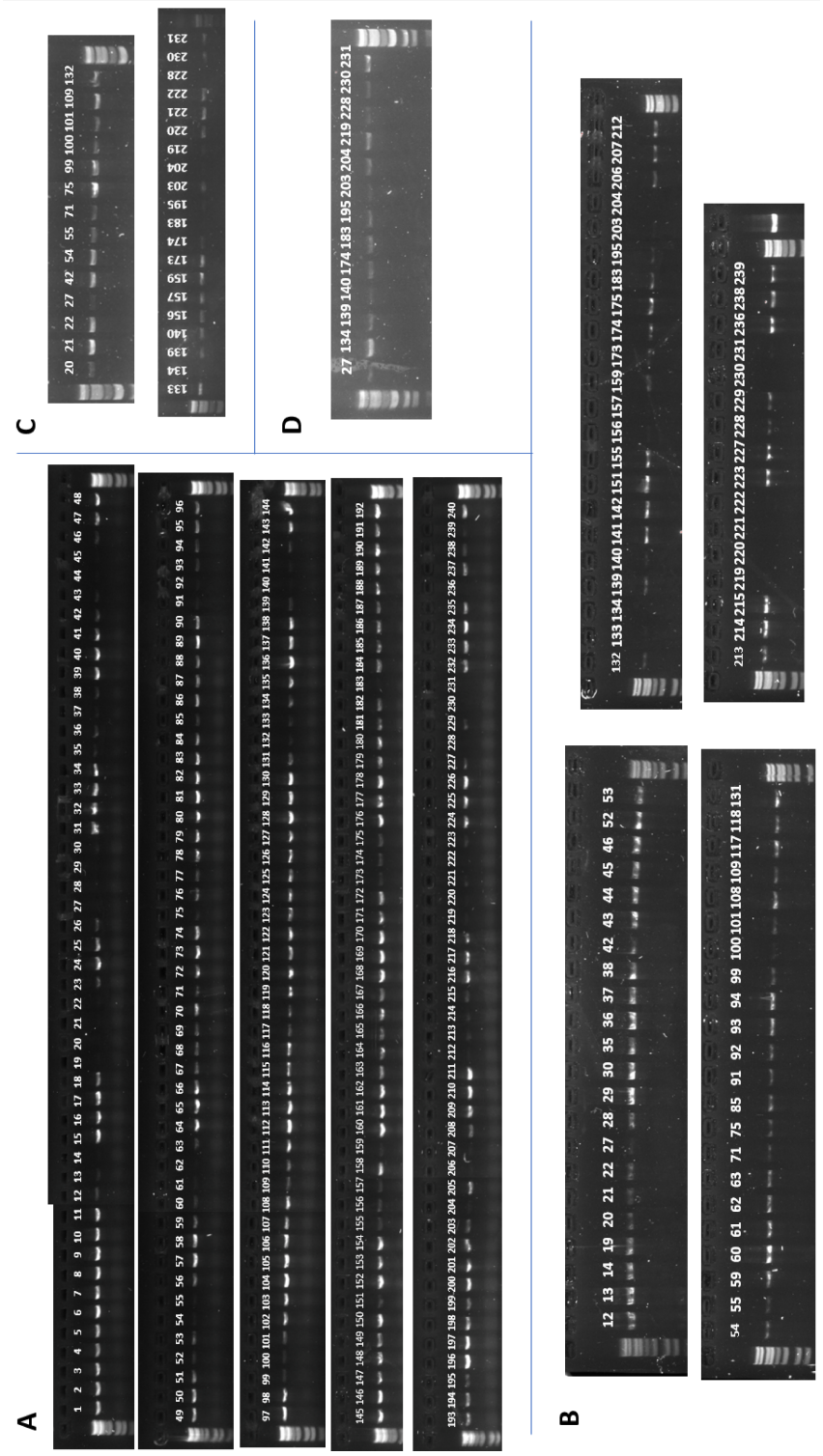


Figure B.2: Neoantigen substitution PCR for p53 A*02:01 library
A 240-element primer library encoding 9-11mer peptides for various p53 hotspot mutations was utilized for inverse PCR. The neoantigen-substituted PCR product bands are depicted in electrophoretic gels, numbered according to peptide ID of the library (**Table N.1-N.2**). After the first attempt with the entire library (A), failed PCR reactions from the prior round are repeated (B-D). The final library is obtained after approximately four rounds. This experiment was conducted using single-codon assignment for each amino acid during primer design.

site provided by the user, and to allow for peptide-encoded primers to be designed in forward or reverse direction. Therefore, its applications extend beyond peptide substitution for SCTs, and should be appropriate for usage in any site-directed mutagenesis experiments where a series of short peptides must be substituted or added into a protein-encoding DNA sequence. As seen in **Figure B.3**, for large libraries of approximately 900 elements, PCR experiments achieved visible product in the first attempt for any given primer pair. Subsequent transformation and sequencing of the plasmids produced approximately 80% success rate within the first round. Additional rounds of sequencing for repeats of the wrong plasmids maintained 80% success rate, such that typically three rounds of sequencing was sufficient to achieve a full plasmid library with correctly sequenced insertions. The main source of error detected in these plasmids occurred at the 3'-terminus of the inserted peptide, typically a single base deletion. Most likely, this sort of deletion occurred during the KL-D step (kinase, ligase, and DpnI treatment to re-circularize plasmid and eliminate initial template plasmid), when the final base of either end of the dsDNA PCR product was phosphorylated and ligated. This single-base deletion was sometimes detected in the first base of the non-peptide side of the dsDNA as well, indicating that the error source arose from the KL-D reaction, and not from a hairpinned peptide-encoded primer during PCR extension. To date, after over 2,000 PCR experiments using peptide-encoded primers designed with probabilistic codon selection and NUPACK-facilitated structural optimization, we have yet to observe a case where a PCR product band failed to form and where subsequent sequencing repeatedly failed to produce the correct sequence. From a practical standpoint, to generate an SCT library of approximately 200-300 plasmids, the total time required, starting from PCR to correctly sequenced plasmid for every library element, is about 3-5 weeks using the original method, while the new script requires only 1-1.5 weeks. It is important to emphasize that not only was there a primary benefit achieved from just obtaining PCR product bands with consistency (compare **Figure B.2** with **Figure B.3**), but also that Sanger sequencing results were significantly improved to drastically reduce turnaround times. The script has been converted into a user-friendly executable file, requiring the user to only provide a list of peptides within a `csv` file and input of the constant binding site. Full documentation of the script is provided in the Jupyter notebook and Python scripts.

B.3 alternative methods

The above two sections chronologically detail how I first started with a rather simplistic codon assignment approach and later attempted to heuristically find the most linearized primer as determined by random selection of codons. These two approaches made use of the inverse PCR method, whereby a single primer per peptide needed to be purchased, while the other primer facilitating PCR may be constant regardless of peptide identity (**Fig. 2.20A**. The only case where the constant primer must be changed is when a different SCT template is selected, particularly one in which the identity of the L1 region is changed (for instance, GGGGS linker to GCGGS linker, thus altering the primer binding site). In this manner then, the inverse PCR approach significantly reduces the cost of primer reagents. For n peptide elements and m unique templates (on the basis of different L1 identities), one requires only $n + m$ total primers to be purchased. This stands in contrast to other methods explored by others. For example, a common alternative approach to substitute short DNA regions is to utilize two primers to encode the desired insert bases, and then to insert the PCR product of the two primers either by Gibson assembly or restriction enzyme ligation (**Fig. 2.20B**) into the template SCT fragment. In the context of Class I SCT production, where the impact of the L1 linker on SCT stability is still being explored (**Section 2.1.4.1**) and has been shown to influence TCR binding efficiency (**Section 2.1.4.2**), such a method is not optimal, as it requires every peptide/L1 combination to have a unique second primer encoding a different L1, resulting in $n + nm$ total primers. Furthermore, these methods are generally more tedious for larger libraries, as they require gel purification of the dsDNA PCR product prior to assembly/digest/ligation into the template plasmid. The primary advantages of this approach, on the other hand, include the fact that they can accommodate substitution of longer peptides, due to the fact that the peptide information is split across two primers, enabling a longer peptide to be encoded within two primers instead of one primers, when primers must be below 60 bp in length to reduce costs. One additional advantage of this approach is that the product plasmid's sequencing result tends to more frequently show correct peptide substitution. With these advantages in mind, this method was more appropriate for the Class II SCT-T design (**Fig. 2.20B**), due to the fact that Class II pMHCs make use of longer peptides that would be cumbersome to entirely encode in a single primer, and that the identity of the peptide linker is fixed for Class II SCTs (at least so far, as we haven't yet begun to explore TCR interactions), therefore obviating the need to account for $n + nm$ complexity. A natural extension of this alternative approach for SCT-T designs then

is to incorporate the same optimizations developed in **Appendix A**. SCT-T primers will be longer, and surely will also encounter similar hairpin issues hampering downstream steps as observed for Class I SCT experiments. Implementation of MFE optimization and randomized codon assignments for the Class II SCT system is currently in-progress.

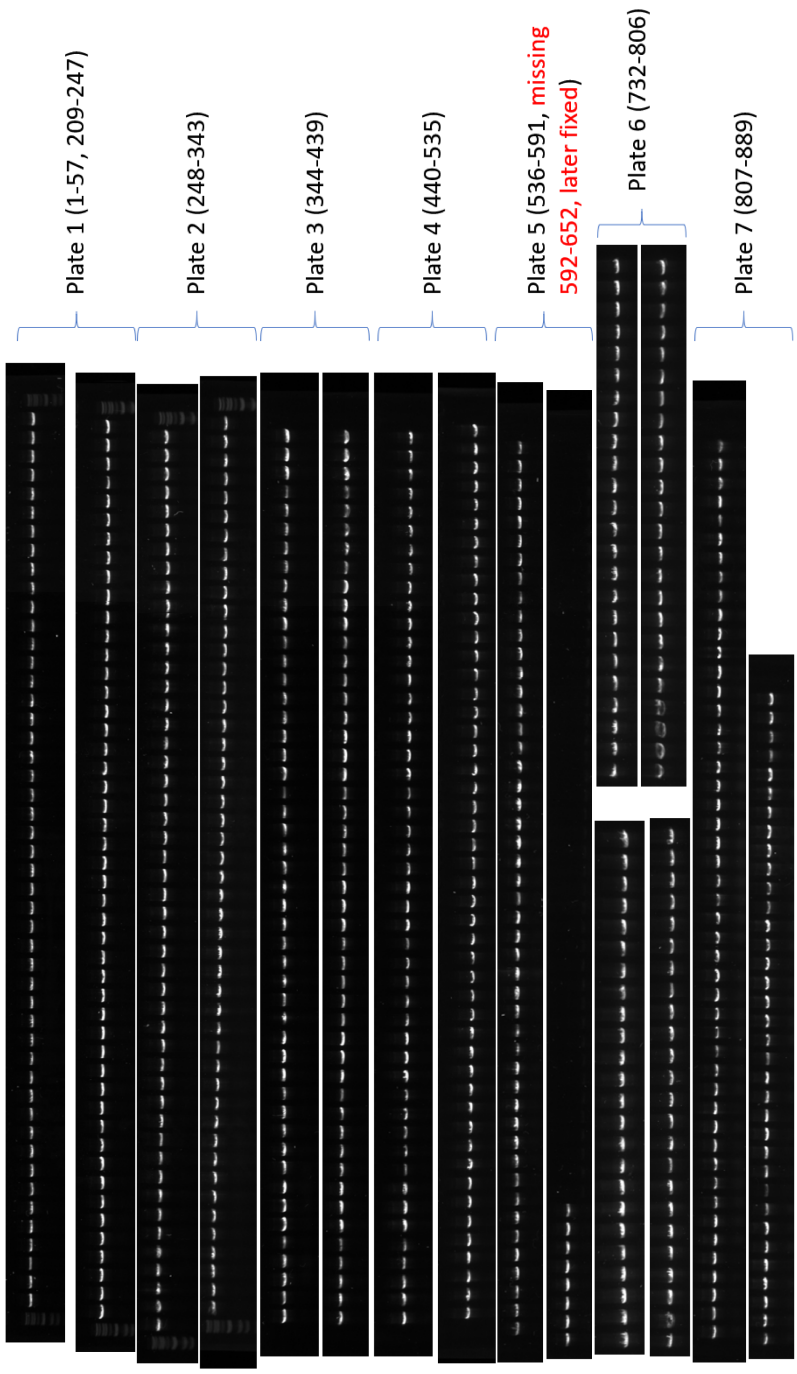


Figure B.3: Antigen substitution PCR for SARS-CoV-2 A*02:01 proteome library An 889-element primer library encoding 9-10mer peptides predicted by netMHC4.0 to favorably bind to A*02:01 was utilized for inverse PCR. The antigen-substituted PCR product bands are depicted in electrophoretic gels (numbers correspond to peptide ID of proteome library in **Table L.1-L.4**), showing that all reactions generate product after one single attempt. Peptides #592-652 were not included in this experiment due to manufacturer mistake during primer production. This experiment was conducted using probabilistic codon assignment and NUPACK MFE optimization for each peptide during primer design.

Appendix C

HIGH-THROUGHPUT METHODS FOR SCT PLASMID PRODUCTION

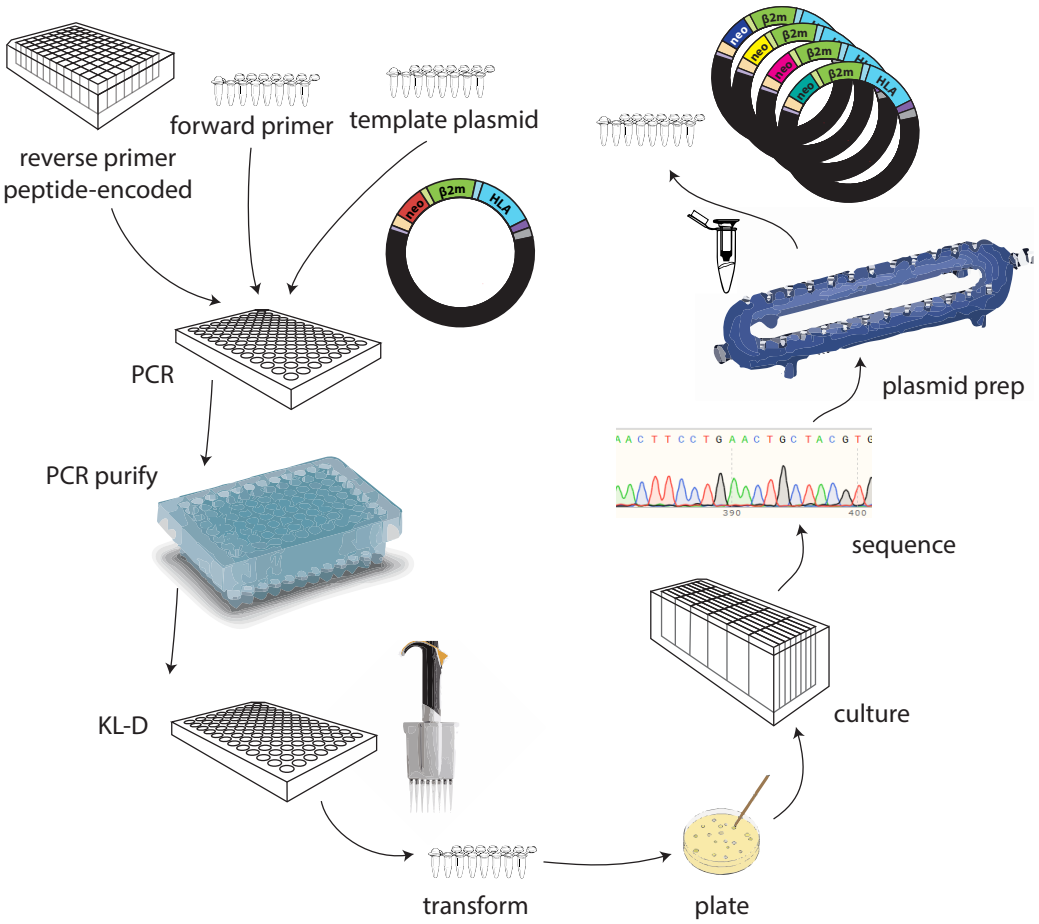


Figure C.1: Plasmid production workflow
Schematic of plasmid preparation workflow, beginning from PCR reagents (top left) and ending with purified, sequence-verified SCT plasmid constructs (top right).

C.1 Transformation

Traditionally, generation of a plasmid library requires transformation of a circularized PCR product, selecting colonies to culture, plasmid purification from the culture, and submission of the plasmid for Sanger sequencing. Given the large size of our library and the need to select multiple colonies per plasmid (due to error rates

associated with site-directed mutagenesis), there were a number of optimizations required to our workflow to accommodate this platform into an academic setting, where automation is generally not common. Our general workflow for plasmid production is shown in [FIGURE]. In general, most processes in plasmid production (e.g. PCR, kinase/ligase/DpnI (KL-D) treatment, etc.) have been commercially streamlined with the use of 96-well plates. The steps following plasmid production, however, where they must next be transformed into TOP10 chemically competent cells, required a significant change to the workflow to enable high-throughput processing. Traditionally, TOP10 chemically competent cells are transformed in single 1.7 ml tubes, and are difficult to work in a large library format due to the delicate time-sensitive steps that would be cumbersome to perform on a per-sample basis. To facilitate high-throughput transformation, heat shock, and immediate transfer of samples onto ice, 3D-printed racks were designed to reduce down-time during each step. TOP10 cells were aliquoted into 8-tube PCR strips to be compatible with this rack design, and, coupled with the use of multi-channel pipettes, allow for the user to achieve transformation and plating of hundreds of samples in one day. Transformed cultures were plated onto agarose plates overnight. Subsequently, colonies were picked and cultured in 48-well plate format overnight in 3 ml media.

C.2 Culture Selection & Submission

The submission of purified plasmids for sequencing when working with error-prone libraries can be expensive if there are many elements in the library, as it will require excessive purification rounds per plasmid and excessive sequencing costs for subsequent rounds. To reduce the labor required for sample prep, we opted to submit bacterial cultures instead for rolling circle amplification and subsequent sequencing. By submitting only an aliquot of the bacterial culture first and freezing down the pellet, the user can wait to receive sequencing results before deciding whether to proceed with plasmid prep of the pellet. Therefore, obtaining correct plasmids through multiple rounds of sequencing can minimize costs, where in each round, only one colony per unique plasmid is cultured, with subsequent rounds filtering down the colony set to include only plasmids which are still incorrect. Plasmid prep therefore is only performed for pellets whose cultures have been sequence-verified.

C.3 Plasmid Sequencing

A third bottleneck in the plasmid production process is sequence verification. Typically, upon receiving DNA chromatographs from a sequencing source, the user must use DNA sequencing software (e.g. Snapgene) to first generate a template against which their sequencing results may be compared. This process is tedious and time-consuming to conduct for large libraries (typically requires 4-6 hours of labor for a 200 element library), and due to its repetitive nature, most likely will result in occasional false positive or negative declarations. To significantly reduce the labor needed for this process, Python scripts have been written to automate DNA template sequence generation, import of chromatography data, and assessment of the peptide-substituted region within any SCT-encoded DNA sequence, requiring less than one minute to conduct all steps. These scripts require the user to provide a reference `csv` file of the plasmids with peptide information, and the sequencing `abs` files from commercial sources (e.g. MCLAB).

C.4 Plasmid Purification

Once the correct plasmid copies have been determined, the next step is to perform plasmid purification. In our experience, we had found commercial 96-well DNA purification kits to be quite variable in terms of yield and volume retention following EB buffer addition. For steps in this workflow where final DNA yield and volume do not matter (e.g. PCR purification of the antigen-substituted SCT template), the use of Qiagen 96-well PCR purification plates was appropriate. However, adaptation of these kits for subsequent plasmid purification generated mediocre results. Significant variations in the expected post-purification volume, as well as large variances in measured DNA absorbance, hindered ease of plasmid normalization for downstream transfection steps. Therefore, plasmid purification was performed using traditional Qiagen MiniPrep tubes, whereby single samples were individually loaded into each tube for batch processing using a 24-sample vacuum manifold. Although this approach was much more tedious, requiring more attention to sample labeling and arrangement, yields were always high enough after purification (> 200 ng/µl) to facilitate re-dilution to normalized concentration, and post-purification volumes were also consistently at approximately 48 µl. Re-visiting this bottleneck to identify a protocol that makes use of high-throughput plates for plasmid purification while also enabling consistent yield retention should be a priority in future attempts to optimize this process.

Appendix D

HIGH-THROUGHPUT METHODS FOR SCT PROTEIN PRODUCTION

The adaptation of conventional transfection methods for high-throughput sample processing a necessary component in streamlining SCT library preparation. Downstream steps, such as sample filtration, concentration, and purification, also required protocol changes to reduce turnaround times for practicality. In this section, I will briefly list some of the major changes to each of these steps from traditional transfection workflows. An accompanying diagram of the workflow is seen in **Figure D.1**.

D.1 Transfection

- Use of 96-well plates and multi-channel pipettes across all transfection reagent preparation steps.
- Final mixtures of plasmid with Expi293 cells may be prepared in 24-well or 96-well plates, incubated using multi-plate shakers or single-plate shakers, respectively.

D.2 Filtration and concentration

The filtration of supernatant collected from transfection medium is required prior to downstream sample concentration steps. Toward this end, there are commercial reagents available, such as 96-well plates equipped with filters, to facilitate the process. However, our experience has shown that these plates often lead to inconsistent volume pass-through during centrifugation, and oftentimes will causing clogging and require pre-filter fluid transfer into another plate to ensure complete filtration through an unclogged filter. Furthermore, at medium-scale transfection capacities (1.25 ml transfection in a 24-well plate), commercial plates are currently equipped to handle only 1 ml maximum, and so will require repeat use to fully filter sample. Therefore, we had to resort to using traditional Luer-Lok syringe tips equipped with filters in order to filter supernatant into Amicon concentration tubes. The syringe tips, although stated to be 1 ml capacity, are capable of handling up to 1.25 ml,

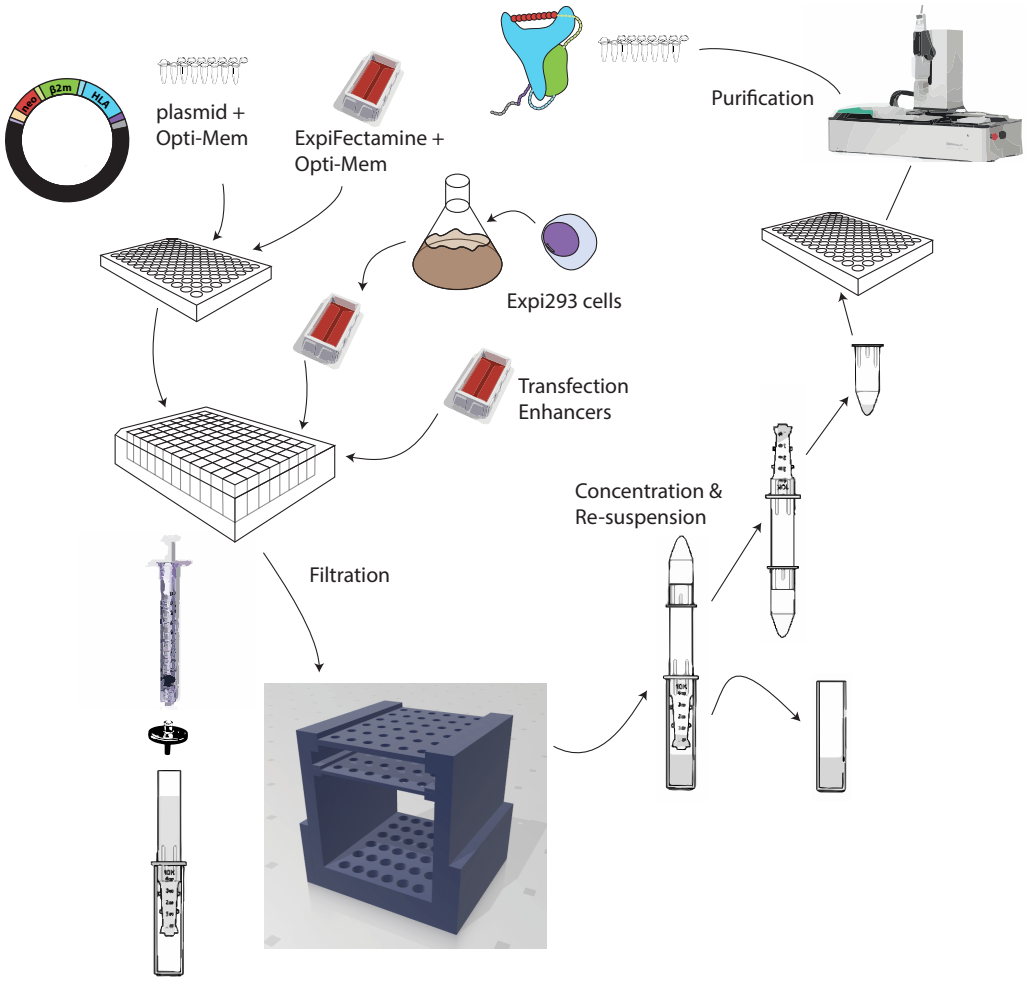


Figure D.1: SCT protein production workflow
SCT-encoded plasmid reagents (top left) are transfected into Expi293 cells in 24 or 96-well plate format. The secreted proteins are collected and individual supernatant is filtered with the assistance of 3D-printed racks designed to hold 1 ml, filter-equipped syringes. The filtered supernatant is concentrated down to 200 µl for the SCT biotinylation reaction. Finally, the SCTs are HisTag-purified with the assistance of a 96-sample Phynexus MEA-2 automated system..

and due to the larger design of syringe-compatible filters (versus plate filters), they generally will not clog. One demerit of using these syringe tips is that they are generally low throughput, allowing for only one sample to be manually processed at a time due to syringe setup and filtration. This was overcome by designing a 3D-printed syringe rack, where the 3-component setup (as depicted in bottom left of [FIGURE]) may be securely attached (syringe+filter attached to upper components of the removable top piece, and Amicon concentration tube inserted into lower rack). With the assistance of a multichannel, width-adjustable pipette (e.g. Integra

6-channel ViaFlow), transfection supernatant may be directly pipetted from 24-well plates into each syringe and immediately filtered into the receiving Amicon tubes. This set-up allows for up to 36 samples to be filtered per rack, with no volume loss to retain maximum protein yield.

D.3 Purification

Traditional approaches to purify HisTag proteins typically involve the use of an FPLC system equipped with a nickel resin column. Such systems are usually appropriate for collecting a large amount of a single type of protein, whereas our workflow and objectives call for a system that enables purification of small amounts of many different proteins. Therefore, we adopted the PhyNexus MEA 2 HisTag purification system. SCTs that have been concentrated in Amicon tubes may be directly transferred into 96-well plates on the MEA 2 system. Up to twelve samples at a time may be purified at once, requiring 16-18 minutes per cycle. The purified samples may be directly stored in the eluted buffer, or re-suspended into PBS/glycerol mixtures with the use of Zeba 96-well spin desalting plates.

Appendix E

SDS-PAGE ANALYSIS

Here I explain the design of a Python-based script to enable gel image analysis for DNA or protein gels. It utilizes several Python packages to facilitate gel band detection, intensity measurement, and standardization against a control lane. These scripts were designed with the intent of streamlining protein expression yield comparison.

