SHARED MINOR HISTOCOMPATIBILITY ANTIGEN DISCOVERY AND TARGETING IN THE CONTEXT OF ALLOGENEIC HEMATOPOIETIC CELL TRANSPLANT FOR HEMATOLOGIC MALIGNANCIES

Kelly Shea Olsen

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Microbiology & Immunology in the UNC School of Medicine

Chapel Hill
2023

Approved by:

Benjamin Vincent

Barbara Savoldo

Paul Armistead

Jonathan Serody

Jenny Ting

ABSTRACT

Kelly Shea Olsen: Shared Minor Histocompatibility Antigen Discovery and Targeting in the context of Allogeneic Hematopoietic Stem Cell Transplant for Hematologic Malignancies
(Under the direction of Benjamin Vincent)

In allogeneic hematopoietic stem cell transplantation (alloHCT), donor-derived T cells that recognize minor histocompatibility antigens (mHAs) are able to eliminate leukemia cells via the “graft versus leukemia effect” (GvL). However, donor T cells can also recognize non-leukemia host antigens and cause inflammatory tissue damage termed “graft versus host disease” (GvHD). Therapies that boost T cell responses can improve alloHCT efficacy, but are limited by concurrent increases in incidence and severity of GvHD. Therapies that prevent GvHD by impairing T cell responses also increase relapse rates. Thus, it is critical to understand the biological differences between GvL and GvHD to develop treatment strategies that separate GvL from GvHD. mHAs with expression restricted to hematopoietic tissue (GvL mHAs) are attractive targets for T cell responses that could drive GvL without causing GvHD. Prior work to identify mHAs has focused on a small set of mHAs or population-level SNP association studies. We sought to broaden the array of known mHAs capable of mediating an anti-leukemia response, including discovery of GvL mHAs that are highly shared across donor-recipient pairs (DRPs). To do this, we applied experimental and computational antigen discovery and validation methods to two alloHCT datasets. We found that the total number of predicted mHAs varied by HLA allele, and number of each class of mHA

significantly differed by recipient genomic ancestry. From the pool of predicted mHAs, we identified the smallest sets of GvL mHAs needed to cover 100% of DRPs with a given HLA allele. We used two different mass spectrometry methods to search for high population frequency GvL mHAs presented by three common HLA alleles. We validated a total of 26 novel predicted GvL mHAs that have a high degree of sharing within DRPs expressing HLA-A*02:01, HLA-B*35:01, and HLA-C*07:02, increasing the number of known GvL mHAs by more than 200%. We confirmed immunogenicity of an example novel mHA via T cell co-culture with peptide-pulsed dendritic cells, and we continue to seek additional mHA-targeting T cell clones against our novel antigens. This work demonstrates that identification of shared mHAs is a feasible and promising technique for expanding GvL mHA-targeting immunotherapies.

ACKNOWLEDGMENTS

I am so grateful to everyone that has supported and inspired me throughout this very long MD/PhD path, including family, friends, colleagues, and the natural world. Thanks to Dr. Angela DiBenedetto, my first PI and a huge role model to me as an incredible female scientist with incredible kindness, compassion, and a rich and wonderful life outside of work. She showed me not only how exciting science can be, but also how important and fulfilling mentorship is, and that being a well-balanced person with passions outside of science and interest in broader academic pursuits like philosophy and literature can make you a batter scientist and human. Thank you also to my PhD PI Dr. Ben Vincent, who has confirmed many of the same lessons for me. I am so lucky to have been continually surrounded by mentors with such rich knowledge and unusually thoughtful approaches to science and life. Ben, I have learned so much from you both in science and in philosophy, literature, the arts, and many other intangibles. My PhD has been a mind-expanding experience and I will always cherish this wonderfully illuminating time in my life with our lab. Thank you to all my colleagues in the Vincent lab and also in Dr. Paul Armistead's lab- I appreciate all of you so incredibly much and are thankful for all you have done for me, and for the times that we have shared together both inside of lab and out. Thank you to Paul as well for taking me under your wing in clinic and sparking my passion for transplantation; I appreciate your guidance and you showing me how wonderfully fulfilling this career path can be. Also, thank you to the MD/PhD program and all the incredible friendships that I have made

through this process. When I entered this program I didn't expect to be handed a premade friend group of some of the most wonderful people I've ever met in my life, but that's exactly what happened and I can't wait for more adventures and to see you all follow your passions. Thanks also to my other friends- whether I've known you for a year or for my whole life, I am more grateful than I can say for your support and love. I cherish every one of you. Thank you to my wonderful pets both current and past for enriching my life and putting a smile on my face every day. Thanks to the natural bounty of the great state of North Carolina that allows me to spend time cultivating a garden that brings me so much joy and provides great opportunities for rock climbing and other outdoor pursuits. Lastly, a massive thanks to my family. To my mom, Cathryn, thank you for encouraging my curiosity and love of the natural world at an early age and opening my eyes to so many opportunities that have gotten me to where I am. To my dad, Garry, thank you for your loving emotional support in good times and bad and always going way above and beyond to help me fulfill my dreams. To Brita, thanks for being the best sister I could ever ask for and I am so happy to have you in NC with us- love you so much. Thanks to Worth for taking care of Brita and being such a fun and hilarious person. Finally, thank you to my wonderful husband (!!) Alex for being my favorite person on earth and for supporting me in more ways I can count. Thanks for moving to NC so I can pursue my dreams, picking up slack when I'm busy, celebrating my victories more than I celebrate them myself, and for building a life with me that makes me happier than I had ever dreamed I could be. I love you.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

LIST OF ABBREVIATIONS

AA	African American
ACT	Adoptive cell therapy
AICD	Activation induced cell death
ALL	Acute Lymphocytic Leukemia
alloHCT	Allogeneic hematopoietic cell transplantation
AML	Acute myeloid leukemia
ANN	Artificial neural network
APC	Antigen presenting cell
ATG	Anti-thymocyte globulin
BEAM	Barcode enabled antigen mapping
BMT	Bone marrow transplant
CAR	Chimeric antigen receptor
CIBMTR	Center for International Bone Marrow Transplant Research
CLIP	Class II-associated invariant chain protein
CML	Chronic Myeloid Leukemia
CR1	Complete response 1
DC	Dendritic cell
DDA	Data-dependent acquisition
DISCOVeRY-BMT	Determining the Influence of Susceptibility Conveying Variants Related to one-Year mortality after BMT
DLI	Donor lymphocyte infusion
DRP	Donor-recipient pair

EA	European American
EBV-LCL	Epstein-Barr virus-transformed lymphoblastic cell line
ER	Endoplasmic reticulum
FACS	Fluorescence-activated cell sorting
FDR	False discovery rate
GTEx	Genotype-Tissue Expression Project
GvHD	Graft versus host disease
GvH	Graft versus host
GvL	Graft versus leukemia
GWAS	Genome-wide association study
HIS	Hispanic
HLA	Human leukocyte antigen
HPLC	High-performance liquid chromatography
HSTF	High-Throughput Sequencing Facility
IFN	Interferon
IL	Interleukin
IMGF	Immune Monitoring and Genomics Facility
IP	Immunoprecipitation
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
MAC	Myeloablative conditioning
MAF	Minor allele frequency
MDS	Myelodysplastic Syndrome
mHA	Minor histocompatibility antigen

MHC	Major histocompatibility complex
MPN	Myeloproliferative Neoplasms
MRD	Matched related donor
mTEC	Medullary thymic epithelial cell
MUD	Matched unrelated donor
NK	Natural killer
NSCLC	Non-Small Cell Lung Cancer
PBMC	Peripheral blood mononuclear cell
PRM	Parallel reaction monitoring
RBC	Red blood cell
RIC	Reduced intensity conditioning
RT-PCR	Reverse transcription polymerase chain reaction
SNP	Single nucleotide polymorphism
SRM	Selected reaction monitoring
TAA	Tumor-associated antigen
TARGET	Therapeutically Applicable Research to Generate Effective Treatments
TBI	Total body irradiation
TCGA	The Cancer Genome Atlas
TCR	T cell receptor
TESLA	Tumor Neoantigen Selection Alliance
TNF	Tumor necrosis factor
Treg	Regulatory T cell

WES	Whole exome sequencing

CHAPTER 1: Introduction to Minor Histocompatibility Antigens

(1.1) Introduction to the immune system

The immune system represents a collection of organs, tissues, cells, and processes that distinguish self from non-self and protect the body from threats. These threats are mostly external pathogens such as viruses, toxins, bacteria, fungi, or parasites, but can also be threats from within (e.g. cancer). The organs that make up the immune system include the thymus, lymph nodes, spleen, and bone marrow. Bone marrow is the site of hematopoiesis, the thymus is where mature T cells are generated, lymph nodes facilitate interactions between antigens and immune cells, and the spleen filters blood and also facilitates antigen-immune cell interactions. The lymph nodes and spleen are also the site of antigen specificity acquisition for B cells. These components are arranged in a parallel circuit such that cellular components of the immune system move through these tissues and throughout the body via the circulatory system and lymphatics. The lymphatic vessels facilitate host defenses by collecting interstitial fluid from body tissues, draining this through the lymph nodes allowing immune cells to encounter antigens from peripheral tissues, and then returning lymph fluid into the peripheral circulation[1]. Hematopoietic stem cells give rise to all blood cells including erythrocytes and platelets, myeloid cells, and lymphoid cells. The immune system is broken into two major divisions: innate and adaptive immunity.

Innate immunity serves as the first line of defense of the immune system. This system generates rapid antigen-independent responses and has no immunologic

memory[2,3]. The innate system includes defensive barriers such as skin and mucous membranes, physiologic defenses such as low stomach pH that kills microbes, circulating proteins that are able to bind and inhibit pathogens, and cells that endocytose foreign molecules and/or produce chemical mediators like cytokines and chemokines that recruit additional immune cells to the site of defense[4]. The immune components that generate these responses include phagocytes such as macrophages, dendritic cells (DCs), and neutrophils, natural killer cells (NK cells), and the complement cascade[5]. Innate immune responses are triggered by engagement of pattern recognition receptors that coordinate immune cells to respond to different pathogenic classes[4]. Phagocytes engulf and kill cells and ingest cellular debris to present antigens to adaptive immune cells, NK cells kill cells that have downregulated normal cell surface molecules and/or upregulated sets of stimulatory molecules, and complement proteins form pores in target cells to kill them[6–8]. These elements together form a system that rapidly responds to a broad range of target pathogens, but does not include antigen-specific responses or elicit immunological memory to these targets.

The adaptive arm of the immune system comprises the lymphocytes (T cells and B cells) and is responsible for antigen-specific responses and memory. It takes 7-10 days to develop primary adaptive responses in humans, though memory recall responses are more rapid[7]. Exogenous antigens from pathogens like viruses and bacteria as well as endogenous antigens like tumor-associated or -specific antigens can be targeted by adaptive responses[9]. B cells drive the humoral side of adaptive immunity by producing antibodies, antigen-specific immunoglobulin proteins that bind the target antigen to drive downstream effects that depend on the antibody isotype and target

antigen. These outcomes include neutralization by antibodies that block portions of a pathogen rendering it unable to interact with host cells, agglutination of clusters of antibodies bound to antigens to mark a target for phagocytosis, or complement fixation to mediate lysis of the foreign cell[10]. T cells are the primary effectors of cell-mediated cytotoxic adaptive immunity. T cells may be divided into several subtypes based on differences in function, including “killer” cytotoxic CD8+ T cells and “helper” CD4+ T cells. CD8+ T cells recognize antigens from infected cells, cancer cells, or other antigens recognized as non-self and induce cell death in cells presenting non-self antigens[11]. There are two primary mechanisms of CD8+ T cell-mediated cytotoxicity. The CD8+ T cells secrete granules containing perforin, granzymes, cathepsin C, and granulysin that fuse with the target cell membrane and form pores in the membrane and induce death[12]. CD8+ T cells also express Fas ligand which ligates Fas receptors on target cells to activate death domains and pathways leading to fragmentation of DNA and eventually cell death[12]. While CD8+ T cells exert direct cytotoxic effects and B cells secrete antibodies, “helper” CD4+ T cells perform a wide variety of supportive and regulatory roles including secretion of cytokines, providing costimulatory or coinhibitory signals to other adaptive immune cells, inhibition of harmful adaptive immune responses, and cytotoxic activity[13]. Helper CD4+ T cells are subdivided into multiple types, including Th1, Th2, Th17, and regulatory T cells (Tregs). Th1 cells secrete interferon (IFN)-γ and tumor necrosis factor (TNF)-β to protect against intracellular pathogens by activating phagocytes and promoting proliferation and cytolytic activity of CD8+ T cells, as well as promoting the development of memory CD8+ T cell and Treg responses[14]. Th2 cells coordinate responses to extracellular parasites and secrete

interleukin (IL)-4, 5, 10, and 13 to activate B cells to promote antibody isotype switching, somatic hypermutation and affinity maturation, and antibody production as well as to induce degranulation of granulocytes and release of proinflammatory mediators[14]. Th2 cells play an important role in asthma and allergic diseases via allergic airway inflammation, histamine release, and activation of eosinophils[15]. Th17 cells generate responses against extracellular bacteria and fungi by secreting IL-6, IL-17, and IL-22 and TNF- α to activate T cells, induce B cell differentiation into plasmocytes and memory B cells, and enhancing mucosal defenses by inducing antimicrobial peptides and activating neutrophils[14]. In addition to these conventional helper T cell types, an additional CD4+ T cell type are regulatory T cells, or Tregs. Tregs protect against immunopathology by suppressing potentially harmful activities of helper T cells through a variety of mechanisms including secreting inhibitory cytokines like IL10 and expressing high levels of CTLA-4 to inhibit proliferation of T and B cells and maintain immunologic tolerance of self-antigens[16,17].

Adaptive immune cells arise in the bone marrow and mature in the thymus (T cells) and bone marrow (B cells). T cells go through a maturation process upon arrival of precursor cells in the thymus that includes differentiation, proliferation, recombination and assembly of the TCR, initiation of expression of T cell-specific cell surface proteins, and positive and negative selection[18]. The T cell receptor is structurally similar to B cell-generated immunoglobulins, and acts as a membrane-bound antigen-specific receptor. The TCR includes a variable region that confers antigen specificity and a constant region. It can be composed of a paired heterodimer of α and β chains, or in the case of the rarer γδ T cells that represent 1-10% of the T cell repertoire, a paired γ chain and δ

chain heterodimer. The TCRα and TCR γ loci contain V and J gene segments, while the TCRβ and TCRδ loci contain D gene segments in addition to V and J. The initial stage of TCR assembly via somatic recombination occurs when the gene segments of TCR-β, TCR-γ, and TCR-δ chains rearrange. Cells commit to the γδ T cell lineage if formation of a productive γδ TCR complex is completed first, or the αβ lineage if they first form a pre-TCR complex of rearranged TCRβ plus the invariant pTα placeholder[19,20]. T cells committed to the αβ lineage then rearrange the TCRα locus, generating complete VJα and VDJβ V-region exons that are transcribed and spliced to constant regions, then translated to produce α and β chains that pair to form a complete TCR heterodimer[21]. Somatic recombination generates TCR diversity due to the combination of germline V(D)J segments, with the TCRα locus having approximately 70 V regions and 61 J regions to select from, while the TCRβ locus has approximately 52 V regions, 2 D regions, and 13 J regions[21]. If selected at random, this gives about 5.8 million potential gene region combinations for an αβ pair, which while large is insufficient to address the vast antigen space that T cells must be capable of responding to. Further TCR diversity is contributed by junctional diversity, in which germline nucleotides can be deleted, and the double stranded breaks made in order to join gene segments can be filled by the random incorporation of P and N nucleotides prior to ligation of the segments[22]. This junctional diversity is estimated to bring the total possible T cell diversity to on the order of 10^{18} possible TCRs[21]. Once a complete TCR αβ pair is assembled, T cells undergo positive and negative selection in the thymus to become mature. T cells (at this stage known as double positive T cells because they are expressing surface proteins CD4 and CD8) first encounter thymic cortical epithelial cells

in the cortex that utilize thymoproteasomes and lysosomal proteases to generate and present peptides derived from the endogenous proteome on Major Histocompatibility Complex (MHC) class I and II[23]. This step is known as positive selection, and generates MHC restriction and ensures that T cells will be capable of productively engaging with peptide-MHC complexes. Cells that recognize peptide-MHC class I complex commit to the CD8 lineage while cells that recognize peptide-MHC class II commit to the CD4 lineage via the activity of two opposing transcriptional regulators, Thpok and Runx3, that initiate repression of the opposite surface protein[24]. About 90% of T cells die of “neglect” via apoptosis at this step due to lack of TCR:peptide-MHC engagement, which generates critical survival signaling[23]. Following positive selection, the remaining T cells that are now single positive for CD4 or CD8 undergo negative selection to remove T cells with high affinity for binding self-peptides or self-MHC, eliminating cells that are potentially self-reactive. Cells undergo directed migration to the thymic medulla via the chemokine receptor CCR7 for negative selection[25]. In the medulla they encounter medullary thymic epithelial cells (mTECs), conventional DCs, and plasmacytoid DCs presenting self-peptides on MHC class I and II[23]. mTECs express the human autoimmune regulator (*Aire*) that controls the expression of over 3,000 downstream genes, increasing gene transcription by interacting with RNA polymerase II and releases it from where it stalls at promoter regions of genes, freeing them to proceed with transcription[26]. *Aire* allows for mTECs to express many more genes than other cells including a large number of peripheral tissue-specific genes that would otherwise not be expressed in the thymus, allowing for presentation of peripheral tissue self-antigens in the thymic medulla to generate immunological tolerance of self-peptides from

throughout the body[26]. This promiscuous gene expression continues throughout the entire time of thymic T cell output[27]. mTEC lifespan is short due to the double stranded breaks generated by the action of *Aire*, leading to rapid turnover[28].

Within the thymus, most T cells that strongly bind self-antigen-MHC complexes are eliminated via repression of the anti-apoptotic protein Bcl2, which represses pro-apoptotic proteins including Bim and Puma[29]. While most cells that bind self-antigens presented on MHC undergo apoptosis, a portion of autoreactive CD4 T cells differentiate into regulatory T cells via agonistic selection[25]. These Tregs are important for maintaining immunological tolerance. In summary, T cells generate unique TCR α and β chains that pair to form a heterodimeric TCR, then T cells traffic through the thymic cortex to remove T cells that do not engage with peptide-MHC in positive selection, then move through the thymic medulla to remove T cells that engage self-peptide-MHC too strongly in negative selection. This pathway results in mature naïve single positive T cells that then traffic throughout the body to secondary/peripheral lymphoid organs that include spleen and lymph nodes in order to make contact with circulating pathogens or antigen-presenting cells (APCs)[9].

T cells become activated upon contact with antigen presented on the appropriate MHC, with CD8+ T cells activated by MHC class I-presented antigens and CD4+ T cells activated by class II-presented antigens. CD8 and CD4 act as coreceptors to the TCR to increase binding affinity to peptide-MHC. T cells need three signals to be optimally activated. Signal 1 is TCR binding to a peptide-MHC complex, initiating intracellular signaling cascades such as the MAPK and NFAT pathways that lead to T cell activation. Signal 2 occurs via binding of costimulatory molecules such as the binding of CD28

from the APC to CD80/86 on the T cell, which completes the signaling necessary for survival, proliferation, and differentiation. Signal 3 is T cell reception of proinflammatory cytokines such as type 1 interferons that are released from the APCs, which lead to continued survival, proliferation, and formation of antigen-specific memory responses[30,31]. If cells receive Signal 1 but no binding of costimulatory molecules for Signal 2 and/or presence of proinflammatory cytokines for Signal 3, this results in T cell anergy and/or tolerance of the cognate antigen[30]. It is also possible for T cells to receive a negative signal at Signal 2, such as CD80/86 binding CTLA-4, preventing CD28 ligation and preventing activation[31]. Cytokines can also serve roles in T cell activation other than directly providing Signal 3 to the T cells. For example, the cytokine IFN-γ enhances antigen presentation to CD8+ T cells by upregulating expression of MHC class I and antigen processing pathways[32–34]. Overall, T cell activation state is determined by a complex aggregation of stimulatory and inhibitory signaling, simplified below (Figure 1.1).

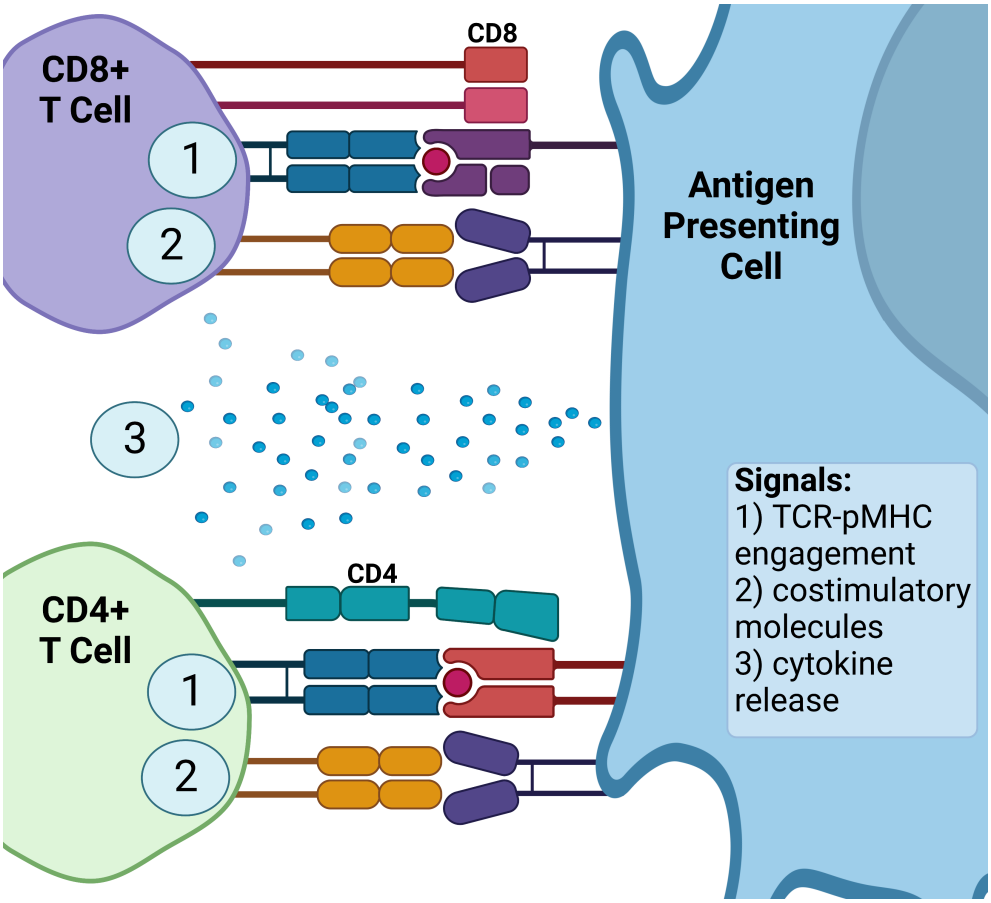


Figure 1.1: T cell activation signals. Figure shows 3 signals necessary to activate CD8+ T cells and CD4+ T cells.

Generation of Signal 1 requires a complex process of antigen processing and presentation on the cell surface of the APC. MHC class I molecules are expressed by all nucleated cells and drive cytotoxic CD8+T cell responses, while class II molecules are presented on “professional” antigen-presenting cells including DCs, macrophages, and B cells and drive helper and regulatory CD4+ T cell responses[35]. For class I MHC proteins, peptides between 8-11 amino acids bind to specific amino acids at the C and N terminus of the MHC protein leading to presentation of amino acids in the center of the peptide for recognition by CD8+ T cells. In contrast, class II MHC proteins are not anchored at the N or C terminus and as a consequence bind peptides that are 13-25

amino acids long[35]. The set of peptides that binds MHC is also determined by the class I and II HLA alleles that are encoded in an individual based on the HLA haplotypes inherited from each parent. Thousands of HLA alleles are known, and alleles and haplotypes vary in prevalence in different ethnic populations[36].

Endogenous antigens are primarily presented by MHC class I, while class II primarily presents exogenous antigens[37,38]. Antigen cross-presentation is also possible, in which antigens presented to CD8+ T cells by MHC class I are exogenously derived. This process is crucial for adaptive immune responses to viral infections and solid tumors because it allows for presentation of tumor or viral antigens by cells other than the directly affected cell to coordinate CD8+ T cell responses. In addition to presentation to different cell types, antigen processing also differs between class I and class II (Figure 1.2).

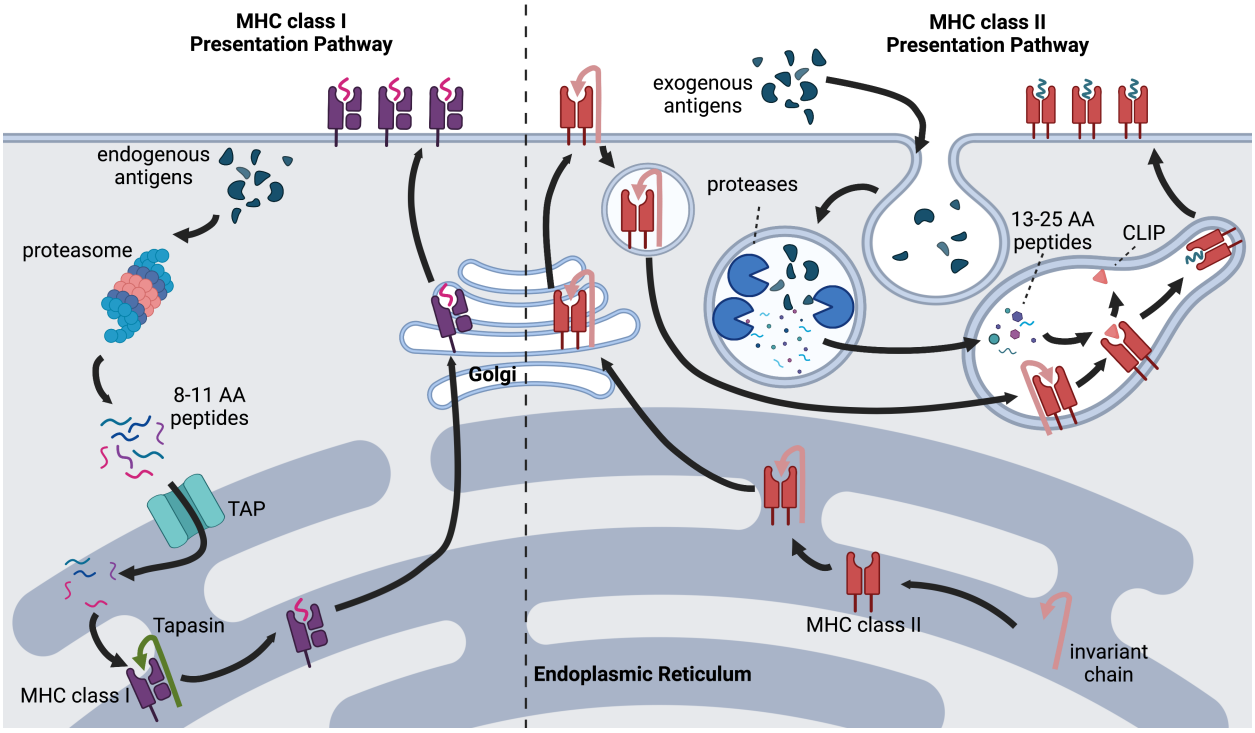


Figure 1.2: MHC class I and II antigen processing and presentation pathways[39,40].

For class I MHC, proteins are degraded in the cytosol and nucleus then routed through the cytosol into the ER by TAP, where MHC class I heterodimer assembly takes place from heavy chain, beta 2 microglobulin, and the peptide[41]. Until peptide binding, MHC class I is stabilized by chaperone proteins including tapasin, calreticulin, and others, and these release when peptides bind MHC class I and the fully assembled complex is freed to exit the ER and move to the cell surface via the Golgi complex[41]. Free peptides are degraded by cytosolic peptidases within seconds, meaning that the number of peptides that bind TAP and ultimately get presented is much lower than the sum of all peptides produced by proteasomal degradation of proteins[42]. For class II antigen presentation, exogenous antigens are endocytosed or phagocytosed and degraded in endosomal-lysosomal compartments or phagolysosomes[43]. MHC class II alpha and beta chains are synthesized and assembled into heterodimers in the ER and bind invariant chain which targets the complex to the peptide-containing endosomal-lysosomal compartments[43,44]. The complex moves from the ER through the Golgi to the plasma membrane, and is endocytosed to then follow the same pathway as the peptides it will present. This invariant chain is degraded and removed in the antigen-processing compartment except for a small remaining component called class II-associated invariant chain peptide (CLIP) that occupies the class II binding groove[45]. The enzyme HLA-DM then removes CLIP to allow a processed peptide to bind the groove, and the completed class II complex moves via transport vesicle to be inserted into the plasma membrane[45].

In summary, the immune system comprises adaptive and innate components, and T cells and B cells are the primary mediators of antigen-specific adaptive immune

responses. CD8+ T cells kill target cells in response to endogenously derived short peptides presented on MHC class I, while helper CD4+ T cells respond to longer and exogenously derived peptides presented on MHC class II. Together, these responses allow for a rapid and dynamically adaptive response to fight infection and disease. This work hereafter primarily focuses on antigens presented by MHC class I to CD8+ T cells to generate cytotoxic responses.

(1.2) Predictors of antigen presentation and immunogenicity

Although the exact determinants of a given TCR/peptide/HLA binding interaction are impossible to perfectly predict, some characteristics of the binding interaction have been discovered that make individual peptides more or less likely to bind HLA and generate a cognate T cell immune response. On the peptide binding side, each HLA allele has a slightly different conformation that leads to binding somewhat overlapping but distinct peptide repertoires[46]. Peptides that are more likely to bind contain specific amino acids at so-called “anchor residues” towards the ends of the peptide which have more contact with the binding groove and are therefore stronger determinants of binding[47–49]. These anchor residues vary based on HLA allele, but are concentrated around peptide side chains[46]. The conformational differences between binding grooves of alleles driving differential binding and therefore different sets of peptides bound and presented by each allele[46]. The peptide C-terminal side chain is one such anchor residue that is highly important for binding and stability of the peptide-MHC complex[49]. Computational scientists have utilized datasets of peptides that are known to bind MHC to train *in silico* models to predict peptide binding to MHC class I and II alleles. These computational tools such as netMHCpan, MHCflurry, and others are available for

researchers to estimate binding prediction values for antigens with HLA alleles of their choice[50,51]. While imperfect, these tools have improved over time as additional training data on antigens validated by mass spectrometry have been generated. The options for predicting which validated HLA-presented antigens will actually elicit antigen-specific T cell responses *in vivo* are even more limited, as the majority of presented antigens do not generate a cognate T cell response and the factors that dictate immunogenicity and the degree to which they matter are unclear[52]. However, studies have established some of the factors that do correlate with endogenous T cell responses such as strong peptide/MHC binding affinity, long half-life (peptide/MHC binding stability), high RNA and protein expression, low agretopicity (ratio of mutant to wild-type binding affinity), and high foreignness (degree of homology to known self-peptides)[52]. Of these, peptide binding affinity, binding stability, and protein expression seem to have the strongest and most reproducible correlations with MHC presentation and elicitation of antigen-specific T cell responses[52]. Peptides with longer MHC interaction half-lives are more likely to come in contact with TAP and be transported for MHC presentation, and strong binding affinity allows for peptide presentation. High expression of the peptide source gene is a strong predictor of an antigen-specific T cell response because high abundance of the protein makes it more likely that the peptide of interest will be produced by proteasomal degradation and presented by MHC, though not all possible peptides from a given protein are generated with equal probability and frequency by the proteasome. The importance of agretopicity and foreignness are less agreed upon, but it appears that mutated peptide variants that are much more likely than the normal form to bind MHC, or peptides with further distance from the normal proteome, may be more likely to be

recognized as foreign[52–56]. Our group and others are working to utilize deep learning and other computational methods to predict antigens that will generate T cell responses, and similarly to binding prediction tools, these approaches will approve as more antigen-specific T cell clones are discovered[57–61].

(1.3) Defining mHAs

Minor histocompatibility antigens (mHAs) are MHC-presented peptide antigens that are relevant in the context of hematopoietic cell transplant. mHAs are derived from nonsynonymous single nucleotide polymorphisms (SNPs) that differ between a transplant donor and recipient, leading to an amino acid change in a peptide such that the recipient peptide is foreign to the donor[62–64]. In allogeneic hematopoietic cell transplantation (alloHCT), any given donor-recipient pair (DRP) has their own individual set of mHAs due to the large number of SNPs that differ between the genomes of any individuals, on the order of thousands to millions based on the estimation of 1 SNP per 1000 bases in the average genome[65]. These SNP-derived peptides are presented by recipient MHC proteins, termed human leukocyte antigen (HLA) proteins in humans (Figure 1.3)[66].

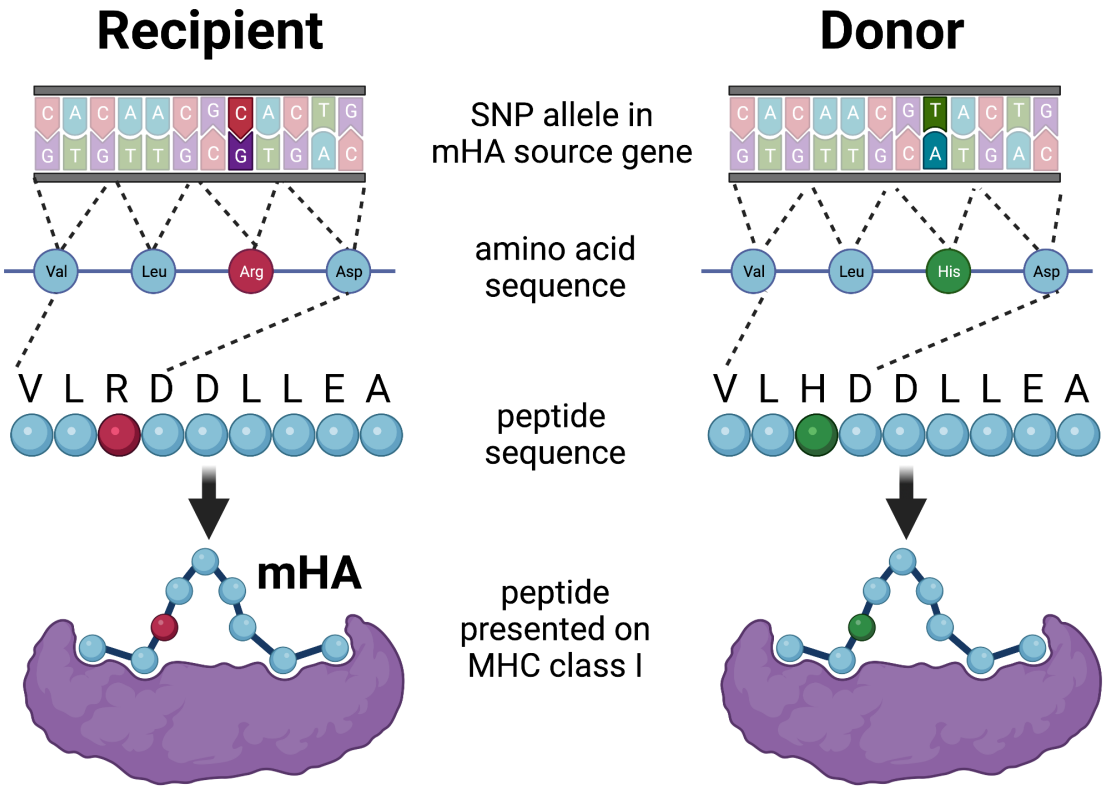


Figure 1.3: Minor histocompatibility antigen presentation on class I MHC. Minor antigen HA-1 (VLRDDLLEA) is used as an example[67].

T cell recognition of these mHAs presented by recipient HLA can lead to T cell-mediated cytotoxicity of recipient cells that express these mHAs. This dissertation will focus on mHAs in alloHCT because these antigens are critical to graft versus host disease (GvHD) and the graft versus leukemia (GvL) effects of transplant[68–70]. Within a single DRP, the T cell recognition of mHAs produces a complex combination of GvL and GvHD. It is hypothesized that the effect that each individual mHA mediates in a transplant recipient depends on the tissue type that the mHA is presented within, with hematopoietic cell expression leading to GvL and expression by epithelial cells leading to GvHD. (Figure 1.4)[71–73].

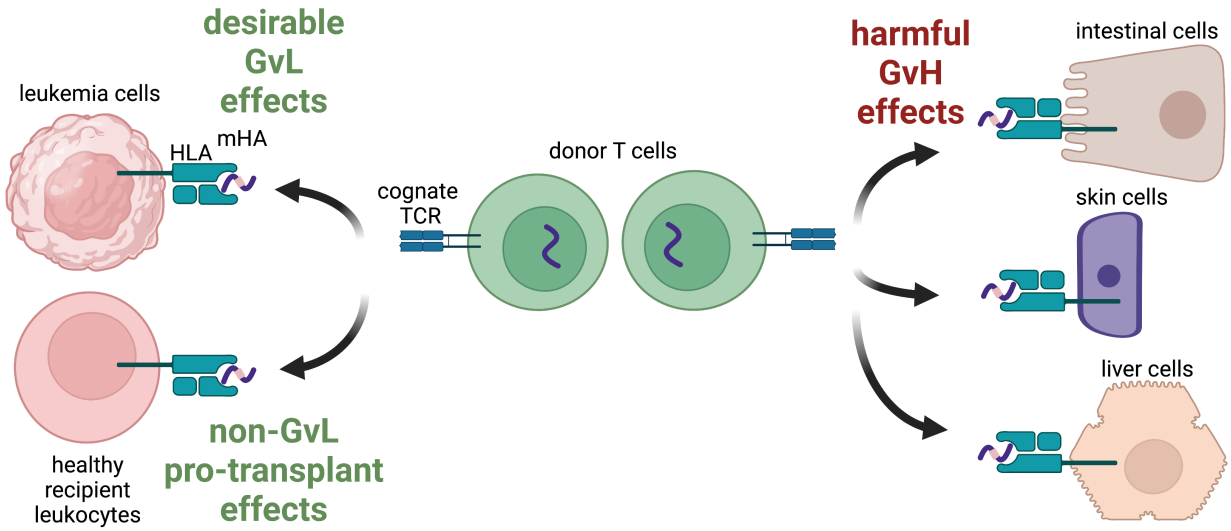


Figure 1.4: GvL and GvH effects of T cells targeting mHAs.

If an mHA is derived from a gene that is expressed in leukemia cells, it is hypothesized that the immune response against that antigen will drive T cell-mediated killing of leukemic cells and therefore the mHA be classified as a GvL mHA. If the mHA is derived from a gene expressed in acute GvHD target organs such as skin, colon, and liver, responses to that mHA will generate graft versus host (GvH) effects. Acute GvHD is characterized by a proinflammatory response mediated by Th1 and/or Th17 cells, while chronic GvHD is a complex pro-fibrotic response mediated by B cells and Th17 cells[74–76]. The pathophysiology of acute GvHD is a three-step process including recipient tissue damage from radiation and chemotherapy included in pre-transplant conditioning, donor T cell activation and clonal expansion in response to tissue damage, and release of cytotoxic effectors of tissue injury[77]. In contrast to the T cell infiltrates seen in acute GvHD, chronic GvHD is characterized by a fibroproliferative signs and acellularity[78]. The clinical symptoms of donor T cells generating immune responses against recipient tissues in acute GvHD include maculopapular rashes, jaundice and

cholestasis, nausea, vomiting, and diarrhea[79]. Symptoms of chronic GvHD can include those similar to acute GvHD as well as dry eyes or mouth, shortness of breath, joint inflammation and discomfort, fatigue, sclerotic skin lesions, vulvovaginitis, or other symptoms of chronic inflammation in a variety of systems[80]. While acute GvHD is almost exclusively mediated by donor T cells, chronic GvHD pathogenesis has been shown to also involve donor B cells[81]. mHAs can also fall into a third category when they are derived from genes that are expressed in healthy circulating or tissue-resident white blood cells (leukocytes). These genes are often also expressed in leukemia cells, making mHAs derived from them classifiable as GvL mHAs, but occasionally an mHA is derived from a gene that is only expressed in healthy leukocytes and not in leukemia. These mHAs are often grouped in with GvL, but T cells targeting these mHAs would not be expected to kill leukemia cells. Instead, recipient healthy leukocytes will be targeted. In the context of alloHCT, this could still be beneficial by eliminating recipient-derived leukocytes and thereby supporting donor hematopoietic stem cell engraftment. Therefore, these non-typical mHAs belong in a third category of mHAs that can lead to a beneficial effect that is not mediated by leukemia targeting.

mHAs can be considered either class I or class II antigens based on whether they are presented on HLA class I or II proteins. We focus on class I mHAs in this work, as these are directly recognized by CD8+ T cells leading to T cell-mediated cytotoxicity, though other studies have shown that class II mHAs can be beneficial targets for GvL as well[82,83].

(1.4) Relevance to allogeneic hematopoietic cell transplantation

mHAs are relevant in the context of alloHCT due to its unique clinical context in which, following engraftment, a donor-derived hematological system is functioning within a completely foreign environment of tissues and antigens. alloHCT has been increasingly used over the last 50 years as a treatment for both malignant and nonmalignant conditions in which therapeutic effect requires a recipient to have donor-derived hematopoiesis[84]. It was first conceptualized as a treatment for radiation-induced bone marrow deficits after WWII, with experiments being performed to reconstitute bone marrow after radiation treatment in mice[84]. The first human alloHCT was performed in 1939 when a woman with aplastic anemia received intravenous injection of bone marrow from her brother along with a large number of transfusions in an attempt to prolong life, though successful alloHCTs were not performed for several more decades[85]. Work progressed to efforts to treat leukemia in mice, canines, and humans throughout the 1960s[84]. Early on, it was determined that endogenous hematopoiesis must be ablated prior to transfer of new hematopoietic materials in a process known as pre-transplant conditioning. This was initially done by administering radiation, but it was soon discovered that conditioning chemotherapy such as cyclophosphamide could be utilized as an alternative[86]. One crucial discovery that allowed for the alloHCT field to progress in humans was the identification of the MHC locus in humans, first discovered by Jean Dausset with Felix Rapaport in 1965[87–92]. After this discovery, the first successful allogeneic (donation from another person, in contrast to autologous transplant in which the donor source is self) transplants were performed with HLA-matched sibling donors for bone marrow failure syndromes and immunodeficiencies in the 1960s-1970s[84]. The first successful curative alloHCTs for malignant conditions were

reported in a 100 patient cohort in 1977, encouraging further development of alloHCT for malignancies and establishing that alloHCT is more effective in patients with early stage disease than with advanced disease[93].

As mentioned previously, mHAs mediate effects of alloHCT by eliciting donor T cell responses against recipient tissues. When mHAs are presented on AML or leukocytes, the corresponding outcome is a desirable GvL effect. The GvL effects of mHAs were first uncovered based on observations of a lack of therapeutic efficacy when alloHCT was performed with identical twin donors compared to HLA-identical non-twin sibling donors[94]. The other major effect of mHA targeting is GvHD, which occurs as discussed when mHAs are presented on healthy recipient tissues. GvHD risk can be reduced via prophylactic treatments including calcineurin inhibitors like cyclosporine or tacrolimus in combination with methotrexate, an antimetabolite that attenuates T cell activation, or mycophenolate mofetil, an inosine monophosphate dehydrogenase inhibitor[95]. Though these are most often used, other methods of prophylaxis include sirolimus, posttransplant cyclophosphamide, or T cell depletion via antithymocyte globulin (ATG)[95]. GvHD can also be treated after it arises using immunosuppressants. Glucocorticoids are the first line agent for acute GvHD, though steroid monotherapy will fail in 35-50% of patients[95,96]. Although it is known that mHAs drive GvHD and GvL, it is currently unknown whether these effects are dependent upon a few immunodominant mHAs that drive the alloantigen-specific T cell response or whether total number of mHAs is important for driving either or both effects[97].

(1.5) Clinical Context of alloHCT for Acute Myeloid Leukemia

In the modern era, alloHCT is utilized for intermediate or high risk cases of Acute Myeloid Leukemia (AML) in patients in Complete Response 1 (CR1) with characteristics that allow for successful alloHCT, as determined by a risk of relapse greater than or equal to 40% if treated with induction chemotherapy alone, and a low risk of treatment related mortality due to age and comorbidities[98]. AML is stratified into risk groups based on cytogenetic and molecular factors. The favorable/low risk group includes patients with NPM1 mutations, biallelic mutation of CEBPA, RUNX1-TUNX1T1 translocations, and other mutations[99]. The adverse/high risk group includes patients with complex cytogenetics or high risk mutations such as BCR-ABL1 translocations or FLT3-ITD with wild-type NPM1[99]. Finally, the intermediate risk group contains patients with cytogenetic and molecular abnormalities that are neither favorable nor adverse. Based on current treatment guidelines, it is suggested that patients with favorable risk factors do not go on to transplant unless they have a c-KIT mutation, while patients with unfavorable risk proceed to transplantation. Intermediate risk patients may or may not go on to transplant depending on specific cytogenetic risk factors and on availability of an HLA-matched donor[100]. Older patients follow generally the same stratification framework, but with individualized considerations of age and comorbidities to determine whether the benefits of transplant will outweigh the risk.

Prior to alloHCT, the first step in treatment of AML patients is to administer induction chemotherapy to induce remission. More than 70% of young patients that undergo induction therapy achieve CR1, with lower rates for older patients[101–103]. If a patient does not enter CR1 after induction therapy, they may receive another round of induction therapy to reach CR1, though this places patients into a higher risk group than

their cytogenetics would indicate[100]. AML patients with high-risk disease may proceed to alloHCT once they achieve CR1. Otherwise, patients with relapsed disease after prior therapeutic regimens may be treated with alloHCT if they achieve remission with further lines of anti-leukemia therapy.

Immediately prior to receiving alloHCT, recipients receive myeloablative conditioning (MAC) or reduced intensity conditioning (RIC) therapy for disease eradication, immunosuppression, and creating space for transplanted cells to engraft[104]. Examples of MAC regimens include cyclophosphamide with total body irradiation (TBI) or busulfan-based treatments such as fludarabine/busulfan[98]. RIC regiments are less taxing and thus beneficial for patients with intermediate to high risk AML who are older or have comorbidities such that MAC regimens would be too harmful[105]. Common RIC regimens include fludarabine with doses of TBI or busulfan that are lower than what is utilized in MAC regimens[105].

Once remission has been achieved and conditioning therapy given, patients can go on to receive an alloHCT. Most alloHCTs have historically been performed with either an HLA matched related donor (MRD) or matched unrelated donor (MUD). Only 30% of patients have a suitable HLA-matched immediate family member to serve as an MRD, so the remainder of patients rely on availability of HLA-matched donors in national donor registries[103]. Due to ethnic and allelic makeup of the donor registry, chance of finding an HLA-matched donor varies greatly by race. White patients have the best chance of finding a matched donor, with a 79% chance of match, while other races have worse chances, with Black patients only having a 29% chance of finding a match in the registry[106]. In cases where a matched donor is unavailable, alternative donor

sources such as umbilical cord blood, haploidentical donor, or mismatched unrelated donor may be considered, but this is beyond the scope of this dissertation. For MUD or MRD alloHCTs, cells will be harvested from the donor via either a bone marrow harvest from the iliac crest or apheresis of stem cells mobilized from peripheral blood. If donating via peripheral blood, donors receive a mobilizing factor such as G-CSF or CXCR4 blockers in order to increase the number of stem cells in the peripheral blood before harvest[107]. Peripheral blood grafts appear to give better engraftment but a slightly increased risk of acute and chronic GvHD[108]. Fresh or cryopreserved stem cells are then infused into the patient and allowed to engraft within their bone marrow and perform hematopoiesis. Prognosis for patients that relapse after receiving alloHCT is poor, but in this case patients can receive additional salvage chemotherapy, donor lymphocyte infusion, or a second alloHCT[99]. Clinical trials with novel therapeutic agents or regimens may be used for patients that fail any of these treatment modalities.

In selecting a donor for alloHCT, though matching at all HLA loci is preferred, matching at certain loci is more important than at others. Matching can be defined at different levels based on the number of alleles that are considered. Mismatching at crucial alleles can lead to graft rejection, GvHD, or worsened patient outcomes including decreased overall survival. The National Marrow Donor Program recommends an 8/8 match for matched alloHCTs, referring to matching at both alleles at HLA-A, B, C, and DRB1 loci[109]. Most European centers consider a 10/10 match, meaning an 8/8 match plus DQB1, to be the gold standard[110]. A 12/12 match refers to matching at these loci plus DPB1. Additional matches beyond 8/8 have been shown in a variety of recent studies to impact outcome as well[109]. Effects of DPB1 matching is a controversial topic,

but there are categories of DPB1 mismatches that are considered permissive versus nonpermissive as some mismatches lead to worse GvHD and mortality outcomes while others have no impact[111–113]. DPB1 is also especially complex because HLA-DR+ leukemias generally also express HLA-DP, so it may mediate GvL effects in addition to GvHD[114]. Even within alleles that are considered a required match including A, B, C, and DQB1, some specific mismatches can be well tolerated while others can lead to transplant rejection or GvHD[115–117]. Another donor source option is haploidentical transplantation, where a patient receives an alloHCT from a relative that is a half match due to unavailability of a matched donor. Recent innovations in GvHD prophylaxis have made this graft type possible, and patients receive posttransplant cyclophosphamide to prevent GvHD and graft rejection[118]. These transplants help to bridge the gap in donor availability by race: while many non-Caucasian patients do not have a matched donor available in the registry, most patients have a haploidentical donor available in the form of a first degree or occasionally even further degree relative. One center reports a greater than 95% success rate in identifying a haploidentical first degree relative for their patients, with an average of 2.7 haploidentical donor sources available per patient[119]. Data on outcomes from MUD versus haploidentical alloHCT are mixed but generally favors MUD where available due to higher incidence of graft failure and GvHD in haplo alloHCTs[118]. Ultimately, determination of what specific donor is best for a recipient is a highly complex topic, and donor availability, clinician judgment, and personal preference play a strong role in donor selection.

Another factor to consider in donor selection is degree of genetic similarity of recipient and donor. As previously mentioned, the lack of graft versus leukemia effects

of transplant with identical twin donors led to discovery of the GvL effects of mHAs[94]. However, it is unknown how the increased genetic distance between donor and recipient in a MUD DRP versus a MRD impacts the number of biologically relevant mHAs in alloHCT. Studies have shown many conflicting results when comparing MUD vs MRD alloHCT, though the largest study to date found no significant difference in survival times based on donor source[120]. Unpublished and published data from our lab have shown that while genetic distance measured via SNP arrays is greater in MUD DRPs than in MRD DRPs, genetic distance does not correlate with total number of predicted mHAs or with number of predicted GvL mHAs per DRP[121]. Overall, we know that some degree of genetic distance is necessary between donor and recipient in order to generate GvL effects and prevent relapse, but genetic differences (mHAs) also generate GvHD. Donor selection and transplant medicine are a delicate balance of reducing transplant-related mortality while also seeking to decrease relapse and prolong survival.

(1.6) History of mHA discovery

The concept of transplant tissue incompatibility arose approximately 80 years ago with the work of Ray Owen, Sir Peter Medawar, and their contemporaries. In 1943, Medawar performed the first systematic examinations of the phenomenon of transplant rejection. He observed a burn patient who received skin grafts both from her own tissues (autograft) and from a donor (allograft), assessed these macroscopically and histologically and found that the allografts were rejected more rapidly than autografts[122]. Skin grafting had been practiced as long ago as 1500 BC but was largely abandoned during the middle ages until reimplementation in the 1800s, with the first successful skin

graft of the modern age performed in 1822[123]. It was widely known that grafts from other individuals would fail, with the exception of grafts from identical twins, but the reasons for this were not understood until the mid-1900s[122]. Medawar's work helped to fill this gap by supporting the hypothesis that rejection of allografts was mediated by the immune system reacting to foreign antigens from the donor. Owen's work added to the concept of transplant compatibility by exploring the idea of immune tolerance. His critical observation took place in 1945 when observing a pair of dizygotic twin calves that had apparently shared circulation during development, leading to antigens from one calf not eliciting an immune response in the other[124]. In 1949, Burnet and Fenner then used Owen's observation to support the idea of immune tolerance and propose the concept of fetal recognition of "self-markers" that lead to tolerance in their book on antibodies[125]. Medawar and Burnet received the 1960 Nobel Prize for the discovery of immunological tolerance, which they had both worked on independently of each other[125].

In addition to immunological tolerance, the discovery of the MHC locus was also a pivotal advance that enabled the progression of discoveries in transplant biology. After several decades of investigations into graft acceptance in mice, the MHC locus was first described in 1948 by George Snell, who named the gene encoding the antigen H2, where the H stands for Histocompatibility, eventually reflected in the MHC name[126,127]. Jean Dausset went on to discover the first MHC antigen in humans in the 1958 by observing that upon transfusion of blood from a single donor into a patient, the recipient patient developed anti-leukocyte antibodies. These antibodies also reacted with the leukocytes of about 50% of tested volunteers, meaning that the volunteers whose

leukocytes did not generate a response must share the antigen with the donor. He named the antigen MAC after the initials of three of his experimental volunteers, though this allele later came to be recognized as HLA-A2[128,129]. As discussed previously, he went on to discover the MHC locus in humans in the mid 1960s. In the 1960s-70s, Baruj Benacerraf performed guinea pig experiments showing that genetic factors related to the MHC antigens facilitate generation and strength of immune reactions[130,131]. Snell, Dausset, and Benacerraf shared the 1980 Nobel Prize in Physiology or Medicine for their work on discovering and characterizing MHC biology.

These discoveries set the stage for uncovering the antigens that mediate tolerance or rejection in the context of transplant between two individuals. An early breakthrough in this field was a 1976 publication by Van Rood and Goulmy describing a case in which a woman with aplastic anemia rejected an HLA-identical alloHCT from her brother[132]. *Ex vivo* analysis of cytotoxicity of patient lymphocytes was performed and cytotoxic T cell responses against only male HLA-A2 positive target cells was found, indicating that a T cell response to an antigen on the Y chromosome drove graft rejection[133]. This observation correlated with earlier observations from the 1950s in mice that skin grafts in females from otherwise identical males failed[134]. These graft failures indicated that there were antigens encoded on the Y chromosome that would be recognized as foreign to female mice, but tolerated and not recognized as non-self by male mice[135]. These human and mouse Y chromosome-encoded antigens are now known as H-Y antigens, a group of mHAs that are specifically relevant in alloHCTs between donors and recipients of disparate sex[136].

A crucial discovery in establishing the mechanism of immune responses to mHAs was made in 1990, when the Rammensee group established that presentation of mHAs is dependent upon MHC, and that the MHC alleles in an individual determine which peptides can be presented[137]. When they compared peptides presented on MHC in mouse spleens from mouse strains that are identical at all genes except for MHC class I, they found that these strains presented different patterns of peptides[137]. Important mHA-related developments continued in the 1990s with the discovery of the first non-Y chromosome mHAs mediating immune responses in mice. Dozens of loci driving immune responses were discovered in the context of transplants between murine strains with the same MHC alleles between the mid 1980s through 1990s[138]. In the early days of mHA discovery, loci that encode mHAs were often identified by linkage analysis or other methods that could narrow down the genomic region that the antigen is derived from, but not identify the actual antigenic peptide. Discovery of the mHA peptide sequences themselves often came later, enabled by technological developments in techniques such as expression cloning or mass spectrometric immunopeptidomics that enable peptide identification. The first autosomal mouse mHA peptide sequence was identified in 1997 after over 50 autosomal mouse mHA loci had been established, with the H13 mHA peptide discovered by the Shastri group via expression cloning[139].

Taken together, this early work established that peptides known as mHAs presented on HLA/MHC can mediate immune responses in individuals that do not recognize these antigens as self. The field progressed in the 1990s in humans in a similar pattern to how mouse mHA work evolved, with discovery of loci that impacted transplant outcomes first, followed by discovery of the specific mHAs corresponding to

the loci. Rammensee and Els Goulmy's groups investigated tissue expression of genes that generated mHAs in 1991 and 1992, finding that distribution of mHA source gene expression corresponded with class I HLA expression. Additionally, some source genes were evenly distributed across tissue types while others were specific for certain tissues, alluding to the modern understanding of mHAs as falling into subgroups based on tissue specificity[140]. Goulmy and her group then went on to discover the first known actual human mHA peptide sequence in 1995 (HA-2) in a hugely labor-intensive experiment. While they had previously identified the HA-2-encoding locus, the peptide epitope was not yet known, and so they purified HLA molecules from EBV-LCLs that expressed HA-2, isolated the HLA-bound peptides, fractionated them by high-performance liquid chromatography (HPLC), and screened fractions for response from T cells via chromium release assay. They then performed mass spectrometry on that fraction, which led to identification of more than 100 peptides. They took these peptides and again fractionated them by microcapillary HPLC and split the resulting fractions into a 96 well plate, all of which they individually screened for immunogenicity of HA-2 reactive T cells by chromium release assay[141]. Her group continued to discover several other antigens in the following years[67,133,141–143]. Her work also established that mHAs are capable of driving graft failure and GvHD, and later went on to establish mHAs as a therapeutic target for augmentation of GvL effects of transplant[144]. Other groups then continued to discover additional human mHA peptides with increasing frequency in the late 1990s-2010[67,73,115,141,145–163]. The rate of novel mHA discovery had somewhat slowed between 2010 and prior to this dissertation work, but innovation in the field has

continued during this time and much work has been done to bring mHA targeting into the clinic.

(1.7) Methods of mHA discovery

Methods used for mHA discovery fall broadly into two categories: forward immunology and reverse immunology. Forward immunology approaches have been the primary method for mHA discovery for most of the history of the field; these generally start with an activated T cell and seek to identify the mHA it recognizes. Reverse immunology approaches are becoming more prominent as technological advances make them more feasible, and generally include a computational prediction component followed by screening for immunogenicity and targeting of predicted mHAs. Only 56 mHAs (including only 12 GvL mHAs) have been discovered prior to the work described in this dissertation, with the majority discovered via forward immunology, though reverse immunology approaches have significantly accelerated discovery in recent years[121,164–168]. The most common forward immunology approach is to isolate an activated T cell clone from a post-transplant alloHCT recipient, then utilize a variety of laborious techniques to identify its cognate antigen[141,145,146,149]. A limitation of this approach is the identification of a mHA from a donor-derived clone gives no information about the mHA itself. Information about an mHA like tissue restriction, allele frequency, or HLA restriction is crucial for translation to therapeutic use and then must be determined by further experiments[169]. mHAs discovered using this forward approach are often not applicable for clinical uses due to expression in off-target tissues or other factors discovered after mHA identification. Also, they are often only applicable to a few DRPs because they are derived from SNPs with minor allele frequencies such that they are

very infrequently disparate and therefore actionable in DRPs, and these forward methods have no way to incorporate allele frequency into discovery.

Due to the logistical difficulty of forward immunology methods for mHA discovery and technological advancements that enable antigen prediction using genomics data, reverse immunology strategies for identifying mHAs have gained popularity in the last decade. In these methods, mHAs are predicted using bioinformatics tools that assess HLA specificity, binding affinity, proteasomal cleavage, and other factors[164]. Researchers then seek to experimentally validate the predicted mHAs using methods such as mass spectrometry of HLA-bound peptides or identification of mHA-targeting T cells using antigen presenting cell co-culture studies[62]. Many interesting twists on this approach are under investigation, including immunoprecipitating peptides from HLA and performing mass spectrometry, then applying mHA prediction algorithms to the identified peptide pool in order to reduce false negatives of prediction[170]. The advent of fluorescent labeled peptide-MHC molecules has also led to advances in reverse immunology approaches, as these can be used to screen T cells from alloHCT samples or healthy donors for specificity towards predicted antigens[171,172]. Combinatorial use of multiple fluorophores and barcoding of peptide-MHC multimers represent other recent technical advances that have moved the field towards higher-throughput identification of mHA-specific T cells[173,174]. On the computational side, most modern approaches to mHA prediction utilize tools such as NetMHCpan and NetMHCIIpan to predict HLA binding of SNP-derived peptides, NetMHCstabpan to predict binding stability, and databases like the Human Protein Atlas to assess tissue expression of source genes[62]. NetMHCpan and NetMHCIIpan predict peptide binding to MHC class I and II alleles with

known sequences, respectively, using artificial neural networks (ANNs) trained on experimental binding affinity and mass spectrometry data[50]. Likewise, NetMHCstabpan uses ANNs trained on peptide-MHC class I dissociation assay data to predict binding stability to class I, though no corresponding predictor exists for class II[175]. Utilizing these predictors allows for elimination of epitopes that are not likely to be immunogenic because they are not predicted to bind MHC, vastly reducing the pool of potential epitopes that must be screened in order to identify immunogenic peptides. However, prediction tools are imperfect and will yield both false negatives and false positives for peptide binding. In addition, the complete landscape of determinants of immunogenicity is unknown and though these tools in combination can identify promising epitopes, the field is currently unable to accurately predict immunogenicity. Thus, wet lab validation of predicted immunogenic epitopes is necessary. A total of 82 human mHAs are now known, including 26 discovered by the work described in this dissertation and 56 known in the field prior to this work (Table 1.1).

Table 1.1: Characteristics of all known minor histocompatibility antigens in humans[62,67,73,73,115,141,145–159,161–163,167–169,176–184].

mHA Name	Gel. mHA	Gene	Chromosome	Peptide	SNP	mHA Nucleotide/Alternate Nucleotide	mHA Amino Acid/ Alternate Amino Acid	HLA Allele	MAF in TOPMED	MAF in AFR	MAF in 1000 GENOMES	Shared mHA	Population Frequency in DISCOVeRY-BMT DRPs with Corresponding HLA Allele(%)	Immunolog y Technique	First Author	PI	Year
HA-2	Yes	MYO1G	7	YIGEVLVSV	rs61739531	C/T	V/M	A*02:01	0.82	0.79	0.86	Yes	4.2	forward	den Haan	Goulmy	1995
HA-1	Yes	HMHA-1	19	VLRDDLLEA	rs1801284	G/A	R/H	A*02:01/B*60	0.58	0.63	0.59	Yes	12.3	forward	den Haan	Goulmy	1998
HB-1H	Yes	HMHB1	5	EEKRGSLHVW	rs161557	C/T	H/Y	A*02:01/B*44:03	0.73	0.76	0.71	No	0	forward	Dolstra	Van De Wiel-Van Kemenade	1999
HA-8	No	KIAA0020	9	RTLDKVLEV	rs2173904	C/G	R/P	A*02:01	0.43	0.45	0.42	No	0	forward	Brickner	Riddell	2001
HB-1Y	No	HMHB1	5	EEKRGSLYVW	rs161557	T/C	Y/H	B*44:02/B*44:03	0.27	0.24	0.29	No	0	forward	Dolstra	Van De Wiel-Van Kemenade	2002
ACC1C	Yes	BCL2A1	15	DYLQCVLQI	rs11383357	C/T	C/Y	A*24:02	0.72	0.74	0.65	Yes	9.2	forward	Akatsuka	Takahashi	2003
ACC1Y	Yes	BCL2A1	15	DYLQYVLQI	rs11383357	T/C	Y/C	A*24:02	0.28	0.26	0.35	Yes	26.5	forward	Akatsuka	Takahashi	2003
ACC2	Yes	BCL2A1	15	KFFDSIINW	rs3826007	T/C	D/G	B*44:03	0.19	0.24	0.2	Yes	5.9	forward	Akatsuka	Takahashi	2003
HA-3	No	AKAP13	15	VTEPGTAQY	rs2061821	C/T	T/M	A*01:01	0.58	0.65	0.62	No	0	forward	Spierings	Engelhard	2003
UGT2B17	No	UGT2B17	5	CVATMFDYI	gene deletion	na	peptide/null	A*02:06	-	-	-	No	0	forward	Murata	Riddell	2003
LRH-1	Yes	P2RX5	17	TPNRORNVC	rs3215407	G/-	peptide/null	B*07:02	0.44	0.43	0.43	No	0	forward	de Rijke	Dolstra	2005
ACC4/ACC5	No	CTSH	15	ATLPLLCAR	rs2289702	T/C	R/G	A*31:01/A*33:03	0.07	0.1	0.08	No	0	forward	Torikai	Takahashi	2006
PANE1	Yes	CENPM	22	RVVDLPGVLK	rs5758511	G/A	R/null	A*03:01	-	0.73	0.76	No	0	forward	Brickner	Warren	2006
SP110	Yes	SP110	2	SLPAGTSTPK	rs1365776	T/C	R/G	A*03:01	0.74	0.62	0.8	No	0	forward	Warren	J Van Den Eynde	2006
ACC6	Yes	HMSD	18	MEIDEVSHF	rs9945924	A/G	peptide/null	B*44:02/B*44:03	0.28	0.23	0.28	No	0	forward	Kawase	Takahashi	2007
LB-ADIR	No	TOR3A	1	SVAPALALFPA	rs2296377	T/C	F/L	A*02:01	0.22	0.29	0.23	No	0	forward	van Bergen	Falkenburg	2007
LB-ECGF-1	No	TYMP	22	RPHAIRRPLAL	rs1127723255	T/C	H/R	B*07:02	-	0.04	0.02	No	0	forward	van Bergen	Falkenburg	2007
TRIM22	No	TRIM22	11	MAVPPCGIGV	rs187416296	T/C	C/R	A*02:01	0.01	0.01	0.004	No	0	forward	Wolfel	Wolfel	2007
C19ORF48	No	C19ORF48	19	CIPPDSLLFPA	rs3745526	A/T	S/T	A*02:01	0.24	0.17	0.33	No	0	forward	Tykodi	Warren	2008
CD19	No	CD19	16	WEGEPPCLP	rs2904880	C/G	L/V	DQA1*05/DQB1*02	0.23	0.29	0.17	No	0	forward	Spaapen	Mutis	2008
LB-P14K2B	No	P14K2B	4	SRSSSAELDRSR	rs313549	T/C	S/P	DQB1*06:03	0.36	0.33	0.38	No	0	forward	Griffioen	Falkenburg	2008
LB-LY75	No	LY75	2	LGITYRNKSLMWF	rs12692566	C/A	K/N	B*07:02	0.14	0.17	0.15	No	0	forward	Stumpf	Griffioen	2009
LB-MR1	No	MR1	1	YFRLGVSDPIRG	rs2236410	G/A	R/H	B*07:02	0.16	0.15	0.19	No	0	forward	Stumpf	Griffioen	2009
LB-MTHFD1	No	MTHFD1	14	SSIIADQIAKL	rs2236225	A/G	Q/R	B*07:02	0.38	0.44	0.34	No	0	forward	Stumpf	Griffioen	2009
LB-PTK2B	No	PTK2B	8	VYMNDDSPLTPEK	rs7531019	C/A	T/K	B*07:02	0.4	0.46	0.36	No	0	forward	Stumpf	Griffioen	2009
SLC19A1	No	SLC19A1	21	RLVCYLCPY	rs1051266	C/T	R/H	DRB1*15:01	0.51	0.56	0.49	No	0	forward	Spaapen	Mutis	2009
SLC1A5	No	SLC1A5	19	AEATANGGLAL	rs3027956	C/G	A/P	B*40:02	0.17	0.2	0.2	No	0	forward	Kamei	Akatsuka	2009
DPH1	No	DPH1	17	SVLPEVDVW	rs35394823	G/C	V/L	B*57:01	0.06	0.08	0.05	No	0	forward	Warren	Riddell	2010
HEATR1	No	HEATR1	1	ISKERAEAL	rs2275687	T/C	E/G	B*08:01	0.43	0.34	0.48	No	0	forward	Bleakley	Riddell	2010
LB-APOBEC3B	No	APOBEC3B	22	KPQYHAEMCF	rs2076109	A/G	K/E	B*07:02	-	0.49	0.36	No	0	forward	van Bergen	Griffioen	2010
LB-ARHGDIB	No	ARHGDIB	12	LPRACWREA	rs4703	C/G	R/P	B*07:02	0.49	0.49	0.48	No	0	forward	van Bergen	Griffioen	2010
LB-BCAT2	No	ARHGDIB	19	QPRRALLFVIL	rs11548193	C/G	R/T	B*07:02	0.14	0.19	0.07	No	0	forward	van Bergen	Griffioen	2010
LB-EBI1	No	EBI1	19	RPRARYMQV	rs4740	A/G	I/V	B*07:02	0.39	0.31	0.45	No	0	forward	van Bergen	Griffioen	2010
LB-ERAP1	No	ERAP1	5	HPRQEQIALLA	rs26653	C/G	R/P	B*07:02	0.38	0.34	0.44	No	0	forward	van Bergen	Griffioen	2010
LB-GEMIN4	No	GEMIN4	17	FPALRFVEV	rs4968104	A/T	V/E	B*07:02	0.18	0.24	0.16	Yes	22.9	forward	van Bergen	Griffioen	2010
LB-PDCD11	No	PDCD11	10	GPDSSKTELCL	rs2986014	T/C	F/L	B*07:02	0.42	0.41	0.48	No	0	forward	van Bergen	Griffioen	2010
LB-PRCP	No	PRCP	11	FMWDVAEDLKA	rs2298668	G/T	D/E	B*07:02	0.15	0.16	0.14	No	0	forward	van Bergen	Griffioen	2010
LB-SSR1	No	SSR1	6	SLAVAQDLT	rs10004	G/A	S/L	B*07:02	0.31	0.27	0.29	No	0	forward	van Bergen	Griffioen	2010
LB-WNK1	No	WNK1	12	RTLSPEITV	rs12828016	T/G	I/M	B*07:02	0.41	0.39	0.42	No	0	forward	van Bergen	Griffioen	2010
P2RX7	No	P2RX7	12	WFHHCHPKY	rs7958311	A/G	H/R	A*29:02	-	0.27	0.24	No	0	forward	Warren	Riddell	2010
T4A1	No	TRIM42	3	GLYTWSAGA	rs9876490	C/A	A/E	A*02:01	0.53	0.44	0.6	No	0	reverse	Armistead	Molldrem	2011
ZAP70	No	ZNF419	19	IPRDSWVVEL	rs2074071	A/G	D/G	B*07:02	0.33	0.31	-	No	0	forward	Broen	Dolstra	2011
LB-NUP133	No	NUP133	1	SEDLLCRL	rs1065674	C/T	R/Q	B*40:01	0.33	0.22	0.3	No	0	forward	Griffioen	Falkenburg	2012
LB-SON	No	SON	21	SETKQRTVL	rs13047599	C/T	R/C	B*40:01	0.29	0.33	0.28	No	0	forward	Griffioen	Falkenburg	2012
LB-SWAP70	No	SWAP70	11	MEQLEQLEL	rs415895	C/G	Q/E	B*40:01	0.4	0.35	0.48	No	0	forward	Griffioen	Falkenburg	2012
LB-TRIP10-1EPC	No	TRIP10	19	GPQDLCTL	rs1049229	A/G	E/G	B*40:01	0.77	0.81	0.76	No	0	forward	Griffioen	Falkenburg	2012
LB-HIVEP1-1S	No	HIVEP1	6	SLPKHSVTI	rs2228220	G/A	S/N	A*02:01	0.12	0.11	0.15	No	0	reverse	Hombrink	Van Veelen	2013
LB-NISCH-1A	No	NISCH	3	ALAPAPAEV	rs887515	C/T	A/V	A*02:01	0.3	0.2	0.27	No	0	reverse	Hombrink	Van Veelen	2013
UTA2-1	Yes	KIAA1551	12	QLLNSVLTL	rs2166807	T/C	L/P	A*02:01	0.18	0.21	0.16	No	0	forward	Oostvogels	Mutis	2013
LB-FUCA2-1V	No	FUCA2	6	RLRQVGSWL	rs2074071	G/A	V/M	B*07:02	0.67	0.69	-	No	0	forward	van Bergen	Osanto	2014
UTDP4-1	No	ZDHHC12	9	RILAHFFCGW	rs11539209	A/T	I/N	DPB1*04	0.06	0.07	0.06	No	0	forward	Oostvogels	Spaapen	2014