E.1 Methods

Existing software for gel analysis is either quite outdated with cumbersome graphical user interfaces or is lacking in parameter flexibility to allow for broad applicability of gel analysis. For example, ImageJ allows for the user to crop areas of interest to identify one band per lane. However, the user is forced to use the same crop area for each subsequent band. This design choice renders sample selection difficult when lanes are too close together (causing crop boxes to overlap and acquire partial gel bands from other lanes in the edge of the box) or when the protein yield in some lanes is so high that it overstretches into the boundaries of nearby lanes. Furthermore, the downstream processing of the data after band selection is cumbersome and difficult to reproduce, as it requires the additional steps of measuring integral of each band peak based upon manually outlining the peak widths. In another instance, other software for gel band detection requires the user to select a large rectangular crop area, in which the bands of interest across all lanes are encapsulated within one selection. This design choice is also non-ideal, as the user must either work with proteins all at approximately the same mass (in order for a crop box to fit all of them), or the user must select a large range of the gel image, essentially the entire area of the gel, to accommodate multiple proteins of variable masses. Furthermore, this design choice will lead to band detection of undesired proteins as well due to their inclusion in the selected area, and if gel detection is based upon normalized thresholds, detection of low-expression protein among undesired, but highly expressing proteins may be problematic.

To address this issue, a Python script was written to facilitate capture and identification of only desired bands from a gel. This script was designed with the purpose of

enabling one to compare relative expression of hundreds of protein samples across multiple gels. To achieve this goal, two important parameters must be experimentally satisfied before running the script.

First, it is strongly recommended to use Bio-Rad Stain-Free gels for sample loading and image acquisition. The use of UV-sensitive, trihalo compounds to measure protein concentration is a highly reproducible process compared to the traditional Coommassie staining methods. The degree of Coommassie staining is typically inconsistent across samples, where protein band intensity can significantly change based on small variations in staining and washing times, stain concentrations, and microwave times, requiring the user to strictly adhere to standardized timings that are unsuitable or unwieldy when preparing many gels simultaneously. Stain-free gels, on the other hand, are designed to allow for fast image acquisition using Bio-Rad imaging machines immediately after electrophoresis is complete. The imaging process is typically conducted in less than one minute, and gel activation and exposure times are automated by the image acquisition software.

Secondly, a key step for stain-free imaging is to use a pure protein standard of known mass and concentration as a reference point for measuring sample yields. In our case, we use a purified single-chain trimer compound representing a peptide-MHC complex of approximately 52 kDa. The inclusion of this sample in one lane of every gel allows for acquired bands from other samples to be normalized against the band intensity of this control, enabling the user to back-calculate expected yield/concentration of each sample in their gel. When the control sample is applied in the same manner to all subsequent gels, all samples are essentially normalized to the same control protein, and the use of Stain-free gels essentially removes all background noise. We have utilized this script on over 200 gels to measure protein expression/yield of over 1,000 SCT proteins.

E.2 Results & Discussion

E.2.1 Object detection thresholds

One important consideration during design of the script was the identification of a compatible thresholding technique for object detection. Using a sample gel with protein bands across a range of concentrations, we focused on the quality of band detection of the lower concentration samples by various threshold techniques. These concentrations represent biologically reasonable levels of protein expression for

SCTs, and therefore required careful imaging calibration for accurate measurement. Among the threshold techniques tested, several either detected too much noise in the background or were unable to accurately capture a region representative of the true band size. Threshold techniques otsu, mean, and triangle were the best candidates. The latter two appeared to overestimate the boundaries of the band regions, especially for those which are at higher concentrations. Therefore, otsu was selected as our thresholding technique of choice for use in the script.

E.2.2 Band intensity correlation with protein concentration

In order to demonstrate that measured band intensities correlate with the true protein concentration of each sample, A*02:01 WT1 SCTs were expressed in Expi293 cells. After four days, the supernatant, containing secreted SCTs, was collected, re-suspended into PBS solution, and concentration was measured using a Nanodrop instrument. Using the absorbance value, aliquots of the SCTs were diluted down into various concentrations across a range representative of biologically-relevant possible levels of SCT expression. This range encapsulates a 55-fold concentration difference, from approximately 0.35 to 20 μM. A sample from each aliquot was then run using Stain-Free gels, imaged, and protein band intensity was calculated using our script. As seen in [FIGURE], there is a 99% correlation between normalized levels of absorbance (as measured by Nanodrop) and band intensity (as measured by script). This indicates that our script is a suitable substitute to traditional Nanodrop instruments for measurement of protein concentration. Whereas the Nanodrop approach requires manual pipeting and measurements are performed in lower throughput (typically single or up to eight samples), band intensity measurement is entirely digital, simply requiring selection of bands by the user, and batch measurements are calculated immediately. Furthermore, the Nanodrop approach typically will be very sensitive to measurement of proteins directly from cell media solutions such as RPMI (due to laser interference by media components), necessitating tedious re-suspension and concentration of the sample into a suitable buffer (e.g. PBS) prior to use. On the other hand, automated measurements of protein band intensity allow for measurement directly using an aliquot of the supernatant media, as it has no interference with image quality in SDS-PAGE. This allows for the user to obtain an immediate concentration measurement directly from the transfection culture without post-processing steps (e.g. re-suspension, concentration, purification) that may distort the actual biological yield of the protein.

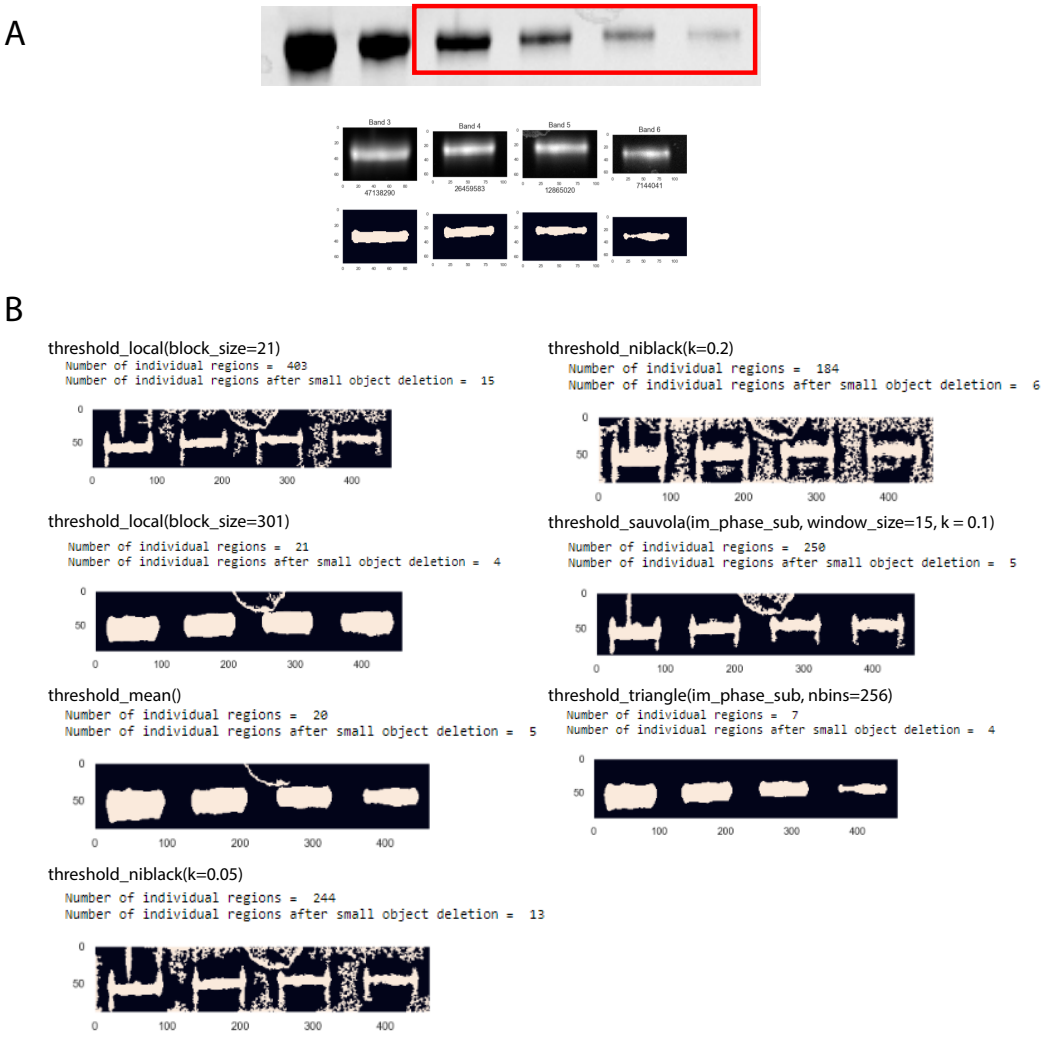


Figure E.1: Band detection optimization
A. SDS-PAGE image of titrated WT1 SCT. Bands within red box (top) serve as reference bands for subsequent intensity calculations. (Middle) Reference bands manually cropped using script. (Bottom) Identification of band areas from references above, using threshold_otsu method. Integers below the middle row represent the relative calculated intensity of each identified band (bottom). B. Band area identification results using various threshold techniques against the reference bands in (A).

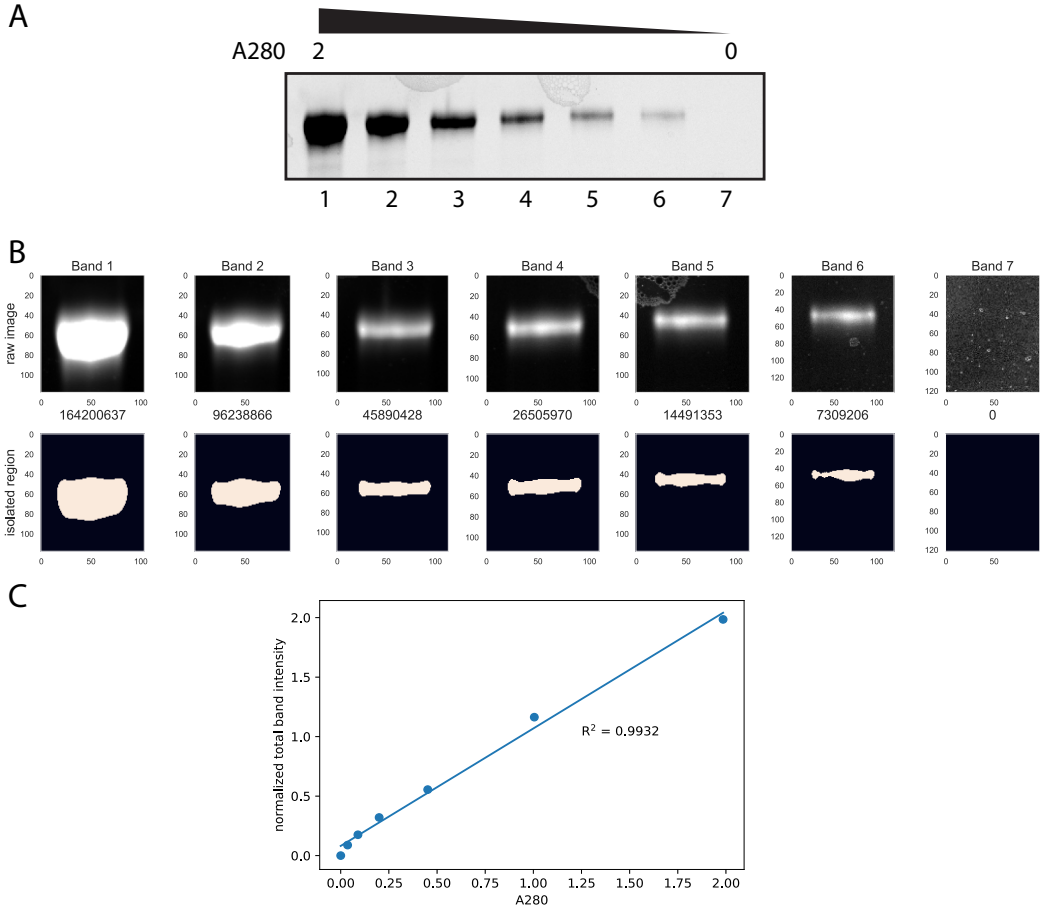


Figure E.2: Protein band intensities are highly correlated with actual concentration A. SDS-PAGE image of titrated WT1 SCTs, titrated across a range of approximately 2.0 to 0 absorbance units (280 nm). B. Manually cropped areas of each band from reference gel (top). Isolated regions of bands identified by script (bottom). Integers below the top row represent the relative calculated intensity of each identified band (bottom). C. Scatterplot showing correlation of A280 values and calculated band intensity from SDS-PAGE for each reference band.

Appendix F

PEPTIDE SOURCES FOR SCT LIBRARY OPTIMIZATION

ID	peptide	protein	PubMed ref ID
1	YMLDLQPET	E7 (PPV-9)	29669936
2	YMLDLQPETTDL	E7 (PPV-9)	27693214
3	LLMGTLGIV	E7 (PPV-9)	26418056
4	TLGIVCPI	E7 (PPV-9)	8617954
5	SLLQHLIGL	MART	22424782
6	VLQELNVTV	Myeloblastin	PMC1534319
7	SVAPALALFPA	LB-ADIR-1F	17234742
8	FLKANLPLL	MTG8b	11122105
9	KLSAMQAHL	Foxp3	23812418
10	LQLPTLPLV	Foxp3	19456276
11	VLHDDLLEA	HA-1/A2	18698046
12	VFEEPEDFL	Foxp3	23812418
13	AIQDLCLAV	nucleophosmin	31291378
14	AIQDLCVAV	nucleophosmin	31291378
15	ALYVDSLFFL	PRAME	PMC2195886
16	RMFPNAPYL	WT1	PMC5722674
17	SLLMWITQV	NY-ESO-1	PMC5941317
18	ELAGIGILTV	MART-1	16860654

Table F.1: Table of peptide sources for use in initial SCT library template optimization studies.

pep ID	peptide	antigen	HLA	yield
1	LLFGYPVYV	HTLV-1 Tax	A0201_B_Y84C	0.35
2	KLVALGINAV	HCV		0.50
7	GLCTLVAML	EBV-BLMF1		0.00
11	WLSLLVPFV	HBV-SAg		0.03
14	YVLDHLIVV	EBV-BRLF1		0.45
19	SITEVECFL	Human polyomavirus 2 protein		0.29
23	FLLSLGIHL	HBV		0.36
24	GILGFVFTL	Flu-M1		0.05
31	SLFNTVATL	HIV gag		0.52
41	YLLFEVFDV	AdV 11 Hexon		0.45
42	LLFEVFDVV	AdV 11 Hexon		0.79
43	YVLFEVFDV	AdV 11 Hexon		0.43
44	FLDKGTYTL	EBV BALF4		0.12
45	YLQQNWWTL	EBV-LMP1-2		0.48
46	YLLEMLWRL	EBV-LMP1-1		0.12
49	FLYALALLL	EBV-LMP1-2		0.10
57	VLEETSVML	CMV-IE1		0.00
62	TLNAWVKVV	HIV-gag		0.08
70	AIMDKNIIL	Influenza NS1		0.33
76	KLIANNTRV	M. tuberculosis Ag85A		0.38
84	ALWALPHAA	varicella-zoster IE62 593-601		0.05
86	NLVPMVATV	CMV-pp65		0.70
87	FMYSDFHFI	Influenza A		0.56
88	YLLPGWKL	Rota-VP3		0.00
89	NMLSTVLGV	Flu-PB1		0.61
90	SLMDPAILTSL	Rota-VP1		1.42
91	TLLANVTAV	Rota-VP6		0.85
92	FMDILTTCVET	CMV-IE1-2		0.23
93	QMWQARLTV	CMV-pp65-2		0.07
94	SLISGMWLL	Rota-VP2-1		0.26
95	LLNYILKSV	Rota-VP7-1		0.14
96	LMNGQQIFL	CMV-pp65-3		0.26
97	FLDSEPHLL	Rota-NSP1		0.15
98	ALWGPDPAAA	Proinsulin precursor 15-24		0.23
99	TLDYKPLSV	EBV BMRF1		0.07
100	CLGGLLTMV	EBV-LMP2A		0.04

pep ID	peptide	antigen	HLA	yield
1	TYFNLGNKF	AdV 11 Hexon (37-45)	A2402_B_Y84C	0.80
2	VYSGSIPYL	AdV 11 Hexon (696-704)		0.18
3	TYFSLNNKF	AdV 5 Hexon (37-45)		0.93
4	DYNFVKQLF	EBV BMLF1 (320-328)		0.84
5	TYPVLEEMF	EBV BRLF1 (198-206)		1.39
6	RYSIFFDYM	EBV EBNA3A (246-254)		0.05
7	TYSAGIVQI	EBV EBNA3B (217-225)		0.84
8	IYVLVMLVL	EBV LMP2 (222-230)		0.00
9	PYLFWLAAI	EBV LMP2 (131-139)		0.00
10	TYGPVFMSL	EBV LMP2 (419-427)		1.13
11	TYGPVFMCL	EBV LMP2 (419-427)		0.64
12	EYLVSFGVW	HBV core (117-125)		0.59
13	KYTSFPWLL	HBV pol (756-764)		0.23
14	QYDPVAALF	HCMV pp65 (341-349)		0.51
15	EYVLLLFLL	HCV E2 (717-725)		0.00
16	PFHCSFHTI	HHV-6B U54 (267-275)		0.00
17	RYLRDQQLL	HIV env gp160 (584-592)		0.19
18	RYLKDQQLL	HIV env (67-75)		0.20
19	RYPLTFGW	HIV nef (134-141)		0.26
20	VYDFAFRDL	HPV16 E6 (49-57)		0.13
21	FFQFCPLIF	HTLV-1 Env (43788)		0.63
22	LFGYPVYVF	HTLV-1 Tax (43819)		0.51
23	PYKRIEELL	HTLV-1 Tax (187-195)		0.00
24	SFHSLHLLF	HTLV-1 Tax (301-309)		0.83
25	YYLEKANKI	Influenza PA (130-138)		0.99
26	SYLIRALTL	Influenza PB1 (216-224)		0.39
27	RYTKTTYWW	Influenza PB1 (430-438)		0.11
28	SYINRTGTF	Influenza PB1 (482-490)		1.02
29	RYGFVANF	Influenza PB1 (498-505)		0.00
30	TYQWIIRNW	Influenza PB2 (549-557)		0.90

Table F.2: Table of viral peptides selected for expression in A*02:01 & A*24:02 templates.

'Yield' column denotes the relative expression of each SCT based on ratio of protein gel band intensity to positive control band (see Fig. S1 for example).

Appendix G

TCR DNA FRAGMENT SEQUENCES FOR CLONING

fragment	DNA
f0_crispr	TTGCTGGGCCTTTTTCCCATGCCTGCCTTTACTCTGCCAGAGTTATATTGCTGGGGTTTTGAAGAAGATCCTATTAAATAAAAG AATAAGCAGTATTATTAAGTAGCCCTGCATTTCAGGTTTCCTTGAGTGGCAGGCCAGGCCTGGCCGTGAACGTTCACTGAAATC ATGGCCTCTTGGCCAAGATTGATAGCTTGTGCCTGTCCCTGAGTCCCAGTCCATCACGAGCAGCTGGTTTCTAAGATGCTATTT CCCGTATAAAGCATGAGACCGTGACTTGCCAGCCCCACAGAGCCCCGCCCTTGTCCATCACTGGCATCTGGACTCCAGCCTGGG TTGGGGCAAAGAGGGAAATGAGATCATGTCCTAACCCTGATCCTCTTGTCCCACAGATATCCAGAACCCTGACCCTGCCTCCGG ATCCGGAGAGGGCAGGGGATCTCTCCTTACTTGTGGCGACGTGGAGGAGAACCCCGGCCCC
f2_lenti_trbc1	GTGGCTGTCTTTGAACCATCAGAAGCCGAGATCTCTCACACCCAAAAAGCCACATTGGTTTGCTTGGCCACTGGATTCTTTCCA GATCACGTTGAGTTGAGTTGGTGGGTGAACGGGAAGGAAGTTCATTCTGGTGTCTGTACCGACCCACAACCCCTTAAAGAACAG CCTGCTCTGAATGATAGTCGGTACTGTCTGAGCTCCAGACTTAGAGTGTCTGCTACTTTCTGGCAGAATCCCCGGAATCACTTC CGGTGTCAAGTCCAATTCTACGGACTGTCCGAGAATGACGAGTGGACACAAGATAGAGCTAAACCTGTCACCCAAATTGTGAGC GCCGAAGCTTGGGGAAGAGCTGATTGTGGCTTCACTAGCGTGAGTTACCAGCAAGGTGTGTTGTCCGCCACAATACTCTACGAG ATTTTGCTCGGCAAGGCCACCTTGTATGCTGTTCTGGTTAGCGCTCTTGTGCTCATGGCTATGGTTAAGAGGAAAGACTTCGGC TCTGGGGCTACCAATTTCAGCCTCCTCAAACAAGCTGGCGACGTTGAAGAGAATCCAGGTCCA
f2_lenti_trbc2	GTGGCTGTCTTTGAACCATCAGAAGCCGAGATCTCTCACACCCAAAAAGCCACATTGGTTTGCTTGGCCACTGGATTCTACCCA GATCACGTTGAGTTGAGTTGGTGGGTGAACGGGAAGGAAGTTCATTCTGGTGTCTGTACCGACCCACAACCCCTTAAAGAACAG CCTGCTCTGAATGATAGTCGGTACTGTCTGAGCTCCAGACTTAGAGTGTCTGCTACTTTCTGGCAGAATCCCCGGAATCACTTC CGGTGTCAAGTCCAATTCTACGGACTGTCCGAGAATGACGAGTGGACACAAGATAGAGCTAAACCTGTCACCCAAATTGTGAGC GCCGAAGCTTGGGGAAGAGCTGATTGTGGCTTCACTAGCGAAAGTTACCAGCAAGGTGTGTTGTCCGCCACAATACTCTACGAG ATTTTGCTCGGCAAGGCCACCTTGTATGCTGTTCTGGTTAGCGCTCTTGTGCTCATGGCTATGGTTAAGAGGAAAGACTCAGGC TCTGGGGCTACCAATTTCAGCCTCCTCAAACAAGCTGGCGACGTTGAAGAGAATCCAGGTCCA
f2_crispr_trbc1	GTGGCTGTCTTTGAACCATCAGAAGCCGAGATCTCTCACACCCAAAAAGCCACATTGGTTTGCTTGGCCACTGGATTCTTTCCA GATCACGTTGAGTTGAGTTGGTGGGTGAACGGGAAGGAAGTTCATTCTGGTGTCAGCACCGACCCACAACCCCTTAAAGAACAG CCTGCTCTGAATGATAGTCGGTACTGTCTGAGCTCCAGACTTAGAGTGTCTGCTACTTTCTGGCAGAATCCCCGGAATCACTTC CGGTGTCAAGTCCAATTCTACGGACTGTCCGAGAATGACGAGTGGACACAAGATAGAGCTAAACCTGTCACCCAAATTGTGAGC GCCGAAGCTTGGGGAAGAGCTGATTGTGGCTTCACTAGCGTGAGTTACCAGCAAGGTGTGTTGTCCGCCACAATACTCTACGAG ATTTTGCTCGGCAAGGCCACCTTGTATGCTGTTCTGGTTAGCGCTCTTGTGCTCATGGCTATGGTTAAGAGGAAAGACTTCGGC TCTGGGGCTACCAATTTCAGCCTCCTCAAACAAGCTGGCGACGTTGAAGAGAATCCAGGTCCA
f2_crispr_trbc2	GTGGCTGTCTTTGAACCATCAGAAGCCGAGATCTCTCACACCCAAAAAGCCACATTGGTTTGCTTGGCCACTGGATTCTACCCA GATCACGTTGAGTTGAGTTGGTGGGTGAACGGGAAGGAAGTTCATTCTGGTGTCAGCACCGACCCACAACCCCTTAAAGAACAG CCTGCTCTGAATGATAGTCGGTACTGTCTGAGCTCCAGACTTAGAGTGTCTGCTACTTTCTGGCAGAATCCCCGGAATCACTTC CGGTGTCAAGTCCAATTCTACGGACTGTCCGAGAATGACGAGTGGACACAAGATAGAGCTAAACCTGTCACCCAAATTGTGAGC GCCGAAGCTTGGGGAAGAGCTGATTGTGGCTTCACTAGCGAAAGTTACCAGCAAGGTGTGTTGTCCGCCACAATACTCTACGAG ATTTTGCTCGGCAAGGCCACCTTGTATGCTGTTCTGGTTAGCGCTCTTGTGCTCATGGCTATGGTTAAGAGGAAAGACTCAGGC TCTGGGGCTACCAATTTCAGCCTCCTCAAACAAGCTGGCGACGTTGAAGAGAATCCAGGTCCA
f4_lenti	TCCAGAACCCTGACCCTGCGGTGTACCAGCTGAGAGACTCTAAATCCAGCGATAAGAGCGTCTGCCTGTTTACAGACTTCGATT CCCAGACCAATGTGTCCCAGTCTAAGGACTCCGATGTCTACATCACCGACAAGTGTGTGCTTGACATGAGGTCCATGGACTTCA AGAGCAACAGCGCTGTCGCATGGTCTAATAAGTCCGACTTCGCTTGTGCCAACGCCTTCAACAATAGCATCATACCCGAGGATA CCTTTTTCCCTTCACCCGAAAGCTCATGTGACGTCAAGCTCGTTGAGAAGAGCTTCGAGACCGACACCAATCTGAACTTCCAGA ACCTGTCAGTGATTGGTTTCCGCATCCTGCTTCTGAAGGTCGCCGGCTTTAACTTGCTTATGACCTTGAGGCTTTGGTCAAGCT GAGTCGACAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGC
f4_crispr	TCCAGAACCCTGACCCTGCGGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATT CTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCA AGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACA CCTTCTTCCCCAGCCCAGGTAAGGGCAGCTTTGGTGCCTTCGCAGGCTGTTTCCTTGCTTCAGGAATGGCCAGGTTCTGCCCAG AGCTCTGGTCAATGATGTCTAAAACTCCTCTGATTGGTGGTCTCGGCCTTATCCATTGCCACCAAAACCCTCTTTTTACTAAGA AACAGTGAGCCTTG

Table G.1: TCR DNA fragment sequences

Appendix H

HLA HAPLOTYPES OF INCOV STUDY PARTICIPANTS

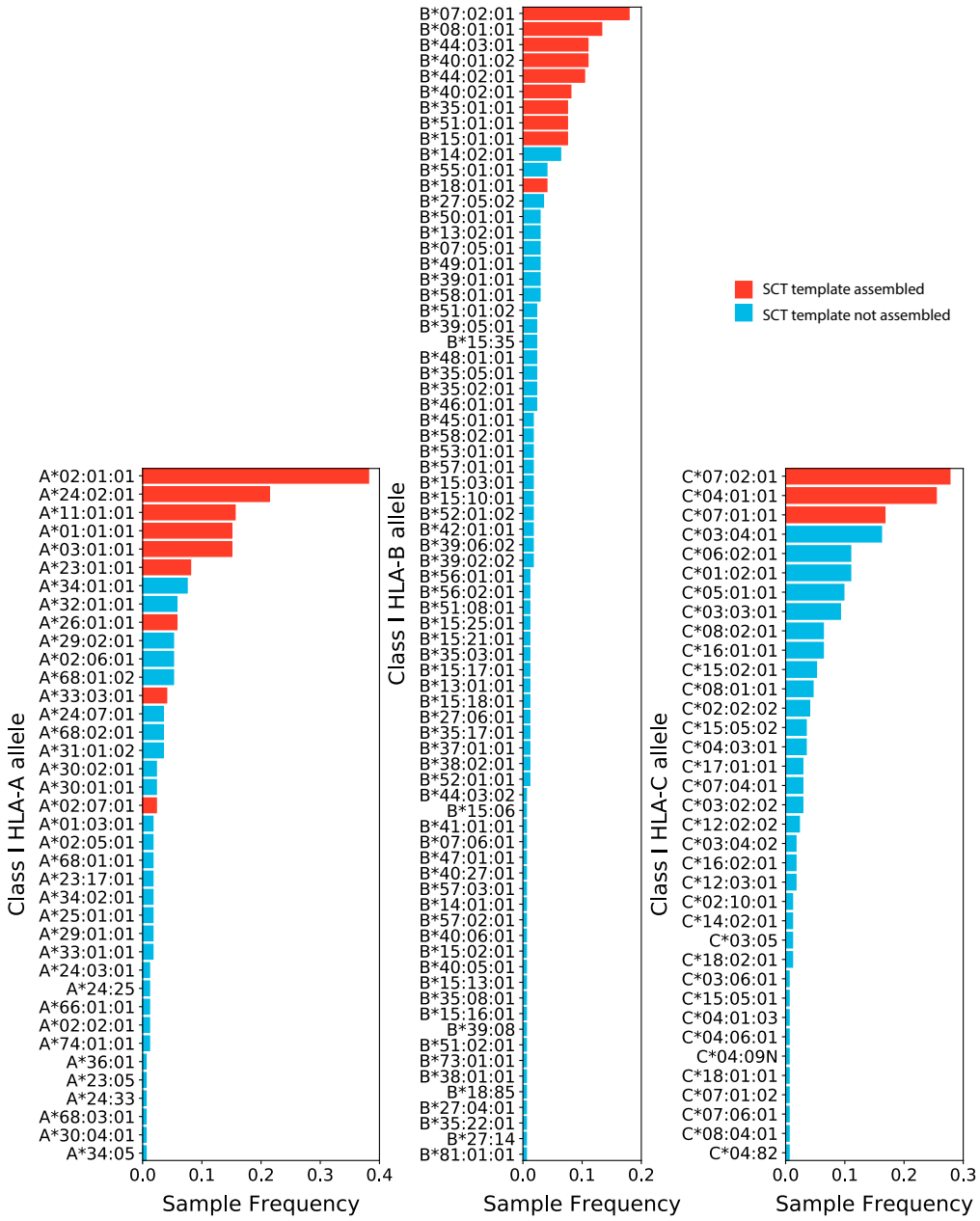


Figure H.1: Prevalence of Class I HLA haplotypes from InCOV study participants Sorted frequencies of Class I HLA-A, HLA-B, & HLA-C haplotypes from InCOV study participants.

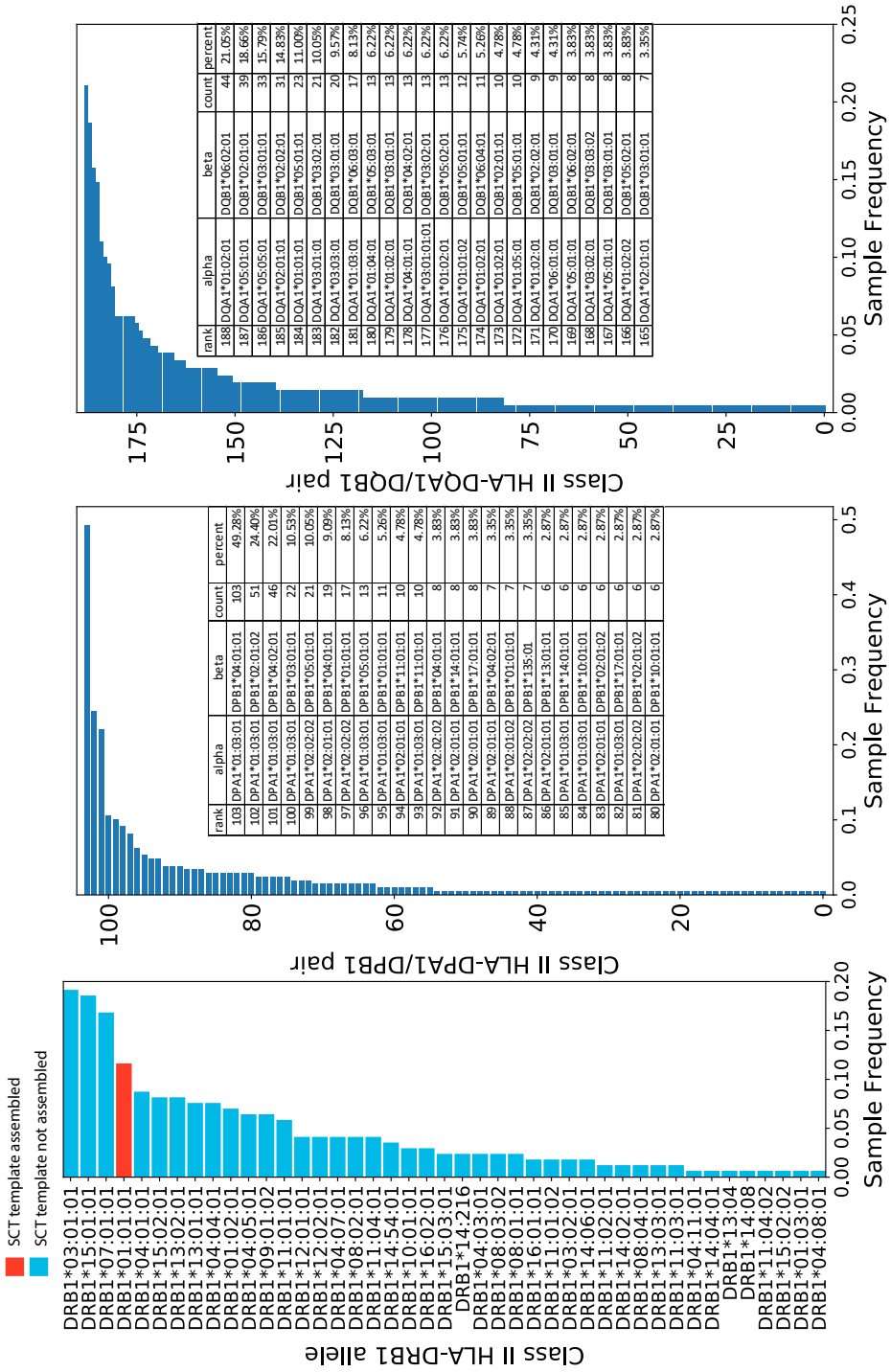
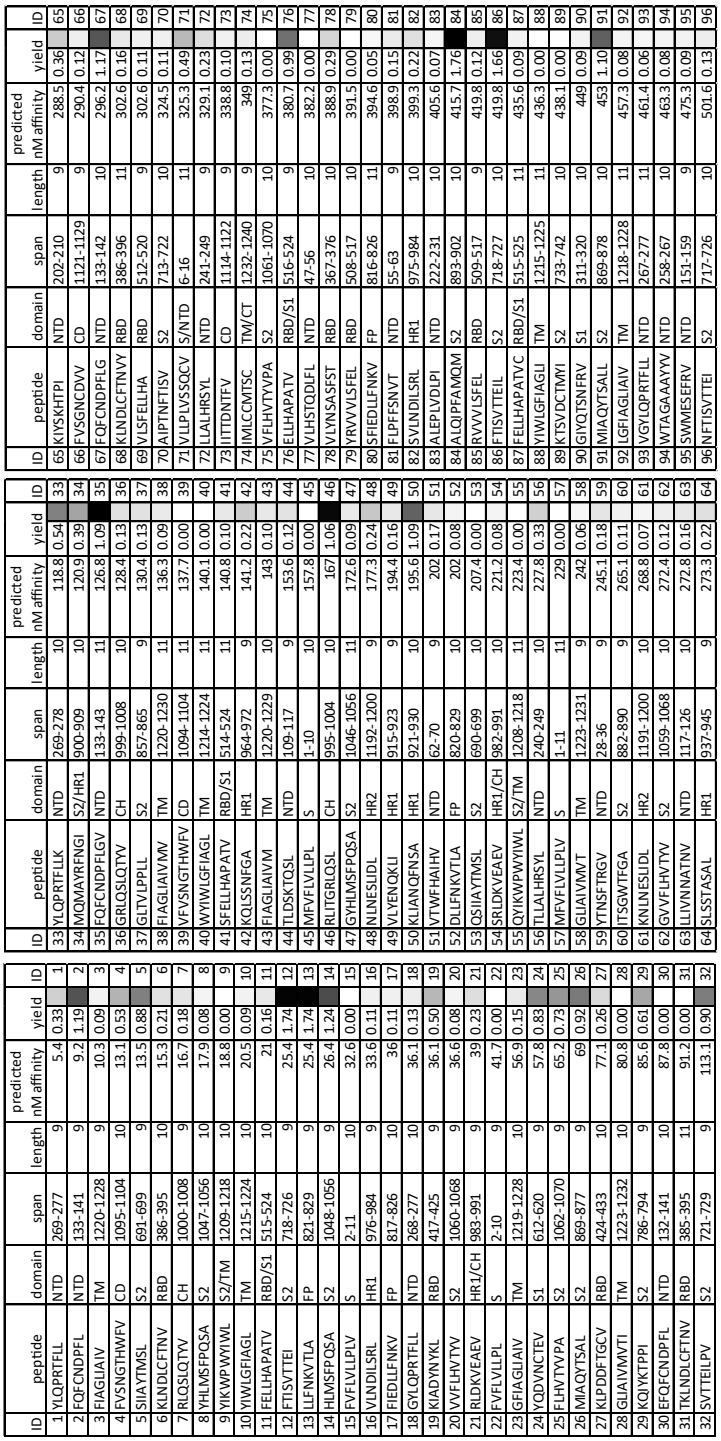


Figure H.2: Prevalence of Class II HLA haplotypes from InCOV study participants Sorted frequencies of Class II HLA-DR, HLA-DP, & HLA-DQ haplotypes from InCOV study participants. For HLA-DQ & HLA-DP haplotypes, due to potential cross-pairing of α and β chains, paired frequencies were assessed instead of individual allelic frequencies, and are reported in the tables.

Appendix I

SCT EXPRESSION OF SARS-COV-2 SPIKE & NSP3 EPITOPES



ID	peptide	domain	span	length	predicted nM affinity	yield	ID
1	YLQPRTFLL	NTD	269-277	9	5.4	0.33	1
2	FQFCNDPFL	NTD	133-141	9	9.2	1.19	2
3	FIAGLIAIV	TM	1220-1228	9	10.3	0.09	3
4	FVSNGTHWFV	CD	1095-1104	10	13.1	0.53	4
5	SIIAYTMSL	S2	691-699	9	13.5	0.88	5
6	KLNDLCFTNV	RBD	386-395	10	15.3	0.21	6
7	RLQSLQTYV	CH	1000-1008	9	16.7	0.18	7
8	YHLMSFPQSA	S2	1047-1056	10	17.9	0.08	8
9	YIKWPWYIWL	S2/TM	1209-1218	10	18.8	0.00	9
10	YIWLGFIAGL	TM	1215-1224	10	20.5	0.09	10
11	FELLHAPATV	RBD/S1	515-524	10	21	0.16	11
12	FTISVTTEI	S2	718-726	9	25.4	1.74	12
13	LFNKVTLA	FP	821-829	9	25.4	1.74	13
14	HLMSFPQSA	S2	1048-1056	9	26.4	1.24	14
15	FVFLVLLPLV	S	2-11	10	32.6	0.00	15
16	VLNDILSRL	HR1	976-984	9	33.6	0.11	16
17	FIEDLLFNKV	FP	817-826	10	36	0.11	17
18	GYLQPRTFLL	NTD	268-277	10	36.1	0.13	18
19	KIADYNYKL	RBD	417-425	9	36.1	0.50	19
20	VVFLHVTYV	S2	1060-1068	9	36.6	0.08	20
21	RLDKVEAEV	HR1/CH	983-991	9	39	0.23	21
22	FVFLVLLPL	S	2-10	9	41.7	0.00	22
23	GFIAGLIAIV	TM	1219-1228	10	56.9	0.15	23
24	YQDVNCTEV	S1	612-620	9	57.8	0.83	24
25	FLHVTYVPA	S2	1062-1070	9	65.2	0.73	25
26	MIAQYTSAL	S2	869-877	9	69	0.92	26
27	KLPDDFTGCV	RBD	424-433	10	77.1	0.26	27
28	GLIAIVMVTI	TM	1223-1232	10	80.8	0.00	28
29	KQIYKTPPI	S2	786-794	9	85.6	0.61	29
30	EFQFCNDPFL	NTD	132-141	10	87.8	0.00	30
31	TKLNDLCFTNV	RBD	385-395	11	91.2	0.00	31
32	SVTTEILPV	S2	721-729	9	113.1	0.90	32
33	YLQPRTFLLK	NTD	269-278	10	118.8	0.54	33
34	MQMAYRFNGI	S2/HR1	900-909	10	120.9	0.39	34
35	FQFCNDPFLGV	NTD	133-143	11	126.8	1.09	35
36	GRLQSLQTYV	CH	999-1008	10	128.4	0.13	36
37	GLITVLPPLL	S2	857-865	9	130.4	0.13	37
38	FIAGLIAIVM	TM	1220-1230	11	136.3	0.09	38
39	VFVSNGTHWFV	CD	1094-1104	11	137.7	0.00	39
40	WYIWLGFIAGL	TM	1214-1224	11	140.1	0.00	40
41	SFELLHAPATV	RBD/S1	514-524	11	140.8	0.10	41
42	KQLSSNFGA	HR1	964-972	9	141.2	0.22	42
43	FIAGLIAIVM	TM	1220-1229	10	143	0.10	43
44	TLDSKTQSL	NTD	109-117	9	153.6	0.12	44
45	MFVFLVLLPL	S	1-10	10	157.8	0.00	45
46	RLITGRLQSL	CH	995-1004	10	167	1.06	46
47	GYHLMSFPQSA	S2	1046-1056	11	172.6	0.09	47
48	NLNESLIDL	HR2	1192-1200	9	177.3	0.24	48
49	VLYENQKLI	HR1	915-923	9	194.4	0.16	49
50	KLIANQFNSA	HR1	921-930	10	195.6	1.09	50
51	VTWFHAIHV	NTD	62-70	9	202	0.17	51
52	DLLFNKVTLA	FP	820-829	10	202	0.08	52
53	QSIIAYTMSL	S2	690-699	10	207.4	0.00	53
54	SRLDKVEAEV	HR1/CH	982-991	10	221.2	0.08	54
55	QYIKWPWYIWL	S2/TM	1208-1218	11	223.4	0.00	55
56	TLLALHRSYL	NTD	240-249	10	227.8	0.33	56
57	MFVFLVLLPLV	S	1-11	11	229	0.00	57
58	GLIAIVMVT	TM	1223-1231	9	242	0.06	58
59	YTNSFTRGV	NTD	28-36	9	245.1	0.18	59
60	ITSGWTFGA	S2	882-890	9	265.1	0.11	60
61	KNLNESLIDL	HR2	1191-1200	10	268.8	0.07	61
62	GVVFLHVTYV	S2	1059-1068	10	272.4	0.12	62
63	LLIVNNATNV	NTD	117-126	10	272.8	0.16	63
64	SLSSTASAL	HR1	937-945	9	273.3	0.22	64
65	KIYSKHTPI	NTD	202-210	9	288.5	0.36	65
66	FVSGNCDVV	CD	1121-1129	9	290.4	0.12	66
67	FQFCNDPFLG	NTD	133-142	10	296.2	1.17	67
68	KLNDLCFTNVY	RBD	386-396	11	302.6	0.16	68
69	VLSFELLHA	RBD	512-520	9	302.6	0.11	69
70	AIPTNFTISV	S2	713-722	10	324.5	0.11	70
71	VLPLVSSQCV	S/NTD	6-16	11	325.3	0.49	71
72	LIALHRSYL	NTD	241-249	9	329.1	0.23	72
73	IITTDNTFV	CD	1114-1122	9	338.8	0.10	73
74	IMLCCMTSC	TM/CT	1232-1240	9	349	0.13	74
75	VFLHVTYVPA	S2	1061-1070	10	377.3	0.00	75
76	ELLHAPATV	RBD/S1	516-524	9	380.7	0.99	76
77	VLHSTQDLFL	NTD	47-56	10	382.2	0.00	77
78	VLYNSASFST	RBD	367-376	10	388.9	0.29	78
79	YRVVVLSFEL	RBD	508-517	10	391.5	0.00	79
80	SFIEDLLFNKV	FP	816-826	11	394.6	0.05	80
81	FLPFFSNVT	NTD	55-63	9	398.9	0.15	81
82	SVLNDILSRL	HR1	975-984	10	399.3	0.22	82
83	ALEPLVDLPI	NTD	222-231	10	405.6	0.07	83
84	ALQIPFAMQM	S2	893-902	10	415.7	1.76	84
85	RVVVLSFEL	RBD	509-517	9	419.8	0.12	85
86	FTISVTTEIL	S2	718-727	10	419.8	1.66	86
87	FELLHAPATVC	RBD/S1	515-525	11	435.6	0.09	87
88	YIWLGFIAGLI	TM	1215-1225	11	436.3	0.00	88
89	KTSVDCTMYI	S2	733-742	10	438.1	0.00	89
90	GIYQTSNFRV	S1	311-320	10	449	0.09	90
91	MIAQYTSALL	S2	869-878	10	453	1.10	91
92	LGFIAGLIAIV	TM	1218-1228	11	457.3	0.08	92
93	VGYLQPRTFLL	NTD	267-277	11	461.4	0.06	93
94	WTAGAAAYYV	NTD	258-267	10	463.3	0.08	94
95	SWMESEFRV	NTD	151-159	9	475.3	0.09	95
96	NFTISVTTEI	S2	717-726	10	501.6	0.13	96

Table I.1: Top HLA-A*02:01 9-mer to 11-mer epitopes from SARS-CoV-2 spike protein predicted by NetMHC4.0

The peptides selected for expression as SCTs are listed in this table, indexed according to binding affinity. 'Yield' column denotes the calculated relative expression of each SCT based on ratio of its protein gel band intensity to positive control band (see Fig. S1 for example).

ID	peptide	domain	span	length	predicted nM affinity	yield	ID
1	SPRRARSVA	S1/S2	680-688	9	4.2	0.25	1
2	SPRRARSVAS	S1/S2	680-689	10	18.4	0.27	2
3	NSPRRARSVA	S1/S2	679-688	10	20.1	0.09	3
4	MIAQYTSAL	S2	869-877	9	28.9	0.36	4
5	YPDKVFRSSV	NTD	38-47	10	50.2	0.60	5
6	IPTNFTISV	S2	714-722	9	95.1	0.47	6
7	FPQSAPHGVV	S2	1052-1061	10	99.6	0.44	7
8	LPVSMTKTSV	S2	727-736	10	100	0.22	8
9	TNSPRRARSVA	S1/S2	678-688	11	104.1	0.12	9
10	SPRRARSVASQ	S1/S2	680-690	11	136.2	0.23	10
11	RARSVASQSI	S1/S2	683-692	10	147	0.21	11
12	QPYRVVVLSF	RBD	506-515	10	163.1	0.00	12
13	LPPAYTNSF	NTD	24-32	9	168	0.32	13
14	FPNITNLCPF	RBD	329-338	10	175.2	0.00	14
15	WPWYIWLGF	TM	1212-1220	9	193.4	0.00	15
16	TPINLVRDL	NTD	208-216	9	202.5	0.59	16
17	VPVAIHADQL	S1	620-629	10	210.8	0.19	17
18	TPCSFGGVSV	S1	588-597	10	249.9	0.35	18
19	YPDKVFRSSVL	NTD	38-48	11	255.4	0.59	19
20	KPFERDISTEI	RBD	462-472	11	286	0.50	20
21	RAAEIRASA	CH	1014-1022	9	317.7	0.08	21
22	SIIAYTMSL	S2	691-699	9	328	0.29	22
23	LPQGFSALEPL	NTD	216-226	11	337.2	0.68	23
24	MIAQYTSALL	S2	869-878	10	350.9	0.41	24
25	NSPRRARSVAS	S1/S2	679-689	11	361.2	0.08	25
26	AIPTNFTISV	S2	713-722	10	381	0.88	26
27	SEPVLKGVKL	CP	1261-1270	10	388.2	0.08	27
28	APGQTGKIA	RBD	411-419	9	407.8	0.38	28
29	FPQSAPHGV	S2	1052-1060	9	409.9	0.45	29
30	FPQSAPHGVVF	S2	1052-1062	11	413.7	0.62	30
31	ITRFQTLLAL	NTD	235-244	10	444.5	0.10	31
32	VVNQNAQAL	HR1	951-959	9	472.6	0.05	32
33	SLSSTASAL	HR1	937-945	9	492.4	0.00	33

Table I.2: Top HLA-B*07:02 9-mer to 11-mer epitopes from SARS-CoV-2 spike protein predicted by NetMHC4.0

The peptides selected for expression as SCTs are listed in this table, indexed according to binding affinity. 'Yield' column denotes the relative expression of each SCT based on ratio of protein gel band intensity to positive control band (see Fig. S1 for example).

ID	peptide	domain	span	length	predicted nM affinity	yield	ID
1	KWPWYIWLGF	TM	1211-1220	10	9	0.16	1
2	LYNSASFSTF	RBD	368-377	10	12.3	0.31	2
3	QYIKWPWYI	TM	1208-1216	9	13.2	0.32	3
4	QYIKWPWYIW	TM	1208-1217	10	14.9	0.00	4
5	VYSTGSNVF	S1	635-643	9	19	0.46	5
6	VYSSANNCTF	NTD	159-168	10	21.9	0.55	6
7	CYFPLQSYGF	RBD	488-497	10	26.6	0.00	7
8	NYNYLYRLF	RBD	448-456	9	28.9	0.49	8
9	RVYSTGSNVF	S1	634-643	10	36.3	0.16	9
10	IKWPWYIWLGF	TM	1210-1220	11	36.3	0.00	10
11	PFAMQMAYRF	S2	897-906	10	38.9	0.00	11
12	YFPLQSYGF	RBD	489-497	9	41.5	0.00	12
13	KWPWYIWLGFI	TM	1211-1221	11	42.1	0.78	13
14	EQYIKWPWYI	TM	1207-1216	10	45.3	0.33	14
15	YEQYIKWPWYI	TM	1206-1216	11	55.8	0.31	15
16	VLYNSASFSTF	RBD	367-377	11	64.3	0.00	16
17	YVVGYLQPRTF	NTD	265-275	11	73.3	0.26	17
18	LYNSASFSTFK	RBD	368-378	11	76.3	0.73	18
19	VYSTGSNVFQ	S1	635-644	10	93.3	0.21	19
20	EQYIKWPWYIW	TM	1207-1217	11	96.2	0.38	20
21	KYEQYIKWPW	TM	1205-1214	10	110.2	0.53	21
22	FAMQMAYRF	S2	898-906	9	111.5	0.54	22
23	QYIKWPWYIWL	TM	1208-1218	11	115.5	0.40	23
24	IYQTSNFRV	S1	312-320	9	124.6	0.00	24
25	TVVPAQEKNF	S2	1066-1075	10	137.3	0.68	25
26	GYLQPRTFLL	NTD	268-277	10	138.1	0.35	26

ID	peptide	domain	span	length	predicted nM affinity	yield	ID
27	RVYSSANNCTF	NTD	158-168	11	143.7	0.17	27
28	RFPNITNLCPF	RBD	328-338	11	149.3	0.13	28
29	KWPWYIWLG	TM	1211-1219	9	152	0.00	29
30	PYRVVVLSF	RBD	507-515	9	152.4	0.00	30
31	VYAWNRKRI	RBD	350-358	9	160.3	0.00	31
32	VYSSANNCTFE	NTD	159-169	11	172.5	0.31	32
33	GNYNYLYRLF	RBD	447-456	10	174.6	0.22	33
34	YYHKNNKSW	NTD	144-152	9	223.5	0.32	34
35	NYNYLYRLFR	RBD	448-457	10	233.2	0.20	35
36	IYSKHTPINL	NTD	203-212	10	233.7	0.00	36
37	IWLGFIAGLI	TM	1216-1225	10	239	0.00	37
38	NCYFPLQSYGF	RBD	487-497	11	243.2	0.16	38
39	IYKTPPIKDF	S2	788-797	10	252.1	0.24	39
40	YNSASFSTF	RBD	369-377	9	252.4	0.61	40
41	WRVYSTGSNVF	S1	633-643	11	274.1	0.00	41
42	CYFPLQSYGFQ	RBD	488-498	11	299.8	0.53	42
43	TFEVVSQPF	NTD	167-175	9	337.1	0.58	43
44	YFPLQSYGFQ	RBD	489-498	10	349.3	0.78	44
45	RVYSTGSNVFQ	S1	634-644	11	371.1	0.32	45
46	CFTNVYADSF	RBD	391-400	10	371.3	0.43	46
47	EYVSQPFLM	NTD	169-177	9	373.2	0.53	47
48	VFVSNGTHWF	CD	1094-1103	10	377.5	0.45	48
49	YLQPRTFLL	NTD	269-277	9	406.7	0.25	49
50	IPFAMQMAYRF	S2	896-906	11	427.8	0.30	50
51	VYYHKNNKSW	NTD	143-152	10	488.9	0.00	51

Table I.3: Top HLA-A*24:02 9-mer to 11-mer epitopes from SARS-CoV-2 spike protein predicted by NetMHC4.0

The peptides selected for expression as SCTs are listed in this table, indexed according to binding affinity. 'Yield' column denotes the relative expression of each SCT based on ratio of protein gel band intensity to positive control band (see Fig. S1 for example).

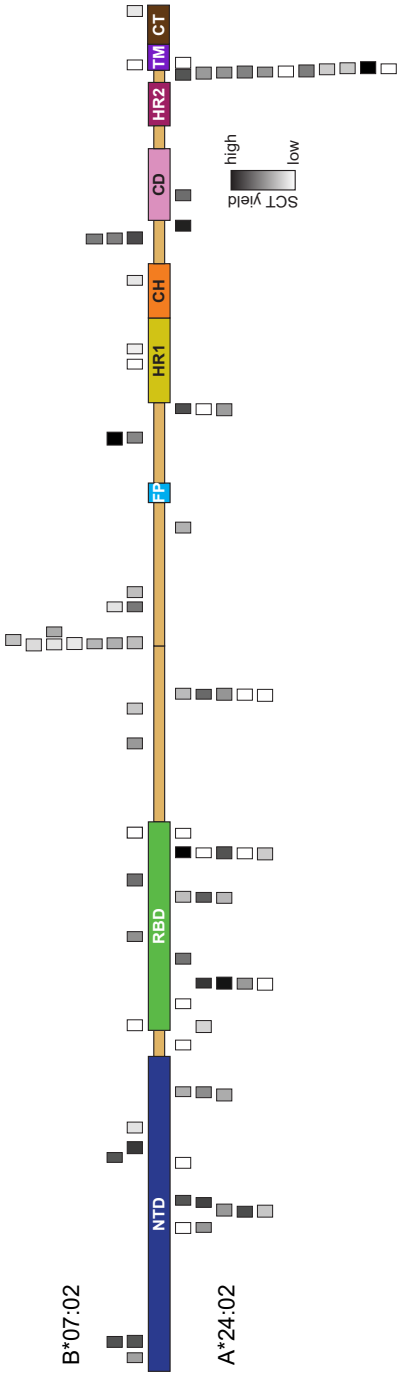


Figure I.1: Expression of SCTs for B*07:02 & A*24:02 SARS-CoV-2 spike protein epitopes

Schematic of the spike protein domains. S, signal sequence; NTD, N-terminal domain; RBD, receptor binding domain; FP, fusion peptide; HR1, heptad repeat 1; CH, central helix; CD, connector domain; HR2, 5 heptad repeat 2; TM, transmembrane domain; CT, cytoplasmic tail; subunits denoted by S1 and S2. Shaded boxes denote relative peptide position and expression yields of SCT proteins (top: B*07:02; bottom: A*24:02).