LB-CLYBL	No	CLYBL	13	SLAAYIPRL	rs17577293	T/G	Y/D	B*07:02	0.02	0.03	0.01	No	0	reverse	Hombrink	Heemskerk	2015
LB-TEP1	No	TEP1	14	APDGAKVASL	rs1760904	A/G	S/P	B*07:02	0.36	0.5	0.31	No	0	reverse	Hombrink	Heemskerk	2015
ITGB2	No	ITGB2	21	GQAGFFPSPF	rs760462	A/G	splice mutant	B*15:01	0.12	0.16	0.11	No	0	forward	Pont	Griffioen	2016
UNC-GRK4-V	Yes	GRK4	4	VLDIEQFSV	rs1801058	T/C	V/A	A*02:01	0.32	0.4	0.31	Yes	22.1	reverse	Lansford	Armistead/Vincent	2018
LB-PIP4K2A-1S	No	PIP4K2A	10	INEGQKIYIDDNNKKVFLE	rs10828317	A/G	N/S	DRB1*03:01	0.77	0.7	0.79	No	0	forward	Kremer	Griffioen	2020
UNC-AHRR-A	Yes	AHRR	5	VVFGQAPPL	rs2292596	G/C	A/P	A*02:01	0.30	0.35	0.38	Yes	21.7	reverse	Olsen	Armistead/Vincent	2022
UNC-ARHGEF18-Q	Yes	ARHGEF18	19	SLICRQLGSA	rs2287918	A/G	Q/R	A*02:01	0.19	0.24	0.17	Yes	23.9	reverse	Olsen	Armistead/Vincent	2022
UNC-BCL2A1-Y	Yes	BCL2A1	15	YRLAQDYLQY	rs1138357	A/G	Y/C	A*02:01	0.28	0.26	0.35	Yes	24.4	reverse	Olsen	Armistead/Vincent	2022
UNC-DPEP1-Q	Yes	DPEP1	16	NLLRVFQA(V)	rs1126464	G/C	E/Q	A*02:01	0.21	,19	0.24	Yes	23.4	reverse	Olsen/Diiorio	Armistead/Vincent	unpublished
UNC-DPP3-H	Yes	DPP3	11	KLIVQPNTHL	rs2305535	A/G	H/R	A*02:01	0.19	0.22	0.21	Yes	23.6	reverse	Olsen	Armistead/Vincent	2022
UNC-FLT3-G	Yes	FLT3	13	ALARGGGQLPL	rs12872889	G/A	G/D	A*02:01	0.35	0.31	0.37	Yes	24.3	reverse	Olsen	Armistead/Vincent	2022
UNC-FPGS-I	Yes	FPGS	9	FLAAASARGI	rs10760502	T/C	I/V	A*02:01	0.28	0.35	0.22	Yes	26.5	reverse	Olsen	Armistead/Vincent	2022
UNC-FPR1-K	Yes	FPR1	19	KVAVAMLTV	rs1042229	G/T	K/N	A*02:01	-	0.45	0.37	Yes	28.3	reverse	Olsen	Armistead/Vincent	2022
UNC-GAA-R	Yes	GAA	17	RRQLDGRVLL	rs1042395	G/A	R/H	A*02:01	0.36	0.28	0.40	Yes	25.4	reverse	Olsen	Armistead/Vincent	2022
UNC-GDPD5-A	Yes	GDPD5	11	ALSQVPSPL	rs571353	G/A	A/T	A*02:01	0.34	0.28	0.33	Yes	23.6	reverse	Olsen	Armistead/Vincent	2022
UNC-GLRX3-S	Yes	GLRX3	10	FLSSANEHL	rs2274217	T/C	S/P	A*02:01	0.21	0.25	0.19	Yes	25.6	reverse	Olsen	Armistead/Vincent	2022
UNC-HEXDC-V	Yes	HEXDC	17	RLHVGCDEV	rs4789773	G/A	V/I	A*02:01	0.45	0.37	0.56	Yes	24	reverse	Olsen	Armistead/Vincent	2022
UNC-HLX-P	Yes	HLX	1	LPAAYHHH	rs12141189	C/T	P/S	B*35:01	0.23	0.23	0.21	Yes	26.6	reverse	Olsen	Armistead/Vincent	2022
UNC-IQCE-V	Yes	IQCE	11	AVLDEAVV	rs2293404	T/C	V/A	A*02:01	0.35	0.29	0.41	Yes	26.4	reverse	Olsen	Armistead/Vincent	2022
UNC-MARCH2-T	Yes	2-Mar	19	GRLLSTVIRTL	rs1133893	T/C	T/A	C*07:02	0.24	0.32	0.20	Yes	26.7	reverse	Olsen	Armistead/Vincent	2022
UNC-NDUFAF1-L	Yes	NDUFAF1	15	ALYPFLGIL	rs3204853	A/C	L/G	A*02:01	0.17	0.24	0.12	Yes	26.4	reverse	Olsen	Armistead/Vincent	2022
UNC-NEK4-A	Yes	NEK4	3	LPAMPRDY	rs1029871	C/G	A/P	B*35:01	0.33	0.39	0.31	Yes	26	reverse	Olsen	Armistead/Vincent	2022
UNC-POLL-W	Yes	POLL	10	HPDGWSHRGIF	rs3730477	T/C	W/R	B*35:01	0.16	0.21	0.10	Yes	27.4	reverse	Olsen	Armistead/Vincent	2022
UNC-RNASE3-R	Yes	RNASE3	14	RYADRPGRRF	rs2073342	G/C	R/T	C*07:02	0.36	0.29	0.36	Yes	25.9	reverse	Olsen	Armistead/Vincent	2022
UNC-SLC25A37-V	Yes	SLC25A37	8	AQYTSVYGA	rs2942194	G/A	V/I	A*02:01	0.21	0.26	0.17	Yes	25.3	reverse	Olsen	Armistead/Vincent	2022
UNC-SLC26A8-M	Yes	SLC26A8	6	FLRCMLTI	rs743923	A/G	M/V	A*02:01	0.30	0.25	0.26	Yes	25	reverse	Olsen	Armistead/Vincent	2022
UNC-SNX19-V	Yes	SNX19	11	FLQPNVRGPLF	rs3751037	C/G	V/L	C*07:02	0.30	0.29	0.27	Yes	25.8	reverse	Olsen	Armistead/Vincent	2022
UNC-TOP1MT-W	Yes	TOP1MT	8	WLLEKLQEQL	rs2293925	T/C	W/R	A*02:01	0.36	0.43	0.46	Yes	21.1	reverse	Olsen	Armistead/Vincent	2022
UNC-TXNDC2-T	Yes	TXNDC2	18	ILSKEDFETSL	rs17732496	G/A	A/T	A*02:01	0.12	0.14	0.11	Yes	20.6	reverse	Olsen/Diiorio	Armistead/Vincent	unpublished
UNC-USP4-V	Yes	USP4	3	KVSFFVPRL	rs35446411	G/T	V/L	A*02:01	0.13	0.16	0.09	Yes	22.4	reverse	Olsen	Armistead/Vincent	2022
UNC-WDR62-L	Yes	WDR62	19	LLGDDDVADGL	rs2285745	T/C	L/S	A*02:01	0.32	0.35	0.35	Yes	26.1	reverse	Olsen	Armistead/Vincent	2022

(1.8) mHA-targeting immunotherapies

mHAs can theoretically be targeted via a variety of immunotherapy strategies, though only a few of these have been tested in patients. Any peptide-targeting therapeutic modality can be applied to mHAs, including peptide vaccines, mHA-targeting T cell receptors (TCRs) on donor T cells, or chimeric antigen receptor (CAR) T cell (using scFv from an anti-peptide/MHC antibody) options. A limited number of mHA-targeting therapies have made the transition to clinical trials thus far, most of these targeting HA-1. The first mHA-targeting trial, in which patients received adoptive transfer of donor-derived, *ex vivo* expanded CD8+ T cells that were specific for minor histocompatibility antigens, unfortunately led to no long term remissions and pulmonary GvHD in 3 of 7 patients, this demonstrating the necessity of careful screening for off-target tissue reactivity (trial NCT00107354)[73,185]. In the strategy used in this trial, mHA-reactive clones were generated by coculturing recipient post-transplant peripheral blood mononuclear cells (PBMCs) with pre-transplant PBMCs followed by limiting dilution and characterization of reactive clones[73]. The clones were used for therapeutic infusion if they lysed recipient Epstein-Barr Virus-transformed lymphoblastic cell lines (EBV-LCLs) but not donor EBV-LCLs or recipient fibroblasts. Clone mHA specificity was established via genome wide association study (GWAS), and tissue expression of source genes established via reverse transcription polymerase chain reaction (RT-PCR)[73]. After several patients developed pulmonary toxicity, it was realized that the mHAs their therapeutics targeted were derived from genes that were also expressed in lung tissues[73]. While this trial was unsuccessful clinically, it established that screening of reactivity and assessment of source gene expression is necessary in order to generate a safe mHA-targeting product[186]. In our view, it also provides a compelling argument for

generating T cells targeting known mHAs rather than infusion of T cells with no prior identified mHA target so one can carefully investigate the possibility for off-target effects prior to treatment.

Shortly after the first mHA-targeting trial, several options for targeting already known mHAs were explored. Though not in humans, the Shlomchik group tested the approach of vaccinating donors against an mHA before mouse alloHCT and found that it augmented antigen-specific CD8+ T cell presence and GvL effects in the recipient, showing that peptide vaccination in donors is a potentially promising approach[187]. A group at Leiden University Medical Center performed optimization of HA-1 targeting T cell generation then administered donor-derived HA-1 specific CD8s to patients with relapsed leukemia in a pilot study with no toxicity observed, but also no clinical response[188–190]. Medigene had initiated a clinical trial with a similar HA-1 targeting TCR in 2020 but discontinued it in 2021 due to a shift in prioritization of solid tumors (trial NCT04464889)[191]. A group in Japan separately generated HA-1 targeting CAR T cells around the same time, though they did not test them in patients and only evaluated them in vitro[192]. An additional HA-1 targeting clinical trial was initiated in 2009 that proposed vaccinating with HA-1 or HA-2 peptide mixed with G-CSF in order to augment GvL responses, but was terminated due to insufficient enrollment with only one patient enrolling (trial NCT00943293)[193].

An intriguing approach to a phase I/II trial was reported in 2017, in which multiple myeloma patients received donor lymphocyte infusion (DLI) with simultaneous host DC vaccination with or without loading the DCs with host mHAs[194]. Seven patients were treated with unloaded DCs, while four patients were treated with DCs loaded with mHAs

including HA-1, HA-2, UTA2-1, ACC-1, ACC-2, or LRH-1 based on which mHAs the DRP had mismatches for[194]. All four patients had T cell responses against a control positive non-mHA peptide that was included, and one patient had a T cell response against ACC-1[194]. This patient was one of the two patients with objective clinical benefit from vaccination, with a partial response persisting for 19 weeks[194]. A decline in ACC-1 specific T cells was observed at the time of relapse[194]. A similar trial is complete but without results reported yet, in which patients were given therapeutic mHA-loaded DC vaccinations, with theoretically increased potency due to siRNA mediated PD-L1/PD-L2 silencing (NCT02528682)[195]. In an adoptive T cell therapeutic approach, the Bleakley group generated a novel engineered HA-1 targeting T cell therapeutic in 2018, containing a TCR targeting HA-1, a CD8+ coreceptor so the construct can function on CD4+ T cells as well as CD8s, an inducible caspase 9 safety switch, and a CD34-CD20 epitope for cell selection and tracking[196]. This intriguing strategy of including CD4+ T cells in transduction was decided on in an attempt to use the CD4+ T cells to boost response by facilitating CD8+ T cell trafficking and expansion and prevention of activation-induced cell death (AICD). This approach appears promising as HA-1 positive leukemia cells were killed *in vitro*[196]. A corresponding clinical trial is currently recruiting relapsed or refractory acute leukemia patients that have undergone alloHCT, and this is the only mHA targeting clinical trial that includes pediatric patients (NCT03326921)[197]. The Falkenburg group recently published a phase I/II trial of multi-antigen specific T cell infusion after T cell depleted alloHCT that included HA-1 as one of the targets in addition to tumor associated antigens (TAAs) and viral antigens[198]. Infusions were well-tolerated, but T cell reactions observed were all in response to the viral antigens rather

than the TAAs and HA-1[198]. One final clinical trial targeting mHAs was initiated in 2017 with no reports of results or what mHAs were targeted as of yet; this trial includes priming of donor T cells against the selected mHAs followed by infusion into recipients with relapsed hematological malignancies after alloHCT (trial NCT03091933)[199]. An overall timeline of mHA discovery and clinical trial applications is shown below (Figure 1.5). These data demonstrate the need for both improved predictors of minor antigens and methods to deliver them to generate anti-tumor immunity without toxicity.

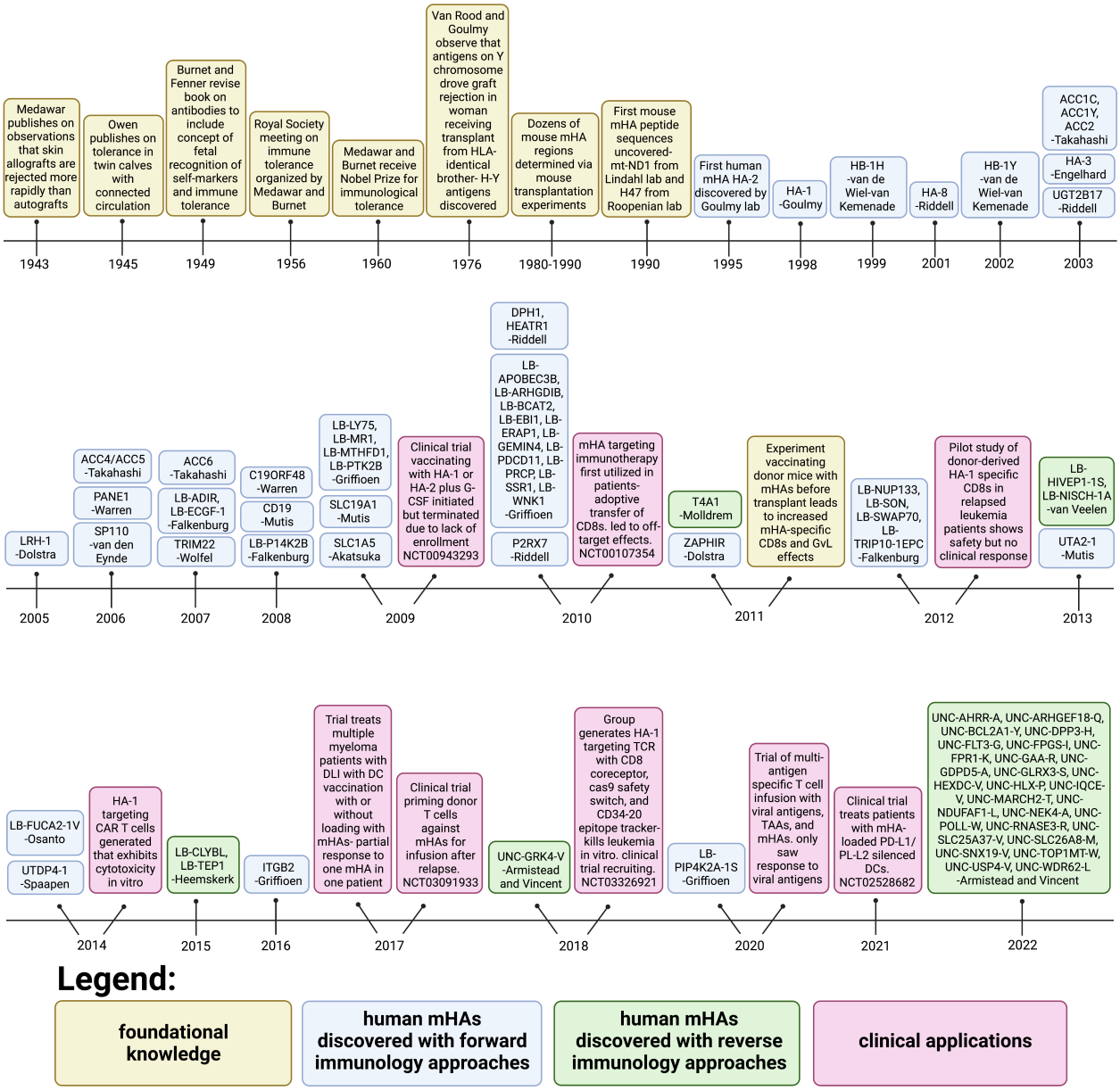


Figure 1.5. Timeline of the history of minor histocompatibility antigens.

(1.9) Shared mHAs

Because mHAs are derived from SNPs that differ between a donor and a recipient and an individual's complement of SNPs throughout their genome is unique, the theoretical set of mHAs for any given DRPs is also unique. mHA discovery methods have traditionally relied upon isolating a reactive T cell clone from a single DRP then

identifying the cognate mHA (forward immunology method). This discovery method is highly likely to uncover mHAs that are only applicable for a single DRP or a low population frequency of DRPs because the single DRP that the mHA was discovered from contains many SNPs that are rarely disparate in the general population of DRPs. Because of this, mHA-targeting therapeutics beyond those targeting HA-1 are often thought of as being useful for personalized immunotherapies rather than ones that are broadly applicable to the patient population of alloHCT recipients. Population sharing of mHAs can be estimated via assessment of the minor allele frequency (MAF) of the SNP allele that generates the mHA. Hardy-Weinberg equilibrium states that

$$\mathrm{p2} + 2\mathrm{pq} + \mathrm{q2} = 1$$

where p and q are each the frequency of a possible allele at a given locus. Using this equation, one can calculate the population frequency of homozygous and heterozygous individuals in a population given the MAF. For a given mHA to be targetable in a DRP, the alloHCT recipient must contain the mHA allele and the mHA allele must be foreign to the donor, so that donor T cells would not be deleted in thymic selection based on self-recognition. This means that the recipient must be either homozygous for the mHA allele (q^2) or heterozygous (2pq), while the donor must be p^2. To have the best chance of finding a donor-recipient pair that have the correct alleles to target an mHA, it is optimal for p^2 to equal 0.5 and $2pq+q^2$ to also equal 0.5, as this would mean that each donor and each recipient have a 50% chance of the appropriate genotype. This gives an ideal MAF of approximately 0.3 for mHA targeting. The MAF of many known mHAs is either much higher or much lower than this, weighting the population towards either more appropriate donors but fewer appropriate recipients or vice versa.

One previously known mHA with a high prevalence of appropriate DRP genotypes is HA-1. Its MAF is greater than optimal, estimated at 0.63 in the ALFA project, which means that the frequency of appropriate recipient genotypes is high at 86%, but the frequency of appropriate donor genotypes is only 14%[200]. We found that overall, the percent of DRPs expressing HLA-A*02:01 in the DISCOVeRY-BMT dataset of 3231 DRPs with HA-1 targetability is 12.3%. Still, these frequencies are high enough where HA-1 is targetable in many patients, and immunotherapies targeting HA-1 have been tested in clinical trials. We would characterize HA-1 as the first known “shared mHA,” or an mHA that is commonly applicable to many DRPs. Though HA-1 ended up being targetable for many patients, the discovery of shared mHAs using forward immunology methods like the one used to discover HA-1 is essentially serendipity, with no knowledge of whether a reactive T cell clone is specific for a common mHA or a rare one until the process is completed. When we investigated the prevalence of the appropriate mismatches to generate all known mHAs prior to 2022 in the DISCOVeRY-BMT dataset of 3231 alloHCT DRPs, we found that only 7 of the previously known 56 antigens are applicable to any of the DRPs. The highest population frequency in DISCOVeRY-BMT was with mHA ACC1Y, which would be targetable in 26.5% of DRPs with its corresponding HLA allele, HLA-A*24:02. In recent years, however, the advent of reverse immunology approaches to mHA discovery has made it possible to specifically seek out mHAs that are shared. These approaches involve predicting mHAs based on sequencing data, then confirming existence of predicted antigens using methods like immunoprecipitation of HLA-bound peptides from cells and mass spectrometry to identify the presented mHAs. If one has access to a dataset of genotyping information

for multiple DRPs (e.g. SNP array or DNA sequencing data), they can both predict mHAs and define the degree of mHA sharing among multiple DRPs. To the best of our knowledge, the Armistead and Vincent groups in collaboration are the first to use this approach to discover novel mHAs, first with UNC-GRK4-V in 2018, and now with 26 novel mHAs in 2022 (published and unpublished data)[62,121]. A particularly important application of shared mHA discovery is that that conventional mHA discovery has been done using samples mostly from Caucasian DRPs, meaning that the majority of previously known mHAs correspond to HLA alleles that are common in Caucasians and less frequent in other ethnic groups. 61% of known class I binding mHAs correspond to either HLA-A*02:01 or HLA-B*07:02, the most common A allele and B allele in Caucasians in the US (Figure 1.6).

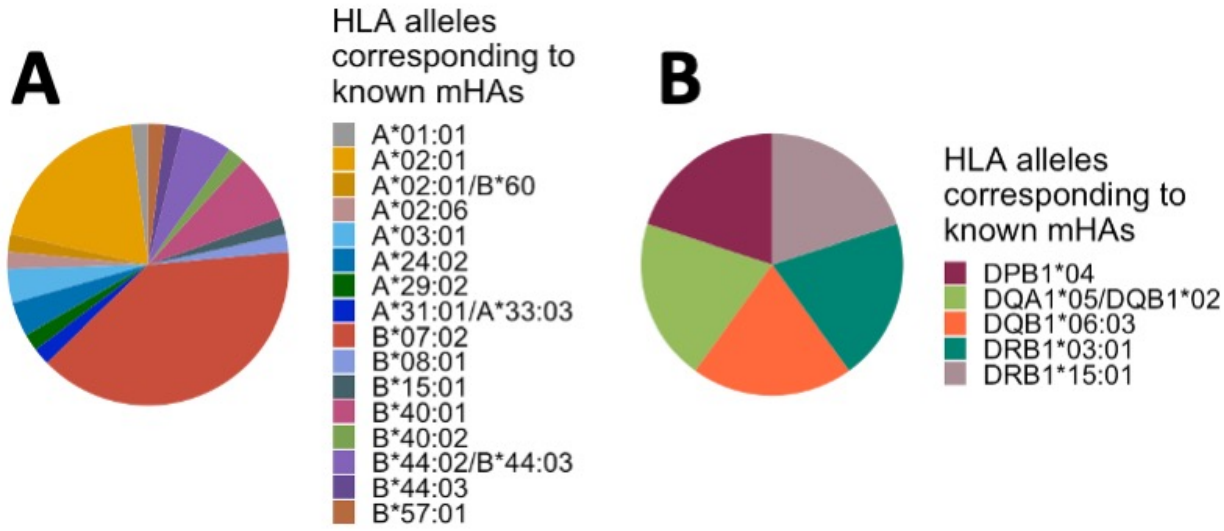


Figure 1.6: Distribution of HLA alleles corresponding to all known mHAs. (A) shows class I mHAs corresponding to class I HLA. (B) shows class II mHAs corresponding to class II HLA.

In fact, 75% of known class I mHAs are specific for HLA alleles that are more prevalent in Caucasians than non-Caucasians. The majority of clinical trials being conducted on mHA targeting are also restricted to patients that express HLA-A*02:01. In addition to HLA alleles that are common in Caucasians being less common in non-Caucasians, non-Caucasian groups also often have more HLA diversity, shown by the higher prevalence of “Other” alleles comprising <5% of the population for each HLA class I allele in non-Caucasians versus Caucasians (Figure 1.7).

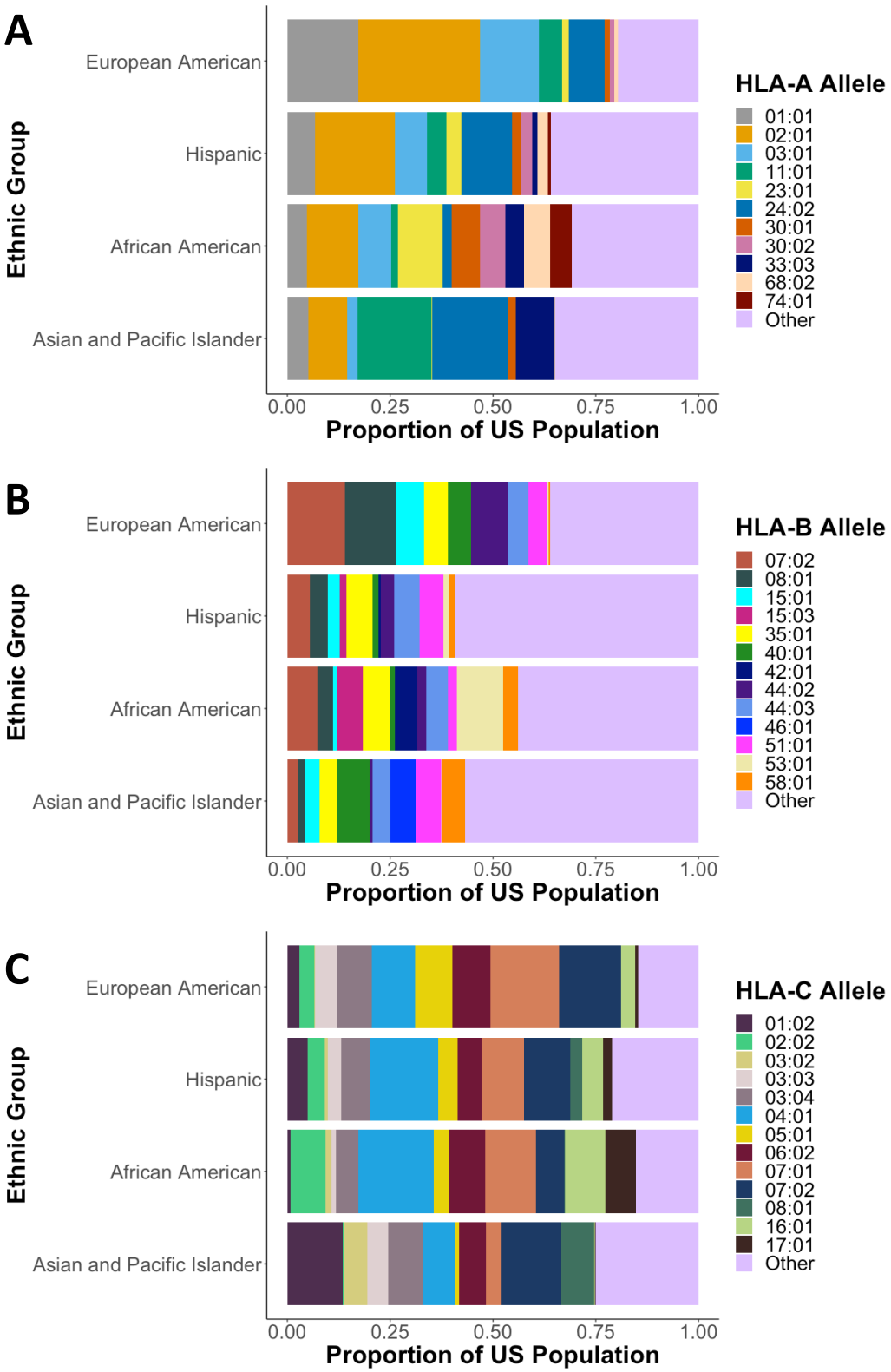


Figure 1.7: HLA allele distribution by ethnic group in the United States. (A) shows HLA-A alleles, (B) shows HLA-B alleles, and (C) shows HLA-C alleles.

This means that even if a forward immunology approach was used with a reactive T cell clone from a non-Caucasian patient to discover a novel mHA, this mHA may be applicable to even fewer patients than ones discovered using this technique in Caucasians because fewer patients even within the same ethnic group express the appropriate HLA allele. As an example, referencing the above Figure 1.4, if an mHA was discovered using an HLA-A*02:01 positive Caucasian patient, this mHA could be applicable for up to 30% of Caucasians, but only 9% of African Americans. If an mHA was discovered in an African American patient expressing HLA-A*23:01, a common allele in this group, it would only be applicable to 11% of African Americans due to high HLA allele diversity in this population. This emphasizes that not only is additional discovery work needed for mHAs present in non-Caucasians, but also the importance of using high-throughput reverse immunology discovery methods such as the one presented in this work for the discovery of large sets of mHAs in HLA alleles presented by diverse ethnic groups. If it requires higher numbers of mHAs to cover all patients with a particular ethnicity, we must utilize the technology available to us to discover as many antigens for these patients as we can.

Non-Caucasians are disadvantaged in many aspects of alloHCT, including matched unrelated donors being much less available in the donor registry for non-Caucasian patients. The expansion of mHA prediction and discovery methods enables prioritization of discovery in any population we choose, provided that DRP genotyping data is available. We are now able to predict and test antigens for any HLA allele, and we can use this ability to prioritize antigen discovery in ethnic groups that are less served by past research and clinical developments.

Despite our specific interest in HLA alleles historically not included in previous research, we chose to include HLA-A*02:01 in our antigen discovery work both to test the concordance of our results with previously known mHAs and because this allele has the most HLA-specific tools available for validation of discovered mHAs (including HLA monomers and other materials for monitoring antigen-specific CD8+ T cell responses). However, once a method had been established with this allele, we shifted our focus to also include alleles that are the most common in other ethnic groups. We included HLA-B*35:01, the most common HLA B allele in Hispanics, the third most common in African Americans, and the fifth most common in Asians and Pacific Islanders[201]. We also included HLA-C*07:02, the most common C allele in Asians and Pacific Islanders, second most common in African Americans, and third most common in Hispanics[201].

(1.10) Concluding remarks and contributions of this work

The studies presented here represent a multidisciplinary approach to minor histocompatibility antigen discovery, combining computational and experimental methods to answer fundamental questions about mHA sharing and discover new targets for translational immunotherapy development. The work discussed here involved many collaborations that I would like to note. We collaborated with the Armistead lab on most of the components of this work. We obtained data from the Sucheston-Campbell lab formerly at Ohio State University as part of a collaboration with them to access data from the Center for International Bone Marrow Transplant Research (CIBMTR). We worked with the UNC Proteomics Core and with Complete Omics Inc on proteomics efforts, with UNC Immune Monitoring and Genomics Facility (IMGF) and High-Throughput Sequencing Facility (HTSF) for sequencing, and the UNC Flow Core for

flow cytometry work. Appendices include work from our lab on a peptide vaccine for SARS-CoV-2, work on clinical outcomes associations of predicted mHAs, a computational platform for antigen prediction, and a prediction tool for splice variant-derived antigens.

CHAPTER 2: Computational Prediction of Graft versus Leukemia and Shared Minor Histocompatibility Antigens[1]

(2.1) Introduction

mHAs can be predicted from paired high-resolution genotyping data from any given DRP. We sought to apply mHA prediction methods to sets of many DRPs in order to assess population prevalence of mHAs and identify novel mHAs that are shared across many DRPs within the population that could serve as targets for future immunotherapies. Current mHA-targeting therapies in clinical trials are only applicable to a small number of patients because only a low percentage of DRPs have the appropriate SNP mismatches in order for the mHAs they target to be relevant. This is a result of the forward immunology approach historically used to discover most of the previously known mHAs, in which an activated T cell clone is isolated from an alloHCT patient post-transplant and the corresponding mHA target is identified via a variety of methods such as library screening or whole genome association studies[164]. Firstly, SNPs with high population frequencies are more likely to be concordant in a DRP, so most SNPs at which donors and recipients differ at are low population frequency SNPs. Therefore, the majority of mHAs for any given DRP will be derived from rare SNPs.

[1] Portions of this chapter are published in *Blood Advances*. Citation: Olsen KS, Jadi O, Dexheimer S, Bortone S, Vensko SP, Bennett SN, Tang H, Diiorio M, Saran T, Dingfelder D, Zhu Q, Wang Y, Haiman CA, Pooler L, Sheng X, Webb A, Pasquini MC, McCarthy PL, Spellman SR, Weimer ET, Hahn T, Sucheston-Campbell LE, Armistead PM*, Vincent B*. Shared graft-vs-leukemia minor histocompatibility antigens in DISCOVeRY-BMT. 2022. *Blood Adv*; bloodadvances.2022008863. I designed and performed experiments, performed computational mHA prioritization, interpreted experimental results, performed statistical testing, generated figures, prepared the manuscript, and made revisions.

Secondly, using strategies that do not deliberately seek out shared mHAs and instead randomly isolate mHA-targeting T cells means that the majority of mHAs discovered will be rare mHAs. In contrast, reverse immunology approaches allow one to prioritize discovery of mHAs that have specific desired characteristics. In reverse immunology methods, mHAs are predicted using computational processing of DRP sequencing data, and mHAs that suit the characteristics selected by the researcher can then be validated by a variety of wet lab methods. The number of mHAs predicted from sequencing data will generally be significantly larger than the number that can be validated using wet lab methods, often by many orders of magnitude if the dataset used for prediction is large, so there is a natural place for a filtering step based on the researcher's priorities in mHA discovery. We believe that this step can be leveraged to generate the greatest benefit for patients by prioritizing 1) mHAs that are shared across many DRPs and 2) mHAs that are presented by HLA alleles less served by traditional antigen discovery work. As part of this, we prioritize identification of mHAs that are predicted to be shared by many DRPs. This means that the mHAs discovered using these methods with a large dataset will have MAFs that are more favorable for these mHAs to be targetable in many patients. As discussed in Chapter 1, the ideal theoretical MAF for mHAs to be targetable in as many DRPs as possible is 0.3, as this gives equal odds of any individual either containing the mHA allele or being homozygous for the alternate allele, and so the odds of finding an appropriate recipient and donor genotype are maximized. In this work, we combine predictions for a single HLA allele that were previously made for a small dataset, and also generate predictions for every HLA class I allele represented in the largest DRP dataset used to date in the field for mHA prediction. For a subset of novel,

shared GvL mHAs, we proceed to mHA validation with three HLA alleles selected for representation of HLA A, B, and C along with high population frequency in different ethnic groups.

Our ultimate translational goal in this line of work is to generate a set of "off-the-shelf" mHA targeting TCR products that together cover 100% of DRPs with HLA alleles common in different ethnic groups in the US. Our vision for integrating this into the typical alloHCT workflow is to genotype patients for the mHAs with off-the-shelf mHA-specific TCRs available, engineer donor T cells to express these TCRs, expand them *ex vivo*, then infuse them post-alloHCT to augment the GvL effects of transplant (Figure 2.1). While the Vincent and Armistead groups are pursuing a clinical trial with mHA-targeting TCRs, these antigens could also be utilized for other targeting methods (e.g. vaccination, peptide/HLA antibody, or CAR-T) in the future either by our groups or by others.