ID	peptide	span	length	nM	yield	ID
1	ILFTRFFVV	1514-1522	9	3.2	0.23	1
2	YILFTRFFV	1513-1522	10	3.8	0.21	2
3	YIFFASFYYV	1566-1575	10	5	0.13	3
4	YLATALLTL	857-865	9	5.1	0.17	4
5	WLMWLIINL	1546-1553	9	6.6	0.00	5
6	YMPYFFTLL	1390-1398	9	7.1	0.21	6
7	FLTENLLLYI	430-439	10	7.2	0.92	7
8	KLINIIWFL	1407-1416	10	7.3	0.00	8
9	QLFSYFAV	1530-1538	9	10	0.09	9
10	LLSAGIFGA	330-338	9	10.1	0.48	10
11	AYLFTRFFV	1512-1522	11	10.5	0.17	11
12	IMQLFSYFAV	1529-1538	10	10.7	0.32	12
13	WLMWLIINLV	1546-1554	10	11	0.11	13
14	YLKLTDNVYI	213-222	10	12.5	0.24	14
15	LMWLIINLV	1546-1554	9	13.3	0.00	15
16	YMPYFFTLLL	1390-1399	10	13.9	0.10	16
17	FLGRYMSAL	824-832	9	14.9	0.32	17
18	KLMPVCVET	569-577	9	15.1	0.90	18
19	FVMMSAPPA	986-994	9	15.1	1.32	19
20	YVWKSYVHV	1574-1582	9	15.2	0.85	20
21	ILFTRFFYVL	1514-1523	10	16	0.10	21
22	QLEMELTPVV	193-202	10	17	0.22	22
23	YLNSTNVTI	1452-1460	9	17.2	0.10	23
24	IIWFLLSV	1412-1420	9	19	0.00	24
25	QLMCQPILL	1745-1753	9	19	0.41	25
26	YLVQDESPFV	978-987	10	19.3	1.31	26
27	QMAPISAMV	1555-1563	9	19.6	0.00	27
28	CYLATALLTL	856-865	10	19.8	0.08	28
29	KLLHKPIVWHV	1166-1176	11	20.5	0.79	29
30	YLFDESGEFKL	88-98	11	22.2	0.74	30
31	YILFTRFFVL	1513-1523	11	22.3	0.11	31
32	MYIFFASFYV	1566-1575	11	22.7	0.09	32
33	IMQLFSYFA	1529-1537	9	23.1	0.47	33
34	ILGTVSWNL	549-557	9	23.6	0.00	34
35	SWLMWLIINL	1544-1553	10	24.9	0.00	35
36	LLHKPIVWHV	1167-1176	10	27	0.26	36
37	ILLLDQALV	1791-1799	9	27.1	0.87	37
38	FISNSWLMWL	1540-1549	10	29	0.00	38
39	KMFDAYVNT	1771-1779	9	30.2	0.88	39
40	SLIYSTAAL	1424-1432	9	30.8	0.93	40
41	YVWKSYVHVV	1574-1583	10	32.5	0.55	41
42	FELDERIDKV	25-34	10	34.8	0.09	42
43	KFLTENLLLYI	429-439	11	36.4	0.22	43
44	SLIKINIIWFL	1406-1416	11	38.4	0.09	44
45	VLLAPLLSA	325-333	9	40.8	0.93	45
46	SQLMCQPILL	1744-1753	10	42.1	0.34	46
47	FLKKDAPYI	460-468	9	42.2	0.18	47
48	TLVTMPLGYV	626-635	10	43.5	0.75	48

ID	peptide	span	length	nM	yield	ID
49	NTFSSTFNV	1778-1786	9	43.6	0.28	49
50	YSQLMCQPILL	1743-1753	11	45.4	0.12	50
51	TLATHGLAAV	1303-1312	10	47.6	0.34	51
52	NYNMYMFTLL	1349-1358	10	48.7	0.12	52
53	AMVRVMYFFA	1361-1570	10	49	0.10	53
54	IMQLFFSYFA	1528-1537	10	51.3	0.10	54
55	KLVSSFLEM	367-375	9	52.5	0.40	55
56	SLDNVLSTFI	1807-1816	10	52.8	0.29	56
57	YTVELGTEV	42-50	9	53.7	0.64	57
58	MLAKALRKV	494-502	9	54.2	0.50	58
59	VLSFISAA	1811-1819	9	55.3	0.13	59
60	YLATALLTLQ	857-866	10	56.4	0.18	60
61	FITESKPSV	396-404	9	57.5	0.48	61
62	LVAEWFLAYI	1505-1514	10	59.3	0.13	62
63	IIIWFLLSV	1411-1420	10	59.9	0.00	63
64	SWLMWLIINLV	1544-1554	11	61.5	0.08	64
65	VQMAPISAMV	1554-1563	10	61.8	0.32	65
66	IMQLFSYFAV	1528-1538	11	67.1	0.07	66
67	QLMCQPILLL	1745-1754	10	67.6	0.43	67
68	FGLVAEWFLA	1503-1512	10	67.7	0.49	68
69	YLDGADVTKI	780-789	10	67.8	0.21	69
70	NLAKHCLHV	291-299	9	68	0.14	70
71	YIFFASFYYVW	1566-1576	11	68.1	0.09	71
72	ILAYCNKTV	896-904	9	68.7	0.17	72
73	LLLSVCLGSL	1416-1425	10	71.4	0.36	73
74	LLDQALVSDV	1753-1762	10	80.1	0.27	74
75	FLKKDAPPVV	460-469	10	80.8	0.21	75
76	LLAPILSAGI	326-335	10	80.9	0.58	76
77	KLINIIWFLL	1407-1417	11	81	0.17	77
78	SLKVPATVSV	650-659	10	85.2	0.11	78
79	KLDGVVCTEI	1065-1074	10	87.1	0.18	79
80	IFFASFYYV	1567-1575	9	87.8	0.07	80
81	YLAVFDKNL	356-364	9	89	0.42	81
82	YMRSLKVPATV	647-657	11	89.2	0.13	82
83	AVIKTLQPV	60-68	9	89.6	0.61	83
84	ALLTLQQIEL	861-870	10	91.7	1.06	84
85	FCLEASFNYL	1391-1400	10	93.3	0.49	85
86	TLNDLNETL	619-627	9	95.6	0.14	86
87	ALQDAYYRA	876-884	9	100.3	0.08	87
88	NYNMYMFTLL	1349-1359	11	101.3	0.10	88
89	FLTENLLLYID	430-440	11	102	1.00	89
90	GLVAEWFLA	1504-1512	9	105.3	0.24	90
91	YVNTFSSTFNV	1776-1786	11	110.1	0.46	91
92	VLGLAAIMQL	1522-1531	10	111.3	0.09	92
93	PLLSAGIFGA	329-338	10	111.6	0.13	93
94	YVDNSSLITI	1281-1289	9	115.3	0.41	94
95	RKLMPVCVET	568-577	10	121.6	0.08	95
96	WLDDDSQQTV	161-170	10	124.2	0.30	96

ID	peptide	span	length	nM	yield	ID
97	ILFTRFFYVLG	1514-1525	11	125.3	0.14	97
98	LINIIWFL	1408-1417	9	126.6	0.08	98
99	NLACEDLKPV	1220-1230	10	128.4	0.22	99
100	KLSHQSDIEV	1838-1848	10	128.5	0.40	100
101	LMWLIINLVQ	1546-1556	10	131.1	0.13	101
102	YYVWKSYVHV	1573-1583	10	131.2	0.15	102
103	SLINTLNDL	615-624	9	133	1.49	103
104	GYLKLTDNVYI	212-223	11	135.4	0.10	104
105	NLAKHCLHVV	291-301	10	137.4	0.19	105
106	TLNDLNETLV	619-629	10	139.8	0.22	106
107	FTRFFVLGL	1516-1526	10	142.2	0.28	107
108	NSWLMWLIINL	1543-1554	11	149.1	0.00	108
109	WLMWLIINLVQ	1546-1556	11	154.2	0.09	109
110	QLFFSYFAVH	1530-1540	10	156.8	0.00	110
111	FLNKVVSTT	1327-1336	9	157	0.32	111
112	GYLNSTNVTI	1451-1461	10	164.2	0.09	112
113	VITAVVIPT	477-486	9	168.2	0.16	113
114	LLLQLCTFT	1357-1366	9	169.2	0.32	114
115	PQLEMELTPVV	192-203	11	170.5	0.21	115
116	VLPNDDTLRV	802-812	10	172.5	0.08	116
117	NCYLATALLTL	855-866	11	173.3	0.13	117
118	SAFYILPSI	533-542	9	179.7	0.13	118
119	TLYCIDGALL	1026-1036	10	180.9	0.10	119
120	LMCQPILLL	1746-1755	9	184.2	0.35	120
121	IIWFLLLSVC	1412-1422	10	188	0.00	121
122	FSYFAVHFI	1533-1542	9	195.1	0.13	122
123	SFIGRYMSAL	823-833	10	196.9	0.07	123
124	MPLGYVTHGL	630-640	10	197.4	0.10	124
125	LILAYCNKTV	895-905	10	199.2	0.15	125
126	MQLFFSYFAVH	1529-1540	11	214.2	0.12	126
127	QLEMELTPV	193-202	9	215.1	0.23	127
128	EILGTVSWNL	548-558	10	217.3	0.13	128
129	PILLLDQALV	1750-1760	10	222.8	0.17	129
130	FLAYILFTRF	1510-1520	10	226.9	0.00	130
131	KYLVQQESPFV	977-988	11	230.2	0.09	131
132	LLTLQQIEL	862-871	9	239.6	0.77	132
133	FISNSWLMWLI	1540-1551	11	240.2	0.00	133
134	TTIQTIVEV	182-191	9	241.3	0.69	134
135	KIAEIPKEEV	384-394	10	247.7	0.19	135
136	YLKSPNFSKL	1399-1409	10	248.8	0.07	136
137	KLMPVCVETKA	569-580	11	249.5	0.82	137
138	YIATNGPLKV	272-282	10	250.5	0.09	138
139	YMRSLKVPA	647-656	9	251.4	0.10	139
140	MELTPVVQTI	196-206	10	255.7	0.11	140
141	TLYCIDGAL	1026-1035	9	267.7	0.08	141
142	IDLDEWSMAT	77-87	10	281.2	0.09	142
143	YLNSTNVTIA	1452-1462	10	289.4	0.09	143
144	MQVESDDYI	265-274	9	294	0.24	144

ID	peptide	span	length	nM	yield	ID
145	EVLLAPLLSA	324-334	10	296.2	0.07	145
146	SIETIQITI	1485-1494	9	297.5	0.25	146
147	GLVAEWFLAYI	1504-1515	11	300.4	0.20	147
148	FLLLSVCLGSL	1415-1425	10	300.8	0.00	148
149	YYVWKSYVHVV	1573-1584	11	302.2	0.00	149
150	LLSVCLGSL	1417-1426	9	305.3	0.11	150
151	VNTFSSTFNV	1777-1787	10	308.6	0.09	151
152	YMPYFFTLLLQ	1350-1361	11	309	0.09	152
153	TGQPVSELL	64-73	9	315	0.32	153
154	FLLLSVCLGSL	1415-1426	11	324.7	0.06	154
155	NLYDKLVSSFL	363-374	11	325.4	0.28	155
156	LLYDINGNL	436-446	10	327.3	0.28	156
157	FGDDTVIEV	7-16	9	328.5	0.30	157
158	RTIKVFTTV	748-757	9	330.4	0.75	158
159	LIYSTAALGV	1425-1435	10	330.8	0.14	159
160	FIEMKSEKQV	372-382	10	334.8	0.09	160
161	LLSAGIFGAD	330-340	10	351.2	0.32	161
162	GSLIYSTAAL	1423-1433	10	358.6	0.94	162
163	YLPDESGEF	88-97	9	360	0.68	163
164	PFVMMSAPPA	985-995	10	363.4	0.20	164
165	TFELDERIDKV	24-35	11	368.3	0.08	165
166	CLGSLIYST	1421-1430	9	375.2	0.07	166
167	KLINIIWF	1407-1416	9	382.9	0.10	167
168	WLINLVQMA	1548-1558	10	395.1	0.00	168
169	LVQMAPISAMV	1553-1564	11	402.1	0.49	169
170	MVRMVFFA	1562-1571	9	404	0.00	170
171	KLMPVCVETK	569-579	10	404.3	0.94	171
172	NLHPDSATL	444-453	9	411.9	0.12	172
173	HFISNSWLMWL	1539-1550	11	415.1	0.00	173
174	VLMSNLGMPS	1434-1444	10	417.1	0.45	174
175	LVQDESPFV	979-988	9	419.2	0.19	175
176	KLTGATTRQV	1923-1933	10	430.8	0.08	176
177	FVCDNIKFA	1112-1121	9	432.8	0.63	177
178	DKLVSSFLEM	366-376	10	447	0.00	178
179	TIVEVQPQL	186-195	9	447.8	0.58	179
180	KSVNITFEL	19-28	9	453.2	0.82	180
181	VLNEKCSAYTV	34-45	11	457.4	0.10	181
182	FLAYILFTR	1510-1519	9	460.6	0.05	182
183	TNYMPYFFTLL	1348-1359	11	467.8	0.00	183
184	NIIWFLLLSV	1410-1421	11	470.4	0.00	184
185	TWSVLFQHA	913-922	9	481.5	0.00	185
186	FPDLNGDVVAI	1142-1153	11	489.5	0.10	186
187	LLDDQALVSDV	1752-1763	11	491.8	1.24	187
188	YPQVNGLTSI	840-850	10	491.9	0.00	188
189	KLASHMKYCS	97-106	9	493.4	0.38	189
190	HNLAKHCLHV	290-300	10	494.6	0.00	190
191	YLVQQESPFVM	978-988	11	497.2	1.48	191

Table I.4: Top HLA-A*02:01 9-mer to 11-mer epitopes from SARS-CoV-2 Nsp3 predicted by NetMHC4.0

The peptides selected for expression as SCTs are listed in this table, indexed according to binding affinity. 'Yield' column denotes the calculated relative expression of each SCT based on ratio of its protein gel band intensity to positive control band (see Fig. S1 for example).

Appendix J

10X SINGLE-CELL TCR SEQUENCING RESULTS

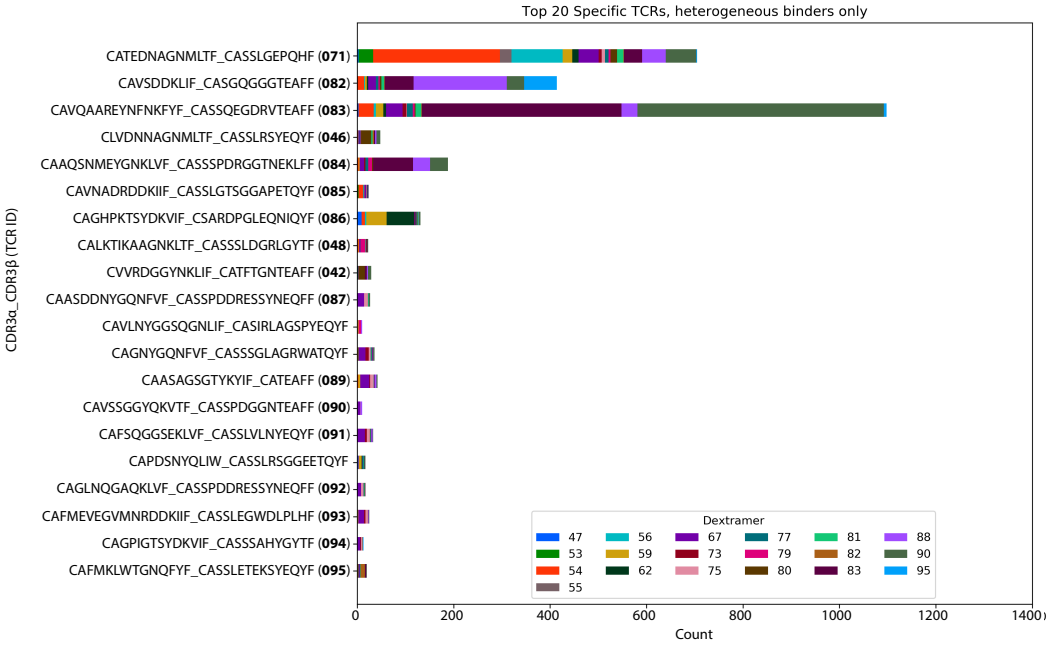


Figure J.1: Enumeration of expanded antigen-specific T cell populations from COVID-19 participants via 10X single-cell sequencing, filtered for heteregeneously binding cells

Frequencies of antigen-specific T cell populations among the top 20 most common detected clonotypes, identified by multiplexed dextramer sorting from expanded T cells for COVID-19 participants. “Dextramer” refers to the ID of the dextramer used in the legend of **Fig. 3.8B**. Compare results with homogeneously binding cells (**Fig. 3.8A**).

Appendix K

CLONED SARS-COV-2-SPECIFIC TCR SEQUENCES

TCR_ID	peptide	HLA	protein	TRAV	CDR3a	TRAJ	TRBV	CDR3b	TRBJ	TRBC	sample	status	expansion
001	LLFNKVTLA	A*02:01	S	9-2	CALNDYKLSF	20	28	CASSLTGGGYEQYF	2-7	2	GB18622	healthy	pep stim
002	RLITGRLQSL	A*02:01	S	39	CAVDANNDMRF	43	12-4	CASSPSYNEQFF	2-1	2	GB18622	healthy	pep stim
009	GYLQPRTFLL	A*24:02	S	24	CAPGSAGNMLTF	39	7-6	CASYGGGLNTEAFF	1-1	1	InCoV_003	T3	tet pool
010	GYLQPRTFLL	A*24:02	S	24	CAPGSAGNMLTF	39	29-1	CSVEVGEGYEQYF	2-7	2	InCoV_003	T3	tet pool
011	GYLQPRTFLL	A*24:02	S	19	CALSDNDYKLSF	20	7-6	CASYGGGLNTEAFF	1-1	1	InCoV_003	T3	tet pool
012	GYLQPRTFLL	A*24:02	S	19	CALSDNDYKLSF	20	29-1	CSVEVGEGYEQYF	2-7	2	InCoV_003	T3	tet pool
013	VYSSANNCTFE	A*24:02	S	12-1	CVVHSHGSSNTGKLIF	37	14	CASSQQTGRIQYF	2-4	2	InCoV_009	T3	tet pool
014	KYEQYIKWPW	A*24:02	S	24	CARNSGGYQKVTF	13	5-6	CASSLMVNEQYF	2-7	2	InCoV_009	T3	tet pool
015	QYIKWPWYI	A*24:02	S	8-2	CVVSLNYQLIW	33	27	CASSQNGETQYF	2-5	2	InCoV_009	T3	tet pool
016	FAMQMAYRF	A*24:02	S	38-2/DV8	CAYRRLEYGNKLVF	47	12-3	CASSFPGEGNIQYF	2-4	2	InCoV_001	T3	tet pool
017	FAMQMAYRF	A*24:02	S	38-2/DV8	CAYRRLEYGNKLVF	47	5-6	CASSLALAGIWSNQPQHF	1-5	1	InCoV_001	T3	tet pool
018	FAMQMAYRF	A*24:02	S	38-2/DV8	CAYRRLEYGNKLVF	47	5-6	CASSLAWYREGESVSYTF	1-2	1	InCoV_001	T3	tet pool
019	FAMQMAYRF	A*24:02	S	4	CLVGETGANNLFF	36	12-3	CASSFPGEGNIQYF	2-4	2	InCoV_001	T3	tet pool
020	FAMQMAYRF	A*24:02	S	4	CLVGETGANNLFF	36	5-6	CASSLALAGIWSNQPQHF	1-5	1	InCoV_001	T3	tet pool
021	FAMQMAYRF	A*24:02	S	4	CLVGETGANNLFF	36	5-6	CASSLAWYREGESVSYTF	1-2	1	InCoV_001	T3	tet pool
022	YYVGYLQPRTF	A*24:02	S	8-4	CAVVPRSDYKLSF	20	20-1	CSARGPTSGREEQYF	2-7	2	InCoV_001	T3	tet pool
023	YYVGYLQPRTF	A*24:02	S	8-4	CAVVPRSDYKLSF	20	19	CASSTHTDNYGYTF	1-2	1	InCoV_001	T3	tet pool
024	YYVGYLQPRTF	A*24:02	S	3	CAVRDSYYGNNRLAF	7	20-1	CSARGPTSGREEQYF	2-7	2	InCoV_001	T3	tet pool
025	YYVGYLQPRTF	A*24:02	S	3	CAVRDSYYGNNRLAF	7	19	CASSTHTDNYGYTF	1-2	1	InCoV_001	T3	tet pool
026	KYEQYIKWPW	A*24:02	S	27	CAGDPSWGKFQF	24	6-1	CASSEAGPAWAFF	1-1	1	InCoV_001	T3	tet pool
027	KYEQYIKWPW	A*24:02	S	27	CAGDPSWGKFQF	24	25-1	CASSEGQGYEQYF	2-7	2	InCoV_001	T3	tet pool
028	KYEQYIKWPW	A*24:02	S	1-2	CAVKDSNYQLIW	33	5-1	CASSFGSSYEQYF	2-7	2	InCoV_001	T3	tet pool
029	KYEQYIKWPW	A*24:02	S	8-6	CAVSDEARLMF	31	5-1	CASSFGSSYEQYF	2-7	2	InCoV_001	T3	tet pool
030	NYNYLYRLFR	A*24:02	S	1-2	CAVMDSNYQLIW	33	9	CASSVGDSLNYGYTF	1-2	1	InCoV_001	T3	tet pool
031	NYNYLYRLFR	A*24:02	S	1-2	CAPMDSNYQLIW	33	9	CASSVGDSLNYGYTF	1-2	1	InCoV_001	T3	tet pool
032	YPDKVFRSSV	B*07:02	S	27	CAGGYNNNDMRF	43	9	CASSVGGGGSTDTQYF	2-3	2	InCoV_006	T3	tet pool
033	RLITGRLQSL	A*02:01	S	13-1	CAASPNNNDMRF	43	3-1	CASSQEQEPHQETQYF	2-5	2	InCoV_002	T3	tet pool
034	RLITGRLQSL	A*02:01	S	10	CVVSPINYGQNFVF	26	3-1	CASSQEQEPHQETQYF	2-5	2	InCoV_002	T3	tet pool
035	RLITGRLQSL	A*02:01	S	17	CATDLKTSYDKVIF	50	3-1	CASSQEQEPHQETQYF	2-5	2	InCoV_002	T3	tet pool
036	RLITGRLQSL	A*02:01	S	12-2	CAGRPYDKIIF	30	3-1	CASSQEQEPHQETQYF	2-5	2	InCoV_002	T3	tet pool
037	RLITGRLQSL	A*02:01	S	13-1	CAASPNNNDMRF	43	20-1	CSARSPGELNIQYF	2-4	2	InCoV_002	T3	tet pool
038	RLITGRLQSL	A*02:01	S	10	CVVSPINYGQNFVF	26	20-1	CSARSPGELNIQYF	2-4	2	InCoV_002	T3	tet pool
039	RLITGRLQSL	A*02:01	S	17	CATDLKTSYDKVIF	50	20-1	CSARSPGELNIQYF	2-4	2	InCoV_002	T3	tet pool
040	RLITGRLQSL	A*02:01	S	12-2	CAGRPYDKIIF	30	20-1	CSARSPGELNIQYF	2-4	2	InCoV_002	T3	tet pool
041	MLAKALRKV	A*02:01	P	14/DV4	CAMRDEDSNYQLIW	33	10-3	CAISESQSGGASETQYF	2-5	2	InCoV_002	T3	tet pool
042	RARSVASQSI	B*07:02	S	8-2	CVVRDGGYNKLIF	4	10-3	CATFTGNTEAFF	1-1	1	InCoV_006	T3	tet pool
043	RARSVASQSI	B*07:02	S	4	CLVDNNAGNMLTF	39	10-3	CATFTGNTEAFF	1-1	1	InCoV_006	T3	tet pool
044	RARSVASQSI	B*07:02	S	9-2	CALGWRSGTDKLIF	34	10-3	CATFTGNTEAFF	1-1	1	InCoV_006	T3	tet pool
045	RARSVASQSI	B*07:02	S	8-2	CVVRDGGYNKLIF	4	7-9	CASSLRSYEQYF	2-7	2	InCoV_006	T3	tet pool
046	RARSVASQSI	B*07:02	S	4	CLVDNNAGNMLTF	39	7-9	CASSLRSYEQYF	2-7	2	InCoV_006	T3	tet pool
047	RARSVASQSI	B*07:02	S	9-2	CALGWRSGTDKLIF	34	7-9	CASSLRSYEQYF	2-7	2	InCoV_006	T3	tet pool
048	LPPAYTNSF	B*07:02	S	16	CALKTIKAAGNKLTF	17	5-5	CASSSLDGRLGYTF	1-2	1	InCoV_006	T3	tet pool
049	LPPAYTNSF	B*07:02	S	8-4	CAVLNYGGSQGNLIF	42	5-5	CASSSLDGRLGYTF	1-2	1	InCoV_006	T3	tet pool

Table K.1: SARS-CoV-2-specific TCRs for cloning

TCR_ID	peptide	HLA	protein	TRAV	CDR3a	TRAJ	TRBV	CDR3b	TRBJ	TRBC	sample	status	expansion
050	LPPAYTNSF	B*07:02	S	16	CALKTIKAAGNKLTF	17	12-3	CASSKKGLSNEQFF	2-1	2	InCoV_006	T3	tet pool
051	LPPAYTNSF	B*07:02	S	8-4	CAVLNYGGSQGNLIF	42	12-3	CASSKKGLSNEQFF	2-1	2	InCoV_006	T3	tet pool
052	FPQSAPHGV	B*07:02	S	27	CAGGYNNNDMRF	43	10-3	CATFTGNTEAFF	1-1	1	InCoV_006	T3	tet pool
054	FPQSAPHGV	B*07:02	S	27	CAGGYNNNDMRF	43	7-6	CASSLAFVSTGLASGANVLTF	2-6	2	InCoV_006	T3	tet pool
055	FPQSAPHGV	B*07:02	S	4	CLVDNNAGNMLTF	39	7-6	CASSLAFVSTGLASGANVLTF	2-6	2	InCoV_006	T3	tet pool
056	FPQSAPHGV	B*07:02	S	27	CAGGYNNNDMRF	43	7-9	CASSLRSYEQYF	2-7	2	InCoV_006	T3	tet pool
058	FPQSAPHGV	B*07:02	S	27	CAGGYNNNDMRF	43	7-9	CASSLETEKSYEQYF	2-7	2	InCoV_006	T3	tet pool
059	FPQSAPHGV	B*07:02	S	4	CLVDNNAGNMLTF	39	7-9	CASSLETEKSYEQYF	2-7	2	InCoV_006	T3	tet pool
060	KQIYKTPPI	A*02:01	S	4	CLVPSEQAGTALIF	15	20-1	CSASSTHSNQPQHF	1-5	1	InCoV_002	T3	tet pool
061	KQIYKTPPI	A*02:01	S	1-2	CAVRDGDSNYQLIW	33	20-1	CSASSTHSNQPQHF	1-5	1	InCoV_002	T3	tet pool
062	KQIYKTPPI	A*02:01	S	24	CALDGAGSYQLTF	28	20-1	CSASSTHSNQPQHF	1-5	1	InCoV_002	T3	tet pool
063	YLQPRTFLLK	A*02:01	S	12-3	CAIHNTDKLIF	34	20-1	CSARSPGELNIQYF	2-4	2	InCoV_002	T3	tet pool
064	YLQPRTFLLK	A*02:01	S	3	CAVRRYGANNLFF	36	20-1	CSARSPGELNIQYF	2-4	2	InCoV_002	T3	tet pool
065	YLQPRTFLLK	A*02:01	S	12-3	CAIHNTDKLIF	34	4-2	CASSPEEGAGELFF	2-2	2	InCoV_002	T3	tet pool
066	YLQPRTFLLK	A*02:01	S	3	CAVRRYGANNLFF	36	4-2	CASSPEEGAGELFF	2-2	2	InCoV_002	T3	tet pool
067	LPPAYTNSF	B*07:02	S	9-2	CALTPYSSASKIIF	3	2	CASTGREQFF	2-1	2	InCoV_022	T3	tet pool
068	LPPAYTNSF	B*07:02	S	8-4	CAVSVAPVNTGFQKLVF	8	2	CASTGREQFF	2-1	2	InCoV_022	T3	tet pool
069	LPPAYTNSF	B*07:02	S	9-2	CALTPYSSASKIIF	3	3-1	CASSKRGGGSSYNEQFF	2-1	2	InCoV_022	T3	tet pool
070	LPPAYTNSF	B*07:02	S	8-4	CAVSVAPVNTGFQKLVF	8	3-1	CASSKRGGGSSYNEQFF	2-1	2	InCoV_022	T3	tet pool
071	RLITGRLQSL	A*02:01	S	17	CATEDNAGNMLTF	39	6-6	CASSLGEPQHF	1-5	1	InCoV_047	T3	tet pool
073	YLQPRTFLLK	A*02:01	S	17	CATDRNTNAGKSTF	27	12-3	CASSLVQYNEQFF	2-1	2	InCoV_005	T3	tet pool
074	YLQPRTFLLK	A*02:01	S	1-2	CAVDKDSNYQLIW	33	12-3	CASSLVQYNEQFF	2-1	2	InCoV_005	T3	tet pool
075	YLQPRTFLLK	A*02:01	S	17	CATDRNTNAGKSTF	27	7-8	CASSLDIPSYNEQFF	2-1	2	InCoV_005	T3	tet pool
076	YLQPRTFLLK	A*02:01	S	1-2	CAVDKDSNYQLIW	33	7-8	CASSLDIPSYNEQFF	2-1	2	InCoV_005	T3	tet pool
077	YLQPRTFLLK	A*02:01	S	24	CACPPGDYKLSF	20	7-8	CASSLDIPSYNEQFF	2-1	2	InCoV_005	T3	tet pool
078	RLITGRLQSL	A*02:01	S	19	CALSEGYNFNKFYF	21	20-1	CSARDPGLEQNIQYF	2-4	2	InCoV_005	T3	tet pool
079	RLITGRLQSL	A*02:01	S	19	CALSEGYNFNKFYF	21	20-1	CSGEGGDTEAFF	1-1	1	InCoV_005	T3	tet pool
080	MLAKALRKV	A*02:01	P	16	CAPDSNYQLIW	33	13	CASSLRSGGEETQYF	2-5	2	InCoV_005	T3	tet pool
081	MLAKALRKV	A*02:01	P	1-2	CAVRDYSGAGSYQLTF	28	13	CASSLRSGGEETQYF	2-5	2	InCoV_005	T3	tet pool
082	LLFNKVTLA	A*02:01	S	12-2	CAVSDDKLIF	34	4-1	CASGQGGGTEAFF	1-1	1	GB17457	healthy	pep stim
083	RLITGRLQSL	A*02:01	S	20	CAVQAAREYNFNKFYF	21	3-1	CASSQEGDRVTEAFF	1-1	1	GB18622	healthy	pep stim
084	RLITGRLQSL	A*02:01	S	22	CAAQSNMEYGNKLVF	47	5-4	CASSSPDRGGTNEKLFF	1-4	1	GB17457	healthy	pep stim
085	MLAKALRKV	A*02:01	P	8-1	CAVNADRDDKIIF	30	11-2	CASSLGTSGGAPETQYF	2-5	2	InCoV_047	T3	tet pool
086	RLITGRLQSL	A*02:01	S	25	CAGHPKTSYDKVIF	50	20-1	CSARDPGLEQNIQYF	2-4	2	InCoV_005	T3	tet pool
087	MLAKALRKV	A*02:01	P	29/DV5	CAASDDNYGQNFVF	26	7-9	CASSPDDRESSYNEQFF	2-1	2	InCoV_002	T3	tet pool
088	MLAKALRKV	A*02:01	P	36/DV7	CAGNYGQNFVF	26	27	CASSSGLAGRWATQYF	2-3	2	InCoV_002	T3	tet pool
089	MLAKALRKV	A*02:01	P	29/DV5	CAASAGSGTYKYIF	40	12-4	CATEAFF	1-1	1	InCoV_002	T3	tet pool
090	MLAKALRKV	A*02:01	P	12-2	CAVSSGGYQKVTF	13	18	CASSPDGGNTEAFF	1-1	1	InCoV_002	T3	tet pool
091	MLAKALRKV	A*02:01	P	38-1	CAFSQGGSEKLVF	57	7-9	CASSLVLNYEQYF	2-7	2	InCoV_002	T3	tet pool
092	MLAKALRKV	A*02:01	P	35	CAGLNQGAQKLVF	54	7-9	CASSPDDRESSYNEQFF	2-1	2	InCoV_002	T3	tet pool
093	MLAKALRKV	A*02:01	P	38-1	CAFMEVEGVMNRDDKIIF	30	11-2	CASSLEGWDLPLHF	1-6	1	InCoV_002	T3	tet pool
094	MLAKALRKV	A*02:01	P	35	CAGPIGTSYDKVIF	50	12-4	CASSSAHYGYTF	1-2	1	InCoV_002	T3	tet pool
095	SPRRARSVA	B*07:02	S	38-1	CAFMKLWTGNQFYF	49	7-9	CASSLETEKSYEQYF	2-7	2	InCoV_006	T3	tet pool

Table K.2: SARS-CoV-2-specific TCRs for cloning (continued)

Appendix L

SCT EXPRESSION OF SARS-COV-2 PROTEOME EPITOPES

ID	peptide	domain	start	len	yield
1	VLQVRDVLV	nsp1	20	9	0.16
2	GLVEVEKGV	nsp1	52	9	0.39
3	VLPQLEQPY	nsp1	60	10	0.13
4	HVMVELVAEL	nsp1	83	10	0.41
5	VMVELVAEL	nsp1	84	9	0.56
6	VMVELVAELE	nsp1	84	10	0.74
7	ETLGVLVPHV	nsp1	102	10	0.24
8	TLGVLVPHV	nsp1	103	9	0.16
9	VLVPHVGEI	nsp1	106	9	0.77
10	CTLSEQLDFI	nsp2	213	10	0.35
11	TLSEQLDFI	nsp2	214	9	0.29
12	FVFPLNSII	nsp2	273	9	1.41
13	FPLNSIIKTI	nsp2	275	10	0.12
14	KLDGFMGRI	nsp2	292	9	0.27
15	YLPQNAVVKI	nsp2	359	10	0.28
16	EGLNDNLLEI	nsp2	444	10	0.11
17	GLNDNLLEI	nsp2	445	9	0.22
18	GLNDNLLEIL	nsp2	445	10	0.17
19	FKLNEEIAII	nsp2	467	10	0.13
20	KLNEEIAII	nsp2	468	9	0.48
21	KLNEEIAIIL	nsp2	468	10	0.47
22	AIILASFSA	nsp2	474	9	0.15
23	ILASFSAST	nsp2	476	9	0.28
24	FSASTSAFV	nsp2	480	9	0.34
25	STSAFVETV	nsp2	483	9	0.37
26	SILSPLYAFA	nsp2	528	10	0.92
27	ILSPLYAFA	nsp2	529	9	0.37
28	SPLYAFASEA	nsp2	531	10	0.11
29	ILDGISQYSL	nsp2	569	10	0.11
30	LRLIDAMMFT	nsp2	578	10	0.40
31	RLIDAMMFTS	nsp2	579	10	1.36
32	RLIDAMMFT	nsp2	579	9	1.22
33	AMMFTSDLA	nsp2	583	9	0.52
34	MMFTSDLAT	nsp2	584	9	1.83
35	SDLATNNLVV	nsp2	588	10	0.08
36	VVMAYITGGV	nsp2	596	10	0.17
37	VMAYITGGVV	nsp2	597	10	0.09
38	VMAYITGGV	nsp2	597	9	0.09
39	YITGGVVQL	nsp2	600	9	0.69
40	WLTNIFGTV	nsp2	612	9	0.70
41	TVYEKLKPV	nsp2	619	9	0.37
42	EFLRDGWEIV	nsp2	640	10	0.09
43	FLRDGWEIV	nsp2	641	9	0.27
44	FISTCACEI	nsp2	651	9	0.12
45	FISTCACEIV	nsp2	651	10	0.10
46	FKLVNKFLAL	nsp2	679	10	0.10
47	KLVNKFLAL	nsp2	680	9	0.33
48	KFLALCADSI	nsp2	684	10	0.10
49	FLALCADSI	nsp2	685	9	0.26
50	FLALCADSII	nsp2	685	10	0.28
51	ALCADSIII	nsp2	687	9	0.20
52	KALNLGETFV	nsp2	701	10	0.49
53	ALNLGETFV	nsp2	702	9	0.10
54	FLEGETLPT	nsp2	741	9	0.24
55	VLKTGDLQPL	nsp2	757	10	0.10
56	MMVTNNTFTL	nsp2	806	10	1.93
57	MVTNNTFTL	nsp2	807	9	1.70
58	FGDDTVIEV	nsp3	825	9	0.30
59	KSVNITFEL	nsp3	837	9	0.82
60	FELDERIDKV	nsp3	843	10	0.09
61	YTVELGTEV	nsp3	860	9	0.64
62	AVIKTLQPV	nsp3	878	9	0.61
63	TLQPVSELL	nsp3	882	9	0.32
64	IDLDEWSMAT	nsp3	895	10	0.09
65	YLFDESGEF	nsp3	906	9	0.68
66	KLASHMFCS	nsp3	915	9	0.38
67	WLDDDSQQTV	nsp3	979	10	0.30
68	TTIQTITVEV	nsp3	1000	9	0.69
69	TTVEVQPQL	nsp3	1004	9	0.58
70	QLEMELTPVV	nsp3	1011	10	0.22
71	QLEMELTPV	nsp3	1011	9	0.23
72	MELTPVVQTI	nsp3	1014	10	0.11
73	YLKLTDNVYI	nsp3	1031	10	0.24
74	MQVESDDYI	nsp3	1083	9	0.24
75	YIATNGPLKV	nsp3	1090	10	0.09
76	HNLAKHCLHV	nsp3	1108	10	0.00
77	NLAKHCLHV	nsp3	1109	9	0.14
78	NLAKHCLHVV	nsp3	1109	10	0.19
79	EVLLAPLLSA	nsp3	1142	10	0.07
80	VLLAPLLSA	nsp3	1143	9	0.93
81	LLAPLLSAGI	nsp3	1144	10	0.58
82	PLLSAGIFGA	nsp3	1147	10	0.13
83	LLSAGIFGA	nsp3	1148	9	0.48
84	LLSAGIFGAD	nsp3	1148	10	0.32
85	YLAVFDKNL	nsp3	1174	9	0.42
86	DKLVSSFLEM	nsp3	1184	10	0.00
87	KLVSSFLEM	nsp3	1185	9	0.40
88	FLEMKSEKQV	nsp3	1190	10	0.09
89	KIAEIPKEEV	nsp3	1202	10	0.19
90	FITESKPSV	nsp3	1214	9	0.48
91	FLTENLLLYI	nsp3	1248	10	0.92
92	LLYIDINGNL	nsp3	1254	10	0.28
93	NLHPDSATL	nsp3	1262	9	0.12
94	FLKKDAPYI	nsp3	1278	9	0.18
95	FLKKDAPYIV	nsp3	1278	10	0.21
96	VLTAVVIPT	nsp3	1295	9	0.16
97	MLAKALRKV	nsp3	1312	9	0.50
98	SAFYIDPSI	nsp3	1351	9	0.13
99	EIIGTVSWNL	nsp3	1366	10	0.13
100	ILGTVSWNL	nsp3	1367	9	0.00
101	RKLMPVCVET	nsp3	1386	10	0.08
102	KLMPVCVET	nsp3	1387	9	0.90
103	KLMPVCVETK	nsp3	1387	10	0.94
104	SLINTINDL	nsp3	1433	9	1.49
105	TLNDLNETL	nsp3	1437	9	0.14
106	TLNDLNETLV	nsp3	1437	10	0.22
107	TLVTMPLGVV	nsp3	1444	10	0.75
108	MPLGVVTHGL	nsp3	1448	10	0.10
109	YMRSLKVPA	nsp3	1465	9	0.10
110	SLKVPATVSV	nsp3	1468	10	0.11
111	RTIKVFTTV	nsp3	1566	9	0.75
112	YLDGADVTKI	nsp3	1598	10	0.21
113	VLPENDDTLRV	nsp3	1620	10	0.08
114	SFLGRYMSAL	nsp3	1641	10	0.07
115	FLGRYMSAL	nsp3	1642	9	0.32
116	YPQVNGLTSI	nsp3	1658	10	0.00
117	CYLATALLTL	nsp3	1674	10	0.08
118	YLATALLTL	nsp3	1675	9	0.17
119	YLATALLTLQ	nsp3	1675	10	0.18
120	ALLTLQQIEL	nsp3	1679	10	1.06
121	LLTLQQIEL	nsp3	1680	9	0.77
122	ALQDAYYRA	nsp3	1694	9	0.08
123	LILAYCNKTV	nsp3	1713	10	0.15
124	ILAYCNKTV	nsp3	1714	9	0.17
125	TMSYLFQHA	nsp3	1731	9	0.00
126	YLVQQESPFV	nsp3	1796	10	1.31
127	LVQQESPFV	nsp3	1797	9	0.19
128	PFVMMSAPPA	nsp3	1803	10	0.20
129	FVMMSAPPA	nsp3	1804	9	1.32
130	TLYCIDGALL	nsp3	1844	10	0.10
131	TLYCIDGAL	nsp3	1844	9	0.08
132	KLDGVVCTEI	nsp3	1883	10	0.18
133	FVCDNIKFA	nsp3	1930	9	0.63
134	LLHKPIVWHV	nsp3	1985	10	0.26
135	NLACEDLKPV	nsp3	2038	10	0.22
136	YVDNSSLTI	nsp3	2099	9	0.41
137	TLATHGLAAV	nsp3	2121	10	0.34
138	FLNKVVSTT	nsp3	2145	9	0.32
139	NYMPYFFTLL	nsp3	2167	10	0.12
140	YMPYFFTLL	nsp3	2168	9	0.21
141	YMPYFFTLLL	nsp3	2168	10	0.10
142	LLLQLCTFT	nsp3	2175	9	0.32
143	FCLEASFNYL	nsp3	2209	10	0.49
144	YLKSPNFSKL	nsp3	2217	10	0.07
145	KLINIIIWF	nsp3	2225	9	0.10
146	KLINIIIWFL	nsp3	2225	10	0.00
147	LINIIIWFL	nsp3	2226	9	0.08
148	IIIWFLLSV	nsp3	2229	10	0.00
149	IIWFLLLSVC	nsp3	2230	10	0.00
150	IIWFLLLSV	nsp3	2230	9	0.00
151	FLLLSVCLGS	nsp3	2233	10	0.00
152	LLLSVCLGSL	nsp3	2234	10	0.36
153	LLSVCLGSL	nsp3	2235	9	0.11
154	CLGSLIYST	nsp3	2239	9	0.07
155	GSLIYSTAAL	nsp3	2241	10	0.94
156	SLIYSTAAL	nsp3	2242	9	0.93
157	LIYSTAALGV	nsp3	2243	10	0.14
158	VLIWSNLGMPS	nsp3	2252	10	0.45
159	GYLNSTNVTI	nsp3	2269	10	0.09
160	YLNSTNVTI	nsp3	2270	9	0.10
161	YLNSTNVTIA	nsp3	2270	10	0.09
162	SLETIQITI	nsp3	2303	9	0.25
163	FGLVAEWFLA	nsp3	2321	10	0.49
164	GLVAEWFLA	nsp3	2322	9	0.24
165	LVAEWFLAYI	nsp3	2323	10	0.13
166	FLAYILFTRF	nsp3	2328	10	0.00
167	FLAYILFTR	nsp3	2328	9	0.05
168	YILFTRFFYV	nsp3	2331	10	0.21
169	ILFTRFFYVL	nsp3	2332	10	0.10
170	ILFTRFFYV	nsp3	2332	9	0.23
171	FTRFFYVLGL	nsp3	2334	10	0.28
172	VLGLAAIMQL	nsp3	2340	10	0.09
173	IMQLFFSYFA	nsp3	2346	10	0.10
174	MQLFFSYFAV	nsp3	2347	10	0.32
175	MQLFFSYFA	nsp3	2347	9	0.47
176	QLFFSYFAV	nsp3	2348	9	0.09
177	QLFFSYFAVH	nsp3	2348	10	0.00
178	FSYFAVHFI	nsp3	2351	9	0.13
179	FISNSWLMWL	nsp3	2358	10	0.00
180	SWLMWLIINL	nsp3	2362	10	0.00
181	WLMWLIINL	nsp3	2363	9	0.00
182	WLMWLIINLV	nsp3	2363	10	0.11
183	LMWLIINLV	nsp3	2364	9	0.00
184	LMWLIINLVQ	nsp3	2364	10	0.13
185	WLIINLVQMA	nsp3	2366	10	0.00
186	YQMAPISAMV	nsp3	2372	10	0.32
187	QMAPISAMV	nsp3	2373	9	0.00
188	AMVRMYMFFA	nsp3	2379	10	0.10
189	MVRMYIFFA	nsp3	2380	9	0.00
190	YIFFASFYYV	nsp3	2384	10	0.13
191	IFFASFYYV	nsp3	2385	9	0.07
192	YVVWKSYVHV	nsp3	2391	10	0.15
193	VVWKSYVHVV	nsp3	2392	10	0.55
194	YVWKSYVHV	nsp3	2392	9	0.85
195	SQLMCQPILL	nsp3	2562	10	0.34
196	QLMCQPILL	nsp3	2563	9	0.41
197	QLMCQPILLL	nsp3	2563	10	0.43
198	LMCQPILLL	nsp3	2564	9	0.35
199	PILLLDQALV	nsp3	2568	10	0.17
200	ILLLDQALV	nsp3	2569	9	0.87
201	LLDQALVSDV	nsp3	2571	10	0.27
202	KMFDAYVNT	nsp3	2589	9	0.88
203	VNTFSSTFNV	nsp3	2595	10	0.09
204	NTFSSTFNV	nsp3	2596	9	0.28
205	SLDNVLSTFI	nsp3	2625	10	0.29
206	VLSTFISAA	nsp3	2629	9	0.13
207	KLSHQSDIEV	nsp3	2656	10	0.40
208	KLTCATTRQV	nsp3	2741	10	0.08
209	KQLIKVTLV	nsp4	2771	9	0.48
210	QLIKVTLVFL	nsp4	2772	10	0.00
211	KVTLVFLFV	nsp4	2775	9	0.00
212	TLVFLFVAAI	nsp4	2777	10	0.00
213	TLVFLFVAA	nsp4	2777	9	0.00
214	LVFLFVAAI	nsp4	2778	9	0.00
215	FLFVAAIFYL	nsp4	2780	10	0.00
216	LFVAAIFYL	nsp4	2781	10	0.00
217	FVAAIFYLI	nsp4	2782	9	0.00
218	FVAAIFYLIT	nsp4	2782	10	0.09
219	AAIFYLITPV	nsp4	2784	10	0.48
220	AIFYLITPVH	nsp4	2785	10	0.14
221	AIFYLITPV	nsp4	2785	9	0.15
222	FYLITPVHVM	nsp4	2787	10	0.10
223	YLITPVHVM	nsp4	2788	9	2.11
224	YTNDKACPL	nsp4	2845	9	0.28
225	LIAAVITREV	nsp4	2853	10	0.11
226	FVVPGLPGTI	nsp4	2864	10	1.43
227	FVVPGLPGT	nsp4	2864	9	1.60
228	HFLPRVFSAV	nsp4	2883	10	0.11
229	FLPRVFSAV	nsp4	2884	9	0.30
230	FLPRVFSAVG	nsp4	2884	10	0.31
231	SKLIEYTDFA	nsp4	2900	10	0.08
232	KLIEYTDFA	nsp4	2901	9	1.58
233	KLIEYTDFAT	nsp4	2901	10	1.62
234	YTDFATSACV	nsp4	2905	10	0.24
235	SIICQFPNTYL	nsp4	2960	10	1.03
236	YLEGSVRVV	nsp4	2968	9	0.42
237	RSLPGVFCGV	nsp4	3012	10	0.33
238	SLPGVFCGV	nsp4	3013	9	1.09
239	SLPGVFCGVD	nsp4	3013	10	1.01
240	GVFCGVDAV	nsp4	3016	9	0.09

Table L.1: Top HLA-A*02:01 9-mer or 10-mer epitopes from SARS-CoV-2 proteome predicted by NetMHC4.0 (#1-240)

Peptides selected for SCT expression are indexed according to relative proteomic position. Epitopes from accessory proteins are listed at the end of table. 'Yield' column denotes the relative expression of each SCT based on ratio of protein gel band intensity to positive control band (see **Appendix E** for example).

ID	peptide	domain	start	len	yield
241	NULTNMFTPL	nsp4	3025	10	0.56
242	LLTNMFTPLI	nsp4	3026	10	0.69
243	LLTNMFTPL	nsp4	3026	9	1.16
244	NMFTPLIQPI	nsp4	3029	10	0.67
245	IVAGGIVAI	nsp4	3047	9	0.17
246	IVAGGIVAIV	nsp4	3047	10	0.14
247	VVAFNTLLFL	nsp4	3077	10	0.10
248	VAFNTLLFL	nsp4	3078	9	0.26
249	TLFLMSFTV	nsp4	3082	10	0.00
250	TLLFLMSFT	nsp4	3082	9	0.00
251	LLFLMSFTV	nsp4	3083	9	0.00
252	LLFLMSFTVL	nsp4	3083	10	0.00
253	LFLMSFTVLC	nsp4	3084	10	0.00
254	FLMSFTVLCL	nsp4	3085	10	0.00
255	FLMSFTVLC	nsp4	3085	9	0.22
256	LMSFTVLCL	nsp4	3086	9	0.09
257	LMSFTVLCLT	nsp4	3086	10	0.00
258	SFTVLCLTPV	nsp4	3088	10	0.08
259	FTVLCLTPV	nsp4	3089	9	0.22
260	FTVLCLTPVY	nsp4	3089	10	0.17
261	LCLTPVYSFL	nsp4	3092	10	0.13
262	CLTPVYSFL	nsp4	3093	9	0.51
263	TPVYSFLPGV	nsp4	3095	10	0.08
264	YSFLPGVYSV	nsp4	3098	10	0.17
265	SFLPGVYSV	nsp4	3099	9	0.18
266	SFLPGVYSV	nsp4	3099	9	0.30
267	FLPGVYSVI	nsp4	3100	9	0.19
268	GVYSVIYLYL	nsp4	3103	10	0.00
269	VIYLYLTPYL	nsp4	3107	10	0.00
270	IYLYLTPYLT	nsp4	3108	10	0.00
271	YLYLTPYLTN	nsp4	3109	10	0.12
272	YLYLTPYLT	nsp4	3109	9	0.16
273	YLTPYLTNDV	nsp4	3111	10	0.40
274	PYLTNDVSFL	nsp4	3114	10	0.09
275	YLTNDVSFL	nsp4	3115	9	1.03
276	YLTNDVSFLA	nsp4	3115	10	1.24
277	LTNDVSFLA	nsp4	3116	9	0.21
278	SFLAHIQWMV	nsp4	3121	10	0.10
279	FLAHIQWMV	nsp4	3122	9	0.25
280	FLAHIQWMVM	nsp4	3122	10	0.19
281	IQWMVMFTPL	nsp4	3126	10	0.00
282	QWMVMFTPLV	nsp4	3127	10	0.00
283	WMVMFTPLV	nsp4	3128	9	0.10
284	WMVMFTPLVP	nsp4	3128	10	0.67
285	FTPLVPFWI	nsp4	3132	9	0.08
286	TPLVPFWITI	nsp4	3133	10	0.00
287	WITIAYIICI	nsp4	3139	10	0.00
288	ITIAYIICI	nsp4	3140	9	0.00

ID	peptide	domain	start	len	yield
289	TFLLNKEMYL	nsp4	3182	10	0.09
290	FLLNKEMYL	nsp4	3183	9	0.69
291	FLLNKEMYLK	nsp4	3183	10	1.11
292	LLNKEMYLKL	nsp4	3184	10	0.09
293	YLKLRSDVLL	nsp4	3190	10	0.15
294	KLRSDVLLPL	nsp4	3192	10	0.12
295	GLWLDDVVYC	nsp5	3292	10	0.09
296	MLNPNYEDLL	nsp5	3312	10	0.83
297	MLNPNYEDL	nsp5	3312	9	0.67
298	FLVQAGNVQL	nsp5	3329	10	0.85
299	SMQNCVLKL	nsp5	3344	9	0.17
300	RIQPGQTFSV	nsp5	3368	10	0.17
301	IQPGQTFSV	nsp5	3369	9	0.18
302	SFLNGSCGSV	nsp5	3402	10	0.07
303	FLNGSCGSV	nsp5	3403	9	0.13
304	FLNGSCGSVG	nsp5	3403	10	0.16
305	YMHHMELPT	nsp5	3424	9	1.19
306	ITVNVLAWL	nsp5	3463	9	0.24
307	VNVLAWLYAA	nsp5	3465	10	0.00
308	NVLAWLYAAV	nsp5	3466	10	0.08
309	NVLAWLYAA	nsp5	3466	9	0.30
310	VLAWLYAAVI	nsp5	3467	10	0.08
311	VLAWLYAAV	nsp5	3467	9	0.10
312	AVINGDRWFL	nsp5	3474	10	0.73
313	VINGDRWFL	nsp5	3475	9	0.10
314	WFLNRFTTTL	nsp5	3481	10	0.11
315	FLNRFTTTL	nsp5	3482	9	0.24
316	FLNRFTTTLN	nsp5	3482	10	0.27
317	TTLNDFNLV	nsp5	3488	9	0.17
318	TLNDFNLVAM	nsp5	3489	10	0.24
319	TLNDFNLVA	nsp5	3489	9	0.17
320	AVLDMCASL	nsp5	3523	9	0.75
321	WLLLTLITSL	nsp6	3582	10	0.00
322	LLLTLITSL	nsp6	3583	9	0.09
323	LLLTLITSLL	nsp6	3583	10	0.00
324	LLTLITSLL	nsp6	3584	9	0.00
325	LLTLITSLLV	nsp6	3584	10	0.11
326	TLLTSLLVL	nsp6	3586	9	0.14
327	TLLTSLLVLV	nsp6	3586	10	0.11
328	LLTSLLVLV	nsp6	3587	9	0.08
329	LLTSLLVLVQ	nsp6	3587	10	0.11
330	SLLVLVQST	nsp6	3590	9	0.45
331	VLVQSTQWSL	nsp6	3593	10	0.49
332	LVQSTQWSL	nsp6	3594	9	0.71
333	STQWSLFFFL	nsp6	3597	10	0.10
334	TQWSLFFFL	nsp6	3598	9	0.11
335	SLFFFLYENA	nsp6	3601	10	0.17
336	FFLYENAFLP	nsp6	3604	10	0.11