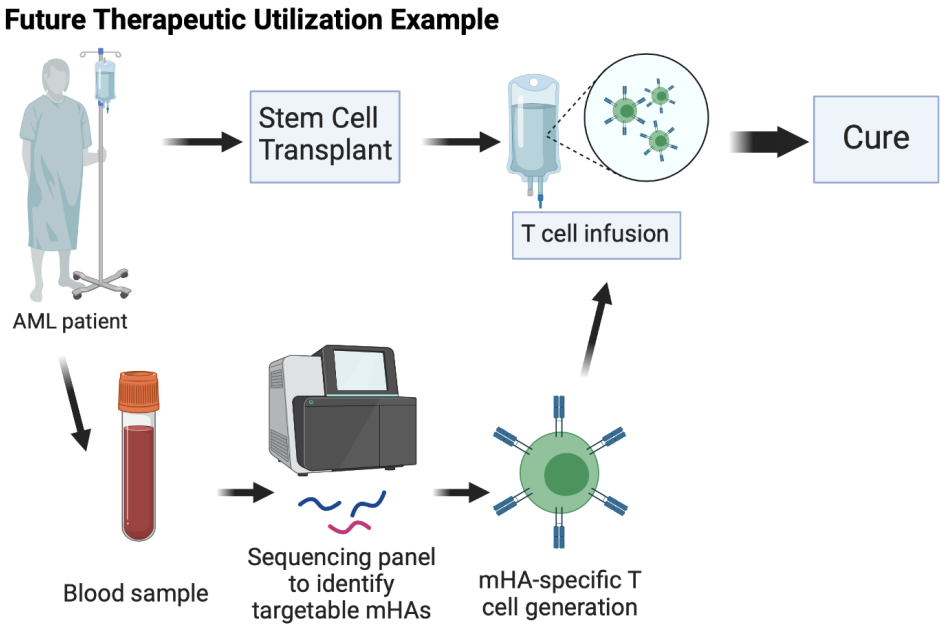


Figure 2.1: Future therapeutic application of shared GvL mHA-targeting TCRs.

In this chapter, I describe the datasets used in this work, the computational methods utilized to predict mHAs from SNP typing data, and the decisions made in mHA prioritization and selection for wet lab validation and targeting efforts described in the next chapters. I also show population-level results on the main dataset utilized in this work (DISCOVeRY-BMT) and the prevalence of different types of mHAs within DRPs in this set.

(2.2) Materials and Methods

Datasets:

Two datasets were used in this work. For the first dataset, including 101 DRPs, mHA predictions were made as part of prior work from the Armistead and Vincent labs and mHA validation from these predictions as part of this work[62,169]. For the second dataset, including 3231 DRPs, mHA predictions were made for the purpose of this work and the dataset was taken all the way through from mHA prediction to immunopeptidomics validation and mHA-specific T cell discovery.

The first dataset includes 101 patients with Acute Myeloid Leukemia (AML), Chronic Myeloid Leukemia (CML), myelodysplastic syndrome (MDS), or myeloproliferative neoplasms (MPN) from MD Anderson Cancer Center who underwent HLA-matched alloHCT[62]. DRPs were genotyped at a panel of 13,917 SNPs via Illumina NS-12 microarrays and the genotyping data are publicly available via the Vincent lab at https://unclineberger.org/vincent/resources. Patient characteristics are shown (Table 2.1). As mHA predictions were already made for this dataset in prior work from the Armistead and Vincent labs, the prediction work in the remainder of the chapter refers

exclusively to predictions for the second dataset. However, validation and targeting work in future chapters includes mHAs predicted from this initial dataset.

Table 2.1: Patient characteristics for 101 patient dataset, adapted from Lansford et al[62].

Category	Recipient Characteristics	Total (n=101)
Age mean and range		48 (21-72)
Sex	M F	64 (63) 37 (37)
Disease	AML CML MDS MPN	61 (60) 25 (25) 14 (14) 1 (1)
Donor type	MRD MUD	72 (72) 29 (28)
Conditioning intensity	Myeloablative Reduced Intensity	71 (71) 30 (29)
Disease	AML ALL MDS	1938 (60) 697 (22) 596 (18)
Sex mismatch	None F donor- M recipient M donor-F recipient	54 (53) 28 (28) 19 (19)
Graft source	Bone marrow Peripheral blood	33 (33) 68 (67)

The second dataset is derived from the DISCOVeRY-BMT (Determining the Influence of Susceptibility Conveying Variants Related to one-Year mortality after BMT) study, reported to CIBMTR from 151 transplant centers within the US[202–205]. Data was obtained via a collaboration between the Armistead and Vincent labs and the Sucheston-Campbell lab formerly at Ohio State University, who obtained sequencing data via CIBMTR. Patients included in this study were treated for Acute Myeloid Leukemia (AML), Acute Lymphocytic Leukemia (ALL), and Myelodysplastic Syndrome

(MDS) with alloHCT. Cohort 1 consists of 2609 10/10 HLA-matched unrelated DRPs treated from 2000-2008, while Cohort 2 consists of 572 10/10 HLA-matched unrelated DRPs treated from 2009-2011 and 351 >=8/8 and <10/10 HLA-matched unrelated DRPs treated from 2000-2011[205]. DRPs were excluded if the grafts were cord blood grafts or T cell-depleted, or SNP data was not available. For antigen prediction, all patients were combined. Patient characteristics are shown below (Table 2.2).

All patients included in the DISCOVeRY-BMT study provided informed consent to be included in the Center for International Blood and Marrow Transplant Research (CIBMTR) registry. Genotyping was performed as previously described using the Illumina HumanOmni Express chip[205–207]. SNP quality control was performed and variants with minor allele frequency (MAF) < 0.005 were removed, leaving 637,655 and 632,823 measured SNPs for Cohort 1 and Cohort 2 respectively[203].

Table 2.2: Patient characteristics of 3231 patient dataset.

Category	Recipient Characteristics	First Cohort (n=2357)	Second Cohort (n=874)	Total (n=3231)
Age, years	<=40 >40	961 (41) 1396 (59)	322 (37) 552 (63)	1283 (40) 1948 (60)
Donor age, years range		18-61	18-60	18-61
Sex	M F	1331 (56) 1026 (44)	480 (55) 394 (45)	1811 (56) 1420 (44)
Donor sex	M F	1577 (67) 780 (33)	620 (71) 254 (29)	2197 (68) 1034 (32)
Disease	AML ALL MDS	1397 (59) 576 (24) 384 (16)	541 (62) 121 (14) 212 (24)	1938 (60) 697 (22) 596 (18)
Year of alloSCT	2000-2002 2003-2005 2006-2008 2009-2011	384 (16) 848 (36) 1125 (48) 0 (0)	33 (4) 83 (9) 119 (14) 639 (73)	417 (13) 931 (29) 1244 (39) 639 (20)
Graft source	Bone marrow Peripheral blood	1504 (64) 853 (36)	624 (71) 250 (29)	2128 (66) 1103 (34)

Ethnicity assessment:

For the DISCOVeRY-BMT dataset, both self-reported ethnicity and SNP typing data were available. From SNP typing data, genomic ancestry was calculated via principal component analysis. Principal components were constructed using a set of independent SNPs in all patients self-declaring White, European, or Caucasian race and Non-Hispanic ethnicity. Mean values for the first three eigenvectors were

determined and individuals with any of the first three eigenvectors greater than two standard deviations from each mean value were excluded. This was repeated for individuals self-declaring Black or African race and Non-Hispanic ethnicity, and for individuals declaring Hispanic ethnicity[206,208]. For this work, three genomic ancestry groups were assessed, including European American (EA), Hispanic (HIS), and African American (AA). Patients that self-reported as Asian American and Native American were included in mHA prediction work but genomic ancestry was not calculated and these patients were excluded from ethnicity analyses due to 1) small patient numbers for these groups and 2) agreements with CIBMTR for maintaining patient privacy. Student's T tests and Chi-squared tests were performed to assess differences in number of predicted mHAs of various types between groups.

Genetic distance calculation:

For both datasets, we calculated pairwise genetic distance for every DRP based on SNP array data[209]. Genetic distance was calculated using the following equation.

$$\text{mean}(1 - 0.5(\text{number of shared alleles between pair at SNP locus}))$$

This measure should theoretically be lower for MRDs and higher for MUDs, as MRDs would be expected to share more SNPs than unrelated pairs. The 101 patient dataset is composed of both MUDs and MRDs, while the 3231 patient dataset only contains MUDs.

mHA prediction:

mHA prediction was performed using the same overall technique for both datasets according to the Vincent lab strategy for antigen prediction. Minor mismatches were called where a SNP allele was present in the transplant recipient (either

homozygous or heterozygous) and was not present in the donor (homozygous for the opposite allele). For example, if a recipient was heterozygous at a SNP locus with one copy each of A and T, and the donor was homozygous for T, this would be labeled a minor mismatch as the A allele is foreign to the donor. Minor mismatches that led to amino acid changes (nonsynonymous mutations) were then selected based on the standard codons encoding amino acids using ANNOVAR[210]. SNPs that led to amino acid differences were then mapped to the genome using the ENSEMBL variant effect predictor with the UCSC Genome Browser database to identify the open reading frames containing these SNPs[211,212]. Based on these open reading frames, all possible 8-11mer peptides were considered for class I HLA binding epitope prediction using antigen.garnish[56]. All possible 15-24mer peptides were also computed for class II-restricted mHAs and processed separately[213]. The work covered in this dissertation exclusively examines the class I mHAs, and only these antigens will be discussed moving forward. We applied netMHCpan-4.1 to the pool of all possible peptides of class I HLA binding length to predict binding affinity of peptides to the class I HLA alleles of the donor that the predicted peptide was derived from[50]. Peptides were selected if their predicted peptide/HLA dissociation constant for the corresponding HLA allele was <500nM. Peptides that met all of these criteria were considered mHAs. Next, tissue expression of the mHA source genes was considered in order to categorize mHAs and identify GvL mHAs that could be clinically beneficial to target. This was done based on RNAseq data obtained from The Cancer Genome Atlas (TCGA) and normal tissue RNA and protein expression data from the Genotype-Tissue Expression (GTEx) project[214]. Peptides were labeled “GvL” if they showed expression levels of >50TPM in AML in

TCGA and <50TPM in healthy GvHD target organs including liver, skin, and colon in GTEx. They were labeled as “GvH” if they showed expression levels of <50TPM in AML and >50TPM in GvHD target organs. Peptides were labeled as “Both” if they showed >50TPM in both AML and GvHD target organs. A final class of mHAs not discussed further here is antigens that are expressed in healthy blood but not in AML or GvHD target organs. Targeting these antigens would give effects that are helpful for transplant by clearing remaining recipient-derived hematopoietic cells to make room for donor-derived cells to engraft and initiate donor-derived hematopoiesis; however, these antigens would not be targets of GvHD or GvL. These mHAs are included in some analyses with GvL mHAs as they give pro-transplant effects, but are not rightfully labeled as GvL mHAs. Our large dataset enables us to be more stringent in our selection criteria, and we excluded these so we would predict only the mHAs that would have the most potential to augment GvL effects of alloHCT. This workflow is summarized below (Figure 2.2).

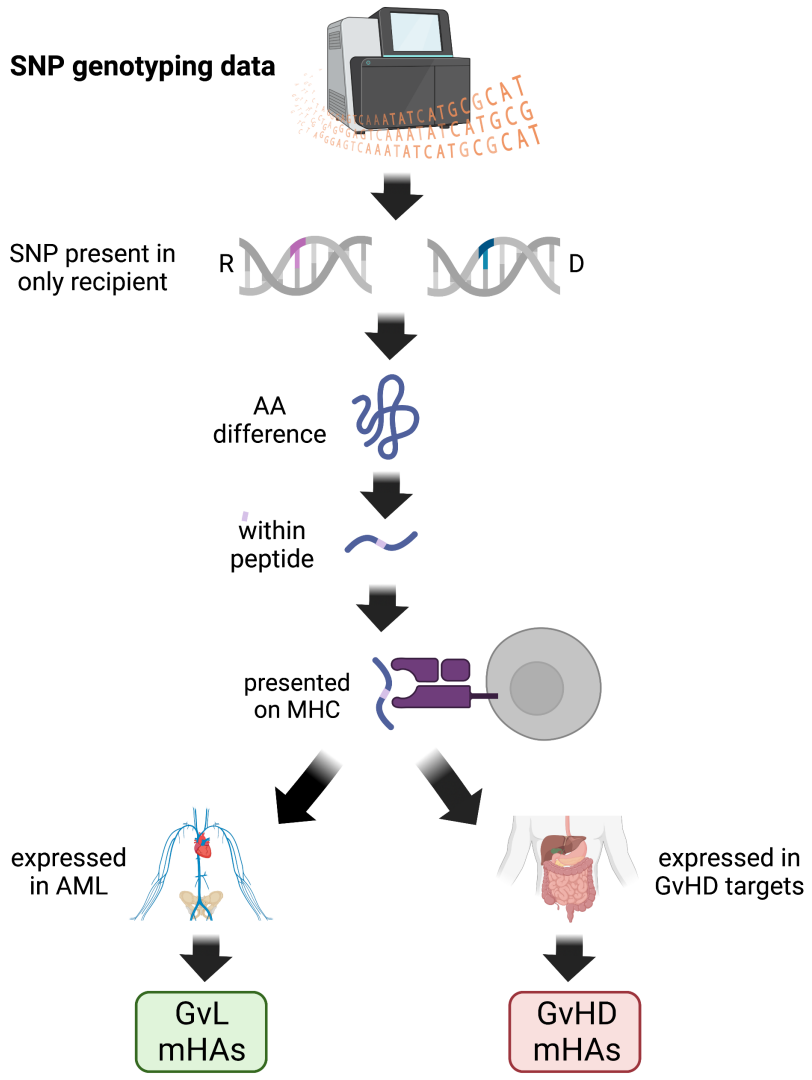


Figure 2.2: mHA prediction workflow from SNP typing data.

This workflow resulted in a small number of predicted mHAs for the 101 patient dataset due to the smaller DRP number as well as the smaller number of SNP loci genotyped. For the 3231 DRP dataset, 1,867,836 total GvL mHAs were predicted. All patients in the 101 patient dataset had myeloid malignancies, and TCGA AML expression data should be general applicable and somewhat similar to their tumor type. However, some patients in DISCOVeRY-BMT did not have myeloid malignancies, instead having ALL. The expression profile of ALL is significantly different than AML[215].

Ideally, for these patients GvL mHAs would have been called based on tissue expression data for ALL instead of TCGA AML. However, ALL is not one of the tumor types covered by TCGA, and at the time of mHA prediction there was no large publicly available dataset of ALL data on which to make these expression calls. We recently identified a public RNAseq data expression set that is more relevant but not directly applicable to this adult patient population: a set of RNAseq data for 110 pediatric ALL patients from phase 3 of the TARGET study via (https://portal.gdc.cancer.gov/repository)[216,217]. After mHA prediction and validation, we analyzed expression of source genes for our validated mHAs in this dataset and found high concordance of gene z-scores between TARGET and TCGA AML (see Chapter 4). However, some predicted mHAs may be less applicable for the ALL patients in the DISCOVeRY-BMT dataset than for the patients with myeloid malignancies.

mHA sharing analyses:

We developed a greedy algorithm implementation of the maximum set coverage solution to generate ranked lists of the most commonly shared mHAs for DRPs with a given HLA allele. This algorithm generates a list of the minimal set of peptides such that every DRP with a given HLA allele in the dataset contains at least one of these mHAs. In short, the algorithm ranks every peptide within a given HLA by the study population frequency in descending order. The peptide with the highest frequency is selected and added to the mHA set, then population frequency of every peptide is recalculated using only DRPs that do not contain an mHA in the set and the new highest frequency peptide is selected. This process is repeated until 100% of DRPs are represented by an mHA in the set. We applied this algorithm to predicted mHAs for all class I HLA alleles

represented by patients in the DISCOVeRY-BMT dataset. Three HLA alleles were selected as representative alleles for analyses based on high frequency in US ethnic groups and to include representative alleles for HLA-A, HLA-B, and HLA-C. HLA-A*02:01 is the most common HLA-A allele among Caucasians, African Americans, and Hispanics within the United States and third most common among Asians and Pacific Islanders, and is found within 28.4% of the total population of the United States[201]. HLA-B*35:01 is the most common HLA-B allele among Asians and Pacific Islanders, is third most common among African Americans, and is fifth most common among Caucasians and Hispanics. It is found within 6.7% of the population of the United States[201]. HLA-C*07:02 is the most common HLA-C allele among Hispanics within the United States, is second most common among Caucasians and Asians and Pacific Islanders, and is seventh most common among African Americans. It is found within 15.4% of the United States population[201].

(2.3) Results

Patient characteristics, ethnicity analyses, and genetic distance analyses

60% of DISCOVeRY-BMT patients had a diagnosis of AML, while the remainder had diagnoses of ALL or MDS. The number of predicted mHAs did not vary by disease type (Figure 2.3).

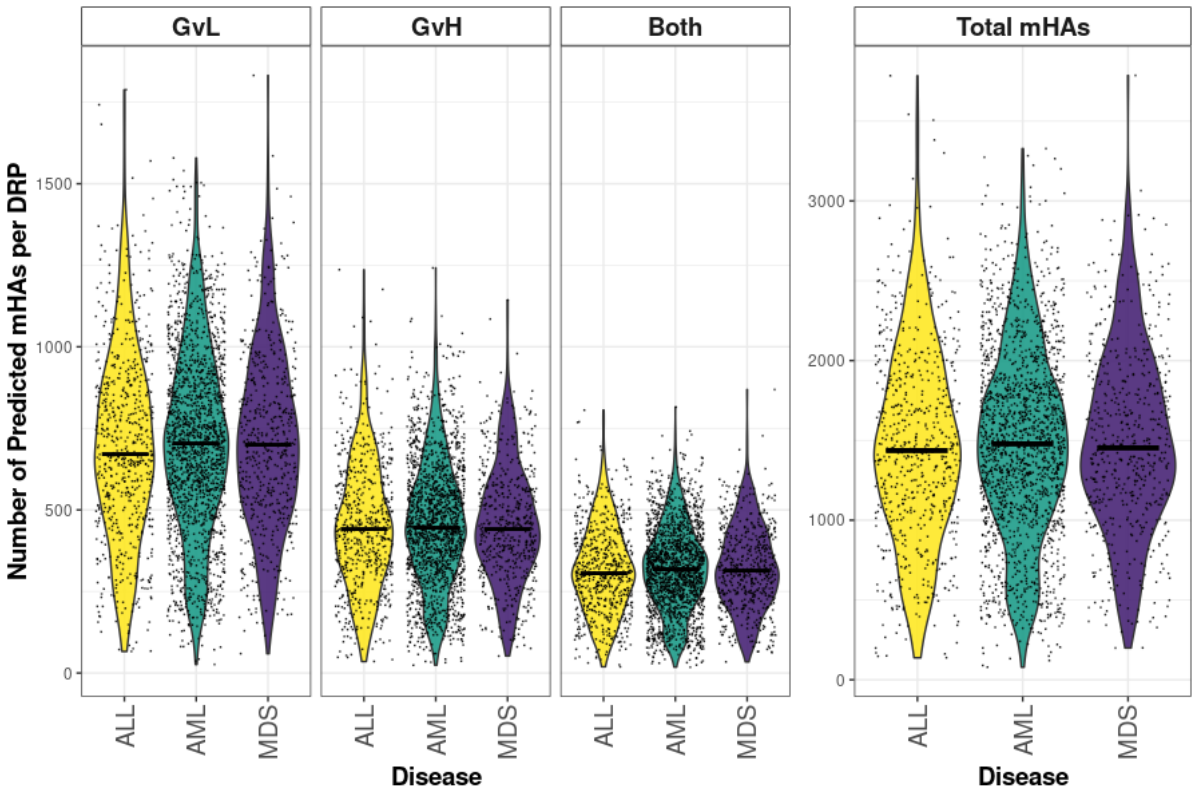


Figure 2.3: Predicted mHAs of each class by disease type in DISCOVeRY-BMT. "GvL" denotes expression in leukemia cells, "GvH" denotes expression in GvH target organs, and "both" denotes expression in both.

60% of recipients in DISCOVeRY-BMT are over 40 years of age, reflecting the general age distribution of AML[218]. The mean age of recipients in this dataset was 42.5. The mean donor age was 33.8, reflecting stringent donor selection criteria that prefer younger and healthier donors. The number of predicted mHAs did not vary by recipient age (r=-0.01, p=.22) or by donor age (r=-0.01, p=0.11) for any class of predicted mHA (Figure 2.4).

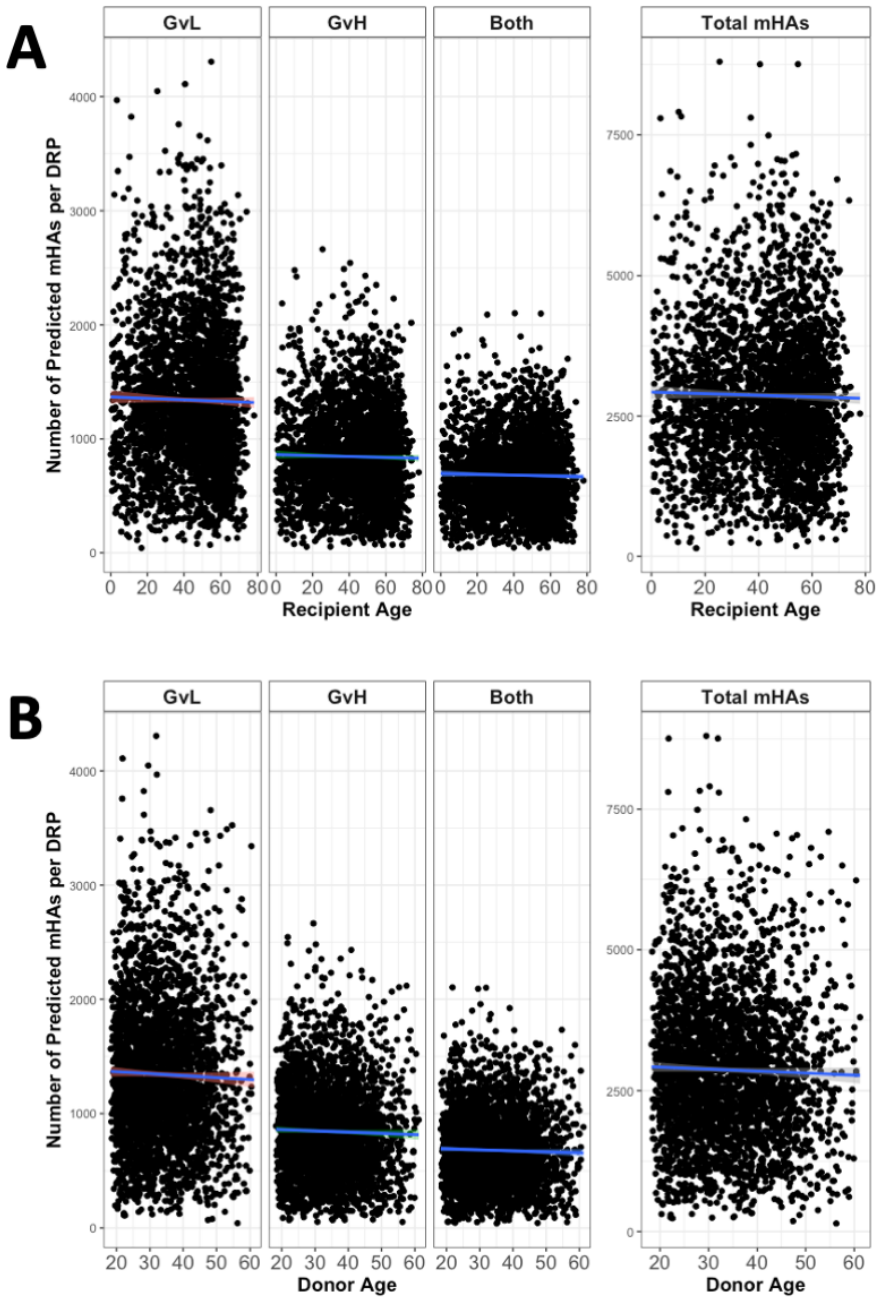


Figure 2.4: Donor and recipient age versus number of predicted mHAs of each mHA class in DISCOVeRY-BMT. (A) shows recipient ages vs number of predicted mHAs. (B) shows donor ages vs number of predicted mHAs. The line of best fit is shown for each group of mHAs.

While no other variables assessed correlated with number of predicted mHAs, investigation of ethnicity uncovered some significant findings. Overall, the self-reported

ethnicity and genomic ancestry of alloHCT recipients in this dataset mirror the general distribution of alloHCT recipients in the US, with an overrepresentation of patients with EA ancestry[219]. A large number of mHAs were predicted for each genomic ancestry group assessed in this study, with 75918 total predicted mHAs for EA, 27557 mHAs for AA, and 39272 mHAs for HIS. mHAs were assigned tags based on expression of the source gene in AML and GvHD target tissues. The mean total predicted mHAs per DRP across all ethnicities was 1476, with a mean of 704 predicted GvL mHAs. Number of predicted mHAs significantly differed by genomic ancestry group, with EA>HIS>AA for number of mHAs labeled as GvL, GvH, and both as well as total mHAs per DRP (Figure 2.5). DRPs had an average of 2152 total mHAs predicted for AA, 2591 for HIS, and 2917 for EA (p=1.23e-5). In the same pattern, DRPs had an average of 1018 GvL mHAs predicted for AA, 1229 for HIS, and 1363 for EA (p=7.26e-5).

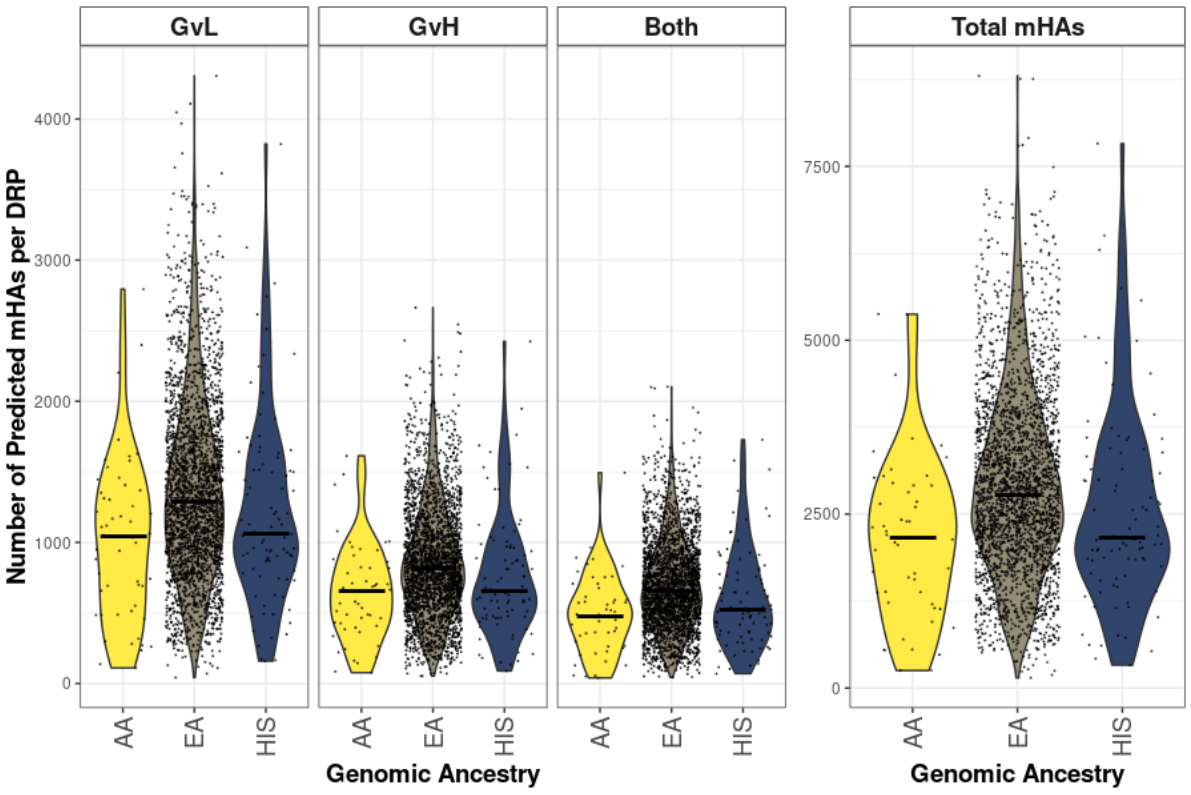


Figure 2.5: Number of predicted mHAs of each class based on genomic ancestry in DISCOVeRY-BMT. Figure shows number of each category of predicted mHA per DRP by genomic ancestry, including patients identifying as European American (EA), African American (AA), or Hispanic (HIS).

We then investigated whether DRPs where self-reported ethnicity was concordant vs discordant between donor and recipient had differing numbers of predicted mHAs. We found that DRPs with self-reported matching ethnicities had higher numbers of predicted mHAs of each class than ethnicity-discordant pairs. Concordant pairs had a mean of 2900 total predicted mHAs, while discordant pairs had 2536 (p=1.45e-5). Concordant pairs also had a mean of 1356 predicted GvL mHAs, while discordant pairs had 1198 (p=.0001). In conclusion, some genetic ancestry groups gave higher numbers of predicted mHAs than others, and if pairs were a self-reported ethnicity match they had higher numbers of predicted mHAs (Figure 2.6).

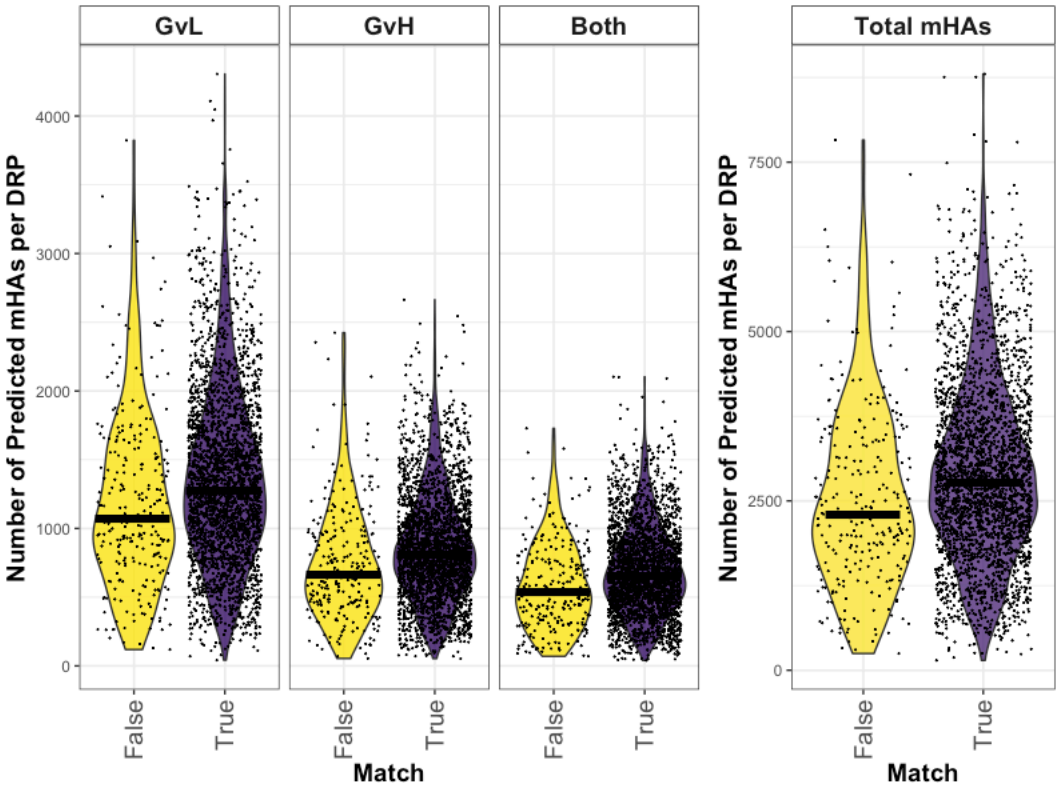


Figure 2.6: Number of predicted mHAs of each class based on self-reported ethnicity concordance for DRPs in DISCOVeRY-BMT. Self-reported ethnicity was utilized for this analysis because genomic ancestry calling was not available for all DRPs. Pairs were labeled “False” if their self-reported ethnicity did not match, and “True” if they did.

Due to the unexpected nature of the finding that numbers of mHAs predicted varies by recipient genomic ancestry and self-reported ethnicity matching in a DRP, we investigated other potential explanations for this result. We considered whether DRPs with recipients corresponding to the genomic groups with more predicted mHAs would have higher pairwise genetic distance values, as a higher number of SNP loci that differ could lead to a higher number of predicted mHAs. We did not find this to be the case, and the EA genomic ancestry group that had the highest number of predicted mHAs actually had the lowest genetic distance values (Figure 2.7). The mean genetic distance for AA was 0.2723, for HIS was 0.2700, and for EA was 0.2620 ($p<2e-16$).

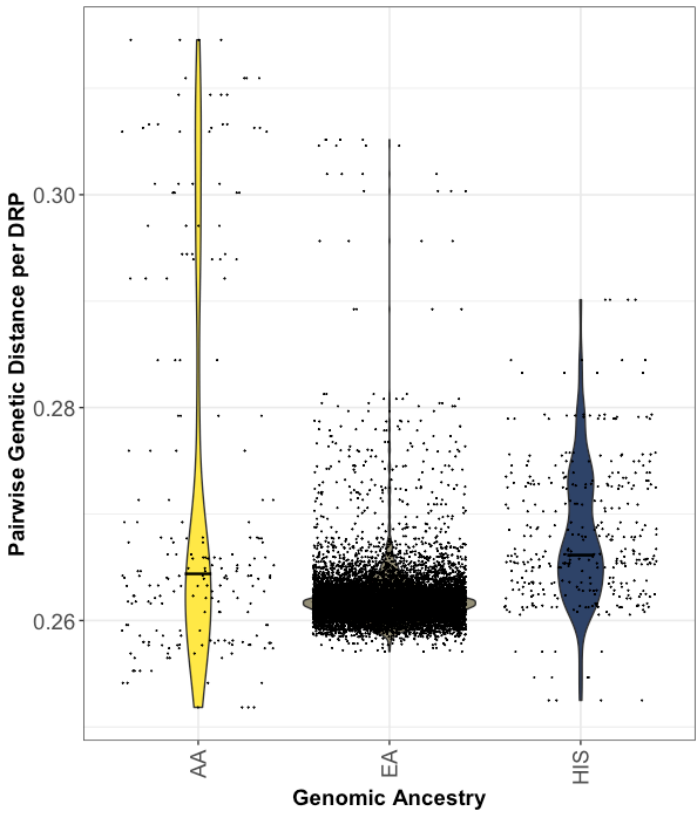


Figure 2.7: Pairwise genetic distance per DRP by recipient genomic ancestry group. Only three genomic ancestry groups were considered as other groups had two few representatives for analyses. For these three groups, n=2960 of the total 3231 in dataset.

Likewise, we investigated the correlations between numbers of predicted mHAs, number of mHA-encoding SNPs per DRP, and genetic distance. We saw a strong positive correlation between total number of mHA-encoding SNPs and number of predicted GvL mHAs (Figure 2.9a). This observation is expected, as more total SNP that mHAs are predicted from will lead to more GvL mHAs. We observed an overall narrow range of pairwise genetic distance across all DRPs within DISCOVeRY-BMT (Figure 2.9b), likely because a large number of rare SNPs were genotyped leading to high denominators of total SNPs and low numerators of SNPs that differ in genetic distance calculations. Still, distance values were consistent with previously reported

data for healthy pairs[209]. For the 101 patient dataset, MUD DRPs had higher genetic distance values than MRD DRPs, which is logically consistent with the nature of this value (Figure 2.8). We utilized DISCOVeRY-BMT to investigate the impact of the genetic distance value and found no correlation between genetic distance and predicted GvL mHAs (Figure 2.9c) or total mHAs (Figure 2.9d). As the DISCOVeRY-BMT cohort did not include MRD DRPs, we were not able to compare MRD to MUD DRPs in DISCOVeRY-BMT.