ID	peptide	domain	start	len	yield
337	FLYENAFLPF	nsp6	3605	10	1.15
338	FLYENAFLP	nsp6	3605	9	1.21
339	FLPFAMGIL	nsp6	3611	9	0.18
340	FLPFAMGILA	nsp6	3611	10	0.13
341	FAMGILAMSA	nsp6	3614	10	0.14
342	GILAMSAFA	nsp6	3617	9	0.15
343	AMSAFAMMFV	nsp6	3620	10	0.22
344	MSAFAMMFV	nsp6	3621	9	0.16
345	FLCLFLPSL	nsp6	3635	10	0.00
346	FLCLFLPS	nsp6	3635	9	0.00
347	LFLLPSLATV	nsp6	3638	10	0.21
348	FLLPSLATV	nsp6	3639	9	0.55
349	FLLPSLATVA	nsp6	3639	10	0.81
350	SLATVAYFNM	nsp6	3643	10	0.38
351	ATVAYFNMV	nsp6	3645	9	0.08
352	NMVYMPASWV	nsp6	3651	10	0.40
353	MVYMPASWV	nsp6	3652	9	0.10
354	YMPASWVMRI	nsp6	3654	10	0.24
355	WVMRIMTWL	nsp6	3659	9	0.10
356	MRIMTWLDMV	nsp6	3661	10	0.10
357	RIMTWLDMV	nsp6	3662	9	0.87
358	RIMTWLDMVD	nsp6	3662	10	0.00
359	TWLDMVDTSL	nsp6	3665	10	0.09
360	WLDMVDTSL	nsp6	3666	9	0.30
361	FKLKDCVMVA	nsp6	3677	10	0.00
362	KLKDCVMVA	nsp6	3678	9	0.09
363	CVMVASAVV	nsp6	3682	9	0.69
364	VMVASAVVLL	nsp6	3683	10	0.12
365	VMVASAVVL	nsp6	3683	9	0.12
366	YASAVVLLI	nsp6	3685	9	0.06
367	LLILMTARTV	nsp6	3691	10	0.03
368	WTLMNVLTLV	nsp6	3709	10	0.13
369	TLMNVLTLV	nsp6	3710	9	0.26
370	TLMNVLTLVY	nsp6	3710	10	0.21
371	MNVLTLVYKV	nsp6	3712	10	0.07
372	NVLTLVYKV	nsp6	3713	9	0.44
373	ALDQAISMWA	nsp6	3726	10	0.09
374	ISMWALLISV	nsp6	3731	10	0.00
375	SMWALLISV	nsp6	3732	9	0.18
376	SMWALLISVT	nsp6	3732	10	0.13
377	MFLARGIVFM	nsp6	3752	10	0.10
378	FLARGIVFM	nsp6	3753	9	0.36
379	FLARGIVFMC	nsp6	3753	10	0.16
380	VFMCVEYCPI	nsp6	3759	10	0.00
381	FMCVEYCPIF	nsp6	3760	10	0.27
382	FMCVEYCPI	nsp6	3760	9	0.35
383	NTLQCIMLV	nsp6	3774	9	0.11
384	FLGYFCTCYF	nsp6	3785	10	0.00

ID	peptide	domain	start	len	yield
385	FCLLNRYFRL	nsp6	3797	10	0.09
386	CLLNRYFRL	nsp6	3798	9	0.68
387	LLNRYFRLTL	nsp6	3799	10	0.08
388	RLTLGVYDYL	nsp6	3805	10	0.11
389	LTLGVYDYLV	nsp6	3806	10	0.77
390	LTLGVYDYL	nsp6	3806	9	0.00
391	TLGVYDYLV	nsp6	3807	9	0.17
392	GVYDYLVST	nsp6	3809	9	0.31
393	FKLNVKLLGV	nsp6	3838	10	0.11
394	KLNIKLLGV	nsp6	3839	9	0.50
395	KMSDVKCTSV	nsp7	3861	10	0.64
396	VVLSVLQQL	nsp7	3870	10	0.00
397	VLSVLQQL	nsp7	3871	9	0.53
398	LLSVLQQLRV	nsp7	3872	10	0.26
399	SKLWAQCVQL	nsp7	3885	10	0.00
400	KLWAQCVQLH	nsp7	3886	10	0.18
401	KLWAQCVQL	nsp7	3886	9	0.17
402	QLHNDILLA	nsp7	3893	9	0.12
403	ILLAKDTTEA	nsp7	3898	10	0.31
404	LLAKDTTEA	nsp7	3899	9	1.03
405	EKMVSLLSVL	nsp7	3909	10	0.11
406	KMVSLLSVL	nsp7	3910	9	0.19
407	KMVSLLSVLL	nsp7	3910	10	0.09
408	VSLLSVLLSM	nsp7	3912	10	0.09
409	SLLSVLLSM	nsp7	3913	9	0.31
410	SLLSVLLSMQ	nsp7	3913	10	0.41
411	VLLSMQGAV	nsp7	3917	9	0.57
412	SLPSYAAFA	nsp8	3950	9	0.36
413	SAMQTMLFTM	nsp8	4027	10	0.18
414	AMQTMLFTM	nsp8	4028	9	0.12
415	AMQTMLFTML	nsp8	4028	10	0.36
416	TMLFTMLRKL	nsp8	4031	10	0.10
417	MLFTMLRKL	nsp8	4032	9	0.22
418	KLDNDALNNI	nsp8	4039	10	0.22
419	ALNNIINNA	nsp8	4044	9	0.15
420	FTYASALWEI	nsp8	4089	10	0.55
421	SALWEIQQVV	nsp8	4093	10	0.81
422	ALWEIQQVV	nsp8	4094	9	0.32
423	ALWEIQQVVD	nsp8	4094	10	0.25
424	SWDNSPNLA	nsp8	4115	9	0.58
425	NLAWPLIVTA	nsp8	4121	10	0.63
426	FVLALLSDL	nsp9	4180	9	0.38
427	ALLSDLQDL	nsp9	4183	9	1.38
428	MVLGSLAATV	nsp9	4241	10	1.48
429	VLGSLAATV	nsp9	4242	9	0.12
430	TVLSFCAFAV	nsp10	4265	10	0.20
431	TVLSFCAFA	nsp10	4265	9	0.30
432	VLSFCAFAV	nsp10	4266	9	0.13

ID	peptide	domain	start	len	yield
433	VLSFCAFAVD	nsp10	4266	10	0.14
434	DVLASGGQDI	nsp10	4282	10	0.12
435	YLASGGQDI	nsp10	4283	9	1.08
436	FTLKNTVCTV	nsp10	4363	10	0.79
437	TLKNTVCTV	nsp10	4364	9	0.09
438	FLNRVCGVSA	nsp12	4399	10	0.34
439	DNLIDSYFVV	nsp12	4455	10	0.11
440	NLIDSYFVV	nsp12	4456	9	1.15
441	NLIDSYFVVK	nsp12	4456	10	0.22
442	YQHEETIYNL	nsp12	4472	10	0.41
443	YNLLKDCPAV	nsp12	4479	10	0.00
444	NLLKDCPAV	nsp12	4480	9	0.24
445	RLTKYTMADL	nsp12	4510	10	0.12
446	KYTMADLVYA	nsp12	4513	10	0.11
447	YTMADLVYAL	nsp12	4514	10	0.00
448	YTMADLVYA	nsp12	4514	9	0.61
449	TMADLVYAL	nsp12	4515	9	0.12
450	TMADLVYALR	nsp12	4515	10	0.14
451	FVENPDILRV	nsp12	4557	10	0.16
452	AMRNAGIVGV	nsp12	4587	10	0.13
453	YSLLMPILTL	nsp12	4630	10	1.13
454	SLLMPILTL	nsp12	4631	9	1.49
455	SLLMPILTLT	nsp12	4631	10	1.01
456	LLMPILTLT	nsp12	4632	9	1.05
457	LLMPILTLTR	nsp12	4632	10	1.03
458	CILHCANFNV	nsp12	4698	10	0.57
459	IILHCANFNVL	nsp12	4699	10	0.19
460	IILHCANFNV	nsp12	4699	9	0.26
461	VLFSTVFPP	nsp12	4707	9	0.51
462	VLFSTVFPPT	nsp12	4707	10	0.61
463	KIFVDGVPFV	nsp12	4724	10	0.39
464	IFVDGVPFVV	nsp12	4725	10	0.59
465	IFVDGVPFV	nsp12	4725	9	0.57
466	FVDGVPFVV	nsp12	4726	9	0.16
467	FVDGVPFVVS	nsp12	4726	10	0.17
468	RLSFKELLV	nsp12	4757	9	0.16
469	LLVYAADPAM	nsp12	4763	10	0.17
470	FSVAALTNNV	nsp12	4788	10	0.22
471	SVAALTNNV	nsp12	4789	9	0.18
472	TMCDIRQLL	nsp12	4854	9	0.39
473	IRQLLFVVEV	nsp12	4858	10	0.03
474	RQLLFVVEV	nsp12	4859	9	0.25
475	RQLLFVVEVV	nsp12	4859	10	0.17
476	QLLFVVEVV	nsp12	4860	9	0.11
477	ALFAYTKRNV	nsp12	4918	10	0.11
478	TQMNLKYAI	nsp12	4932	9	0.18
479	KLLKSIAAT	nsp12	4966	9	1.05
480	NMLKTVYSDV	nsp12	4992	10	0.48

Table L.2: Top HLA-A*02:01 9-mer or 10-mer epitopes from SARS-CoV-2 proteome predicted by NetMHC4.0 (#241-480)

Peptides selected for SCT expression are indexed according to relative proteomic position. Epitopes from accessory proteins are listed at the end of table. 'Yield' column denotes the relative expression of each SCT based on ratio of protein gel band intensity to positive control band (see **Appendix E** for example).

ID	peptide	domain	start	len	yield
481	AMPNMLRIMA	nsp12	5017	10	0.15
482	NMLRIMASL	nsp12	5020	9	0.69
483	NMLRIMASLV	nsp12	5020	10	0.25
484	MLRIMASLVL	nsp12	5021	10	0.14
485	LRIMASLVLA	nsp12	5022	10	0.11
486	RIMASLVLA	nsp12	5023	9	0.19
487	SLSHRFYRL	nsp12	5039	9	0.20
488	YRLANECAQV	nsp12	5045	10	0.10
489	RLANECAQV	nsp12	5046	9	0.19
490	MVMCGGSLVV	nsp12	5058	10	0.50
491	VMCGGSLVV	nsp12	5059	9	0.25
492	VMCGGSLVVK	nsp12	5059	10	0.21
493	SVFNICQAV	nsp12	5084	9	0.73
494	DFVNEFYAYL	nsp12	5132	10	0.09
495	FVNEFYAYL	nsp12	5133	9	0.30
496	FVNEFYAYLR	nsp12	5133	10	0.27
497	YLRKHFSMMI	nsp12	5140	10	0.09
498	FSMMILSDDA	nsp12	5145	10	0.12
499	SMMILSDDAV	nsp12	5146	10	1.39
500	SMMILSDDA	nsp12	5146	9	0.34
501	MMILSDDAVV	nsp12	5147	10	1.91
502	MMILSDDAV	nsp12	5147	9	1.82
503	MILSDDAVV	nsp12	5148	9	0.96
504	SVLYYQNNV	nsp12	5176	9	0.76
505	VLYYQNNVFM	nsp12	5177	10	0.29
506	FMSEAKCWT	nsp12	5185	9	0.38
507	MLVKQGDDV	nsp12	5210	10	0.37
508	KDGDDVVYL	nsp12	5213	9	0.15
509	YLPYPDPSRI	nsp12	5220	10	0.22
510	RILGAGCFV	nsp12	5228	9	0.25
511	TLMIERFVSL	nsp12	5245	10	0.63
512	LMIERFVSL	nsp12	5246	9	1.00
513	LMIERFVSLA	nsp12	5246	10	0.79
514	VSLAIDAYPL	nsp12	5252	10	0.24
515	SLAIDAYPL	nsp12	5253	9	0.48
516	YADVFHLYL	nsp12	5269	9	0.29
517	HMLDMYSVM	nsp12	5290	9	1.20
518	HMLDMYSVML	nsp12	5290	10	1.27
519	MLDMYSVML	nsp12	5291	9	1.08
520	MLDMYSVMLT	nsp12	5291	10	1.05
521	VLCNSQTSL	nsp13	5330	9	0.31
522	KLVSLSVNPYV	nsp13	5364	10	0.79
523	LVSLSVNPY	nsp13	5365	9	1.27
524	LVSLSVNPYVC	nsp13	5365	10	0.95
525	YLGGMSYYC	nsp13	5388	9	0.17
526	ILANTCTERL	nsp13	5445	10	0.14
527	KLFAAETLKA	nsp13	5455	10	0.59
528	FKLSYGIATV	nsp13	5469	10	0.14
529	KLSYGIATV	nsp13	5470	9	0.48
530	KLSYGIATVR	nsp13	5470	10	0.59
531	VLSDRELHL	nsp13	5481	9	0.21
532	YKLNVGDYFV	nsp13	5541	10	0.13
533	KLNVGDYFVL	nsp13	5542	10	0.44
534	KLNVGDYFV	nsp13	5542	9	0.47
535	VLTSHTVMPL	nsp13	5550	10	0.53
536	LTSHTVMPL	nsp13	5551	9	0.45
537	VMPLSAPTL	nsp13	5556	9	0.25
538	VMPLSAPTLV	nsp13	5556	10	0.24
539	TLVPQEHYV	nsp13	5563	9	1.09
540	YVRITGLYPT	nsp13	5570	10	0.09
541	IGLALYYPSA	nsp13	5617	10	0.09
542	GLALYYPSA	nsp13	5618	9	0.15
543	ALYYPSARV	nsp13	5620	10	0.20
544	ALYYPSARI	nsp13	5620	9	0.19
545	YTACSHAAV	nsp13	5630	9	0.23
546	ALCEKALKYL	nsp13	5640	10	0.16
547	YLPIDKCSRI	nsp13	5648	10	0.09
548	TLEQVVFCTV	nsp13	5675	10	0.12
549	QYVFCTVNAL	nsp13	5678	10	0.00
550	YVFCTVNAL	nsp13	5679	9	0.32
551	ALPETTADI	nsp13	5686	9	0.15
552	ALPETTADIV	nsp13	5686	10	0.16
553	VVFDEISMA	nsp13	5695	9	0.40
554	SMATNYDLSV	nsp13	5701	10	0.60
555	MATNYDLSV	nsp13	5702	9	0.37
556	YDLSVVNARL	nsp13	5706	10	0.11
557	RLRAKHYVI	nsp13	5714	10	0.00
558	TLEPEYFNSV	nsp13	5740	10	0.13
559	IVDTVSALV	nsp13	5772	9	0.24
560	AVASKILGL	nsp13	5844	9	0.14
561	KILGLPTQTV	nsp13	5848	10	1.47
562	ILGLPTQTV	nsp13	5849	9	0.15
563	RDLYDKLQFT	nsp13	5903	10	0.08
564	KLQFTSLEI	nsp13	5908	9	0.70
565	GLFKDCSKV	nsp14	5931	9	0.16
566	GLCVDIPGI	nsp14	5962	9	0.19
567	RLISMMGFKM	nsp14	5978	10	0.34
568	MMGFKMNYQV	nsp14	5982	10	0.30
569	YQVNGYPNM	nsp14	5989	9	0.16
570	LQLGFSTGV	nsp14	6032	9	1.22
571	NLVAVPTGYV	nsp14	6041	10	0.32
572	HLPLMYKGL	nsp14	6073	10	0.39
573	LMYKGLPWNV	nsp14	6077	10	0.87
574	GLPWNVVRI	nsp14	6081	9	0.12
575	QMLSDTLKNL	nsp14	6093	10	0.09
576	MLSDTLKNL	nsp14	6094	9	0.80
577	KNLSDRVVFV	nsp14	6100	10	0.11
578	NLSDRVVFVL	nsp14	6101	10	0.29
579	NLSDRVVFV	nsp14	6101	9	0.33
580	FVLWAHGFEL	nsp14	6108	10	0.76
581	VLWAHGFEL	nsp14	6109	9	0.16
582	VLWAHGFELT	nsp14	6109	10	0.15
583	FELTSMKYFV	nsp14	6115	10	0.14
584	ELTSMKYFV	nsp14	6116	9	0.19
585	YVVNPFMIDV	nsp14	6160	10	1.03
586	FMIDVQQWGF	nsp14	6165	10	1.43
587	FMIDVQQWG	nsp14	6165	9	1.07
588	DAIMTRCLAV	nsp14	6198	10	0.17
589	AIMTRCLAV	nsp14	6199	9	0.75
590	RCLAVHECFV	nsp14	6203	10	0.09
591	CLAVHECFV	nsp14	6204	9	0.23
592	RVDWTIEYPI	nsp14	6214	10	0.15
593	AALLADKFPV	nsp14	6244	10	1.36
594	ALLADKFPV	nsp14	6245	10	1.41
595	LLADKFPV	nsp14	6245	9	1.34
596	LLADKFPVL	nsp14	6246	9	0.82
597	KQLPFFYYS	nsp14	6365	9	0.77
598	QVVSDIDYV	nsp14	6383	9	0.11
599	YRLYLDAYNM	nsp14	6416	10	0.12
600	RLYLDAYNM	nsp14	6417	9	0.35
601	LYLDAYNMMI	nsp14	6418	10	0.18
602	YLDAYNMMI	nsp14	6419	9	0.23
603	YLDAYNMMIS	nsp14	6419	10	0.23
604	NMMISAGFSL	nsp14	6424	10	0.77
605	MMISAGFSLW	nsp14	6425	10	1.56
606	MMISAGFSL	nsp14	6425	9	1.17
607	MISAGFSLWV	nsp14	6426	10	0.20
608	YNLWNTFTRL	nsp14	6442	10	0.19
609	NLWNTFTRL	nsp14	6443	9	0.39
610	SLENVAFNV	nsp15	6453	9	0.37
611	SLENVAFNVV	nsp15	6453	10	0.38
612	KVDGVDVEL	nsp15	6486	9	0.25
613	TLPVNVAFEL	nsp15	6500	10	0.20
614	SQMEIDFLEL	nsp15	6659	10	0.69
615	QMEIDFLELA	nsp15	6660	10	0.09
616	FLELAMDEFI	nsp15	6665	10	0.25
617	KLEGYAFEHI	nsp15	6678	10	0.14
618	SQLGGGLHLL	nsp15	6695	9	0.58
619	DLLDDFVEI	nsp15	6748	10	0.10
620	LLDDFVEI	nsp15	6749	9	1.35
621	LLDDFVEII	nsp15	6749	10	1.60
622	LDDFVEII	nsp15	6750	9	0.40
623	FMLWCKDGHV	nsp15	6781	10	0.73
624	MLWCKDGHV	nsp15	6782	9	0.21
625	NLYKMQRMLL	nsp16	6811	10	0.00
626	TLPKGIMMNV	nsp16	6833	10	0.26
627	IMMNVAKYT	nsp16	6838	9	0.43
628	QLCQYLNTL	nsp16	6847	9	0.44
629	QYLNTLTLAV	nsp16	6850	10	0.13
630	YLNTLTLAV	nsp16	6851	9	0.42
631	YLNTLTLAVP	nsp16	6851	10	0.53
632	TLTLAVPYNM	nsp16	6854	10	0.27
633	TLAVPYNMRV	nsp16	6856	10	0.40
634	QWLPTGTLLV	nsp16	6885	10	0.09
635	WLPTGTLLV	nsp16	6886	9	0.64
636	FVSDADSTL	nsp16	6901	9	0.72
637	STLIGDCATV	nsp16	6907	10	0.47
638	TLIGDCATV	nsp16	6908	9	1.13
639	TLIGDCATVH	nsp16	6908	10	1.22
640	KLALGGSVAI	nsp16	6958	10	0.61
641	KITEHSWNA	nsp16	6968	9	0.65
642	KLMGHFAWWT	nsp16	6980	10	0.65
643	LMGHFAWWTA	nsp16	6981	10	0.32
644	LMGHFAWWT	nsp16	6981	9	0.23
645	FLIGCNYLG	nsp16	7003	9	0.38
646	YIFWRNTNPI	nsp16	7026	10	0.17
647	SLFDMSKFPL	nsp16	7041	10	1.03
648	FDMSKFPLKL	nsp16	7043	10	0.11
649	KLRGTAVMSL	nsp16	7051	10	0.13
650	GQINDMILSL	nsp16	7063	10	0.52
651	QINDMILSL	nsp16	7064	9	0.33
652	VVISSDVLV	nsp16	7086	9	0.85
653	MFVFLVLLPL	S	7097	10	0.00
654	FVFLVLLPLV	S	7098	10	0.00
655	FVFLVLLPL	S	7098	9	0.00
656	YTNSFTRGV	S	7124	9	0.21
657	VLHSTQDLFL	S	7143	10	0.00
658	FLPFFSNVT	S	7151	9	0.14
659	VTWFHAIHV	S	7158	9	0.15
660	TLDSKTQSL	S	7205	9	0.13
661	LLVNNATNV	S	7213	10	0.18
662	EFQFCNDPFL	S	7228	10	0.00
663	FQFCNDPFL	S	7229	9	1.21
664	FQFCNDPFLG	S	7229	10	1.20
665	SWMESEFRV	S	7247	9	0.09
666	KYSKHTPI	S	7298	9	0.38
667	ALEPLVDLPI	S	7318	10	0.07
668	TLLALHRSYL	S	7336	10	0.34
669	LLALHRSYL	S	7337	9	0.25
670	WTAGAAAYYV	S	7354	10	0.08
671	GYLQPRTFLL	S	7364	10	0.13
672	YLQPRTFLL	S	7365	9	0.33
673	YLQPRTFLLK	S	7365	10	0.55
674	GIYQTSNFRV	S	7407	10	0.09
675	VLYNSASFST	S	7463	10	0.28
676	KLNDLCFTNV	S	7482	10	0.21
677	KIADYNYKL	S	7513	9	0.50
678	KLPDDFTGCV	S	7520	10	0.26
679	YRVVVLSFEL	S	7604	10	0.00
680	RVVVLSFEL	S	7605	9	0.11
681	VLSFELLHA	S	7608	9	0.10
682	FELLHAPATV	S	7611	10	0.16
683	ELLHAPATV	S	7612	9	1.01
684	YQDVNCTEV	S	7708	9	0.86
685	QSIIAYTMSL	S	7786	10	0.00
686	SIIAYTMSL	S	7787	9	0.88
687	AIPTNFTISV	S	7809	10	0.10
688	FTISVTTEI	S	7814	9	1.73
689	FTISVTTEIL	S	7814	10	1.66
690	SVTTEILPV	S	7817	9	0.90
691	KTSVDCTMYI	S	7829	10	0.00
692	KQIYKTPPI	S	7882	9	0.65
693	FIEDLLFNKV	S	7913	10	0.11
694	DLLFNKVTLA	S	7916	10	0.00
695	LLFNKVTLA	S	7917	9	1.75
696	GILTVLPPIL	S	7953	9	0.13
697	MIAQVTSAL	S	7965	9	0.96
698	MIAQVTSALL	S	7965	10	1.08
699	ITSGWTFGA	S	7978	9	0.12
700	ALQIPFAMQM	S	7989	10	1.74
701	MQMAYRFNGI	S	7996	10	0.39
702	VLYENQKLI	S	8011	9	0.15
703	KLIANQFNSA	S	8017	10	1.02
704	SLSSTASAL	S	8033	9	0.25
705	KQLSSNFGA	S	8060	9	0.21
706	SVLNDILSRL	S	8071	10	0.22
707	VLNDILSRL	S	8072	9	0.11
708	SRLDKVEAEV	S	8078	10	0.11
709	RLDKVEAEV	S	8079	9	0.24
710	RLITGRLQSL	S	8091	10	1.06
711	GRLQSLQTYV	S	8095	10	0.12
712	RLQSLQTYV	S	8096	9	0.18
713	YHLMSFPQSA	S	8143	10	0.08
714	HLMSFPQSA	S	8144	9	1.22
715	GVVFLHVTYV	S	8155	10	0.14
716	VVFLHVTYV	S	8156	9	0.08
717	VFLHVTYVPA	S	8157	10	0.00
718	FLHVTYVPA	S	8158	9	0.73
719	FVSNGTHWFV	S	8191	10	0.53
720	IITTDNTFV	S	8210	9	0.10

Table L.3: Top HLA-A*02:01 9-mer or 10-mer epitopes from SARS-CoV-2 proteome predicted by NetMHC4.0 (#481-720)

Peptides selected for SCT expression are indexed according to relative proteomic position. Epitopes from accessory proteins are listed at the end of table. 'Yield' column denotes the relative expression of each SCT based on ratio of protein gel band intensity to positive control band (see **Appendix E** for example.