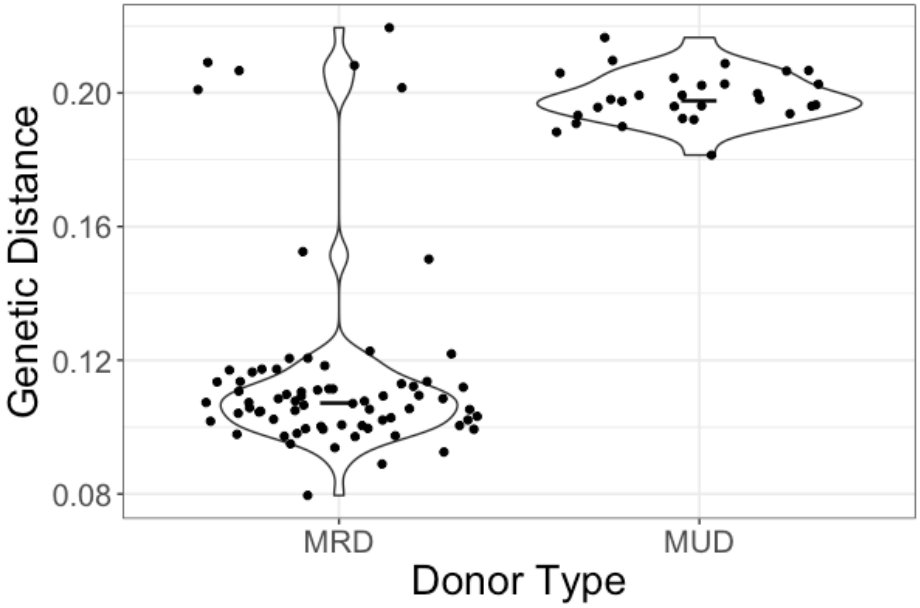


Figure 2.8: Pairwise genetic distance values for matched related donor (MRD) and matched unrelated donor (MUD) DRPs in the 101 patient dataset.

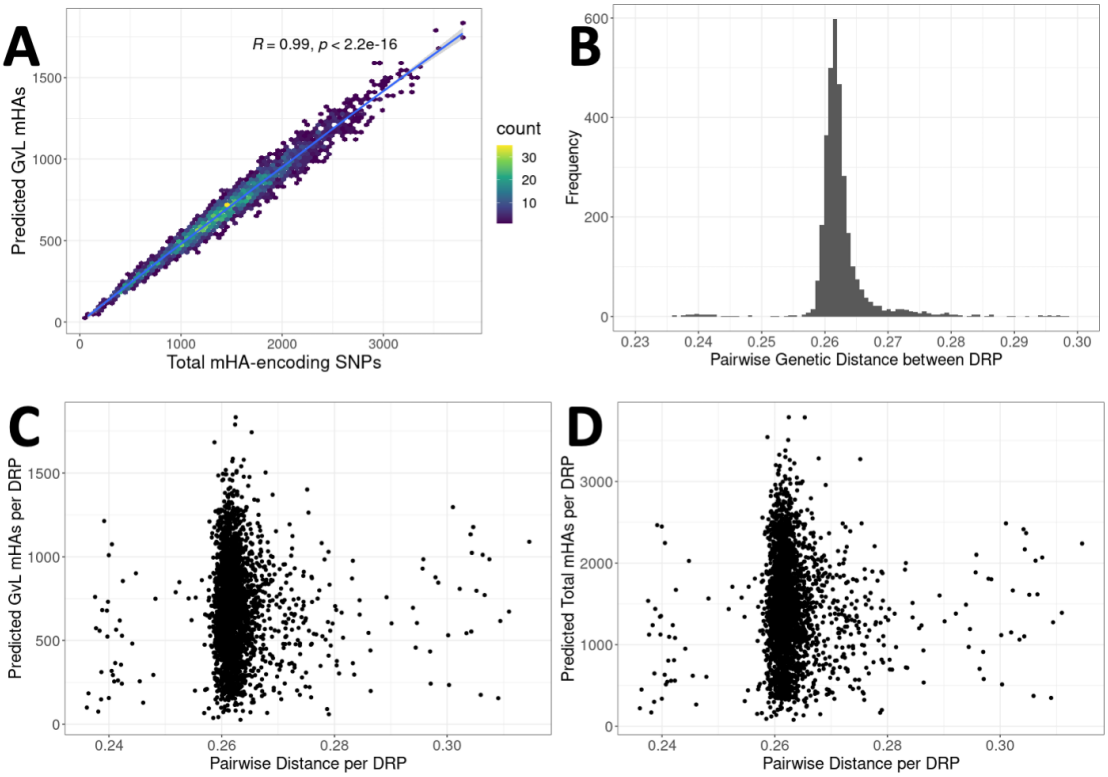


Figure 2.9: Degree of genetic distance versus number of predicted GvL mHAs by DRP in the DISCOVeRY-BMT dataset. (A) shows number of total SNPs that differ and are predicted to lead to an mHA versus number of predicted GvL mHAs per patient. (B) shows distribution of pairwise distance values for every DRP in the DISCOVeRY-BMT dataset. Pairwise genetic distance value is calculated as the mean of (1-.5(number of shared alleles at SNP locus)) for every genotyped SNP locus for a DRP. (C) shows pairwise genetic distance versus number of predicted total mHAs per DRP. (D) shows pairwise genetic distance versus number of predicted GvL mHAs per DRP.

mHA predictions by HLA allele

A total of 23 HLA-A alleles, 26 HLA-B alleles, and 7 HLA-C alleles were represented in DISCOVeRY-BMT. The total number of predicted mHAs that bind each allele varied widely: from 82 to 11017 for HLA-A alleles, 19 to 8585 for HLA-B alleles, and 946 to 7537 for HLA-C alleles (Figure 2.10A/B/C). However, our method predicted GvL mHAs for every HLA allele represented within DISCOVeRY-BMT. Next, we looked at the proportion of mHAs classified as GvL, GvH, or both for each HLA allele. GvL

mHA comprised approximately half of all predicted mHAs for each HLA allele (Figure 2.10D/E/F).

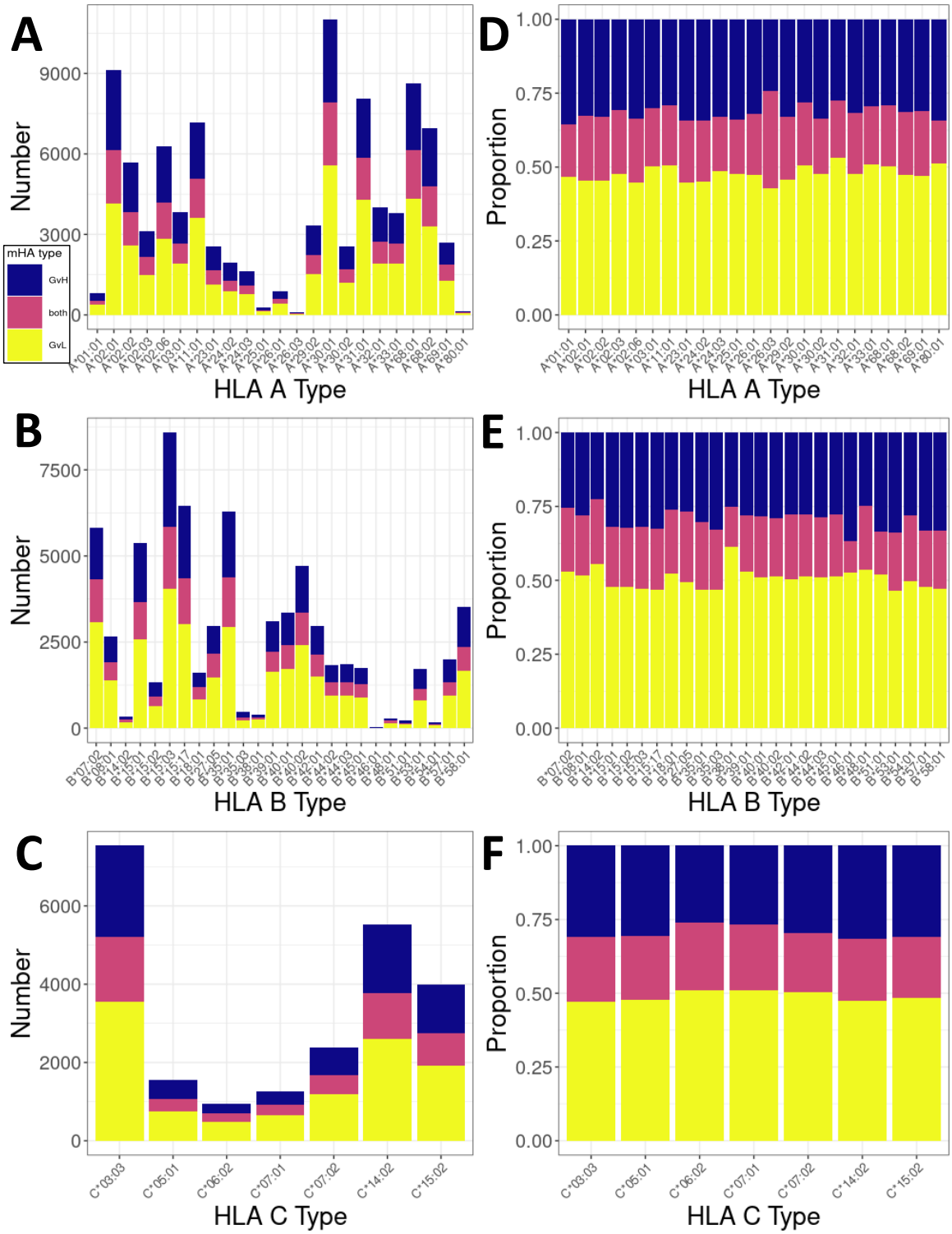


Figure 2.10: Number and proportion of predicted mHAs by HLA allele within study population. mHAs classed as "GvL" broadly represent mHAs that are desirable to target for anti-leukemia effects with minimal GvHD. mHAs classed as "GvH" represent mHAs that are undesirable to target as they are predicted to correspond to GvHD and

no GvL effects. “Both” category represents peptides that are predicted to lead to both GvL and GvH effects. (A) shows counts of each predicted class of mHA for HLA-A alleles represented in patient dataset. (B) shows counts for HLA-B alleles represented in patient dataset. (C) shows counts for HLA-C alleles represented in patient dataset. (D) shows proportion of predicted mHAs corresponding to each mHA class for HLA-A alleles. (E) shows proportion for HLA-B alleles. (F) shows proportion of HLA-C alleles.

We evaluated sharing of predicted mHAs within the DISCOVeRY-BMT cohort. Of our predicted mHAs, the majority were found within less than ten DRPs (<0.3% population frequency). However, 38.7% of our predicted mHAs were shared by 1% or more of the study population, and 4% were shared by 10% or more of the study population (Figure 2.11A). Next, we assessed sharing of mHAs within individual HLA alleles. For the three HLA alleles focused on in this work, the population frequency of predicted mHAs shows a bimodal distribution. Most mHAs are unshared, but a group of mHAs covers approximately 20-30% of patients (Figure 2.11B-D).

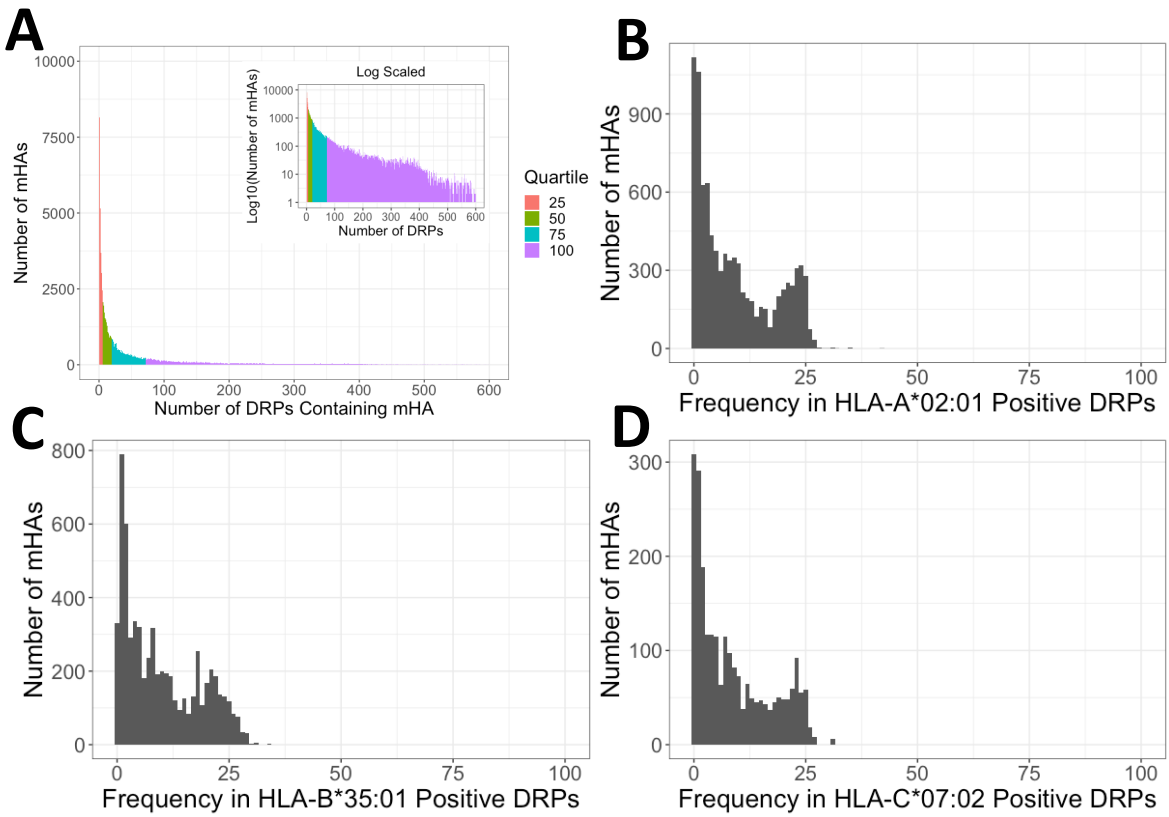


Figure 2.11: Degree of sharing of predicted mHAs across study population. (A) shows distribution of predicted mHAs by the number of patients in the DISCOVeRY-BMT cohorts that possess them. Most mHAs are shared by ten or fewer patients. Inlaid is the same data with log transformed y axis to highlight the tail of the distribution. Data are colored by quartile of number of patients for each mHA. (B) shows distribution of predicted HLA-A*02:01 mHAs by population frequency in DRPs with HLA-A*02:01. (C) shows distribution of predicted HLA-B*35:01 mHAs by population frequency in DRPs with HLA-B*35:01. (D) shows distribution of predicted HLA-C*07:02 mHAs by population frequency in DRPs with HLA-C*07:02.

Finally, we assessed predicted mHA frequency across all HLA alleles represented by greater than 0.5% of DISCOVeRY-BMT patients. The same bimodal distribution of mHA population frequency was observed across most HLA alleles (Figure 2.12).

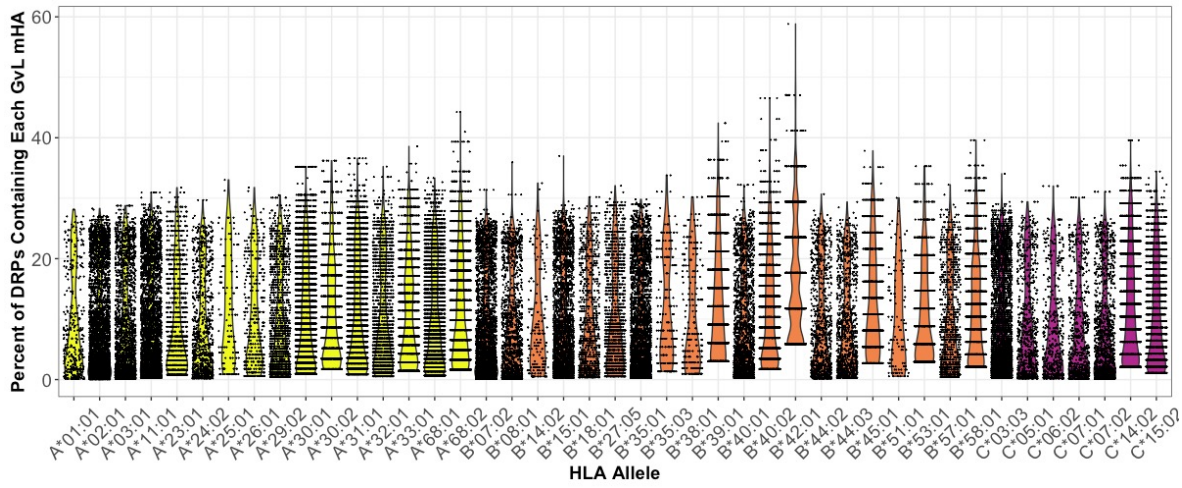


Figure 2.12: Degree of sharing of all predicted mHAs for most HLA alleles. Figure shows percentage of DISCOVeRY-BMT cohort with each HLA allele covered by each predicted GvL mHA that binds that HLA allele, for all HLA alleles representing greater than 0.5% of DISCOVeRY-BMT patients.

We selected three HLA alleles to generate minimal mHA sets *in silico*. Together, HLA-A*02:01, HLA-B*35:01, and HLA-C*07:02 represent a set of common alleles within the US population and within the major ethnic groups found in the DISCOVeRY-BMT population. For the most common HLA allele in the US, HLA-A*02:01, a set of fifteen GvL mHAs are needed to ensure that every DRP with this HLA allele has at least one of the fifteen (Figure 2.13A). Only seven peptides are needed to reach 90% coverage. The non-cumulative population frequencies for each of these top fifteen peptides range from 19.4% to 28.3%. We obtained similar results with HLA-B*35:01: eleven peptides are needed to reach 100% population coverage and six peptides are needed to reach 90%, with non-cumulative population frequencies between 20.9% and 29.3% (Figure 2.13B). HLA-C*07:02 also showed similar results, with fourteen peptides needed to reach 100% population coverage and seven peptides needed to reach 90%. Noncumulative frequencies ranged from 19.3% to 31.1% (Figure 2.13C). A total of 40 peptides gives

100% population coverage of three HLA alleles that are among the most common in major ethnic groups in the United States.

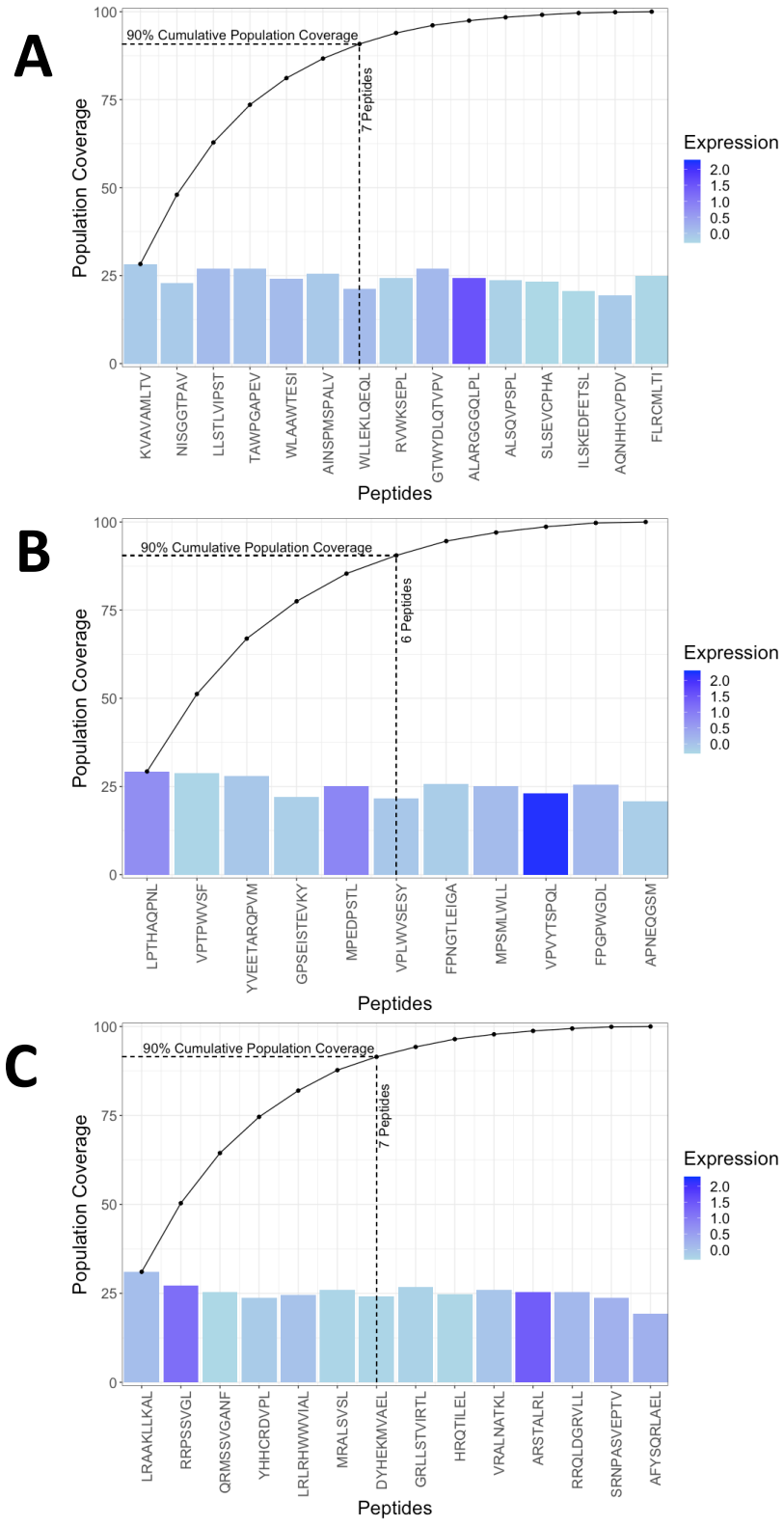


Figure 2.13: Patient population cumulative coverage by shared GvL mHAs. (A) Coverage of DISCOVeRY-BMT patients with HLA-A*02:01 allele with predicted GvL

mHAs. Noncumulative independent population frequencies of each of the top fifteen peptides within the HLA-A*02:01 population range from 19.4% to 28.3%, shown as bar heights. Colors of bars show z-scores of expression for the genes that contain each peptide from The Cancer Genome Atlas AML sample expression data (TCGA_AML). Cumulative population coverage by the fifteen predicted GvL mHAs needed to reach 100% corresponding coverage is shown as an overlaid line graph. Dotted lines indicate seven peptides needed to reach 90% population coverage. (B) shows coverage of DISCOVeRY-BMT patients with HLA-B*35:01 allele with predicted GvL mHAs. Eleven predicted GvL mHAs correspond to 100% cumulative population coverage and six correspond to 90% coverage for this HLA allele. Noncumulative coverage by the top 11 peptides for this HLA allele range from 20.9% to 29.3%. (C) shows coverage of DISCOVeRY-BMT patients with HLA-C*07:02. Fourteen predicted GvL mHAs correspond to 100% cumulative population coverage and seven correspond to 90% for this allele. Noncumulative coverage for these mHAs range from 19.3% to 31.1%.

(2.4) Discussion

Discovery and characterization of novel mHAs may be crucial for enhancing immune monitoring in alloHCT, predicting clinical outcomes based on donor and recipient genetics, and improving outcomes by optimizing donor selection and/or specifically targeting GvL mHAs. We built upon previous work to perform the first population level survey of mHA peptides, taking a new approach by predicting mHAs common among recipients with diverse HLA alleles. This ensures that therapeutics targeting our newly identified mHAs would be applicable to as broad of a recipient population as possible.

We evaluated mHAs for the total of 56 HLA-A, -B, and -C alleles called in 3231 DISCOVeRY-BMT recipients. We found that multiple factors influence the number of mHAs we called for each HLA allele, including the total number of DRPs with that specific HLA allele and the recipient genomic ancestry group. An unexplained finding was the significant difference in number of mHAs predicted per DRP by genomic ancestry group, with EA>HIS>AA. Adding to this unexpected finding, when we assessed the pairwise genetic distance for each DRP we found the opposite trend in genomic

ancestry group, with AA>HIS>EA for pairwise genetic distance. When we assessed the number of mHAs predicted per DRP versus genetic distance, we found no correlation for any class of mHA. That makes this trend of higher numbers of predicted mHAs for groups with lower genetic distance values more unexpected, and we have no clear explanation for this finding other than potential technical limitations with the SNP arrays used in DISCOVeRY-BMT as discussed below. We also found that DRPs with matching self-reported ethnicities have higher numbers of predicted mHAs than pairs that do not match. This was also an unexpected finding, as we expected that ethnicity-discordant pairs might have more SNPs that differ between them due to prevalence of different SNPs in different ethnic populations. One proposed explanation for these observations is that the design of the SNP panel used for genotyping of the DISCOVeRY-BMT dataset plays a role in number and qualities of the SNPs identified and therefore mHAs predicted. SNP panels are designed based on SNPs that are relatively high population frequency, focusing on genotyping sites with known polymorphism rather than genotyping loci where any difference is extremely rare[220,221]. Scientific research as a whole and particularly genetic sequencing has traditionally been done with a strong focus on Caucasian and European American individuals, so SNPs selected for inclusion in a SNP panel are likely based on which SNPs are commonly polymorphic in individuals of this ethnicity. Therefore, SNP panels may be more likely to uncover SNP differences in Caucasians as the panels are designed to detect SNPs in that population[222,223]. Loci that are more often polymorphic in non-Caucasian individuals may not be included in SNP typing panels, therefore artificially making the number of SNPs that non-Caucasian individuals have differences at look less frequent. If we are searching more loci that are

common for EA individuals, that could explain why we predicted more mHAs for EA individuals and less for other groups. Previous work has shown that ascertainment bias, in which SNPs are ascertained in one population and then applied to genotyping of other populations, can lead to erroneous conclusions. Some studies have found that assessing heterozygosity of SNPs ascertained in European populations can lead to falsely concluding that European populations have greater variation than non-European populations[224,225]. Likewise, due to the high prevalence of Caucasian donors in the donor registry, most donors in this study are Caucasian. This means that most Caucasian recipients in our dataset received transplants from Caucasian donors, and the “Match” category in ethnicity analyses is predominated by Caucasian DRPs. Therefore, if SNP panels are designed to identify more mHAs in Caucasians, the category with higher numbers of Caucasian DRPs would also have more mHAs, and this would explain why we found more mHAs for ethnicity-matched pairs than mismatched. These findings merit more investigation with additional whole genome sequencing to capture genetic variants that are less common in Caucasians and more common in other ethnicities. In the future, we would like to repeat mHA prediction with a dataset with a higher prevalence of non-Caucasian DRPs if such a set becomes available. We also feel that SNP typing panels specifically targeting loci with polymorphisms higher in a variety of ethnic backgrounds is warranted if the high cost of whole genome sequencing persists and we hope that sequencing technology continues to make strides towards equitable availability of genetic research resources.

Despite large differences in total number of mHAs per HLA allele, approximately 50% of predicted mHAs for each HLA allele are GvL. Therefore, we expect that every

HLA allele will present a set of GvL mHAs. We also expect that if we had equal numbers of individuals with each HLA allele, the total number of mHAs predicted for each allele would have been roughly similar, and the variance in numbers of total mHAs predicted was due to the relative abundance of alleles in the dataset. Some number of GvL mHAs were predicted for every DRP in the dataset regardless of HLA allele. The majority of GvL mHAs are shared among fewer than 10 patients in the dataset, highlighting the largely private nature of the mHA landscape. However, though shared mHAs are less common than private mHAs, there are some shared mHAs that cover large percentages of the population. 38.7% of mHAs were shared by 1% or more of the population, and 4% were shared by 10% or more, indicating the low but present frequency of shared mHAs in the population. 1602 total predicted mHAs were shared by 10% or more of the population, which is more than we foresaw finding given the field's prior expectation that shared mHAs were very rare. We found the presence of these antigens that are broadly shared across many mHAs encouraging. We proceeded to assess how many GvL mHAs it would take to cover all DRPs with a given HLA allele within our dataset. We found that a small set of shared antigens is sufficient to cover all individuals within an HLA allele. For each HLA allele, we predicted a small number of highly shared mHA expressed by 20-25% of the recipient population. For all three of the HLA alleles we selected for further analyses, 6-8 mHA peptides would cover >80% of recipients that express that allele, and 11-15 mHAs would cover 100% of recipients. Therefore, we conclude that targeting a small number of shared GvL mHAs could treat a majority of alloHCT recipients regardless of race or ethnicity. We envision a set of TCRs targeting mHAs that are mostly highly shared in a population, such that any patient presenting with an HLA allele

common in any major ethnic group in the US would have at least one targetable mHA and we would be able to treat all of these patients with off-the-shelf therapeutics.

CHAPTER 3: Custom Mass Spectrometry Method Generation for Minor Histocompatibility Antigen Validation[2]

(3.1) Introduction

While mHA prediction methods have advanced significantly since the discovery of the first mHA in 1976, development of the tools that enable a reverse immunology approach to mHA discovery is still an active field of research[133]. While individual tools such as netMHCpan predict peptide/HLA binding affinity and other mHA qualities fairly well and generally have had a lower rate of false positives with every new software release, the combined false positive rate from fusing a series of imperfect tools to predict mHAs is both high and not perfectly estimable. Every group doing antigen prediction work has a slightly (or in some cases drastically) different approach to their prediction algorithm, with varying degrees of successful identification of actual antigens that generate immune responses *in vivo*. Some studies have attempted to quantify the efficacy of different prediction approaches via “bake off” style competitions in which labs are invited to submit antigen predictions and the predictions are centrally tested to establish which method had the most success. A recent study set up a consortium called the Tumor Neoantigen Selection Alliance (TESLA) and challenged teams from academia, nonprofits, and industry to predict neoantigens from tumor and normal whole exome sequencing (WES), RNAseq, and clinical HLA typing from three patients with non-small cell lung cancer (NSCLC) and three patients with metastatic melanoma[52]. 25

[2] This chapter includes unpublished data that are part of a planned future manuscript from the Vincent and Armistead labs. I performed experiments, analyzed results, generated figures, and wrote the text included here.

teams submitted between 7 to 81,904 predicted peptides for each sample with very little overlap between teams, with a median of 13% overlap between pairs of teams and a maximum overlap of 62%[52]. In addition to the final peptide predictions, the prediction methods used by each team varied widely. The study group selected 608 peptides from the top-ranked predicted peptides across groups and tested them for immunogenicity and found that only 6% of these most agreed upon predicted peptides were immunogenic[52]. This work illustrates that regardless of the quality of your peptide predictions, with the current resources we have for this work, one must experimentally validate their predicted peptides and not simply assume that all peptides predicted to be immunogenic 1) are actually presented via MHC and 2) generate a cognate T cell response *in vivo*. The two broad categories of options for mHA validation in a reverse immunology approach are to identify the presence of the predicted peptide bound by HLA on the cell surface, or to identify a T cell clone that targets the predicted peptide. Identifying a T cell clone gives you one level of information beyond finding the antigen presented on the cell surface and is more directly applicable for downstream work including antigen targeting. However, identification of antigen-specific T cells is extremely laborious, time consuming, and sample dependent. We chose an approach where we would first validate our predicted mHAs by mass spectrometry and identify peptides that are presented *in vivo* as our main confirmational approach, then separately take these validated mHAs forward to T cell targeting work for downstream applications. In the work described in this chapter, we sought to optimize mass spectrometry methods for mHA validation in collaboration with the UNC Proteomics Core.

Multiple different modalities of mass spectrometry are available, and selection of the appropriate sampling method for your analyte of interest is critical. For immunopeptidomics, where the target of identification is not a full size protein but a set of HLA-bound peptides, this selection is even more important, as identification of peptides is technically difficult[226]. The standard method of proteomics data acquisition for the last 15 years has been liquid chromatography-tandem mass spectrometry (LC-MS/MS) in data-dependent acquisition (DDA) mode[227]. In this method, proteins are extracted from a tissue of interest, enzymatically digested into tryptic peptides, and fractionated based on their hydrophobicity via liquid chromatography prior to spraying into the mass spectrometer for protein identification[228]. The name of DDA can be somewhat misleading for investigators unfamiliar with mass spectrometry, as it is an untargeted “shotgun” method of mass spectrometry. In shotgun methods, one obtains a broad view of the proteins presented and relative abundance of the proteins observed, not inputting any data on specific proteins that should be searched for[229]. These methods are often used in traditional proteomics to assess the broad scale differences in protein expression between two conditions, such as healthy versus cancer[229]. In DDA, the peptides fractionated by LC are electrosprayed into the tandem mass spectrometer via an electrospray needle coupled to the LC column, the ions enter the mass spectrometer, and are separated according to mass-to-charge ratio (m/z) by the mass analyzer[230]. This initial scan is known as the MS1 survey scan, and the most intense ions observed in the MS1 scan are selected for MS2 scans[231]. In the MS2 scan process, the mass spectrometer isolates and fragments the ions selected by the MS1 scan, giving a set of MS2 high quality spectra for the most abundant ions in the

sample[231]. Shotgun methods applied to immunopeptidomics have some drawbacks, including that most algorithms used to identify peptides based on MS output spectra are designed to process data from trypsin digested proteins, and immunopeptidomics studies peptides are isolated directly from HLA molecules rather than using trypsin to digest larger proteins[232]. The C termini of HLA-bound peptides vary because different HLA alleles bind peptides with different sequences at the termini, while tryptic peptides will always have the same C terminus because trypsin cleaves at specific sites[232]. However, recent methods advances such as hybrid fragmentation and changes to the way that spectral data are acquired have enabled better quality of spectra from peptides eluted from HLA[233,234].

So, while shotgun methods are improving, they are less than ideal for identification of HLA-bound peptides. They are especially unsuited for confirmation of specific predicted peptides as these methods only give data on the most abundant spectra and are not suited to searching for particular peptides of low-to-moderate abundance. One can overcome this limitation using targeted proteomics, where the sample spectra are compared to spectra from known standards to obtain information on specific proteins or peptides of interest[235]. The classic method for targeted MS is selected reaction monitoring (SRM), in which samples are put into a triple quadrupole mass spectrometer, with a precursor ion selected in quadrupole 1 based on predetermined m/z ratio for the target of interest, fragmentation of selected precursor ions in quadrupole 2, and additional mass filtering once more in quadrupole 3[235]. SRM approaches are costly and fairly low-throughput due to the necessity of generation of isotope-labeled peptide standards for each peptide or protein of interest[236]. One way to

increase throughput is by utilizing a parallel reaction monitoring (PRM) method, in which the third quadrupole is replaced with an Orbitrap[237]. Less method optimization and development time is needed for PRM compared to SRM because the parallel monitoring means that there is no need to pre-determine target peptide transitions[235].

We sought to combine shotgun and targeted approaches to peptide discovery and validation in order to establish functional immunopeptidomics in-house, as well as to establish a workflow that allowed for stepwise selection and evaluation of a few peptides of interest from a large list of predicted peptides. We first ran DDA on samples and crosschecked the obtained spectra against a large pool of predicted peptides, then we obtained labeled peptide standards for these identified peptides and validated their presence using a targeted PRM approach (Figure 3.1).

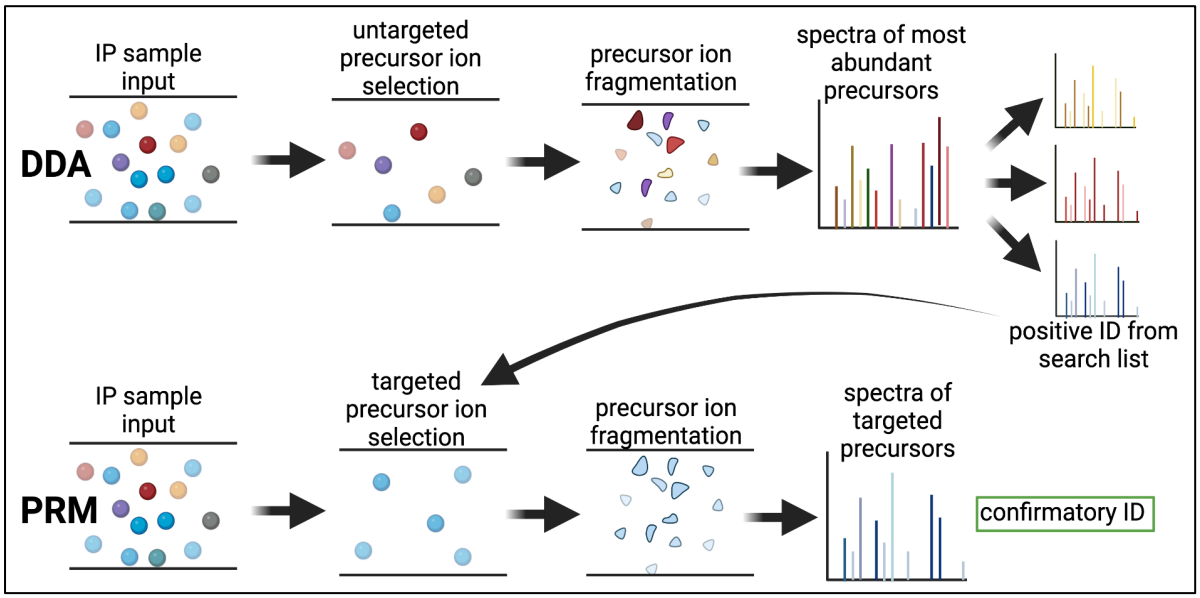


Figure 3.1: Mass spectrometry approach for predicted mHA identification and validation by shotgun and targeted mass spectrometry.

This combined approach allowed us to select peptides of highest interest from our large list of predicted antigens to then further analyze, eliminating the need to obtain

heavy labeled standards and spend time on methods developments for peptides that are ultimately less likely to be positively identified. Using our combined approach, we validated two novel HLA-A*02:01 binding mHAs. These peptides are UNC-DPEP1-Q (NLLRVFQAV), and UNC-TXNDC2-T (ILSKEDFETSL).

(3.2) Methods

mHA selection for validation:

mHA predictions from both the 101 DRP dataset and the 3231 DRP dataset were utilized for validation attempts described in this chapter. We utilized a list of 31 total peptides for these analyses that bound HLA-A*02:01 as the only HLA allele-specific antibody available for mass spectrometry was for HLA-A*02 and we wished to develop this method with as few variables as possible, including ensuring that all immunoprecipitated peptides were bound to an HLA allele of interest.

Peptide immunoprecipitation from cells

We sought to immunoprecipitate HLA-bound peptides from leukemia cells to test for mass spectrometric validation. We used the leukemia cell line U937A2 that expressed the HLA allele of interest. This cell line was U937 cells (ATCC CRL-1593.2) transfected to stably express HLA-A*02:01 by the Dotti lab at UNC, and we thank Dr. Dotti for his contribution in sharing this cell line with us. Cells were grown in Gibco RPMI 1640 (Thermo Fisher 11875119) with 10% Hyclone fetal bovine serum (FBS) (VWR 76237-676) , 100 U/mL penicillin/streptomycin (Thermo Fisher 15140122) and 2 mM L-glutamine (Thermo Fisher 25030081). We created an optimized protocol for immunoprecipitation (IP) of HLA-bound peptides by modifying previous IP protocols from compiled literature[238–240]. We expanded the U937A2 cells to 150-500 million cells

depending on mass spectrometric needs. We then pelleted the cells, washed them three times in 4C Gibco Dulbecco's phosphate buffered saline (DPBS) (Thermo Fisher 14040117), and lysed for 1 hour on ice in 6mL of IP lysis buffer, which is comprised of 1x DPBS, 1% Triton X-100 (Millipore Sigma T8787), 5mM EDTA (Thermo Fisher 17892), 1x HALT protease inhibitor (Thermo Fisher 78430), 1mg/mL leupeptin (R&D Systems 1167), and 1mM phenylmethylsulfonyl fluoride (PMSF) (Millipore Sigma P7626), vortexing every fifteen minutes during the lysis. Samples were then centrifuged for 20 minutes at 12,000 rpm at 4C. We then added Protein A/G Ultralink Resin (Thermo Fisher 53135) to the lysate and incubated on a rotator at 4C for thirty minutes to preclear lysates of any materials that would have bound resin with no crosslinked antibody. The samples were then centrifuged and precipitate discarded to remove nonspecifically bound proteins. HLA molecules with bound peptide were then immunoprecipitated from the precleared lysates via incubation for three hours on a rotator at 4C with HLA-A*02 specific BB7.2 antibody (BioLegend 343307) crosslinked to Protein A/G Ultralink Resin[39,241]. Resin-bound HLA was then washed twice with 50mM Tris pH 8.0 (Thermo Fisher AM9855G), resuspended in Tris, and transferred to a 2mL Pierce centrifuge column (Thermo Fisher 89896). Peptides were finally eluted from the HLA molecules by gravity flow in 5 fractions with one column volume per fraction of ice cold 10% glacial acetic acid (Millipore Sigma A6283). Based on our experience, fractions 2-4 contained almost the entirety of the eluted peptides, so these fractions were pooled as the final immunopeptide samples and stored at -80C until analysis. An overview of the IP workflow is shown below (Figure 3.2). While dozens of IPs were performed for methods development purposes, 11 final samples were split and utilized

for DDA and PRM. One adjustment was made in the procedure for samples 10 and 11, with pipette tips cut to create a broader opening when pipetting samples containing Ultralink Resin to avoid shearing the samples off of the resin.

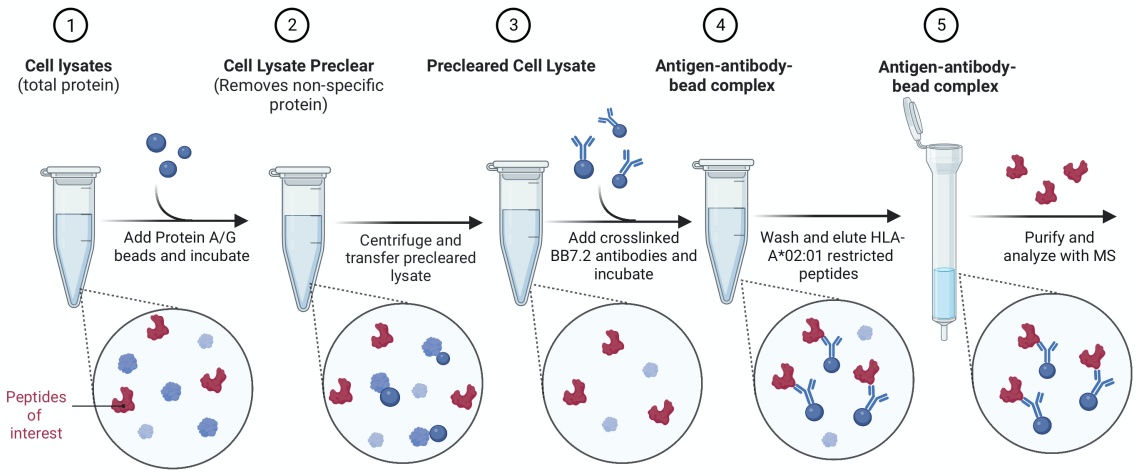


Figure 3.2: HLA-bound peptide immunoprecipitation workflow. Figure generated by Marisa Diiorio.

LC-MS/MS Sample Preparation

Immunoprecipitated peptides were lyophilized and desalted using 1cc/100mg tC18 SEP-PAK cartridges (Waters WAT036820) using 30% or 50% acetonitrile (Millipore Sigma 34851) and 0.1% trifluoroacetic acid (Millipore Sigma 302031). A C18 cleanup was then performed on NEST Silica C18 UltraMicroSpin columns (Harvard Apparatus 74-7206).

LC-MS/MS Data Acquisition: DDA

For DDA, post-cleanup samples were loaded onto a 75um ID x 25cm 2um particle size EasySpray PepMap RSLC C18 column (Thermo Fisher ES802). They were eluted over a 110 minute acquisition with a separation gradient of 3-60% B at a

250nL/min flow rate with mobile phase A comprised of 0.1% formic acid, and mobile phase B comprised of 0.1% formic acid and 80% acetonitrile. Precursor scan (m/z range of 300 to 1700) resolution was set to 60,000 with a target value of 3x10e6 ions and 50ms maximum ion injection time (IT)[242]. Dynamic exclusion was set to 5s, peptide match set to preferred, and precursors with a charge state of greater than or equal to five or unknown charge were excluded. MS/MS scan resolution was set to 15,000 with a target value of 1x10e5 ions and 120ms maximum ion IT and normalized HCD collision energy set to 27%.

LC-MS/MS Data Acquisition: PRM

Peptide standards (Peptide 2.0 Inc) were obtained in lyophilized form and resuspended at 5mM in DMSO, then diluted to 250nM in pooled HLA-A*02:01-binding peptide from immunoprecipitation with 0.1% formic acid. Peptide standard precursor m/z ratios and collision energies were assessed and optimized by direct infusion and nLC-MS/MS using a Thermo Easy nLC 1000 coupled to a QExactive HF. PRM samples were eluted from the EasySpray C18 column using the same 110 minute, 3-60% B gradient utilized for DDA. Precursor scan (m/z range of 400 to 1200) resolution was set to 60,000 with a target value of 3x10e6 ions and 200ms maximum ion IT. MS/MS scan resolution was set to 15,000 with a target value of 1x10e6 ions, 200ms maximum ion IT, normalized HCD collision energy set to 27%, loop count 25, and isolation windows of 2 m/z.

Data Analysis

DDA data were analyzed in ProteomeDiscoverer 2.4 (Thermo Fisher). PRM data were analyzed using Skyline. An overview of the entire workflow from mHA prediction to data analysis and mHA validation is shown below (Figure 3.3).

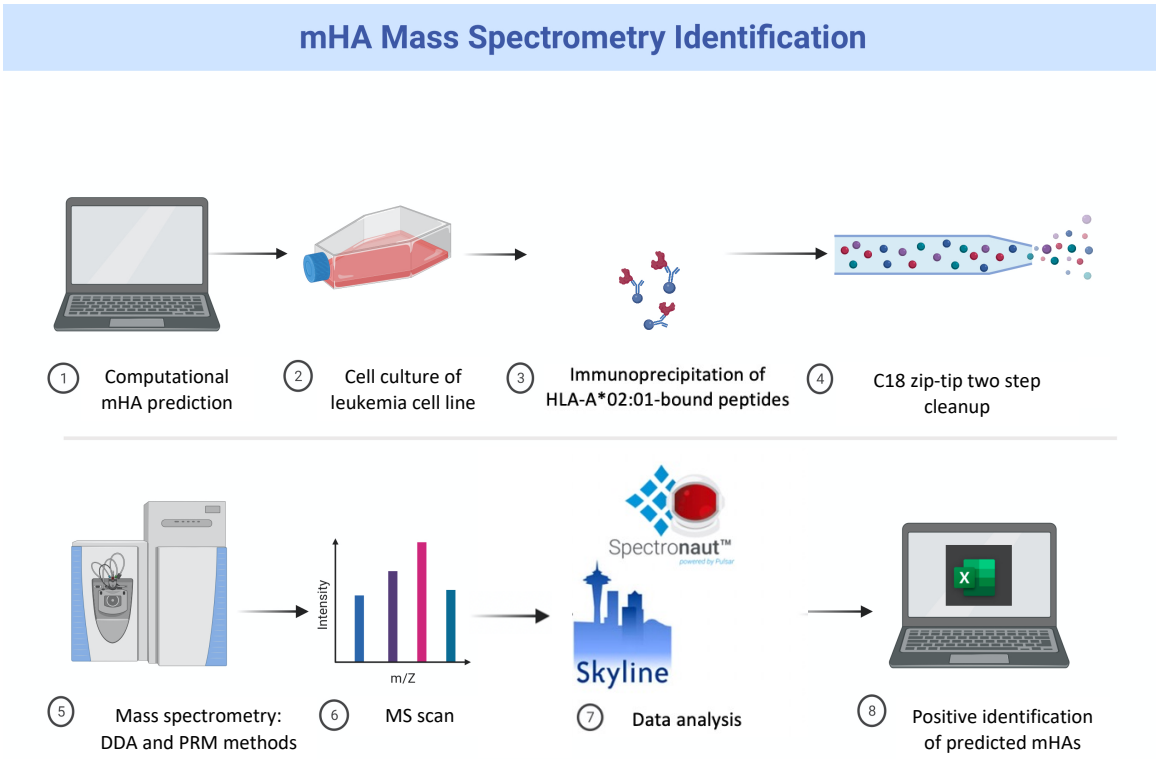


Figure 3.3: Overall mHA prediction and MS validation workflow. Figure generated by Marisa Diiorio.

As a standard for assessing if MS results successfully identified peptides, we assessed the presence of a known peptide that should be present in these samples called CG1 (amino acid sequence FLLPTGAEA). This is a cathepsin G-derived HLA-A*02:01-presented peptide that we have seen presence of in every successful MS run in past work, due to cathepsin G's high expression levels in AML cells[243].

(3.3) Results

DDA Results

We first assessed the difference in 30% vs 50% acetonitrile for IP sample elution prior to C18 cleanup. In 11 samples tested with 30% elutions, our standard peptide CG1 was detected in nine, while with 50% CG1 gave a strong signal in all samples. We therefore selected 50% acetonitrile for our final method. The total number of peptides identified across 11 samples using this method was 5919, with 57.13% peptides between 8-12 amino acids long, indicating some impurities in immunoprecipitation as this sample should have contained only HLA class I-binding peptides due to the HLA-A*02-specific antibody that was used. Only 2.84% of peptides detected were over 25 amino acids, indicating that many of the peptides detected were between 13-24 amino acids long, making them likely to be class II binding peptides that were aberrantly included in the IP samples[244]. Overlap between the 11 samples was relatively low, which is to be expected due to the complexity and variability of immunopeptidomics samples and conforms to previous observations about low consistency of immunoprecipitated peptides from HLA (Figure 3.4)[245–247]. Samples 10 and 11 contained the highest number of peptides identified per sample, with 1605 and 558 peptides identified respectively. This is likely due to the protocol adjustment made for these two samples in which pipette tips were cut, probably leading to less shearing of peptides from the resin when pipetting.

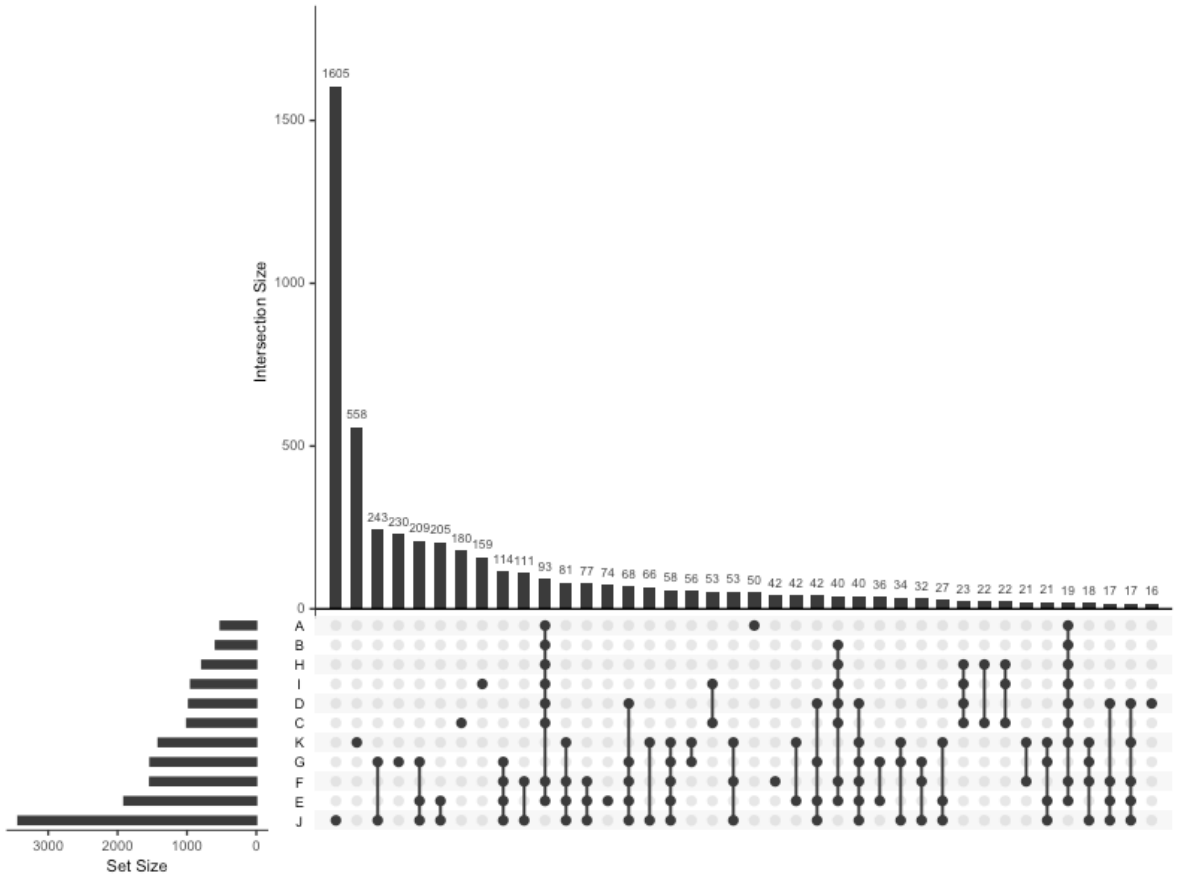


Figure 3.4: UpSetR plot indicating overlap between 11 MS samples analyzed by DDA.

While CG1 was detected in eleven of eleven samples, consistent with its remarkably high expression and consistent presence that we have observed in prior experiments, we detected three novel peptides in one sample. Peptide NLLRVFQAV, predicted from the 101 DRP dataset, was identified with high confidence as determined by the ProteomeDiscoverer software, with high confidence indicating false discovery rate (FDR) below a cutoff of 0.01. Peptide NLLRVFQA, a length variant of the previous peptide, was identified with medium confidence, indicating FDR below a cutoff of 0.05. Peptide ILSKEDFETSL, predicted from the 3231 DRP dataset, was detected with

medium confidence. They were identified with m/z of 530.3, 480.8, and 641.3 respectively, and retention times of 62.6, 50.1, and 48.1 minutes respectively.

PRM Results

We then sought to confirm the peptides we identified in samples via DDA with targeted PRM analysis. We used the same eleven samples eluted with 50% acetonitrile as in the DDA runs. We made a targeted method looking exclusively for the three peptides identified via DDA as well as CG1. We identified CG1 in all eleven samples as expected. NLLRVFQA was identified in all samples as well via PRM, illustrating the increased sensitivity of targeted methods over shotgun methods. Additionally, NLLRVFQAV and ILSKEDFETSL were identified in one sample each. Therefore, all three novel peptides identified via DDA were confirmed via PRM. Example spectra for peptide standards vs IP sample are shown below (Figure 3.5).

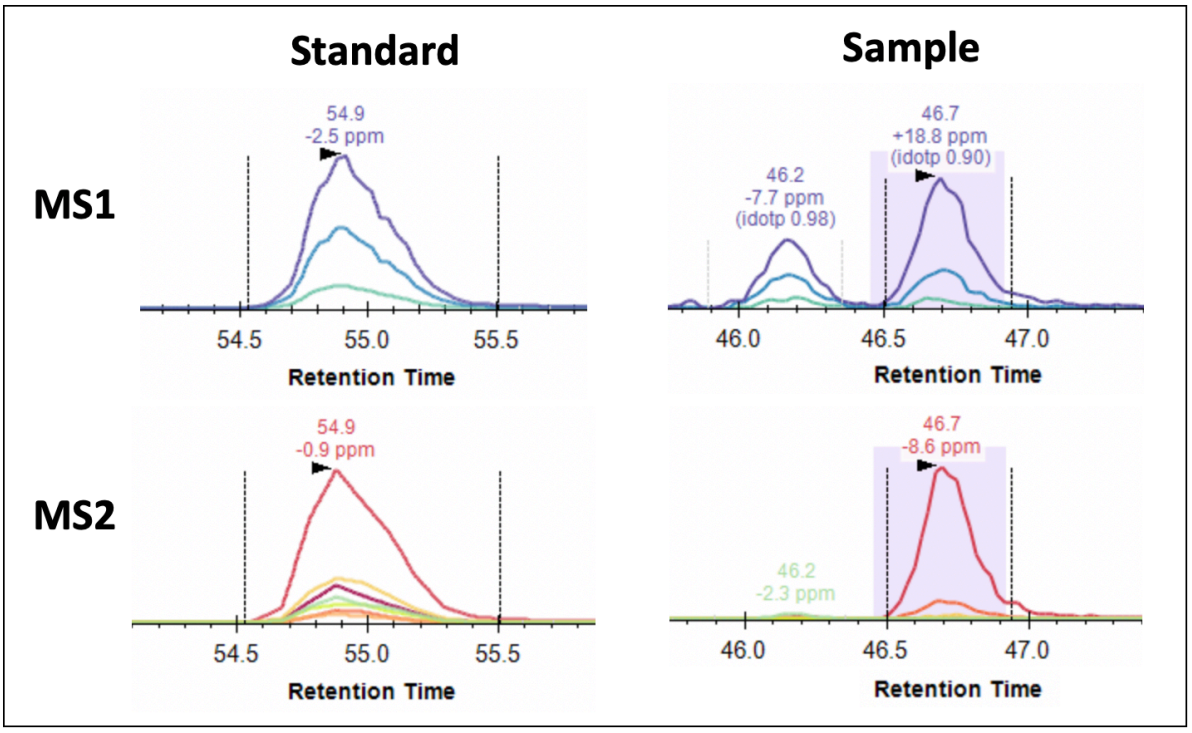


Figure 3.5: Example PRM MS1 and MS2 spectra for one novel identified peptide, NLLRVFQA.

We consider these three novel peptides to be two novel mHAs total since two were length variants derived from the same SNP. We named these two novel mHAs UNC-DPEP1-Q (NLLRVFQA(V)) and UNC-TXNDC2-T (ILSKEDFETSL).

(3.4) Discussion

Using a custom MS workflow including shotgun DDA MS for initial identification of candidate peptides and targeted PRM MS for validation of candidates, we identified two novel mHAs. In the methods development process, we created a new IP method for eluting class I HLA-bound peptides from cell lines that gave better results including higher numbers of identified peptides and more frequent identification of the peptide standard CG1 than previously reported IP protocols. We also assessed different IP cleanup methods and different MS methods, and report here that elution with 50% acetonitrile, C18 cleanup, and DDA and PRM mass spectrometry using the specifications noted here gave the best results. Immunopeptidomics is an exceedingly difficult technique due to the small size of the desired targets and difficulty of identifying specific amino acid sequences versus their single amino acid variants in peptides as small as 8 amino acids long. Despite these challenges, we successfully developed a method that enabled validation of predicted mHAs, adding two new GvL mHAs to the previously known 12 GvL mHAs in the field. Though these mHAs were predicted from different sources, with UNC-DPEP1-Q predicted from the 101 patient dataset and UNC-TXNDC2-T from the 3231 patient dataset, both were found In a large number of patients within the larger dataset. We screened this set for both mHAs as it has a larger number of DRPs and is more representative of the general population of alloHCT DRPs. Both mHAs are specific for HLA-A*02:01, so this restricts the pool of applicable DRPs to

1569 DRPs in DISCOVeRY-BMT with this HLA allele. Of these DRPs, 20.6% have the appropriate mismatches for UNC-TXNDC2-T to be applicable, and 23.5% have the appropriate mismatches for UNC-DPEP1-Q to be applicable. These two mHAs have a higher known applicability than any previously known GvL mHAs other than ACC1Y, and UNC-GRK4-V for UNC-DPEP1-Q only. Overall, though this mHA validation pathway was successful, it was extremely labor-intensive and time-consuming, and for future MS work we moved to a working with a company called Complete Omics Inc that has a more established and less time and resource prohibitive pathway for immunopeptidomics work, described in the next section.

CHAPTER 4: Mass Spectrometry Validation of Shared Minor Histocompatibility Antigens from DISCOVeRY-BMT[3]

(4.1) Introduction

In this chapter, we sought to validate shared GvL mHAs predicted from the DISCOVeRY-BMT dataset using targeted heavy-labeled peptide standards via the Complete Omics method. Heavy labeled peptide standards were cost-prohibitive to produce for in-house mHA validation via the Proteomics Core, but enable enhanced detection of peptides in immunopeptidomics MS assays and other MS work where there are technical difficulties for detection of the analyte of interest[248–250]. Therefore, we transitioned to targeted peptide validation with through a company with more cost-effective access to heavy-labeled peptides, extensive experience with immunopeptidomics, and a unique in-house method for peptide immunoprecipitation. We validated 24 novel GvL mHAs, increasing the number of total known GvL mHAs by 200% compared to the 12 reported in the literature, and bringing our total contribution of 26 novel GvL mHAs to the field up to 217% compared to the prior 12 known antigens.

[3] Portions of this chapter are published in *Blood Advances*. Citation: Olsen KS, Jadi O, Dexheimer S, Bortone S, Vensko SP, Bennett SN, Tang H, Diiorio M, Saran T, Dingfelder D, Zhu Q, Wang Y, Haiman CA, Pooler L, Sheng X, Webb A, Pasquini MC, McCarthy PL, Spellman SR, Weimer ET, Hahn T, Sucheston-Campbell LE, Armistead PM*, Vincent B*. Shared graft-vs-leukemia minor histocompatibility antigens in DISCOVeRY-BMT. 2022. *Blood Adv*; bloodadvances.2022008863. I designed and performed experiments, performed computational mHA prioritization, interpreted experimental results, performed statistical testing, generated figures, prepared the manuscript, and made revisions.

(4.2) Methods

Computational Methods

mHAs selected for validation in this chapter were predicted from the DISCOVeRY-BMT dataset exclusively, as this dataset is most broadly applicable to the general population of alloHCT recipients. mHA validation was performed for HLA-A*02:01, HLA-B*35:01, and HLA-C*07:02, the representative HLA alleles for A, B, and C selected to give high coverage of a variety of ethnic groups as discussed in Chapter 2. We chose to search for 40 predicted peptides per MS run as this was a feasible number from a cost perspective and from an MS methods development perspective. One MS run each was performed for HLA-B*35:01 and HLA-C*07:02, for a total of 40 predicted peptides assessed per allele. Two MS runs were performed for HLA-A*02:01. New RNAseq data was generated for the cell line utilized for A*02:01 between the two MS runs, slightly shifting the top 40 peptides of interest for this allele. Therefore, we searched two overlapping but non-identical sets of peptides for HLA-A*02:01, totaling 67 searched peptides.

The list of mHAs for validation for each HLA allele was selected by compiling a list of the mHAs selected by the greedy algorithm to give 100% population coverage, followed by mHAs that were not selected by the greedy algorithm in descending order of population frequency within DISCOVeRY-BMT DRPs with the appropriate HLA allele. This list was then filtered to exclude peptides with zero coverage of the source gene by RNAseq in the cell line being utilized for validation in that HLA allele. The top 40 peptides on the remaining list were then selected for each search.

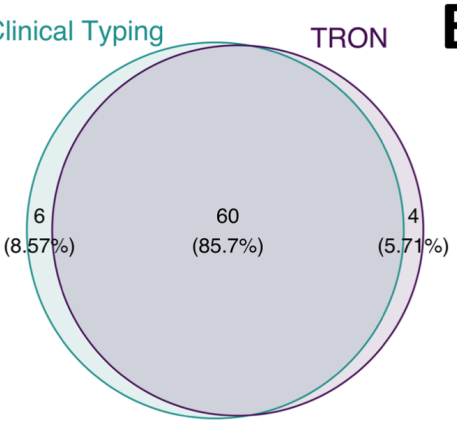
Cell Line RNAseq

RNAseq was performed to assess gene expression in multiple AML cell lines. RNA was isolated from 5 million cells per cell line using the Qiagen AllPrep DNA/RNA extraction kits (Qiagen, 80004). Nucleic acid quantification and quality detection were performed and RNA sequencing libraries were generated via the Illumina Stranded mRNA library preparation protocol. Paired-end sequencing was run on a NovaSeq SP 2x150 with the NovaSeq 6000 SP Reagent Kit v1.5 (NovaSeq, 20028400) and aligned to reference genome GRCh37, as SNP locations were called using this reference genome. Differential gene expression analysis was performed using the DESeq2 R package[251]. Expression levels were calculated relative to the average expression in the five tested cell lines. Several cell lines for each of the three studied HLA alleles were examined and one selected for each allele based on ease of expansion to needed number of cells for mass spectrometry.

Cell Lines

The AML cell lines selected for MS were U937A2, the U937 cell line stably transfected to express HLA-A*02:01 as described in the previous chapter, NB4, which endogenously expresses HLA-B*35:01, and MONOMAC1, which endogenously expresses HLA-C*07:02[252]. Cell line HLA expression data was downloaded from the TRON cell lines portal and validated by the Clinical HLA Typing Laboratory at the University of North Carolina Hospitals, with differences as reported (Figure 4.1)[252,253].

A



B

Caller	Cell Line	Allele
Clinical Typing	OCIM2	A*01:01
Clinical Typing	U937	A*31:01
Clinical Typing	Nomo1	B*55:02
Clinical Typing	OCI-AML3	B*53:XX
Clinical Typing	OCIM1	C*02:10
Clinical Typing	OCIM2	C*07:01
TRON	HL60	A*24:15
TRON	OCI-AML3	B*53:01
TRON	OCIM1	C*02:02
TRON	OCIM2	C*07:06

C

Cell Line	A Alleles	B Alleles	C Alleles
BV173	02:01, 30:01	15:10, 18:01	03:04, 12:03
F36P	24:02, 33:03	40:06, 58:01	03:02, 08:01
HL60	01:01	57:01	06:02
MONOMAC1	03:01	07:02, 51:01	07:02, 15:02
NB4	11:01	35:01, 40:01	03:04, 04:01
Nomo1	24:02, 26:03	54:01	01:02, 07:02
OCI-AML2	02:01	15:01, 18:01	03:03, 07:01
OCI-AML3	02:01, 23:01	44:02	04:01, 05:01
OCIM1	02:01, 03:01	14:02, 15:03	08:02
OCIM2	02:01	08:01, 44:02	05:01
U937	03:01	18:01, 51:01	01:02, 07:01
U937A2	02:01, 03:01	18:01, 51:01	01:02, 07:01

Figure 4.1: HLA typing for AML cell lines. (A) shows Venn diagram of number of typed HLA alleles agreed upon and disagreed upon by HLA typing on TRON database versus performed by UNC Hospitals Clinical Laboratories. (B) shows alleles called by one source only. (C) shows alleles agreed upon by both HLA typing sources.

Where discrepancies were found between the HLA haplotype on TRON and clinical typing, the clinical typing result was used. Cell lines were maintained in culture with RPMI 1640 +10% FBS +1% penicillin-streptomycin +1% L-glutamine[243].

Immunoprecipitation and Mass Spectrometry

Cell lines were expanded to 1-5x10^8 per sample. Cells were centrifuged and washed with PBS, followed by treatment with 1x cOmplete Mini EDTA-free Protease Inhibitor Cocktail tables prepared in PBS (Roche, 11836170001). Cells were

centrifuged, supernatant was removed, and cell pellets were snap frozen in liquid nitrogen and placed at -80C. Frozen pellets were sent to Complete Omics Inc for immunoprecipitation and antigen validation and quantification by mass spectrometry through Valid-Neo® platform[254]. Pellets were processed into single-cell frozen powder and then lysed. Peptide-HLA complexes were immunoprecipitated using Valid-NEO® neoantigen enrichment column pre-loaded with anti-human HLA-A, B, C antibody clone W6/32 (Bio-X-Cell). After elution, dissociation, filtration and clean up, peptides were lyophilized before further analysis. Transition parameters for each epitope peptide were examined and curated through Valid-NEO® method builder, an AI-based biostatistical pipeline. Ions with excessive noise due to co-elution with impurities were further optimized or removed. To boost detectability, a series of computational recursive optimizations of significant ions were conducted. Each mHA sequence was individually detected and quantified in a high-throughput manner through a Valid-NEO® modified mass spectrometer.

SNP Typing and Sanger Sequencing

PCR primers for the SNP loci encoding the novel mHAs that were validated by mass spectrometry were designed using Primer3 and Geneious. Sequencing primer T7 pro sequence was added to the 5' end of each forward primer for ease of sequencing. Genomic DNA was extracted from frozen PBMC pellets from alloSCT recipient pretransplant and post-transplant D90 peripheral blood draw samples from samples expression HLA-A*02:01, HLA-B*35:01, or HLA-C*07:02, obtained from the UNC Lineberger Comprehensive Cancer Center Tissue Procurement Facility, using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, 69504) and purified using the Monarch

DNA Cleanup Kit (NEB, T1030S). D90 samples were taken to represent alloSCT donor genetic information, as patient samples were selected for successful engraftment, meaning that at this time point the sample is >95% donor as transplant has engrafted and hematopoiesis is fully donor in origin. PCR amplification was performed using the FastStart High Fidelity PCR System (Roche, 3553400001). Samples were directly sent for Sanger Sequencing by Eton Biosciences using T7pro sequencing primer.

Outcomes Analyses

Patient data were separated out for the HLA alleles corresponding to novel mHAs. Presence or absence of corresponding SNP alleles in each DRP was tabulated for each mHA. Competing risk regression was performed to assess associations of each mHA with relapse at 1 year via the cmprsk R package. Datasets for each HLA allele were then combined and the total number of novel mHAs across all three HLA alleles was tabulated per DRP. Cox proportional hazard ratio was calculated to assess association of total number of mHAs with relapse at 1 year via the survival R package.

Source Gene Expression Analysis for ALL

RNAseq read quantifications for 110 pediatric ALL patients were accessed from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative Phase 3, phs000218 (https://ocg.cancer.gov/programs/target). The data used for this analysis are available at https://portal.gdc.cancer.gov/projects. We analyzed expression of source genes of mHAs validated in this chapter. Available data were centered so no direct comparisons of TPM were possible, so z-scores for mean expression values across all ALL patients were calculated and compared to z-scores of

the TCGA_AML gene expression data that were utilized for mHA prediction and categorization.

(4.3) Results

We utilized RNAseq to identify gene expression in AML cell lines both to identify which mHAs to select for validation for a particular HLA allele based on which mHA source genes had nonzero expression in the cell line utilized for that allele. As part of this RNAseq data analysis, we identified the top set of genes that were differentially upregulated or downregulated compared to the mean expression across all the cell lines identified (Figure 4.2). Overall, we found that source gene expression for our mHAs was generally fairly similar across cell lines.