ID	peptide	domain	start	len	yield
721	FVSGNCDVV	S	8217	9	0.12
722	KNLNESLIDL	S	8287	10	0.09
723	NJNESLIDL	S	8288	9	0.23
724	YIKWPWYIWL	S	8305	10	0.00
725	YIWLGFIAGL	S	8311	10	0.10
726	GFIAGLIAIV	S	8315	10	0.10
727	FIAGLIAIV	S	8316	9	0.11
728	FIAGLIAIVM	S	8316	10	0.08
729	GLIAIVMVTI	S	8319	10	0.00
730	GLIAIVMVT	S	8319	9	0.09
731	IMLCCMTSC	S	8328	9	0.14
732	MDLFMRIFTI	ORF3a	8370	10	0.22
733	DLFMRIFTI	ORF3a	8371	9	0.15
734	FMRIFTIGTV	ORF3a	8373	10	0.11
735	RIFTIGTVTL	ORF3a	8375	10	0.40
736	FVRATATIPI	ORF3a	8397	10	0.14
737	SIPFGWLIV	ORF3a	8409	9	0.17
738	GWLIVGVALL	ORF3a	8413	10	0.00
739	WLIVGVALL	ORF3a	8414	9	0.08
740	WLIVGVALLA	ORF3a	8414	10	0.00
741	LIVGVALLAV	ORF3a	8415	10	0.22
742	VALLAVFQSA	ORF3a	8419	10	0.14
743	ALLAVFQSA	ORF3a	8420	9	0.49
744	FQSASKIITL	ORF3a	8425	10	0.20
745	LASLKGVHFV	ORF3a	8440	10	0.37
746	ALSKGVHFV	ORF3a	8441	9	0.22
747	ALSKGVHFVC	ORF3a	8441	10	0.16
748	FVCNLLLLFV	ORF3a	8448	10	0.00
749	CNLLLLFVTV	ORF3a	8450	10	0.11
750	NLLLLFVTV	ORF3a	8451	9	0.13
751	NLLLFVTVY	ORF3a	8451	10	0.17
752	LLFVTVYSHL	ORF3a	8454	10	0.16
753	FVTVYSHLL	ORF3a	8456	9	0.56
754	VTVYSHLLLV	ORF3a	8457	10	0.16
755	TVYSHLLLV	ORF3a	8458	9	0.29
756	SHLLLVAAGL	ORF3a	8461	10	0.17
757	HLLLVAAGL	ORF3a	8462	9	0.50
758	GLEAPFLYL	ORF3a	8469	9	0.14
759	FLYLVALVYF	ORF3a	8474	10	0.10
760	LYLVALVYFL	ORF3a	8475	10	0.00
761	YLYALVYFLQ	ORF3a	8476	10	0.12
762	YLYALVYFL	ORF3a	8476	9	0.11
763	YALVYFLQSI	ORF3a	8478	10	0.11
764	ALVYFLQSI	ORF3a	8479	9	0.16
765	YFLQSINFV	ORF3a	8482	9	0.37
766	FLQSINFVRI	ORF3a	8483	10	0.62
767	PLLYDANYFL	ORF3a	8507	10	0.13
768	LLYDANYFLC	ORF3a	8508	10	0.51
769	LLYDANYFL	ORF3a	8508	9	0.66
770	YQLYSTQLST	ORF3a	8581	10	0.38
771	FVSEETGTL	E	8648	9	0.65
772	TLVNSVLL	E	8655	9	1.27
773	LVNSVLLFL	E	8656	10	0.13
774	IVNSVLLFL	E	8657	9	0.17
775	NSVLLFLAFV	E	8659	10	0.00
776	SVLLFLAFV	E	8660	10	0.12
777	SVLLFLAFV	E	8660	9	0.00
778	VLLFLAFVVF	E	8661	10	0.00
779	VLLFLAFVV	E	8661	9	0.11
780	LLFLAFVVFL	E	8662	10	0.00
781	LFLAFVVFLL	E	8663	10	0.00
782	FLAFVVFLL	E	8664	9	0.00
783	FLAFVVFLLV	E	8664	10	0.00
784	AFVVFLLVTL	E	8666	10	0.00
785	FVVFLLVTL	E	8667	9	0.00
786	FVVFLLVTLA	E	8667	10	0.00
787	VFLLVTLAIL	E	8669	10	0.00
788	FLLVTLAIL	E	8670	9	0.11
789	FLLVTLAILT	E	8670	10	0.11
790	LLVTIAILTA	E	8671	10	0.15
791	VTLAILTAL	E	8673	9	0.14
792	LRICAYCCNI	E	8681	10	0.00
793	RLCAYCCNI	E	8682	9	0.15
794	RLCAYCCNIV	E	8682	10	0.17
795	VSLVKPSFVV	E	8693	10	0.15
796	SLVKPSFVV	E	8694	9	1.22
797	SLVKPSFVVV	E	8694	10	1.37
798	KKLLEQWNLV	M	8733	10	0.16
799	KLLEQWNLV	M	8734	10	1.20
800	KLLEQWNLV	M	8734	9	1.28
801	NLVIGFLFL	M	8740	9	0.11
802	NLVIGFLFLT	M	8740	10	0.11
803	LVIGFLFLT	M	8741	9	0.18
804	GFLFLTWICL	M	8744	10	0.00
805	FLFLTWICL	M	8745	10	0.00
806	FLFLTWICL	M	8745	9	0.13
807	FLYIIKLIFL	M	8764	10	0.12
808	YIIKLIFLWL	M	8766	10	0.13
809	IIKLIFLWLL	M	8767	10	0.14
810	LIFLWLLWPV	M	8770	10	0.00
811	IFLWLLWPV	M	8771	9	0.00
812	IFLWLLWPVT	M	8771	10	0.08
813	FLWLLWPVTL	M	8772	10	0.15
814	FLWLLWPVT	M	8772	9	0.13
815	WLLWPVTLA	M	8774	9	0.37
816	LLWPVTLAC	M	8775	9	0.76
817	VTLACFVLAA	M	8779	10	0.13
818	TLACFVLAAV	M	8780	10	0.15
819	TLACFVLAA	M	8780	9	0.00
820	CFVLAAVYRI	M	8783	10	0.13
821	FVLAAVYRI	M	8784	9	0.67
822	AIAMACLVGL	M	8800	10	0.15
823	IAMACLVGL	M	8801	9	0.15
824	AMACLVGLM	M	8802	9	0.13
825	VGLMWLSYFI	M	8807	10	0.00
826	GLMWLSYFI	M	8808	9	0.11
827	GLMWLSYFIA	M	8808	10	0.14
828	LMWLSYFIA	M	8809	9	0.17
829	LMWLSYFIAS	M	8809	10	0.12
830	FIASFRLFA	M	8815	9	0.66
831	FRLFARTRSM	M	8819	10	0.12
832	RSMWSFNPET	M	8826	10	0.13
833	SMWSFNPET	M	8827	9	1.06
834	ILLNVPLHGT	M	8837	10	2.02
835	RPLLESELVI	M	8850	10	0.15
836	LLESELVIGA	M	8852	10	0.27
837	FHLVDFQVTI	ORF6	8943	10	0.09
838	HLVDFQVTI	ORF6	8944	9	1.02
839	TIAEILLII	ORF6	8951	9	0.15
840	LLIIMRTFKV	ORF6	8956	10	0.00
841	LIIMRTFKV	ORF6	8957	9	0.68
842	SIWNLDYII	ORF6	8966	9	0.78
843	WNLDYIINLI	ORF6	8968	10	0.12
844	NLDYIINLI	ORF6	8969	9	0.11
845	YIINLIIKNL	ORF6	8972	10	0.19
846	QLDEEQPMEI	ORF6	8992	10	0.10
847	MKILFLALI	ORF7a	9003	10	0.00
848	KILFLALI	ORF7a	9004	9	0.00
849	ILFLALITL	ORF7a	9005	10	0.00
850	ILFLALITL	ORF7a	9006	9	0.00
851	LFLALITLA	ORF7a	9006	10	0.00
852	LFLALITLAT	ORF7a	9007	10	0.10
853	FLALITLATC	ORF7a	9008	10	0.15
854	FLALITLAT	ORF7a	9008	9	0.12
855	ALITLATCEL	ORF7a	9010	10	0.57
856	CELYHYQECV	ORF7a	9017	10	0.08
857	ELYHYQECV	ORF7a	9018	9	0.34
858	ALTCFSTQFA	ORF7a	9057	10	0.14
859	FAFACPDGV	ORF7a	9065	9	0.16
860	PKLFIRQEEV	ORF7a	9086	10	0.09
861	KLFIRQEEV	ORF7a	9087	9	1.05
862	QELYSPIFLI	ORF7a	9096	10	0.09
863	ELYSPIFLIV	ORF7a	9097	10	0.39
864	ELYSPIFLI	ORF7a	9097	9	0.90
865	FLIVAAIVFI	ORF7a	9103	10	0.16
866	LIVAAIVFI	ORF7a	9104	9	0.00
867	IVAAIVFITL	ORF7a	9105	10	0.00
868	IVFITLCFTL	ORF7a	9109	10	0.00
869	FLVFLGIITT	ORF8	9126	10	0.00
870	LVFLGIITTV	ORF8	9127	10	0.00
871	FLGIITTVA	ORF8	9129	9	0.29
872	FLGIITTVAA	ORF8	9129	10	0.20
873	YVVDDPCPI	ORF8	9154	9	0.86
874	QYIDIGNYTV	ORF8	9195	10	0.00
875	YIDIGNYTV	ORF8	9196	9	0.97
876	YIDIGNYTVS	ORF8	9196	10	1.02
877	YTVSCLPFTI	ORF8	9202	10	0.40
878	YTVSCLPFT	ORF8	9202	9	0.81
879	TVSCLPFTI	ORF8	9203	9	0.17
880	DFLEYHDVRV	ORF8	9230	10	0.12
881	FLEYHDVRV	ORF8	9231	9	0.00
882	FLEYHDVRVV	ORF8	9231	10	0.25
883	YLGTGPEAGL	N	9356	10	0.52
884	ALALLLLDRL	N	9462	10	0.15
885	LLLLDRLNQL	N	9465	10	0.63
886	LLLDRLNQL	N	9466	9	0.81
887	FGMSRIGMEV	N	9559	10	0.77
888	GMSRIGMEV	N	9560	9	0.18
889	WLTYTGAIKL	N	9574	10	0.22
890	MMPTIFFAGI	ORF3b		10	0.33
891	TIFFAGILIV	ORF3b		10	0.05
892	ILIVTTIVYL	ORF3b		10	0.05
893	LIVTTIVYL	ORF3b		9	0.13
894	VYLTIVLLQL	ORF3b		10	0.04
895	YLTIVLLQL	ORF3b		9	0.05
896	VILQLSLLQV	ORF3b		10	0.09
897	LLQLSLLQV	ORF3b		9	0.38
898	SLLQVMAQQV	ORF3b		10	0.34
899	LLQVMAQQV	ORF3b		9	0.18
900	QVMAQQVLFL	ORF3b		10	0.04
901	VMAQQVLFL	ORF3b		9	0.07
902	VLFLNMTTRL	ORF3b		10	0.16
903	LFLNMTTRLV	ORF3b		10	0.15
904	FLNMTTRLV	ORF3b		9	0.13
905	FLNMTTRLV	ORF3b		9	0.09
906	TVLYYTVTSL	ORF3b		10	0.09
907	VLYYTVTSL	ORF3b		9	0.16
908	YTVTSLQTI	ORF3b		9	0.53
909	TQLNVQTLV	ORF3b		9	0.43
910	NMLPSSSTI	ORF3b		9	0.60
911	MLPSSSTIKL	ORF3b		10	0.23
912	STIKLLWSL	ORF3b		9	0.38
913	ELLIQWNQFM	ORF3b		10	0.04
914	LLIQWNQFM	ORF3b		9	0.86
915	LLIQWNQFMM	ORF3b		10	0.84
916	FMMNRRRLL	ORF3b		9	0.27
917	FMMNRRRLLA	ORF3b		10	0.34
918	LLACLCKHKL	ORF3b		10	0.05
919	KLWSTNLCT	ORF3b		9	0.67
920	IELSLIDFYL	ORF7b		10	0.06
921	ELSLIDFYL	ORF7b		9	0.45
922	SLIDFYLCFL	ORF7b		10	0.33
923	SLIDFYLCF	ORF7b		9	0.36
924	LIDFYLCFL	ORF7b		9	0.04
925	LIDFYLCFLA	ORF7b		10	0.04
926	FYLCFLAFLL	ORF7b		10	0.04
927	FYLCFLAFL	ORF7b		9	0.03
928	YLCFLAFLL	ORF7b		9	0.04
929	YLCFLAFLLF	ORF7b		10	0.05
930	CFLAFLLFLV	ORF7b		10	0.06
931	FLAFLLFLV	ORF7b		9	0.05
932	FLAFLLFLVL	ORF7b		10	0.05
933	AFLLFLVLIM	ORF7b		10	0.05
934	FLLFLVLIML	ORF7b		10	0.04
935	FLLFLVLIM	ORF7b		9	0.05
936	LLFLVLIMLI	ORF7b		10	0.05
937	LLFLVLIML	ORF7b		9	0.04
938	LFLVLIMLI	ORF7b		10	0.06
939	FLVLIMLII	ORF7b		9	0.05
940	IMLIFWFSL	ORF7b		10	0.05
941	IMLIFWFS	ORF7b		9	0.06
942	MLIIFWFSL	ORF7b		9	0.06
943	MLIIFWFSLE	ORF7b		10	0.05
944	LIIFWFSLEL	ORF7b		10	0.04
945	IIFWFSLEL	ORF7b		9	0.06
946	IIFWFSLELQ	ORF7b		10	0.08
947	KISEMHPAL	ORF9b		9	0.39
948	RLVDPQIQL	ORF9b		9	0.65
949	KVYPIILRL	ORF9b		9	0.37
950	SLEDKAFQL	ORF9b		9	0.12
951	ILRLGSPLSL	ORF9b		10	0.08

Table L.4: Top HLA-A*02:01 9-mer or 10-mer epitopes from SARS-CoV-2 proteome predicted by NetMHC4.0 (#721-951)

Peptides selected for SCT expression are indexed according to relative proteomic position. Epitopes from accessory proteins are listed at the end of table. 'Yield' column denotes the relative expression of each SCT based on ratio of protein gel band intensity to positive control band (see **Appendix E** for example).

Appendix M

MULTI-HLA SCT EXPRESSION OF KRAS G12X HOTSPOT MUTATION EPITOPES

SCT ID	pep ID	peptide	HLA	EL_rank	len	yield
001	016	VVGADGVGKSA	A*01:01	13.0	11	0.00
002	057	GADGVGKSAL	A*02:01	8.5	10	0.00
003	058	VGADGVGKSA	A*02:01	13.4	10	0.00
004	096	GADGVGKSA	A*02:01	7.4	9	0.00
005	016	VVGADGVGKSA	A*02:01	17.7	11	0.00
006	020	YKLVVVGADGV	A*02:01	19.3	11	0.00
007	027	VVGAVGVGKSA	A*02:01	17.0	11	0.00
008	031	YKLVVVGAVGV	A*02:01	12.5	11	0.00
009	045	GGVGKSALTI	A*02:01	17.8	10	0.00
010	052	KLVVVGAGGV	A*02:01	6.1	10	0.32
011	057	GADGVGKSAL	A*02:01	14.6	10	0.00
008	062	KLVVVGADGV	A*02:01	5.2	10	0.54
013	065	VGVGKSALTI	A*02:01	18.4	10	0.00
014	067	GAVGVGKSAL	A*02:01	18.4	10	0.00
015	072	KLVVVGAVGV	A*02:01	2.9	10	0.71
016	075	RGVGKSALTI	A*02:01	14.8	10	0.00
017	082	KLVVVGARGV	A*02:01	5.4	10	0.00
018	091	LVVVGAGGV	A*02:01	11.7	9	0.00
019	096	GADGVGKSA	A*02:01	17.1	9	0.00
020	100	LVVVGADGV	A*02:01	7.6	9	0.00
021	104	AVGVGKSAL	A*02:01	11.4	9	0.00
022	105	GAVGVGKSA	A*02:01	17.7	9	0.00
023	109	LVVVGAVGV	A*02:01	3.1	9	0.00
024	118	LVVVGARGV	A*02:01	7.7	9	0.00
025	005	VVGAGGVGKSA	A*02:01	18.0	11	0.00
026	009	YKLVVVGAGGV	A*02:03	13.2	11	0.00
027	016	VVGADGVGKSA	A*02:03	17.0	11	0.00
028	020	YKLVVVGADGV	A*02:03	18.5	11	0.00
029	027	VVGAVGVGKSA	A*02:03	14.5	11	0.00
030	031	YKLVVVGAVGV	A*02:03	12.6	11	0.00
031	038	VVGARGVGKSA	A*02:03	15.3	11	0.19
032	042	YKLVVVGARGV	A*02:03	11.8	11	0.00
033	045	GGVGKSALTI	A*02:03	17.2	10	0.23
034	048	VGAGGVGKSA	A*02:03	18.6	10	0.00
035	052	KLVVVGAGGV	A*02:03	3.2	10	0.38
036	057	GADGVGKSAL	A*02:03	16.9	10	0.00
037	058	VGADGVGKSA	A*02:03	15.3	10	0.00
038	062	KLVVVGADGV	A*02:03	3.3	10	0.28
039	065	VGVGKSALTI	A*02:03	17.0	10	0.00
040	066	AVGVGKSALT	A*02:03	11.9	10	0.19
041	067	GAVGVGKSAL	A*02:03	13.9	10	0.00
042	068	VGAVGVGKSA	A*02:03	18.1	10	0.31
043	072	KLVVVGAVGV	A*02:03	2.3	10	0.31
044	075	RGVGKSALTI	A*02:03	14.9	10	0.00
045	077	GARGVGKSAL	A*02:03	14.8	10	0.00
046	082	KLVVVGARGV	A*02:03	2.7	10	0.00
047	087	GAGGVGKSA	A*02:03	15.3	9	0.00
048	091	LVVVGAGGV	A*02:03	4.1	9	0.30
049	092	KLVVVGAGG	A*02:03	18.8	9	0.00
050	096	GADGVGKSA	A*02:03	16.7	9	0.00
051	100	LVVVGADGV	A*02:03	6.0	9	0.00
052	104	AVGVGKSAL	A*02:03	7.2	9	0.00
053	105	GAVGVGKSA	A*02:03	8.2	9	0.00
054	109	LVVVGAVGV	A*02:03	2.5	9	0.00
055	110	KLVVVGAVG	A*02:03	17.5	9	0.30
056	114	GARGVGKSA	A*02:03	11.6	9	0.00
057	118	LVVVGARGV	A*02:03	3.4	9	0.44
058	119	KLVVVGARG	A*02:03	17.0	9	0.00
059	016	VVGADGVGKSA	A*02:07	14.7	11	0.00
060	031	YKLVVVGAVGV	A*02:07	17.4	11	0.00
061	052	KLVVVGAGGV	A*02:07	12.1	10	0.00
062	057	GADGVGKSAL	A*02:07	6.5	10	0.00
063	062	KLVVVGADGV	A*02:07	11.4	10	0.00
064	065	VGVGKSALTI	A*02:07	18.9	10	0.00
065	072	KLVVVGAVGV	A*02:07	7.0	10	0.00
066	075	RGVGKSALTI	A*02:07	19.8	10	0.00
067	082	KLVVVGARGV	A*02:07	10.8	10	0.00
068	091	LVVVGAGGV	A*02:07	18.1	9	0.00
069	096	GADGVGKSA	A*02:07	14.0	9	0.00
070	100	LVVVGADGV	A*02:07	10.8	9	0.00
071	104	AVGVGKSAL	A*02:07	11.6	9	0.00
072	109	LVVVGAVGV	A*02:07	4.5	9	0.00
073	118	LVVVGARGV	A*02:07	11.0	9	0.00
074	016	VVGADGVGKSA	A*02:11	18.1	11	0.17
075	020	YKLVVVGADGV	A*02:11	18.8	11	0.15
076	024	AVGVGKSALTI	A*02:11	19.9	11	0.13
077	031	YKLVVVGAVGV	A*02:11	9.9	11	0.00
078	042	YKLVVVGARGV	A*02:11	16.9	11	0.14
079	052	KLVVVGAGGV	A*02:11	4.7	10	0.68
080	057	GADGVGKSAL	A*02:11	15.6	10	0.13
081	062	KLVVVGADGV	A*02:11	4.5	10	0.60
082	065	VGVGKSALTI	A*02:11	19.2	10	0.17
083	072	KLVVVGAVGV	A*02:11	2.6	10	0.75
084	075	RGVGKSALTI	A*02:11	16.3	10	0.18
085	082	KLVVVGARGV	A*02:11	4.0	10	0.43
086	091	LVVVGAGGV	A*02:11	11.3	9	0.31
087	096	GADGVGKSA	A*02:11	14.5	9	0.13
088	100	LVVVGADGV	A*02:11	8.6	9	0.31
089	104	AVGVGKSAL	A*02:11	9.5	9	0.14
090	109	LVVVGAVGV	A*02:11	3.2	9	0.31
091	118	LVVVGARGV	A*02:11	7.5	9	0.21
092	006	VVGAGGVGKS	A*03:01	8.6	11	0.12
093	007	LVVVGAGGVGK	A*03:01	2.4	11	0.00
094	017	VVVGADGVGKS	A*03:01	11.0	11	0.00
095	018	LVVVGADGVGK	A*03:01	4.0	11	0.00
096	028	VVVGAVGVGKS	A*03:01	8.3	11	0.12
097	029	LVVVGAVGVGK	A*03:01	2.8	11	0.00
098	039	VVVGARGVGKS	A*03:01	8.6	11	0.26
099	040	LVVVGARGVGK	A*03:01	2.9	11	0.00
100	044	TEYKLVVVGAR	A*03:01	18.0	11	0.00
101	049	VVGAGGVGKS	A*03:01	7.6	10	0.00
102	050	VVVGAGGVGK	A*03:01	0.7	10	0.27
103	059	VVGADGVGKS	A*03:01	15.6	10	0.00
104	060	VVVGADGVGK	A*03:01	0.9	10	0.00
105	069	VVGAVGVGKS	A*03:01	6.5	10	0.00
106	070	VVVGAVGVGK	A*03:01	0.5	10	0.00
107	075	RGVGKSALTI	A*03:01	16.0	10	0.00
108	079	VVGARGVGKS	A*03:01	8.3	10	0.00
109	080	VVVGARGVGK	A*03:01	0.6	10	0.39
110	084	EYKLVVVGAR	A*03:01	18.3	10	0.00
111	089	VVGAGGVGK	A*03:01	0.3	9	0.00
112	098	VVGADGVGK	A*03:01	1.0	9	0.00
113	104	AVGVGKSAL	A*03:01	19.2	9	0.00
114	107	VVGAVGVGK	A*03:01	0.2	9	0.00
115	116	VVGARGVGK	A*03:01	0.3	9	0.00
116	006	VVGAGGVGKS	A*11:01	8.0	11	0.50
117	007	LVVVGAGGVGK	A*11:01	2.2	11	0.38
118	017	VVVGADGVGKS	A*11:01	9.5	11	0.59
119	018	LVVVGADGVGK	A*11:01	3.0	11	0.32
120	027	VVGAVGVGKSA	A*11:01	19.3	11	0.08
121	028	VVVGAVGVGKS	A*11:01	7.5	11	0.34
122	029	LVVVGAVGVGK	A*11:01	2.8	11	0.22
123	039	VVVGARGVGKS	A*11:01	9.2	11	0.51
124	040	LVVVGARGVGK	A*11:01	3.5	11	0.27
125	044	TEYKLVVVGAR	A*11:01	14.2	11	0.00
126	049	VVGAGGVGKS	A*11:01	7.5	10	0.12
127	050	VVVGAGGVGK	A*11:01	0.6	10	0.47
128	059	VVGADGVGKS	A*11:01	10.8	10	0.11
129	060	VVVGADGVGK	A*11:01	0.8	10	0.67
130	069	VVGAVGVGKS	A*11:01	6.1	10	0.10
131	070	VVVGAVGVGK	A*11:01	0.4	10	0.40
132	079	VVGARGVGKS	A*11:01	10.8	10	0.19
133	080	VVVGARGVGK	A*11:01	0.7	10	0.87
134	084	EYKLVVVGAR	A*11:01	17.9	10	0.10
135	089	VVGAGGVGK	A*11:01	0.2	9	0.27
136	098	VVGADGVGK	A*11:01	0.5	9	0.12
137	104	AVGVGKSAL	A*11:01	19.3	9	0.11
138	107	VVGAVGVGK	A*11:01	0.1	9	0.10
139	116	VVGARGVGK	A*11:01	0.4	9	0.20
140	065	VGVGKSALTI	A*23:01	18.6	10	1.39
141	074	EYKLVVVGAV	A*23:01	9.9	10	0.91
142	075	RGVGKSALTI	A*23:01	18.8	10	1.50
143	084	EYKLVVVGAR	A*23:01	19.6	10	1.05
144	104	AVGVGKSAL	A*23:01	19.5	9	1.85

Table M.1: Top 9-mer to 11-mer epitopes from KRAS G12X hotspot mutations predicted by NetMHCpan 4.1 against various HLA alleles (SCT ID #001-144). 'SCT ID' corresponds to numbered lanes of **Figures 4.16 & 4.17**. 'EL_rank' denotes the predicted percentile output for binding score compared to a set of random natural peptides. 'Yield' column denotes the relative expression of each SCT based on ratio of protein gel band intensity to positive control band (see **Appendix E** for example).