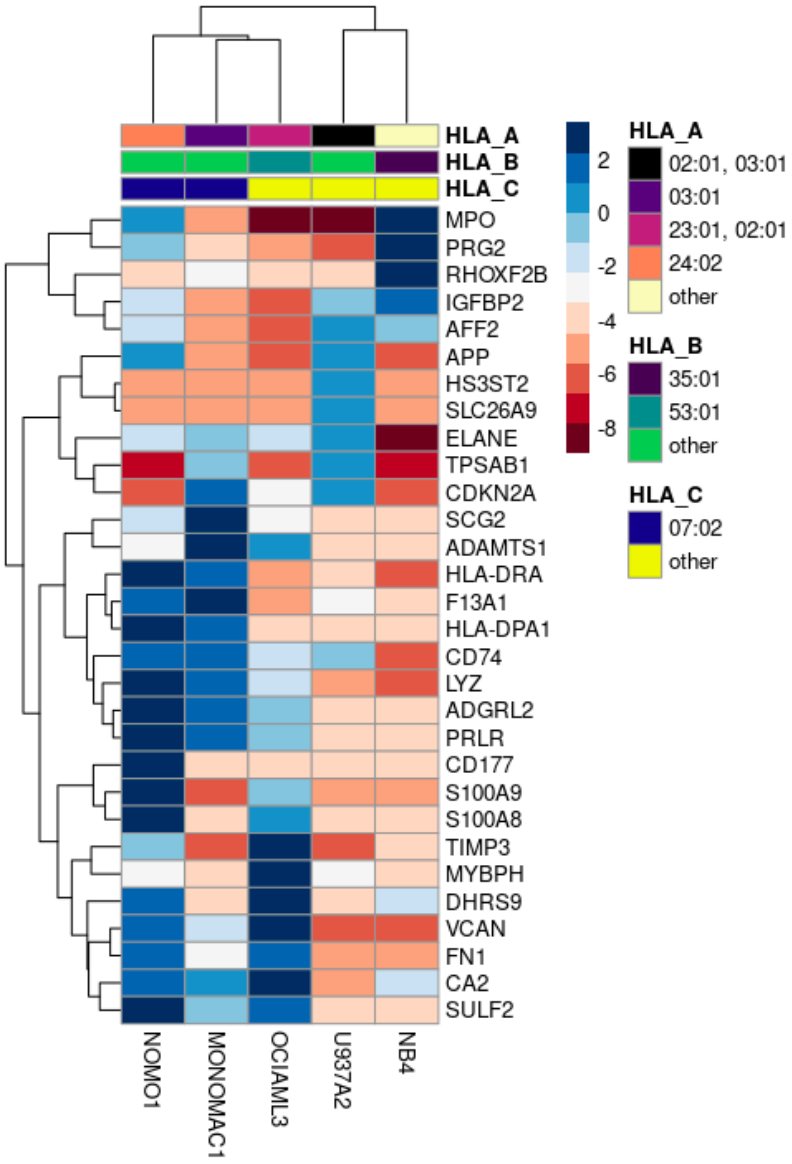


Figure 4.2: RNAseq Differential Gene Expression analysis and HLA alleles expressed by AML cell lines evaluated for use. Top 30 genes that are differentially expressed compared to the mean across the 5 cell lines are shown.

We employed mass spectrometry to validate HLA presentation of predicted GvL mHAs. Of the 67 searched peptides for HLA-A*02:01 across two U937A2 cell line samples, we positively identified 17 peptides. Of the 40 searched for HLA-B*35:01 using an NB4 cell line we identified three peptides, and of the 40 searched for C*07:02 using a MONOMAC1 cell line we identified five peptides. Representative spectra are shown

for a heavy labeled peptide standard and endogenous identified peptide from an immunoprecipitated NB4 cell sample (Figure 4.3).

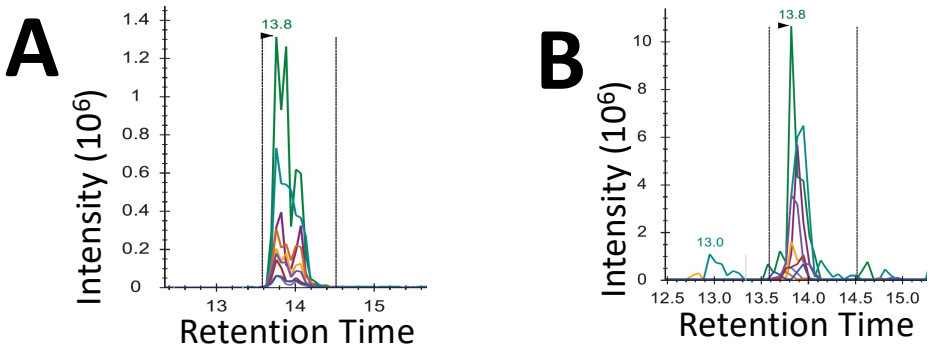


Figure 4.3: Example mass spectra of an MS validated novel mHA. (A) shows representative spectra for heavy-labeled peptide standard for HLA C*07:02-binding mHA LPAAYHHH. (B) shows endogenous LPAAYHHH peptide identified from immunoprecipitated peptide sample from cell line MONOMAC1.

From the list of 17 validated peptides for HLA-A*02:01, peptide VLDIEQFSV is also known as UNC-GRK4-V and was previously identified by our group as a GvL mHA using the U937A2 cell line[62,165]. Mass spectrometry analysis was blinded to the peptide's status as previously identified. As this peptide is previously known, a total of sixteen novel HLA-A*02:01 binding mHAs were discovered. These sixteen novel HLA-A*02:01-binding mHAs cumulatively cover 98.8% of HLA-A*02:01-positive patients in the DISCOVeRY-BMT dataset, with individual peptide population frequencies between 21.1% and 28.3% (Figure 4.4A). The three novel HLA-B*35:01 binding mHAs cover 60.7% of the HLA-B*35:01 positive DISCOVeRY-BMT population with population frequencies of 26.0-27.6% (Figure 4.4B). The five novel HLA-C*07:02 binding mHAs give cumulative HLA-C*07:02 positive DISCOVeRY-BMT patient coverage of 78.9%,

with independent frequencies of 24.4-26.7% (Figure 4.4C). Characteristics of all novel mHAs are shown (Table 4.1)[211,255].

A

Validated Peptides	Noncumulative Population Coverage	Cumulative Population Coverage
KVAVAMLTV	28.3	28.3
ALYPFLGIL	26.4	47.7
FLAAASARGI	26.5	62.0
AVLDEAVV	26.4	72.5
ALARGGGQLPL	24.3	79.5
FLRCMLTI	25.0	85.0
SLICRQLGSA	23.9	89.1
ALSQVPSPL	23.6	92.0
KLIVQPNTHL	23.6	94.1
FLSSANEHL	25.6	95.6
AQYTSVYGA	25.3	96.7
KVSFFVPRL	22.4	97.6
RLHVGCDEV	24.0	98.1
LLGDDDVADGL	26.1	98.3
VVFGQAPPL	21.7	98.6
WLLEKLQEQL	21.1	98.8

B

Validated Peptides	Noncumulative Population Coverage	Cumulative Population Coverage
HPDGWSHRGIF	27.4	27.4
LPAAYHHH	26.6	48.2
LPAMPRDY	26.0	60.7

C

Validated Peptides	Noncumulative Population Coverage	Cumulative Population Coverage
GRLLSTVIRTL	26.7	26.7
RRQLDGRVLL	25.4	46.1
RYADRPGRRF	25.9	60.6
FLQPNVRGPLF	25.8	71.1
YRLAQDYLQY	24.4	78.9

Figure 4.4: Population frequencies of 24 novel shared GvL mHAs. (A) shows all novel identified peptides from cell line U937A2 sample. “Noncumulative population coverage” shows percentage of DRPs expressing HLA-A*02:01 within the DISCOVeRY-BMT dataset where the recipient expresses the mHA allele and donor does not. “Cumulative population coverage” shows output from the maximal set coverage solution calculating total population coverage by each peptide and the ones preceding it, with a total of 98.8% population coverage by the ten peptides. (B) shows all identified peptides from cell line NB4 sample, with 60.7% cumulative coverage of DRPs expressing HLA-B*35:01 within the dataset by the three peptides. (C) shows all identified peptides from cell line MONOMAC1, with 78.9% cumulative coverage of HLA-C*07:02-expressing DRPs within the dataset.

Table 4.1: Novel mHA characteristics. 16 novel GvL mHAs that bind HLA-A*02:01, three that bind B*35:01, and five that bind C*07:02 and were validated by mass spectrometry are shown.

mHA name	mHA	HLA Allele	Gene	Chromosome	rsID of SNP	Donor Amino Acid	Recipient Amino Acid	Major Allele	Minor Allele	MAF in TOPMED	MAF in ALFA	MAF in 1000 Genomes	Peptide Length	Binding Affinity	Frequency in DISCOVeRY-BMT Patients with Corresponding HLA
UNC-IQCE-V	AVLDEAVV	A*02:01	IQCE	11	rs2293404	A	V	C	T	0.35	0.29	0.41	8	412.7421	26.4
UNC-GLRX3-S	FLSSANEHL	A*02:01	GLRX3	10	rs2274217	P	S	C	T	0.21	0.25	0.19	9	14.9364	25.6
UNC-SLC25A37-V	AQYTSVYGA	A*02:01	SLC25A37	8	rs2942194	I	V	A	G	0.21	0.26	0.17	9	45.2204	25.3
UNC-ARHGEF18-Q	SLICRQLGSA	A*02:01	ARHGEF18	19	rs2287918	R	Q	G	A	0.19	0.24	0.17	10	367.3796	23.9
UNC-DPP3-H	KLIVQPNTHL	A*02:01	DPP3	11	rs2305535	R	H	G	A	0.19	0.22	0.21	10	218.2962	23.6
UNC-HEXDC-V	RLHVGCDEV	A*02:01	HEXDC	17	rs4789773	I	V	A	G	0.45	0.37	0.56	9	357.9789	24
UNC-TOP1MT-W	WLLEKLQEQL	A*02:01	TOP1MT	8	rs2293925	R	W	C	T	0.36	0.43	0.46	10	8.8962	21.1
UNC-USP4-V	KVSFFVPRL	A*02:01	USP4	3	rs35446411	L	V	T	G	0.13	0.16	0.09	9	445.1234	22.4
UNC-AHRR-A	VVFGQAPPL	A*02:01	AHRR	5	rs2292596	P	A	C	G	0.30	0.35	0.38	9	311.398	21.7
UNC-FPR1-K	KVAVAMLTV	A*02:01	FPR1	19	rs1042229	N	K	T	G	-	0.45	0.37	9	178.6526	28.3
UNC-FLT3-G	ALARGGGQLPL	A*02:01	FLT3	13	rs12872889	D	G	A	G	0.35	0.31	0.37	11	257.8616	24.3
UNC-GDPD5-A	ALSQVPSPL	A*02:01	GDPD5	11	rs571353	T	A	A	G	0.34	0.28	0.33	9	94.9686	23.6
UNC-SLC26A8-M	FLRCMLTI	A*02:01	SLC26A8	6	rs743923	V	M	G	A	0.30	0.25	0.26	8	100.0366	25
UNC-FPGS-I	FLAAASARGI	A*02:01	FPGS	9	rs10760502	V	I	C	T	0.28	0.35	0.22	10	23.9163	26.5
UNC-NDUFAF1-L	ALYPFLGIL	A*02:01	NDUFAF1	15	rs3204853	R	L	C	A	0.17	0.24	0.12	9	164.2741	26.4
UNC-WDR62-L	LLGDDDVADGL	A*02:01	WDR62	19	rs2285745	S	L	C	T	0.32	0.35	0.35	11	182.9111	26.1
UNC-POLL-W	HPDGWSHRGIF	B*35:01	POLL	10	rs3730477	R	W	C	T	0.16	0.21	0.10	11	31.3362	27.4
UNC-HLX-P	LPAAYHHH	B*35:01	HLX	1	rs12141189	S	P	T	C	0.23	0.23	0.21	8	285.0737	26.6
UNC-NEK4-A	LPAMPRDY	B*35:01	NEK4	3	rs1029871	P	A	G	C	0.33	0.39	0.31	8	39.4681	26
UNC-MARCH2-T	GRLLSTVIRTL	C*07:02	MARCH2	19	rs1133893	A	T	C	T	0.24	0.32	0.20	11	175.8924	26.7
UNC-GAA-R	RRQLDGRVLL	C*07:02	GAA	17	rs1042395	H	R	A	G	0.36	0.28	0.40	10	375.2756	25.4
UNC-RNASE3-R	RYADRPGRRF	C*07:02	RNASE3	14	rs2073342	T	R	C	G	0.36	0.29	0.36	10	147.8419	25.9
UNC-SNX19-V	FLQPNVRGPLF	C*07:02	SNX19	11	rs3751037	L	V	G	C	0.30	0.29	0.27	11	161.7018	25.8
UNC-BCL2A1-Y	YRLAQDYLQY	C*07:02	BCL2A1	15	rs1138357	C	Y	G	A	0.28	0.26	0.35	10	188.9083	24.4

One interesting finding to note is that one of our novel mHAs, UNC-BCL2A1-Y, is derived from the same SNP as the previously identified mHA ACC1Y described in 2013 by the Takahashi group as one of the first identified and validated mHAs[147]. These two mHAs overlap by five amino acids, and ACC1Y binds HLA-A*24:02, while our novel UNC-BCL2A1-Y binds HLA-C*07:02. Identification of both of the previously discovered GvL mHA (UNC-GRK4-V) and a GvL mHA overlapping with another mHA was reassuring that our results are consistent with work previously done in the field.

Next, we were interested in whether our identified mHAs gave high cumulative population frequencies due to chance or because our mHA selection method was such that any discovered mHAs would result in high cumulative coverage. Essentially, we wanted to identify whether it was possible that a different set of the same number of identified antigens from our list of predicted shared GvL mHAs would have given much lower cumulative coverage because all the mHAs were found in a more overlapping set of DRPs, or whether any set of this size from our list would result in high population coverage. To evaluate the generalizability of our discovery process, we calculated the range of cumulative coverage that would be obtained with a subset of the number of peptides that we validated from the searched lists. For each HLA allele, 1000 random sets of peptides were selected from the searched peptide list and cumulative coverage by each set was calculated. The range of cumulative coverage by the 1000 random sets of 16 HLA-A*02:01 peptides was 97.4-99.7%, by the 1000 random sets of three HLA-B*35:01 peptides was 42.8-66.4%, and by the 1000 random sets of five HLA-C*07:02 peptides was 65.4-80.7% (Figure 4.5). This indicates that our selection methods for mHA lists to validate results in exclusively sets of peptides that would give high

cumulative coverage, regardless of which individual peptides are validated within each set.

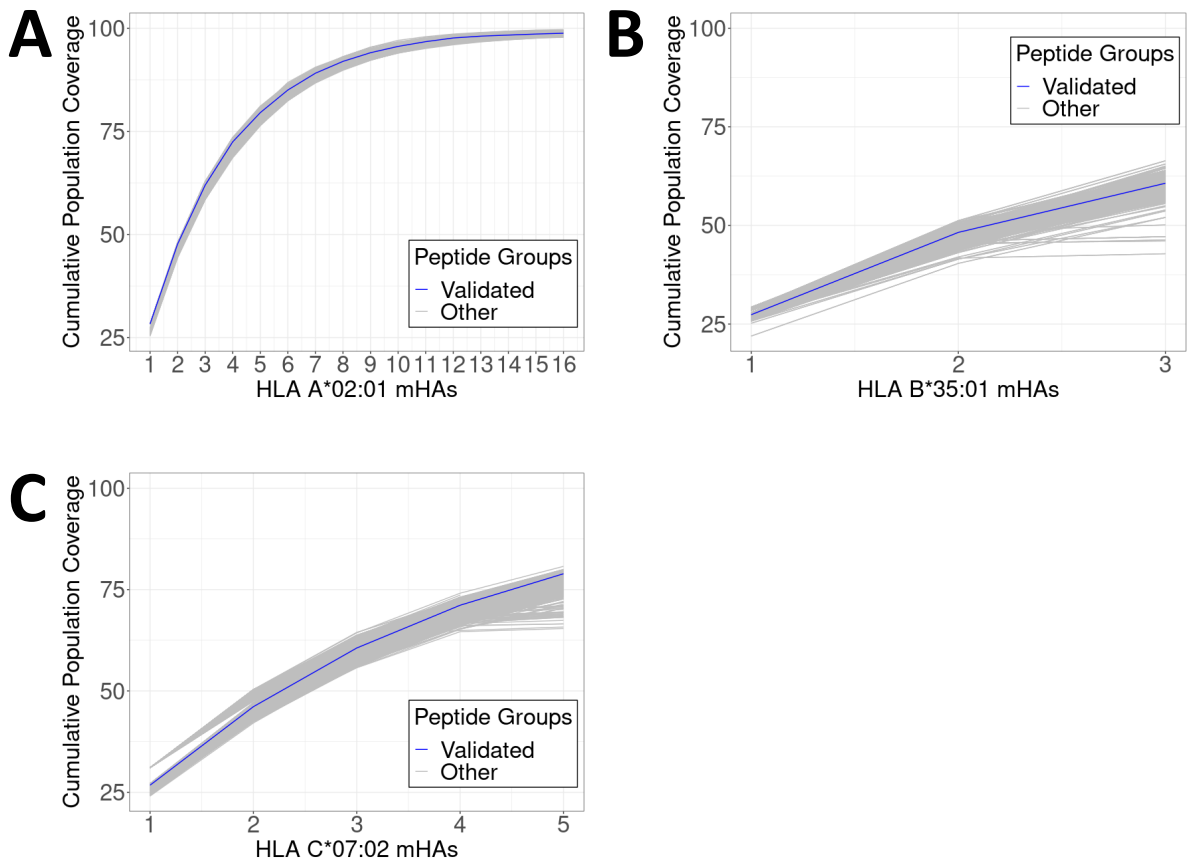


Figure 4.5: Cumulative population coverage of simulated mHA sets. (A) shows cumulative coverage by the 16 novel confirmed HLA-A*02:01-binding mHAs and 1000 simulated sets of 16 peptides from the set of mHAs searched by mass spectrometry. Cumulative coverage by confirmed peptides is shown in blue, while each simulated run is shown as an individual gray line. (B) shows cumulative coverage for the three confirmed HLA-B*35:01-binding mHAs and 1000 simulated sets of three peptides. (C) shows cumulative coverage for the five confirmed HLA-C*07:02-binding mHAs and 1000 simulated sets of five peptides.

We genotyped seven DRPs from the UNC Lineberger Cancer Center Tissue Procurement Facility expressing HLA-A*02:01, one expressing HLA-B*35:01, and four expressing HLA-C*07:02 for the majority of the novel mHAs for the corresponding HLA alleles (Figure 4.6A-C). We found appropriate minor antigen mismatches for potential

utilization of these mHAs in 58% of the genotyped DRPs, highlighting their utility for future work. This does not align perfectly with predicted coverage of DISCOVeRY-BMT patients with these mHAs, but it is likely explained by the small patient count and different patient population. However, most of the patients genotyped could use treatments targeting these mHAs. We also genotyped the seven HLA-A*02:01 positive DRPs for the previously known GvL mHAs HA-1 and UTA2-1 and discovered they were targetable in 0% of these DRPs (Figure 4.6A). As illustrated by these example frequencies from a relevant DRP population, our newly discovered shared mHAs will be applicable to many more patients than previously known mHAs. As discussed in Chapter 1, our novel mHAs also have much more favorable minor allele frequencies within larger general populations than previously known mHAs as shown in Table 1.1, explaining the higher frequency of targetability due to Hardy-Weinberg Equilibrium.

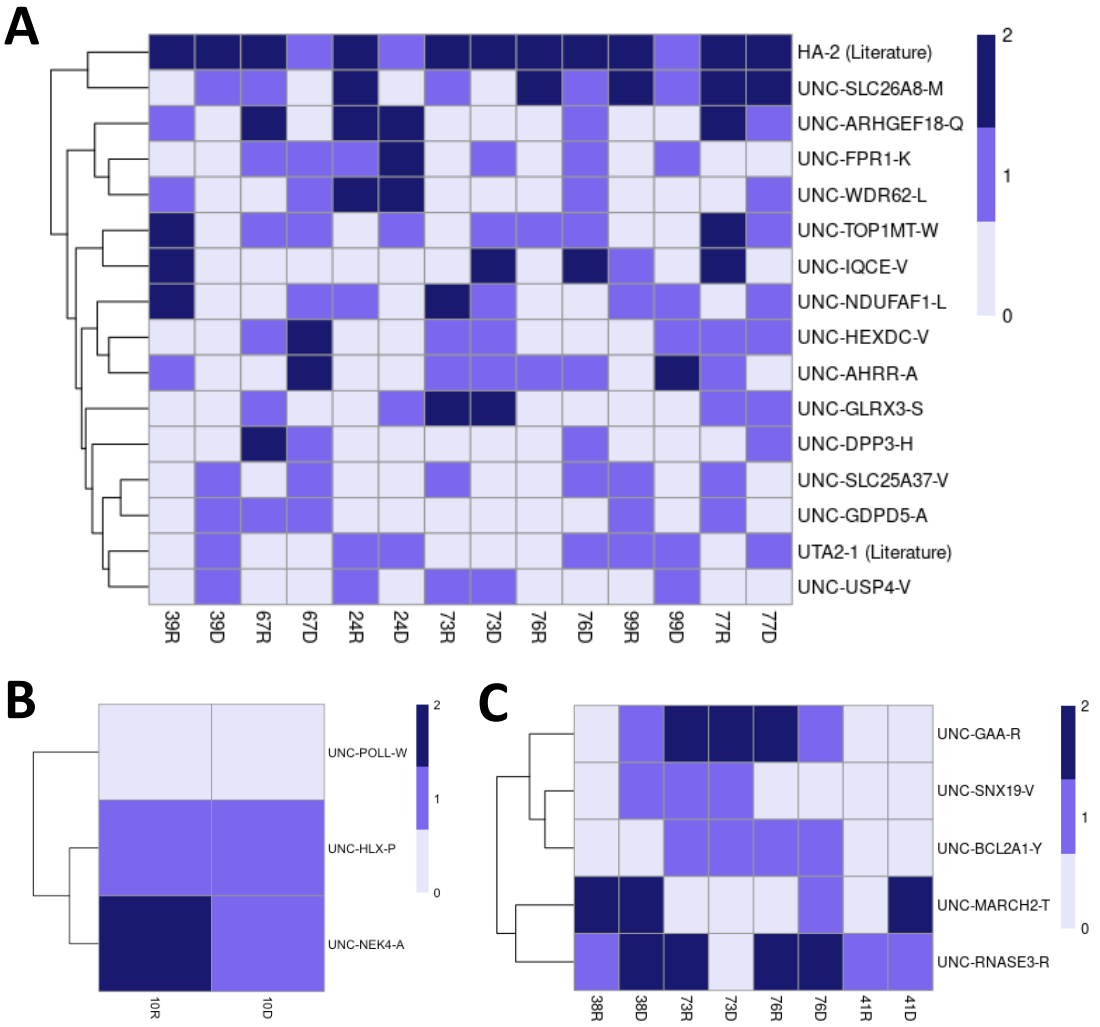
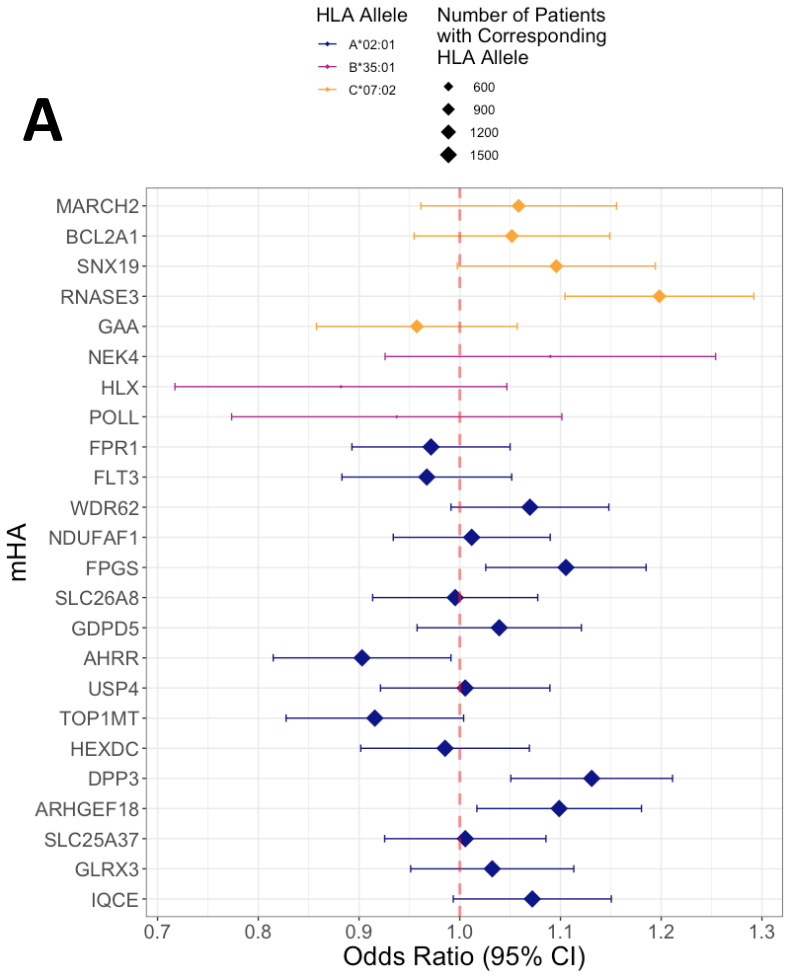


Figure 4.6: Assessment of presence or absence of SNPs encoding validated predicted GvL mHAs. (A) shows number of copies of alleles encoding fourteen novel HLA-A*02:01 mHAs in seven DRPs from alloSCT for AML. Two novel mHAs were excluded due to difficulty of appropriate primer design. Two previously known HLA-A*02:01-binding mHAs from the literature, HA-2 and UTA2-1, are also shown. (B) shows number of copies of alleles encoding three novel HLA-B*35:01 mHAs in one DRP. (C) shows number of copies of alleles encoding five novel HLA-C*07:02 mHAs in four DRPs.

In outcomes assessments with our novel validated mHAs, we found no significant association between any individual mHA and relapse at 1 year (Figure 4.7A/B). We also saw no significant correlation between total number of novel mHAs

per DRP and 1 year relapse via Cox proportional hazards ratio (Figure 4.7C). This finding is consistent with the hypothesis that alloimmunity is driven by T cell responses to multiple antigens rather than single antigens. While some antigens may be immunodominant and drive a disproportionately large share of the immune response, there is still a complex combination of antigens driving immune responses in any given patient[256,257]. There is also no guarantee that a shared GvL mHA would also be an endogenously immunodominant mHA. Few individual mHAs have shown direct associations with positive outcomes such as overall survival or relapse-free survival. Most studies examining specific mHAs have found no outcomes associations, though one study has shown positive outcomes for DRPs with >1 known mHA difference versus 0[258,259]. Even the most well-studied mHA, HA-1, has shown conflicting results on associations with outcomes, with two studies showing no association with relapse but two showing an association with lower leukemia relapse rates in the context of CML[150,260–262]. Associations of individual mHAs with outcomes are much more common on the GvHD side, with a variety of individual mHAs shown to correlate with clinical outcomes[141,213,263–266]. Therefore, we were overall unsurprised to find no significant outcomes associations with our individual mHAs, though one HLA-C*07:02-binding mHA, UNC-RNASE3-R, was close to significance with a p value of 0.054 and warrants future investigation in other datasets (Figure 4.7).

A



B

mHA	Odds Ratio	95% CI	p value
MARCH2	1.06	0.96,1.16	0.560
BCL2A1	1.05	0.95,1.15	0.600
SNX19	1.10	1.00,1.19	0.350
RNASE3	1.20	1.10,1.29	0.054
GAA	0.96	0.86,1.06	0.660
NEK4	1.09	0.93,1.25	0.600
HLX	0.88	0.72,1.05	0.450
POLL	0.94	0.77,1.10	0.690
FPR1	0.97	0.89,1.05	0.710
FLT3	0.97	0.88,1.05	0.690
WDR62	1.07	0.99,1.15	0.390
NDUFAF1	1.01	0.93,1.09	0.880
FPGS	1.11	1.03,1.18	0.210
SLC26A8	1.00	0.91,1.08	0.960
GDPD5	1.04	0.96,1.12	0.640
AHRR	0.90	0.81,0.99	0.250
USP4	1.01	0.92,1.09	0.950
TOP1MT	0.92	0.83,1.00	0.320
HEXDC	0.99	0.90,1.07	0.860
DPP3	1.13	1.05,1.21	0.130
ARHGEF18	1.10	1.02,1.18	0.250
SLC25A37	1.01	0.93,1.09	0.950
GLRX3	1.03	0.95,1.11	0.690
IQCE	1.07	0.99,1.15	0.380

C

1 Year Relapse

Characteristic	HR[1]	95% CI[1]	p-value
number_novel_mHAs	1.02	0.99, 1.04	0.2

[1] HR = Hazard Ratio, CI = Confidence Interval

Figure 4.7: Outcomes analyses for novel GvL mHAs. (A) shows odds ratio with 95% CI for each individual mHA with relapse at 1 year. Dot sizes denote size of patient population with the HLA allele corresponding to the mHA. n=1568 for HLAA*02:01, n=369 for HLA-B*35:01, and n=1034 for HLA-C*07:02. (B) shows statistics corresponding to the same analysis. (C) shows hazard ratio and 95% CI for association of total number of novel GvL mHAs per DRP with relapse at 1 year.

Finally, we assessed the applicability of these novel mHAs for ALL patients. As discussed in Chapter 2, ALL patients are included in DISCOVeRY-BMT but mHA

classification was performed based on TPM in TCGA_AML due to the lack of TCGA data for ALL. We compared source gene expression for the 24 validated mHAs in TCGA_AML versus a dataset of 110 pediatric ALL patients as this was the largest publicly available ALL RNAseq dataset. Z-scores of gene expression in each dataset were positively correlated (r=0.84, p=2.93e-7), indicating that these mHAs may be applicable for ALL patients, though this dataset does not suffice for definitively determining this (Figure 4.8). This analysis can be repeated when RNAseq data for adult ALL patients undergoing alloHCT becomes publicly available.

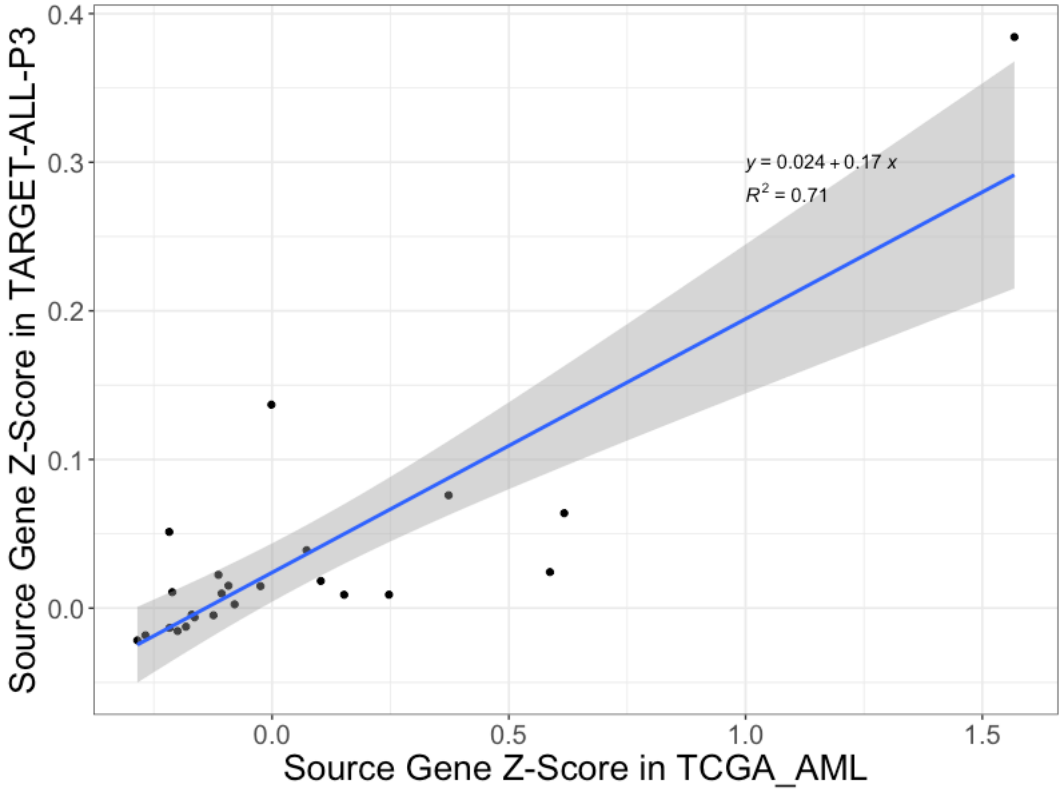


Figure 4.8: Comparison of mHA source gene expression z-score in TCGA_AML and TARGET-ALL-P3. Expression z-scores for the 24 source genes of mHAs validated in this chapter are shown.

(4.4) Discussion

Discovery and characterization of novel mHAs may be crucial for enhancing immune monitoring in alloHCT, predicting clinical outcomes based on donor and recipient genetics, and improving outcomes by optimizing donor selection and/or specifically targeting GvL mHAs. We built upon previous work to perform the first population level survey of mHA peptides, taking a new approach by predicting mHAs common among recipients with diverse HLA alleles. This ensures that therapeutics targeting our newly identified mHAs would be applicable to as broad of a recipient population as possible. As discussed in Chapter 2, we predicted mHAs for 56 class I HLA alleles represented in the DISCOVeRY-BMT dataset and found that approximately 50% of the mHAs predicted were designated as GvL mHAs. Additionally, we found that shared mHAs are more common than expected, with a distinct subgroup of mHAs found in approximately 20-30% of the DRP population with the relevant HLA allele for most alleles. We found that the number of mHAs needed to target in order for each DRP with a given HLA allele to contain at least one targetable mHA is smaller than expected, with as few as 11 mHAs per HLA allele covering all DISCOVeRY-BMT DRPs with that allele. We next sought to validate our predicted shared GvL mHAs.

Using mass spectrometry, we validated a total of 24 novel GvL mHAs in this chapter in addition to the 2 novel GvL mHAs validated in the previous chapter, representing a large increase from the 12 class I GvL mHAs that have been discovered since Els Goulmy et al reported the first discovered GvL mHA, HA-1, in 1983[62,144,164,166–168]. The 16 novel GvL mHAs found for HLA-A*02:01 together cover 98.8% of HLA-A*02:01-positive patients in the DISCOVeRY-BMT dataset, the three for HLA-B*35:01 cover 60.7% of HLA-B*35-:01-positive DISCOVeRY-BMT patients, and the five for HLA-

C*07:02 cover 78.9% of HLA-C*07:02-positive DISCOVeRY-BMT patients. We expect that these novel mHAs will serve as future targets for antigen-directed therapeutics.

Our study is limited in important ways. We biologically validated predicted GvL mHAs for three HLA alleles that were selected based on their high frequency of expression within diverse ethnic groups. In the future, mHAs for additional HLA alleles should be validated. Also, we validated GvL mHAs in a single AML cell line for each HLA allele. This is sufficient to establish that the mHAs are capable of being presented; however, antigen expression, HLA expression and antigen presentation efficiency will be heterogeneous across patient samples. Further studies of primary AML samples will be required to estimate the frequency of expression of each GvL mHA in AML. We validated more mHAs for HLA-A*02:01 than the other HLA alleles, which is likely not only due to running two samples for this allele but also because the cell line U937A2 is engineered to express HLA-A*02:01 and presents larger quantities of it on its cell surface than endogenously expressed HLA alleles. NB4 endogenously expresses HLA-B*35:01 and MONOMAC1 endogenously expresses HLA-C*07:02. In addition, while we validated mHAs to be presented by MS, this does not guarantee actual targeting by T cells nor assess differential frequency of mHA-specific T cells or immunogenicity. In the next chapter, we discuss T cell targeting work to identify, isolate, and perform functional assays of mHA-specific T cells to use for future therapeutics and to better understand determinants of GvL mHA immunogenicity.

CHAPTER 5: Generation of mHA-targeting T cells and TCR sequencing[4]

(5.1) Introduction

While validating 26 novel shared GvL mHAs serves as a major contribution to the field, we also seek to confirm immunogenicity of predicted GvL mHAs and target these mHAs with T cells. Establishing T cell targeting of GvL mHAs represents a further step towards clinical translation of TCR-based therapies. While we could currently produce a peptide vaccine-type therapeutic with our antigens, we believe that directly targeting mHAs with antigen-specific T cells will avoid several potential roadblocks for therapeutic efficacy. Peptide-based vaccines have multiple barrier for successful deployment, including needing to modify short peptides to avoid rapid enzymatic digestion, dysfunctional T cell responses with poor adjuvant efficacy, and difficulty of epitope selection due to imperfect immunogenicity prediction algorithms[267,268]. Peptide vaccines to date have shown to be safe but have limited efficacy thus far in humans; however, one recent study analyzing multiple clinical trials showed that peptide-specific T cells persist in patients that survive long-term after treatment with peptide vaccines[269,270]. Recent work has suggested some potential improvements for generating more substantial immune responses to vaccination also increase clinical response rates, so this is an approach we would consider if we are unable to generate satisfactory preclinical results with TCR-based therapeutics either in vitro or in vivo[271,272]. We have chosen to develop

[4] This chapter includes unpublished data that are part of a planned future manuscript from the Vincent and Armistead labs. I performed experiments, analyzed results, generated figures, and wrote the text included here.