SCT ID	pep ID	peptide	HLA	EL_rank	len	yield
145	113	ARGVGKSAL	A*23:01	16.6	9	0.00
146	065	VGVGKSALTI	A*24:02	19.5	10	1.36
147	074	EYKLVVVGAV		11.2	10	0.63
148	075	RGVGKSALTI		16.7	10	1.35
149	084	EYKLVVVGAR		19.5	10	0.60
150	104	AVGVGKSAL		19.8	9	1.48
151	113	ARGVGKSAL		15.9	9	1.32
152	044	TEYKLVVVGAR	A*26:01	15.4	11	0.00
153	050	VVVGAGGVGK		12.8	10	0.00
154	055	DGVGKSALTI		14.6	10	0.00
155	060	VVVGADGVGK		14.0	10	0.00
156	067	GAVGVGKSAL		12.6	10	0.00
157	070	VVVGAVGVGK		11.3	10	0.00
158	077	GARGVGKSAL		19.5	10	0.00
159	080	VVVGARGVGK		15.0	10	0.00
160	084	EYKLVVVGAR		16.7	10	0.00
161	089	VVGAGGVGK		19.8	9	0.00
162	091	LVVVGAGGV		10.0	9	0.00
163	100	LVVVGADGV		17.4	9	0.00
164	104	AVGVGKSAL		13.4	9	0.00
165	105	GAVGVGKSA		17.3	9	0.00
166	107	VVGAVGVGK		15.7	9	0.00
167	109	LVVVGAVGV		10.0	9	0.00
168	118	LVVVGARGV		9.4	9	0.00
169	007	LVVVGAGGVGK	A*26:01	13.8	11	0.00
170	018	LVVVGADGVGK		17.4	11	0.00
171	029	LVVVGAVGVGK		13.0	11	0.00
172	040	LVVVGARGVGK		13.0	11	0.00
173	043	EYKLVVVGARG		3.8	11	0.10
174	044	TEYKLVVVGAR		2.7	11	0.00
175	050	VVVGAGGVGK		7.9	10	0.00
176	060	VVVGADGVGK		9.2	10	0.00
177	070	VVVGAVGVGK		6.0	10	0.00
178	074	EYKLVVVGAV		12.7	10	0.00
179	080	VVVGARGVGK		6.0	10	0.00
180	084	EYKLVVVGAR		0.1	10	0.15
181	089	VVGAGGVGK		6.7	9	0.00
182	098	VVGADGVGK		11.1	9	0.00
183	107	VVGAVGVGK		4.4	9	0.00
184	116	VVGARGVGK		4.7	9	0.00
185	120	YKLVVVGAR		5.6	9	0.00
186	004	VGAGGVGKSAL	B*07:02	8.0	11	0.00
187	015	VGADGVGKSAL		8.0	11	0.00
188	026	VGAVGVGKSAL		11.0	11	0.00
189	036	GARGVGKSALT		7.6	11	0.00
190	037	VGARGVGKSAL		6.2	11	0.00
191	038	VVGARGVGKSA		8.9	11	0.00
192	047	GAGGVGKSAL		5.1	10	0.00
193	057	GADGVGKSAL		2.7	10	0.00
194	067	GAVGVGKSAL		3.1	10	0.00
195	075	RGVGKSALTI		15.7	10	0.00
196	077	GARGVGKSAL		0.5	10	0.00
197	078	VGARGVGKSA		6.9	10	0.00
198	086	AGGVGKSAL		2.9	9	0.00
199	087	GAGGVGKSA		10.7	9	0.00
200	095	ADGVGKSAL		6.7	9	0.00
201	096	GADGVGKSA		7.4	9	0.00
202	104	AVGVGKSAL		0.8	9	0.00
203	105	GAVGVGKSA		7.7	9	0.00
204	113	ARGVGKSAL		5.2	9	0.10
205	114	GARGVGKSA		1.9	9	0.00
206	004	VGAGGVGKSAL	B*08:01	18.8	11	0.25
207	015	VGADGVGKSAL		18.2	11	0.24
208	026	VGAVGVGKSAL		18.1	11	0.21
209	037	VGARGVGKSAL		16.3	11	0.40
210	047	GAGGVGKSAL		19.6	10	0.12
211	055	DGVGKSALTI		8.4	10	0.16
212	057	GADGVGKSAL		14.5	10	0.15
213	067	GAVGVGKSAL		14.9	10	0.32
214	074	EYKLVVVGAV		10.1	10	0.00
215	077	GARGVGKSAL		5.9	10	0.13
216	086	AGGVGKSAL		5.7	9	0.14
217	094	DGVGKSALT		18.9	9	0.16
218	095	ADGVGKSAL		12.1	9	0.11
219	104	AVGVGKSAL		3.5	9	0.23
220	111	YKLVVVGAV		11.0	9	0.14
221	113	ARGVGKSAL		4.7	9	0.33
222	114	GARGVGKSA		11.4	9	0.17
223	047	GAGGVGKSAL	B*15:01	18.8	10	0.00
224	050	VVVGAGGVGK		18.1	10	0.00
225	067	GAVGVGKSAL		9.5	10	0.00
226	077	GARGVGKSAL		6.5	10	0.00
227	086	AGGVGKSAL		18.2	9	0.00
228	104	AVGVGKSAL		5.8	9	0.00
229	105	GAVGVGKSA		13.6	9	0.00
230	107	VVGAVGVGK		18.4	9	0.00
231	110	KLVVVGAVG		13.9	9	0.00
232	113	ARGVGKSAL		17.3	9	0.00
233	114	GARGVGKSA		10.2	9	0.00
234	011	TEYKLVVVGAG	B*18:01	18.8	11	0.12
235	033	TEYKLVVVGAV		4.1	11	0.11
236	044	TEYKLVVVGAR		6.6	11	0.12
237	095	ADGVGKSAL		14.5	9	0.08
238	096	GADGVGKSA		17.7	9	0.09
239	111	YKLVVVGAV		17.4	9	0.00
240	120	YKLVVVGAR		17.5	9	0.00
241	057	GADGVGKSAL	B*35:01	10.8	10	0.00
242	067	GAVGVGKSAL		11.1	10	0.08
243	077	GARGVGKSAL		17.5	10	0.00
244	087	GAGGVGKSA		19.1	9	0.00
245	096	GADGVGKSA		10.0	9	0.00
246	104	AVGVGKSAL		18.2	9	0.00
247	105	GAVGVGKSA		10.1	9	0.00
248	114	GARGVGKSA		13.7	9	0.00
249	011	TEYKLVVVGAG	B*40:01	15.7	11	0.00
250	033	TEYKLVVVGAV		1.8	11	0.00
251	044	TEYKLVVVGAR		9.1	11	0.00
252	057	GADGVGKSAL		6.9	10	0.00
253	067	GAVGVGKSAL		10.3	10	0.00
254	077	GARGVGKSAL		15.0	10	0.00
255	087	ADGVGKSAL		17.2	9	0.00
256	095	ADGVGKSAL		2.9	9	0.00
257	096	GADGVGKSA		8.8	9	0.00
258	104	AVGVGKSAL		11.9	9	0.00
259	105	GAVGVGKSA		11.9	9	0.00
260	111	YKLVVVGAV		15.2	9	0.00
261	113	ARGVGKSAL		7.4	9	0.00
262	114	GARGVGKSA		12.2	9	0.00
263	011	TEYKLVVVGAG	B*40:02	8.6	11	0.11
264	013	ADGVGKSALTI		18.7	11	0.08
265	015	VGADGVGKSAL		18.4	11	0.09
266	022	TEYKLVVVGAD		12.9	11	0.15
267	033	TEYKLVVVGAV		1.5	11	0.15
268	044	TEYKLVVVGAR		7.5	11	0.11
269	057	GADGVGKSAL		11.7	10	0.09
270	067	GAVGVGKSAL		18.3	10	0.08
271	077	GARGVGKSAL		17.1	10	0.00
272	086	AGGVGKSAL		16.3	9	0.08
273	095	ADGVGKSAL		2.0	9	0.09
274	096	GADGVGKSA		13.1	9	0.09
275	104	AVGVGKSAL		16.6	9	0.10
276	105	GAVGVGKSA		12.9	9	0.11
277	111	YKLVVVGAV		10.7	9	0.07
278	113	ARGVGKSAL		9.0	9	0.14
279	114	GARGVGKSA		10.5	9	0.09
280	011	TEYKLVVVGAG	B*44:02	18.0	11	0.00
281	013	ADGVGKSALTI		19.7	11	0.00
282	033	TEYKLVVVGAV		5.2	11	0.00
283	044	TEYKLVVVGAR		8.5	11	0.00
284	057	GADGVGKSAL		16.1	10	0.00
285	095	ADGVGKSAL		5.6	9	0.00
286	096	GADGVGKSA		13.7	9	0.00
287	104	AVGVGKSAL		15.9	9	0.00
288	105	GAVGVGKSA		15.8	9	0.00

Table M.2: Top 9-mer to 11-mer epitopes from KRAS G12X hotspot mutations predicted by NetMHCpan 4.1 against various HLA alleles (SCT ID #145-288). 'SCT ID' corresponds to numbered lanes of **Figures 4.16 & 4.17**. 'EL_rank' denotes the predicted percentile output for binding score compared to a set of random natural peptides. 'Yield' column denotes the relative expression of each SCT based on ratio of protein gel band intensity to positive control band (see **Appendix E** for example).

SCT ID	pep ID	peptide	HLA	EL_rank	len	yield
337	114	GARGVGKSA	C*07:01	8.6	9	0.00
338	037	VGARGVGKSAL	C*07:02	17.9	11	0.00
339	057	GADGVGKSAL		12.9	10	0.00
340	077	GARGVGKSAL		19.9	10	0.00
341	089	VVGAGGVGK		19.5	9	0.00
342	111	YKLVVVGAV		10.4	9	0.00
343	113	ARGVGKSAL		2.0	9	0.00
344	114	GARGVGKSA		16.2	9	0.00
345	045	GGVGKSALTI	A*24:02	33.3	10	1.26
346	055	DGVGKSALTI		24.6	10	1.37
347	054	EYKLVVVGAG		37.0	10	0.72
348	064	EYKLVVVGAD		45.5	10	0.87
349	086	AGGVGKSAL		30.5	9	1.36
350	095	ADGVGKSAL		33.0	9	1.30

SCT ID	pep ID	peptide	HLA	EL_rank	len	yield
289	113	ARGVGKSAL	B*44:02	12.7	9	0.00
290	114	GARGVGKSA	B*44:03	19.9	9	0.00
291	011	TEYKLVVVGAG		14.2	11	0.00
292	022	TEYKLVVVGAD		17.2	11	0.00
293	033	TEYKLVVVGAV		4.5	11	0.00
294	044	TEYKLVVVGAR		7.8	11	0.00
295	057	GADGVGKSAL		18.5	10	0.00
296	095	ADGVGKSAL		7.5	9	0.00
297	096	GADGVGKSA		16.9	9	0.00
298	105	GAVGVGKSA		15.5	9	0.00
299	113	ARGVGKSAL		11.3	9	0.00
300	055	DGVGKSALTI	B*51:01	1.6	10	0.00
301	057	GADGVGKSAL		18.3	10	0.00
302	058	VGADGVGKSA		16.1	10	0.00
303	065	VGVGKSALTI		10.4	10	0.00
304	067	GAVGVGKSAL		16.9	10	0.00
305	068	VGAVGVGKSA		17.9	10	0.00
306	075	RGVGKSALTI		15.4	10	0.00
307	078	VGARGVGKSA		19.9	10	0.00
308	091	LVVVGAGGV		18.9	9	0.00
309	094	DGVGKSALT		9.9	9	0.00
310	096	GADGVGKSA		10.0	9	0.00
311	100	LVVVGADGV		19.0	9	0.00
312	105	GAVGVGKSA		11.4	9	0.00
313	109	LVVVGAVGV		8.9	9	0.00
314	111	YKLVVVGAV		10.3	9	0.00
315	114	GARGVGKSA		11.8	9	0.00
316	118	LVVVGARGV		8.7	9	0.00
317	015	VGADGVGKSAL	C*04:01	13.7	11	0.00
318	057	GADGVGKSAL		2.7	10	0.00
319	058	VGADGVGKSA		17.1	10	0.00
320	067	GAVGVGKSAL		15.8	10	0.00
321	077	GARGVGKSAL		13.1	10	0.00
322	086	AGGVGKSAL		11.6	9	0.00
323	095	ADGVGKSAL		14.4	9	0.00
324	096	GADGVGKSA		8.9	9	0.00
325	104	AVGVGKSAL		8.0	9	0.00
326	113	ARGVGKSAL		7.4	9	0.00
327	035	ARGVGKSALTI	C*07:01	17.4	11	0.15
328	037	VGARGVGKSAL		15.0	11	0.15
329	057	GADGVGKSAL		17.7	10	0.00
330	075	RGVGKSALTI		19.6	10	0.00
331	077	GARGVGKSAL		14.1	10	0.00
332	096	GADGVGKSA		17.0	9	0.00
333	104	AVGVGKSAL		18.4	9	0.00
334	105	GAVGVGKSA		17.9	9	0.00
335	111	YKLVVVGAV		16.0	9	0.00
336	113	ARGVGKSAL		1.0	9	0.00

Table M.3: Top 9-mer to 11-mer epitopes from KRAS G12X hotspot mutations predicted by NetMHCpan 4.1 against various HLA alleles (SCT ID #289-350). 'SCT ID' corresponds to numbered lanes of **Figures 4.16 & 4.17**. 'EL_rank' denotes the predicted percentile output for binding score compared to a set of random natural peptides. 'Yield' column denotes the relative expression of each SCT based on ratio of protein gel band intensity to positive control band (see **Appendix E** for example).

Appendix N

A*02:01 SCT EXPRESSION OF P53 HOTSPOT MUTATION EPITOPES

ID	mutation	peptide	len	nM	yield
001	R175H	RHCPHHERC	9	41551.3	0.16
002	R175H	VRHCPHHER	9	41751.4	0.11
003	R175H	VVRHCPHHE	9	35769.9	0.11
004	R175H	EVVRHCPHH	9	40640.3	0.11
005	R175H	TEVVRHCPH	9	39589.1	0.12
006	R175H	MTEVVRHCP	9	32072.6	0.14
007	R175H	HMTEVVRHC	9	7826.1	0.34
008	Y220C	PCEPPEVGS	9	42356.1	0.18
009	Y220C	VPCEPPEVG	9	38213.7	0.10
010	Y220C	VVPCEPPEV	9	681	0.15
011	Y220C	VVVPCEPPE	9	31857.1	0.10
012	Y220C	SVVVPCEPP	9	25057.7	0.00
013	Y220C	HSVVVPCEP	9	29588.1	0.11
014	Y220C	RHSVVVPCE	9	34735.6	0.00
015	G245S	GSMNRRPIL	9	21588.2	0.15
016	G245S	MGSMNRRPI	9	32048.6	0.00
017	G245S	CMGSMNRRP	9	32158.7	0.00
018	G245S	SCMGSMNRR	9	34531.7	0.00
019	G245S	SSCMGSMNR	9	33958.2	0.00
020	G245S	NSSCMGSMN	9	40724.3	0.00
021	G245S	CNSSCMGSM	9	29767	0.00
022	G245D	GDMNRRPIL	9	36301.3	0.12
023	G245D	MGDMNRRPI	9	29633.6	0.00
024	G245D	CMGDMNRRP	9	30395.7	0.00
025	G245D	SCMGDMNRR	9	36017.3	0.00
026	G245D	SSCMGDMNR	9	35081.2	0.00
027	G245D	NSSCMGDMN	9	40491.9	0.00
028	G245D	CNSSCMGDM	9	31994.6	0.00
029	R248Q	NQRPILTII	9	12813.2	0.14
030	R248Q	MNQRPILTI	9	19719.3	0.16
031	R248Q	GMNQRPILT	9	4775	0.00
032	R248Q	GGMNQRPIL	9	33911.5	0.10
033	R248Q	MGGMNQRPI	9	29727	0.12
034	R248Q	CMGGMNQRP	9	27001.7	0.11
035	R248Q	SCMGGMNQR	9	34323.1	0.11
036	R248W	NWRPILTII	9	21658.4	0.20
037	R248W	MNWRPILTI	9	7504	0.15
038	R248W	GMNWRPILT	9	1968	0.13
039	R248W	GGMNWRPIL	9	22084.4	0.15
040	R248W	MGGMNWRPI	9	20228.2	0.08
041	R248W	CMGGMNWRP	9	13639.8	0.09
042	R248W	SCMGGMNWR	9	30367.1	0.12
043	R249S	RSPILTIIT	9	20992.6	0.12
044	R249S	NRSPILTII	9	23762	0.10
045	R249S	MNRSPILTI	9	20555.8	0.12
046	R249S	GMNRSPILT	9	4459.4	0.14
047	R249S	GGMNRSPIL	9	23569.4	0.14
048	R249S	MGGMNRSPI	9	30682.5	0.12
049	R249S	CMGGMNRSP	9	28491.6	0.00
050	R273C	VCVCACPGR	9	31770.4	0.11
051	R273C	EVCVCACPG	9	26134.3	0.00
052	R273C	FEVCVCACP	9	16598.4	0.00
053	R273C	SFEVCVCAC	9	16410	0.00
054	R273C	NSFEVCVCA	9	3333.3	0.00
055	R273C	RNSFEVCVC	9	31993.2	0.00
056	R273C	GRNSFEVCV	9	21379.3	0.13
057	R273H	VHVCACPGR	9	36000.6	0.12
058	R273H	FEVHVCACP	9	27335.1	0.04
059	R273H	FEVHVCACP	9	21882.9	0.14
060	R273H	SFEVHVCAC	9	23740.7	0.00
061	R273H	NSFEVHVCA	9	7518	0.16
062	R273H	RNSFEVHVC	9	35245.9	0.12
063	R273H	GRNSFEVHV	9	20043.7	0.13
064	R282W	DWRTEEENL	9	39861.2	0.00
065	R282W	RDWRTEEEN	9	41576.5	0.11
066	R282W	GRDWRTEEE	9	41735.1	0.13
067	R282W	PGRDWRTEE	9	43885.8	0.13
068	R282W	CPGRDWRTE	9	41213.3	0.13
069	R282W	ACPGRDWRT	9	32101.7	0.14
070	R282W	CACPGRDWR	9	38926.5	0.11
071	R175H	RHCPHHERCS	10	45108.1	0.11
072	R175H	VRHCPHHERC	10	43074.8	0.12
073	R175H	VVRHCPHHER	10	38296.5	0.10
074	R175H	EVVRHCPHHE	10	39959.2	0.10
075	R175H	TEVVRHCPHH	10	40157.7	0.10
076	R175H	MTEVVRHCPH	10	32605.4	0.13
077	R175H	HMTEVVRHCP	10	16852.8	0.18
078	R175H	QHMTEVVRHC	10	15558.1	0.12
079	Y220C	PCEPPEVGSD	10	43910	0.38
080	Y220C	VPCEPPEVGS	10	39635.8	0.11
081	Y220C	VVPCEPPEVG	10	12861.8	0.19
082	Y220C	VVVPCEPPEV	10	1510.9	0.29
083	Y220C	SVVVPCEPPE	10	32753.6	0.12
084	Y220C	HSVVVPCEPP	10	26844.7	0.12
085	Y220C	RHSVVVPCEP	10	32866.1	0.11
086	Y220C	FRHSVVVPCE	10	22982.4	0.10
087	G245S	GSMNRRPILT	10	15779.7	0.32
088	G245S	MGSMNRRPIL	10	23166.2	0.11
089	G245S	CMGSMNRRPI	10	9100.3	0.11
090	G245S	SCMGSMNRRP	10	36881.3	0.00
091	G245S	SSCMGSMNRR	10	36482.1	0.11
092	G245S	NSSCMGSMNR	10	37705.7	0.10
093	G245S	CNSSCMGSMN	10	37582.3	0.10
094	G245S	MCNSSCMGSM	10	16480.8	0.12
095	G245D	GDMNRRPILT	10	18129.1	0.12
096	G245D	MGDMNRRPIL	10	31938.2	0.13
097	G245D	CMGDMNRRPI	10	5704.2	0.11
098	G245D	SCMGDMNRRP	10	35295.5	0.11
099	G245D	SSCMGDMNRR	10	36904.1	0.12
100	G245D	NSSCMGDMNR	10	37831.5	0.10
101	G245D	CNSSCMGDMN	10	38624.8	0.10
102	G245D	MCNSSCMGDM	10	18956.7	0.11
103	R248Q	NQRPILTIIT	10	22523.6	0.09
104	R248Q	MNQRPILTII	10	10883.1	0.11
105	R248Q	GMNQRPILTI	10	1324.5	0.11
106	R248Q	GGMNQRPILT	10	14068.6	0.11
107	R248Q	MGGMNQRPIL	10	32925.2	0.09
108	R248Q	CMGGMNQRPI	10	3276.9	0.11
109	R248Q	SCMGGMNQRP	10	32457.6	0.09
110	R248Q	SSCMGGMNQR	10	36717.3	0.11
111	R248W	NWRPILTIIT	10	27550.7	0.12
112	R248W	MNWRPILTII	10	9132.1	0.10
113	R248W	GMNWRPILTI	10	181	0.12
114	R248W	GGMNWRPILT	10	8616.8	0.13
115	R248W	MGGMNWRPIL	10	22224.6	0.10
116	R248W	CMGGMNWRPI	10	178.5	0.13
117	R248W	SCMGGMNWRP	10	19475	0.09
118	R248W	SSCMGGMNWR	10	33401	0.10
119	R249S	RSPILTIITL	10	3605.6	0.11
120	R249S	NRSPILTIIT	10	26116.7	0.13
121	R249S	MNRSPILTII	10	20617.3	0.13
122	R249S	GMNRSPILTI	10	996.1	0.12
123	R249S	GGMNRSPILT	10	15368.7	0.11
124	R249S	MGGMNRSPIL	10	25028.4	0.12
125	R249S	CMGGMNRSPI	10	3678.8	0.10
126	R249S	SCMGGMNRSP	10	32647.8	0.13
127	R273C	VCVCACPGRD	10	37188.7	0.00
128	R273C	EVCVCACPGR	10	31493.1	0.00
129	R273C	FEVCVCACPG	10	12158.9	0.00
130	R273C	SFEVCVCACP	10	19828.8	0.00
131	R273C	NSFEVCVCAC	10	8353.9	0.13
132	R273C	RNSFEVCVCA	10	4340.8	0.00
133	R273C	GRNSFEVCVC	10	32692	0.00
134	R273C	LGRNSFEVCV	10	22872	0.00
135	R273H	VHVCACPGRD	10	39154.6	0.00
136	R273H	EVHVCACPGR	10	32868.6	0.00
137	R273H	FEVHVCACPG	10	15089.7	0.00
138	R273H	SFEVHVCACP	10	26146.7	0.00
139	R273H	NSFEVHVCAC	10	13482	0.00
140	R273H	RNSFEVHVCA	10	11567.9	0.00
141	R273H	GRNSFEVHVC	10	33460.7	0.00
142	R273H	LGRNSFEVHV	10	22047.1	0.00
143	R282W	DWRTEEENLR	10	43338.4	0.10
144	R282W	RDWRTEEENL	10	32912.8	0.00

Table N.1: Top 9-mer to 11-mer epitopes from p53 hotspot mutations predicted by NetMHC 4.0 against HLA-A*02:01 (SCT ID #001-144).

'ID' corresponds to numbered lanes of **Figure 4.19**. 'nM' denotes the predicted nanomolar binding affinity. 'Yield' column denotes the relative expression of each SCT based on ratio of protein gel band intensity to positive control band (see **Appendix E** for example).

ID	mutation	peptide	len	nM	yield
145	R282W	GRDWRTEEEN	10	43118.6	0.00
146	R282W	PGRDWRTEEE	10	42437.3	0.00
147	R282W	CPGRDWRTEE	10	42048.8	0.00
148	R282W	ACPGRDWRTE	10	40509.9	0.00
149	R282W	CACPGRDWRT	10	30711.1	0.00
150	R282W	VCACPGRDWR	10	40372.5	0.00
151	R175H	RHCPHHERCSD	11	45821.9	0.12
152	R175H	VRHCPHHERCS	11	44864.7	0.09
153	R175H	VVRHCPHHERC	11	39258.5	0.11
154	R175H	EVVRHCPHHER	11	41049.1	0.00
155	R175H	TEVVRHCPHHE	11	41439.5	0.00
156	R175H	MTEVVRHCPHH	11	34604.7	0.00
157	R175H	HMTEVVRHCPH	11	28604.7	0.27
158	R175H	QHMTEVVRHCP	11	29554.2	0.00
159	R175H	SQHMTEVVRHC	11	22376.6	0.00
160	Y220C	PCEPPEVGSDC	11	42642.6	0.18
161	Y220C	VPCEPPEVGSD	11	42742.3	0.08
162	Y220C	VVPCEPPEVGS	11	30681.9	0.14
163	Y220C	VVVPCEPPEVG	11	18368.6	0.37
164	Y220C	SVVVPCEPPEV	11	10286.8	0.14
165	Y220C	HSVVVPCEPPE	11	34633.9	0.09
166	Y220C	RHSVVVPCEPP	11	32249	0.09
167	Y220C	FRHSVVVPCEP	11	28167.4	0.08
168	Y220C	TFRHSVVVPCE	11	29467	0.13
169	G245S	GSMNRRPILTI	11	9222	0.17
170	G245S	MGSMNRRPILT	11	20992.3	0.09
171	G245S	CMGSMNRRPIL	11	15034	0.09
172	G245S	SCMGSMNRRPI	11	22014.2	0.09
173	G245S	SSCMGSMNRRP	11	39031.9	0.08
174	G245S	NSSCMGSMNRR	11	39034	0.06
175	G245S	CNSSCMGSMNR	11	37787.4	0.12
176	G245S	MCNSSCMGSMN	11	31670.5	0.09
177	G245S	YMCNSSCMGSM	11	3144.8	0.91
178	G245D	GDMNRRPILTI	11	25902	0.09
179	G245D	MGDMNRRPILT	11	20937	0.09
180	G245D	CMGDMNRRPIL	11	11663.7	0.08
181	G245D	SCMGDMNRRPI	11	19541.5	0.09
182	G245D	SSCMGDMNRRP	11	38702.2	0.08
183	G245D	NSSCMGDMNRR	11	39162.6	0.09
184	G245D	CNSSCMGDMNR	11	37900.8	0.11
185	G245D	MCNSSCMGDMN	11	32713.6	0.12
186	G245D	YMCNSSCMGDM	11	5350.3	0.98
187	R248Q	NQRPILTITTL	11	6737.4	0.00
188	R248Q	MNQRPILTITT	11	25710.5	0.07
189	R248Q	GMNQRPILTIT	11	6910.8	0.07
190	R248Q	GGMNQRPILTI	11	8515.2	0.09
191	R248Q	MGGMNQRPILT	11	17103.9	0.09
192	R248Q	CMGGMNQRPIL	11	14036.4	0.10
193	R248Q	SCMGGMNQRPI	11	15170.6	0.12
194	R248Q	SSCMGGMNQRP	11	37825.4	0.09
195	R248Q	NSSCMGGMNQR	11	39113.9	0.11
196	R248W	NWRPILTITTL	11	11409.7	0.08
197	R248W	MNWRPILTITT	11	20715.7	0.08
198	R248W	GMNWRPILTII	11	3895.7	0.07
199	R248W	GGMNWRPILTI	11	1919	0.16
200	R248W	MGGMNWRPILT	11	11412.3	0.09
201	R248W	CMGGMNWRPIL	11	4245	0.14
202	R248W	SCMGGMNWRPI	11	2421.1	0.12
203	R248W	SSCMGGMNWRP	11	30202.7	0.07
204	R248W	NSSCMGGMNWR	11	36218.6	0.08
205	R249S	RSPILTITTLE	11	13412.5	0.00
206	R249S	NRSPILTITTL	11	9487.9	0.08
207	R249S	MNRSPILTITT	11	27619.9	0.09
208	R249S	GMNRSPILTII	11	4999.1	0.08
209	R249S	GGMNRSPILTI	11	8112.5	0.12
210	R249S	MGGMNRSPILT	11	18399.8	0.09
211	R249S	CMGGMNRSPIL	11	6386.6	0.08
212	R249S	SCMGGMNRSPI	11	16049.6	0.11
213	R249S	SSCMGGMNRSP	11	38377.8	0.08
214	R273C	VCVCACPGRDR	11	38192.7	0.00
215	R273C	EVCVCACPGRD	11	35832.7	0.00
216	R273C	FEVCVCACPGR	11	24426.5	0.00
217	R273C	SFEVCVCACPG	11	21616.3	0.00
218	R273C	NSFEVCVCACP	11	18674.2	0.09
219	R273C	RNSFEVCVCAC	11	15347.7	0.08
220	R273C	GRNSFEVCVCA	11	12638.3	0.00
221	R273C	LGRNSFEVCVC	11	32781.6	0.00
222	R273C	LLGRNSFEVCV	11	1665.9	0.12
223	R273H	VHVCACPGRDR	11	40438.1	0.00
224	R273H	EVHVCACPGRD	11	36394.9	0.09
225	R273H	FEVHVCACPGR	11	26735.5	0.09
226	R273H	SFEVHVCACPG	11	24285.2	0.09
227	R273H	NSFEVHVCACP	11	24477.3	0.13
228	R273H	RNSFEVHVCAC	11	23228.5	0.10
229	R273H	GRNSFEVHVCA	11	24547.8	0.08
230	R273H	LGRNSFEVHVC	11	33995.7	0.08
231	R273H	LLGRNSFEVHV	11	1473.9	0.21
232	R282W	DWRTEEENLRK	11	42033.8	0.11
233	R282W	RDWRTEEENLR	11	40924	0.08
234	R282W	GRDWRTEEENL	11	38306.4	0.11
235	R282W	PGRDWRTEEEN	11	43021.2	0.11
236	R282W	CPGRDWRTEEE	11	42275.5	0.09
237	R282W	ACPGRDWRTEE	11	41984.7	0.10
238	R282W	CACPGRDWRTE	11	39558.7	0.11
239	R282W	VCACPGRDWRT	11	37382.7	0.00
240	R282W	RVCACPGRDWR	11	37545.3	0.00

Table N.2: Top 9-mer to 11-mer epitopes from p53 hotspot mutations predicted by NetMHC 4.0 against HLA-A*02:01 (SCT ID #145-240).

'ID' corresponds to numbered lanes of **Figure 4.19**. 'nM' denotes the predicted nanomolar binding affinity. 'Yield' column denotes the relative expression of each SCT based on ratio of protein gel band intensity to positive control band (see **Appendix E** for example).