TCR-based therapies because of the better results with adoptive cell therapy (ACT) compared to peptide vaccines, and some evidence of efficacy specifically with TCR ACTs[273]. We sought to generate T cells targeting our novel mHAs, isolate them, and sequence their TCRs. Once the TCR sequences are obtained, we can clone these sequences into a lentiviral vector to insert them into donor primary T cells. We can then test these cells to assess their cytotoxicity against AML cell lines and primary AML cell samples expressing the target antigen. We will also screen T cells for off-target effects against other cell types as a preclinical screening for potential safety. Eventually, we hope to translate the TCRs that exhibit potent antigen-specific cytotoxic effects to a clinical trial.

(5.2) Methods

Overall approach and sample selection

We used a healthy donor T cell priming approach to generate mHA-specific T cells, adapted from the Wolfl et al method[274]. We attempted to generate T cells for our HLA-A*02:01-binding mHAs and our HLA-C*07:02-binding mHAs, but not HLA-B*35:01-binding mHAs due to reagent limitations. No HLA monomers are commercially available for HLA-B*35:01 at present.

Non-HLA typed human donor Leukopaks (Gulf Coast Regional Blood Center) were used for HLA-A*02:01 due to high population prevalence of this allele in the donor population from the blood source. HLA-C*07:02 is less common in the donor population, so for this allele, HLA-typed cryopreserved PBMCs were used (Cytologics). From these donor sources, we generated monocyte-derived DCs, genotyped samples for our mHAs of interest, pulsed them with target mHAs applicable for that donor, then co-cultured

them with naïve CD8+ T cells from the same donor. We then used ELISpot and mHA-specific tetramer staining to confirm the presence of antigen-specific T cells, and we used fluorescent-activated cell sorting (FACS) to isolate antigen-specific activated T cells for single cell TCR sequencing (Figure 5.1).

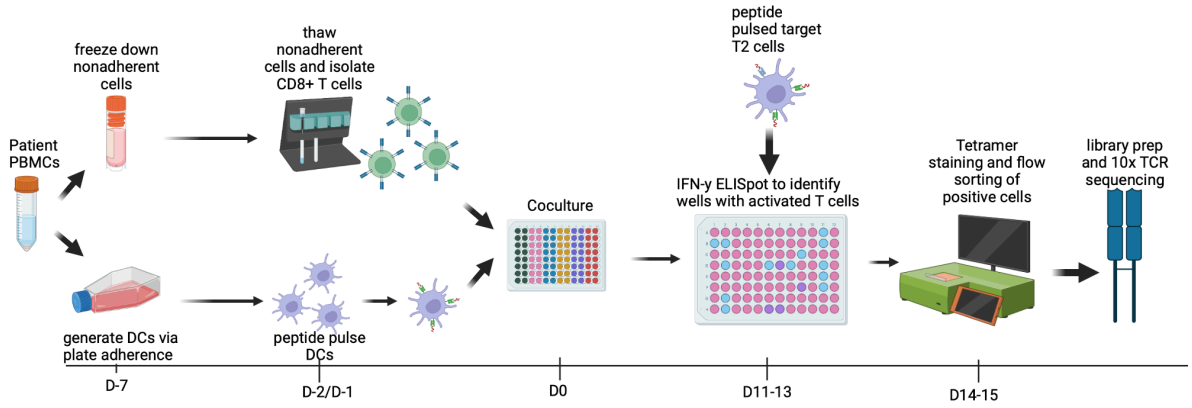


Figure 5.1: Overall workflow of T cell priming mHA-specific T cell isolation method.

Cell processing and DC generation

Donor Leukopaks were genotyped for HLA-A*02 via flow cytometry staining with purified anti-human HLA-A2 antibody clone BB7.2 (BioLegend). HLA-A*02 positive samples were selected and PBMCs isolated by layering over Ficoll-Paque PLUS (VWR) then centrifuging at 2200rpm for 20 minutes with low brake. PBMCs were then pulled off using transfer pipettes, and cells were washed then remnant red blood cells (RBCs) lysed using ACK lysis buffer (Gibco). 1 million PBMCs were set aside for gDNA isolations. Then, remaining PBMCs were plated in T75 flasks and incubated on their side for 90 minutes in order for monocytes to adhere to the plate. Nonadherent cell fraction was then removed and cells vigorously washed, then media replaced for an additional 60 minute incubation. Finally, additional nonadherent cells were removed and

cryopreserved in Cryostor (Sigma Aldrich). The adherent cells remaining in the flask were then cultured to generate monocyte-derived DCs by adding 1000 IU/mL GM-CSF and 1000 IU/mL IL-4 and culturing for 4-6 days undisturbed. Two days before co-culture initiation, DCs were matured by adding 10ng/mL LPS and 100 IU/mL IFNγ and incubating for 16 hours. After incubation, DCs were phenotyped for confirmation and peptide pulsed (Peptide 2.0 Inc) with mHAs, negative control mHAs that are endogenous to the donor and should not generate an immune response, or positive control melanA or a known immunodominant Influenza A Virus $M1_{58\text{-}66}$ HLA-A*02:01-binding influenza peptide (referred to hereafter as "flu")[275].

mHA SNP typing

DNA was extracted from reserved PBMCs from donors using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, 69504) and purified using the Monarch DNA Cleanup Kit (NEB, T1030S). SNP typing was performed as described in Chapter 4. Up to four peptides that the donor has zero alleles encoding were selected for pooling for DC pulsing and co-culture, and one mHA endogenous to the donor was selected as the negative control peptide.

CD8+ T cell isolation

Cryopreserved nonadherent cell fractions were thawed and cells washed. Naïve CD8+ T cells were then isolated using the EasySep Isolation Kit (StemCell Technologies). Briefly, cells were incubated with Isolation Cocktail of antibodies, then with RapidSpheres. Cells were then placed in flow tubes on an EasySep magnet for cell separation. Once separated, the enriched naïve CD8+ T cell fraction was pipetted off, counted, and placed in flasks with 5ng/mL IL7.

DC-T cell co-culture

Peptide pulsed DCs were harvested and resuspended at 0.5x10e6 cells/mL in media. Naïve CD8+ T cells were harvested and resuspended at 2x10e6 cells/mL in media with 60 ng/mL IL21. Co-cultures were then established by combining 100uL peptide pulsed DCs and 100uL CD8+ T cells per well of a 96 well plate. Cells were then fed every 2-3 days until day 9 after co-culture establishment.

ELISpot

IFNg ELISpots were performed to test T cell activation and IFNγ secretion on days 9-11 after establishment of co-culture. Each well was tested against negative control peptide as well as against pooled peptides of interest, and flu or melanA positive wells are also tested against their cognate peptide. First, plates were coated by prewashing with 35% ethanol, then incubating with anti-human IFNγ mHA 1-D1K overnight at 4C. T2 target cells were then pulsed with each necessary peptide, and peptide-pulsed T2s plated in the ELISpot plate after antibody incubation. Small fractions of each co-culture well were then added to the ELISpot plate and the plate is incubated overnight. Anti-human IFNγ mHA 7-B6-1 Biotin secondary antibody was then added and incubated for 2 hours, followed by Vectastain Elite ABC Kit (Novus Biologicals) and then AEC solution (BD Biosciences) for developing. Plates were washed, dried, and read on an ELISpot plate reader. Wells with ELISpot positives and control peptide negative corresponding wells were then taken on to tetramer staining for T cell isolation.

Monomer peptide exchange and tetramer generation

Peptides loaded on their corresponding HLA monomers were generated using Flex-T HLA monomers (BioLegend, 280003/280004), which come preloaded with UV-

labile placeholder peptides. Monomer and peptide were mixed and incubated on ice directly under 366 nm UV light for 30 minutes. Peptide exchange was then evaluated using HLA ELISA. Briefly, plates are coated with purified streptavidin and incubated overnight, then blocked with dilution buffer. HLA control standards were prepared with dilution buffer, and controls and samples were plated and incubated for 1 hour. Plates were washed, then HRP-conjugated antibody was added and incubated for 1 hour. Plates were washed, then Substrate Solution with ABTS (240uL ABTS stock in 12 mL of substrate buffer containing 0.1M citric acid and 0.1M trisodium citrate dihydrate) was added to develop the plates, followed by Stop Solution to halt the reaction when appropriate color is reached and plates were read at 414nm. If peptide exchange was successful, monomers were then tetramerized. 2uL fluorophore-conjugated streptavidin was combined with 18uL peptide-exchanged monomer and incubated for 30 mins in the dark, then 1.6uL blocking solution (1.6uL 50mM D-Biotin + 6uL 10% sodium azide + 192.4uL PBS) was added and tetramer solution was incubated overnight.

Tetramer staining and cell sorting

A small sample of ELISpot positive wells ware first stained and assessed via flow cytometry for tetramer positivity. If positive, we proceeded to tetramer staining and FACS sorting of positive cells. Briefly, 50nM Dasatinib was added to tetramer positive wells and incubated for 30 minutes. Cells were then washed and stained with FVS700 live-dead (BD Biosciences), followed by staining with tetramer and CD8-BV421 (BioLegend, Clone: SK1), and resuspended in FACS buffer and filtered. Cells that were doubly positive for CD8+ and mHA-specific tetramer were sorted on a FACS Aria and

given to UNC Immune Monitoring and Genomics Facility for single cell TCR sequencing library preparation. Gating schema is shown (Figure 5.2).

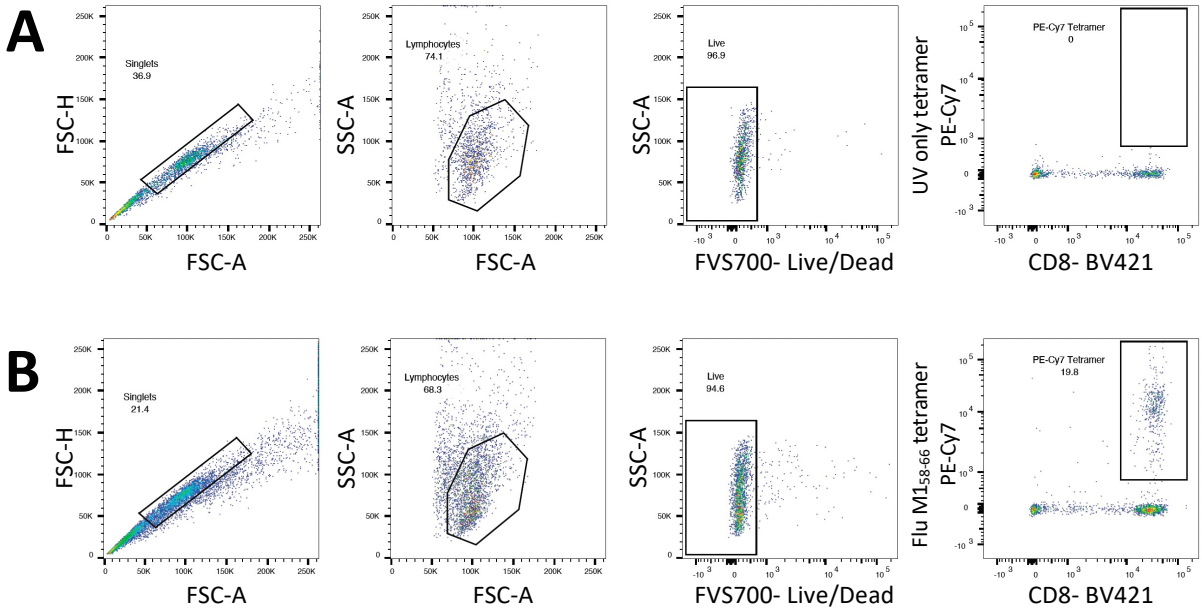


Figure 5.2: Gating strategy for mHA immunogenicity assessment. Gating was performed on single cells, lymphocytes, and live cells. (A) shows negative control cells stained with tetramer exposed to UV light with no peptide. (B) shows positive control cells from CD8+ T cells co-culture with flu-M1$_{58-66}$ pulsed DCs, stained with fluM1$_{58-66}$ tetramer.

Sequencing and data processing

A library was prepared for 10x single cell TCR sequencing from a minimum of 10,000 sorted CD8+ tetramer positive cells using the SMARTer Human TCR a/b Profiling Kit (Takara). If the number of tetramer positive cells sorted was less than 10,000, the sorted cell population was supplemented with T2 cells. These APCs do not present TCR, so they add cell bulk to the sorted fraction to increase manual ease of library preparation without adding additional TCRs to the sequencing pool[276]. VDJ TCRa/b 10x single cell sequencing was then performed on a NovaSeq instrument, with demultiplexing done via Illumina BCL Convert version 3.8.2-12. Sequencing quality

control was performed using FastQC[277]. Final TCRa/b sequences were determined using either MiXcR or using the 10x Genomics Cloud Server[278].

(5.3) Results

To date, we have identified and isolated a T cell clone specific for one of our novel mHAs, UNC-HEXDC-V, via tetramer staining of CD8+ T cells co-cultured with mHA-pulsed DCs. We sorted this tetramer positive clone, prepared a sequencing library, and performed 10x single cell TCRa/b sequencing. We found one clear predominant clone with high quality sequencing data, and we are currently engaged in cloning this TCR into a lentiviral vector in order to transduce large numbers of T cells and test efficacy of this clone via proliferation, killing, and cytokine release assays. Flow cytometry analysis of negative control, positive flu control, and our identified UNC-HEXDC-V specific T cell population is shown (Figure 5.3).

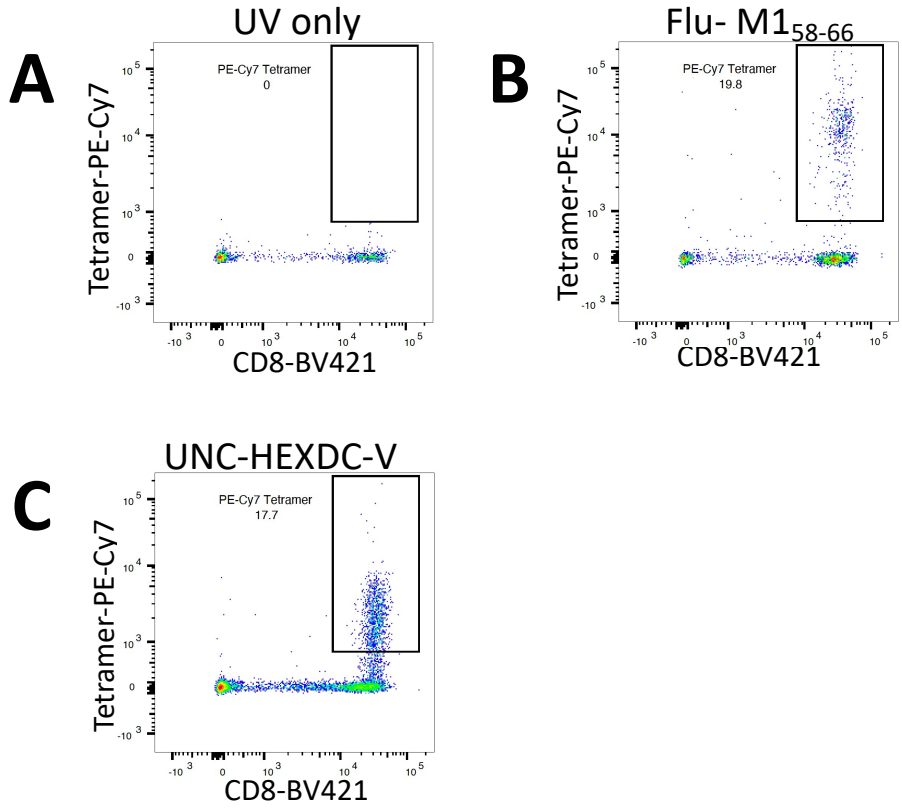


Figure 5.3: mHA targeting co-culture flow cytometry. (A) "UV only" shows negative control stained with tetramer exposed to UV light with no peptide. (B) "Flu-M15$_{8-66}$" shows CD8+ T cells co-cultured with M1$_{58-66}$ pulsed DCs, stained with M1$_{58-66}$ tetramer. (C) "UNC-HEXDC-V" shows CD8+ T cells co-cultured with novel mHA UNC-HEXDC-V stained with UNC-HEXDC-V tetramer.

(5.4) Discussion

By identifying and isolating a TCR clone specific for one of our novel shared GvL mHAs, we confirmed immunogenicity of this peptide and generated a TCR product for potential future clinical use. We actively continue to seek T cells targeting additional antigens from our work. This process is laborious and lengthy particularly due to the low precursor frequency of antigen-specific T cells within the naïve CD8+ T cell fraction from a healthy donor. The frequency of T cells specific for a given antigen is estimated to be 1 in 1 million[279–281]. We obtain roughly 300-700 million PBMCs per Leukopak on

average, which gives a yield of between 1-10 million naïve CD8+ T cells. Therefore, we can expect that an average round of co-culture experiments would be done with a sample that contains somewhere between one and ten T cells of interest per mHA used for peptide pulsing. There are many places in the experimental protocol that these cells can be lost, including use within a control well, for cell counting, for the ELISpot assays instead of FACS, or in routine cell culture procedures over the course of several weeks. While we are investigating more high-throughput approaches for GvL mHA-specific T cell isolation as discussed in the next chapter, we continue to strive to identify additional TCRs via this method. Our long-term goal is to translate these TCRs to clinic by engineering donor T cells to express TCRs specific for GvL mHAs expressed by the recipient leukemia in parallel with the regular alloHCT process then infusing them into the patient to augment GvL effects (Figure 5.4).

CHAPTER 6: Discussion and Future Directions

(6.1) Summary

mHAs in the context of alloHCT for hematologic malignancies can serve as beneficial targets of anti-tumor T cell responses. Targeting GvL mHAs can address the critical “transplanter’s dilemma”: the link between increases in transplant efficacy with increases in harmful GvHD[78,108,282–291]. Discovery of new mHAs is essential for future clinical applications because only 12 GvL mHAs were known by the field prior to this work, and most of these were applicable to only a few patients. The work we present here represents a major step forward in the major histocompatibility antigen field, establishing targeting of shared mHAs as a viable approach for the majority of transplant recipients and identifying 26 novel mHAs. We discovered that shared GvL mHAs are much more prevalent than previously believed, with a small number of GvL mHAs covering 20-30% of DRPs with a given HLA allele.

Prior to the work described here, the mean population prevalence of known mHAs in the largest relevant dataset, DISCOVeRY-BMT, was 1.84%. This indicates why mHAs have traditionally been viewed as targets for personalized immunotherapy and why mHA-targeting clinical trials have had fairly low recruitment. Therapies targeting mHAs that are very rarely discordant in a DRP are applicable to very few people, so the efficacy of mHA-targeting therapies has been limited thus far. We believe that targeting shared GvL mHAs will vastly expand the utility of targeting these antigens. The mean population prevalence of our newly discovered mHAs was 25.10%, meaning

that many times more patients could be treated with any therapeutic agent targeting these. If we were to develop therapeutic agents that targeted each of our novel mHAs, we could treat 99% of HLA-A*02:01 DRPs, 61% of HLA-B*35:01 DRPs, and 79% of HLA-C*07:02 DRPs, representing a massive improvement in the fraction of patients that mHA-targeting therapeutics are available for. This also expands the future availability of mHA-targeting therapies to ethnic groups beyond Caucasians. In this final chapter, we discuss technical advances that could benefit this work in the future, the relevance of our work to foundational transplant immunobiology, and future clinical applications.

(6.2) Future methodological advances

While we have discovered one GvL mHA-targeting TCR clone against our new targets thus far, we will need to identify more mHA-targeting clones to translate these targets to clinic with the method that we view as most promising: adoptive cell transfer of donor T cells with GvL mHA-targeting TCRs transgenically inserted. To accomplish this, we have multiple options for TCR identification, including continuing with co-culture experiments, utilizing the new Barcode Enabled Antigen Mapping (BEAM) technology from 10X Genomics, or utilizing the ATRAP method from the Hadrup and Nielsen labs[292]. The BEAM-T method enables single cell immune profiling of antigen-specific T cells by utilizing multiplexed antigen screening with barcoded and fluorescently labeled peptide-MHC monomers followed by flow sorting of antigen-specific cells. The ATRAP method is mostly similar, but requires in-house barcoding of peptide-MHC monomers using dextran molecules complexed with streptavidin and fluorescent molecules to create barcoded peptide-MHC fluorescent dextramers[292]. The main benefit of BEAM-T over ATRAP is that the peptide-MHC molecules are pre-barcoded for ease of staining

and T cell isolation. While barcoding dextramers in-house is more laborious and introduces an additional failure mode to the experiment, it also allows for testing as many antigens as desired rather than being limited to the 15 antigens plus positive control that the BEAM-T kit allows. In addition, the ATRAP method is built to enable screening of multiple samples simultaneously using Hashing antibodies so the experiment can be multiplexed on two different dimensions. We will continue with TCR discovery via co-culture while also exploring these options for making our discovery pathway more high-throughput. At present, BEAM-T is available for HLA-A*02:01 but not HLA-B*35:01 and HLA-C*07:02, so we plan to utilize BEAM-T to identify HLA-A*02:01-binding antigen-specific T cell clones, then move into utilizing the ATRAP method for the other two alleles of interest.

In addition to these two methods screening for T cell clones specific for known target antigens, additional high throughput methods could be used for more rapid testing of predicted antigens in the future. One such method is T-Scan, a novel library screen method using a fluorescent reporter of granzyme B activity to isolate activated T cells, followed by TCR sequencing[293]. Target cells are engineered to express a genome-wide library of candidate antigens that are then processed and presented endogenously on MHC. After co-culture with T cells, target cells undergoing binding of peptide/MHC by TCR are identified via a fluorescent reporter that is activated by granzyme B binding, enabling them to be FACS sorted and antigens identified via next generation sequencing of library barcodes[293]. Another new method we could utilize is RAPTR. This method expands upon the limitation of methods by not only identifying antigens that lead to T cell activation, but giving paired information on TCR sequence

and antigen[294]. In this system, 293T cells are transduced with a library of pMHC constructs and a fluorescently labeled barcode. They are cultured with a T cell population, and upon antigen recognition the viral-packaged barcode and fluorescent label from the APC enters the T cell via the TCR, after which cells can then be sorted and single cell sequenced to obtain a pMHC-TCR pair[294]. While this method would require incorporating several new techniques into our current workflow, if it performs as reported it would be an extremely efficient method to identify paired antigens and their cognate TCRs.

Finally, in addition to improved TCR identification efficiency and improved library screens for mHAs that generate immune responses, advances in mass spectrometry are also a potential improvement point. A recent major advance in shotgun immunopeptidomics was reported by Ehx et al in 2021. This work described a novel way to build MS databases for shotgun searches by RNAseq of the tumor sample, computational division of RNAseq reads into fixed length k-mers, and subtraction of k-mers that are also found in RNAseq of a healthy control to generate a database of cancer-specific k-mers to screen by mass spectrometry[295,296]. They can then identify the k-mers that are overrepresented in tumor versus normal and assemble these k-mers into contigs, followed by all-frame translation to identify expected peptides for the MS search. This allows for a streamlined pre-filtering for peptide candidates most likely to appear in MS searches, leading to more usable information and positive peptide identifications from each MS run[295]. They successfully used this approach to identify dozens of tumor specific and associated antigens from AML samples as proof of concept[295]. On the targeted MS side, the technique of scans with heavy-labeled

standards is the current optimal strategy for targeted immunopeptidomics, but future innovations in the field can be incorporated into antigen validation as they become available.

Ultimately, incorporating higher throughput antigen discovery and validation methods into our experimental workflow could increase the speed of future TCR discoveries. We will evaluate the strengths and weaknesses of these techniques in our hands to determine which of these technical advances will be most beneficial to propel us towards our eventual goal of translating these antigens to clinic.

(6.3) Relevance to foundational transplant immunology

The ability to identify mHA-specific T cell responses will enhance our understanding of foundational transplant immunobiology. Previous investigations in mice and with human cells *in vitro* have indicated that immune recognition and response is focused on a small set of immunodominant antigens[297–300]. The current low number of known mHA-targeting TCR sequences means that it is difficult to investigate the differences between immunodominant and non-dominant mHAs. Identification of additional mHA-targeting T cells in patient samples in future work will allow for assessment of what sets mHAs that drive measurable T cell responses apart from those that do not. T cells targeting our validated GvL mHAs can be utilized to investigate immunodominance in the context of graft versus leukemia effects. This analysis could be extended to immunodominance in GvHD by experimentally validating GvH mHAs predicted in this work and identifying their cognate T cells. Understanding of antigens that drive these two major effects of alloHCT would help us to better understand the factors influencing transplant outcomes. They would also allow us to develop better

therapeutics, as future targeting could be focused on antigens that have displayed immunodominance, or designed to avoid antigens that are the dominant drivers of GvHD. Knowledge of antigen-specific T cell clones could also be used to track these cells within patients as a predictive factor for prognosis and as a piece of data to guide implementation of GvHD prophylaxis or treatments, or additional GvL therapeutics if antileukemic effects are insufficient. Previous work has found an association between viral and tumor-associated antigen-specific immune responses and GvHD, and the authors theorized that this could be utilized for characterizing GvHD risk. We could pursue this same avenue of risk assessment with greater resolution using GvH mHAs[301]. While characterizations of immunodominance are difficult with a limited number of antigens, they will become easier as the lab extends this work to discover additional T cell clones mapped to mHA. For a more complete picture of the immunodominance hierarchy in a system with less complicating factors, these mHA discovery efforts could be repeated in one or more syngeneic mouse models. This would represent a system in which the full immunogenicity determining landscape could be assessed. These discovered mHAs can also be utilized in future experiments in our lab for the assessment of mechanisms of resistance to T cell killing of leukemia. Knowledge of a large number of antigenic peptides and their cognate TCR sequences enables experiments in which dynamics of mixed antigen-specific T cell populations can be intensively probed upon exposure to pools of antigens. Complete characterization of antigen-specific T cells that mediate versus do not mediate killing of cancer cells will represent a foundational science advance in the field and guide future antigen selection for immunotherapies.

The antigens and concepts uncovered by this work can also be applied to other transplant methodologies in the future, including haploidentical transplants. The different immunobiology of haplo transplants and the immunosuppression that is necessitated by HLA mismatch leads to a unique immune landscape in these transplants, requiring special considerations for adoptive cell therapies post-haplo transplant[302,303]. T cells must be depleted to avoid alloreactivity to mismatched HLA and the TCR repertoire replenished via naïve T cells passing through the recipient thymus over several years[302]. In addition to the changed T cell landscape in haploidentical transplants, the mHA landscape also differs due to the shared and unshared HLA alleles in the DRP. mHA identification can be completed as usual for the HLA alleles shared by donor and recipient, with the mHA needing to be foreign to the donor in order for targeting by donor-derived T cells. However, for the HLA alleles that the recipient has and donor does not, there does not need to be any consideration of donor mHA alleles. This is because the donor cells do not express the HLA allele corresponding to the mHAs so these mHAs will not be presented on donor cells, eliminating the possibility of fratricide of donor cells. Overall, haploidentical transplants may have more potential targetable mHAs, though differing T cell dynamics may make targeting more complicated. Future work should utilize a large sequencing dataset of haploidentical DRPs to extend this work, assessing applicability of our validated GvL mHAs to this population and assessing whether mHA population frequencies are similar for haploidentical DRPs.

(6.4) Potential pitfalls

Though we feel optimistic about future applications of our identified shared GvL mHAs, there are several potential limitations to mHA-targeting immunotherapies that we

must keep in mind as this work progresses. Firstly, we must learn from history and rigorously test for potential off-target effects (used here to mean targeting of tissues other than the intended leukemia target) of our mHA-targeting T cells in order to avoid serious GvHD effects. A previous attempt at treating relapsed leukemia after alloHCT at the Fred Hutchinson Cancer Research Center with mHA-targeting CD8 T cell clones that were believed to be tissue-restricted gave pulmonary toxicity in 4 of 7 patients requiring intubation in one[73]. This represented a serious cautionary tale for the field and emphasizes the need to screen T cells against a broader panel of normal tissues, as this group defined “hematopoietic specific” mHAs exclusively based upon recognition of recipient hematopoietic cells and not recipient dermal fibroblasts[73]. This screening was insufficient to thoroughly examine the potential for off-target effects. Prior to applying our mHA-targeting T cells to clinic, we will screen them against cells derived from a wide variety of tissue types for reactivity. In addition, we have a potentially greater ability to avoid off-target effects due to our knowledge of specific mHAs and their tissue restriction patterns. In approaches such as that of the Fred Hutchinson group and others, the exact target of the T cells used is unknown, and is inferred to be tissue-specific mHAs based on reactivity to recipient and not donor-derived EBV-LCL transformed cells. Information on target cells is exclusively derived from reactivity testing against whatever cells the investigator chooses to test against, whereas in our approach, because the target mHA is known we have the advantage of being able to vigorously assess source gene tissue expression and other factors. If concerns about off-target effects are high, a safety switch could be incorporated into the T cell construct for rapid irreversible inactivation of the T cell population[304].

In addition to worries of off-target effects, another potential pitfall of our approach is imperfection of immunogenicity predictions as mentioned previously. We as a field have knowledge of some predictors of immunogenicity and can utilize tools such as netMHCpan to eliminate peptides that are unlikely to generate an immune response. However, the sum of the known factors that influence immunogenicity cannot fully predict epitopes that generate T cell responses robust enough to drive clinical effects. Experimental data that confirms that the peptides of interest are presented on MHC, such as our validation data of mHA presentation on AML cell lines, brings the probability of these mHAs generating an immune response higher than if they were only predicted epitopes, but it is likely that not all of our validated mHAs will generate a measurable T cell response *in vivo*. However, due to our focus on shared mHAs, any of our validated mHAs that are targetable will be applicable to a large number of DRPs. Moreover, now that we have established and confirmed the functionality of this mHA prediction and validation workflow, we can incorporate new and improved immunogenicity prediction tools as they are developed in the future by our group and others. An additional benefit to the reverse immunology approach to mHA identification is its flexibility. We can fine-tune our prediction pathways for the mHA qualities we deem potentially beneficial, in contrast to forward immunology approaches in which you cannot pre-specify the qualities of the mHAs you seek to discover.

A third potential limitation of TCR-engineered T cell therapy targeting mHAs is immune and cancer cell factors that limit the functionality of cellular immunotherapies in general. Immune evasion by the cancer could interfere with function of mHA-targeting T cells via a variety of mechanisms. For example, the leukemia could downregulate

expression of the source gene of the targeted mHA. This possibility is why we feel that future applications of this project should include targeting of multiple mHAs simultaneously, as the possibility of immune evasion of one T cell antigen specificity is high. Loss of heterozygosity at the MHC locus is another concern as this is frequent in some cancers, leads to loss of presentation of antigens on either the maternal or paternal HLA haplotype, and has been observed (at least in the case of HLA class II loss of heterozygosity) to be associated with leukemia relapse after allogeneic stem cell transplantation[305–308]. This would lead to loss of mHA presentation on half of the recipient's HLA class I alleles, which could strongly limit the applicability of mHA-targeting therapies. Because of this, we believe that it is important to not only target multiple mHAs in a single patient, but also target mHAs that bind different HLA alleles. If mHAs that bind different alleles are targeted, the possibility of antigen presentation being lost with loss of heterozygosity of HLA is reduced. However, antigen presentation on HLA class I as a whole can also occur via downregulation of beta 2 microglobulin, leading to complete loss of antigen presentation on HLA class I[305]. In this case, HLA class II presentation would still be functional, so class II-binding mHAs would still be theoretically targetable. We have also generated class II-binding mHA predictions and seek to validate these predicted mHAs in the future to potentially integrate into our immunotherapy strategy. A further limiter of target availability is the decrease in antigen presentation as leukemia killing occurs[309]. In mouse experiments in a model where GvL is exclusively driven by T cells recognizing the mHA H60, various strategies helped to combat this loss of efficacy, including H60 vaccination, administration of anti-CD40 at the time of transplant, and PD1 blockade[309]. These strategies could be incorporated into

the clinical plan if we are concerned about loss of efficacy as leukemia cells are eliminated. T cell exhaustion could also limit treatment efficacy but could also be abrogated by interventions at multiple stages of generation of the T cell product, including modulating T cell phenotype via cytokine administration, metabolic or epigenetic programming, or genetic engineering, or by administering immunotherapies such as PD1 blockade to the patient after T cell administration[310,311]. If initial applications of our mHA-targeting T cells show safety but not efficacy, these mechanisms for augmenting T cell function can be explored.

(6.5) Translational applications

Once more TCRs are identified and screened rigorously for both on-target and off-target effects, we will pursue a clinical trial utilizing T cells with transgenically inserted mHA-specific TCRs to enhance GvL effects of transplant. We would also like to continue discovering additional GvL mHAs for even more HLA alleles to keep expanding this field beyond the traditional one or two HLA alleles in which mHA discovery is usually performed. In order to discover which antigens would be most beneficial to target, we must discover the determinants of immunodominance in the GvL mHA landscape. We expect that some mHAs are more likely to generate a robust immune response than others, but do not know for sure and thus far do not know why. Future work in our lab including intensive profiling of mHA-specific T cells will enable determination of factors that lead to more or less robust immune responses. We can then further tailor our mHA discovery methods to incorporate the features that lead to robust responses, so future mHAs we discover will be even more likely to be therapeutically applicable.

The mHAs we describe in this work and in future efforts can also be useful for clinical efforts outside of targeting them. If we are able to identify immunodominant mHAs and/or mHAs that correlate with clinical outcomes, it would be logical to utilize these mHAs to guide donor selection for transplant. If presence of a specific mHA in a DRP correlates with worse outcomes, such as the mHAs identified by Vincent lab member Othmane Jadi that correlate with increased GvHD death, decreased leukemia-free survival, and increased disease-related mortality, recipients and potential donors could be genotyped prior to donor selection to ensure that these mHAs are not present[213]. Conversely, if we identify immunodominant GvL mHAs or ones that correlate with better clinical outcomes, donors could be selected for appropriate mismatches with the recipient at this mHA. Monitoring populations of mHA-specific T cells for these mHAs that drive desirable or undesirable clinical effects could also represent a way to monitor efficacy of transplant. This could perhaps give early indicators of whether transplant will be successful at eradicating cancer, or whether a patient will need intensified GvHD prophylaxis or treatment. Taken together, our novel mHAs could be used in two major ways for benefit of patients: as targets for novel immunotherapies, or as prognostic indicators to guide donor selection and give insights into progress of treatment.

Our ultimate dream in the mHA space would be to expand this work beyond using it to augment alloHCT and actually be able to replace alloHCT entirely with infusion of mHA-specific T cells. While this is not technologically possible due to the low number of discovered antigens, even lower number of known TCRs corresponding to them, and requirement for tumor-specific mHA for this strategy to work, we foresee this

eventually becoming possible as mHA discovery and targeting progresses over the next decade. Work in the immunodominance space will also need to progress such that we can know how many antigens must be targeted so that the cancer can no longer evade elimination by the immune system. If we know this number and have enough mHAs discovered to target that many to make immune evasion impossible, every patient could receive a bespoke combination of T cells with TCRs specific for their cancer and eliminate their cancer by the action of these T cells alone. This could eliminate the need for transplant and accompanying factors such as conditioning chemotherapy that generate side effects and decrease patient quality of life. While this is our holy grail of mHA targeting work, we will be very happy with any translational applications of our mHAs, including the clinical trials we hope to launch in the next few years. Ultimately, our work will enable novel therapeutics to be developed for AML and other hematologic malignancies, and hopefully improve survival and outcomes for these patients.

APPENDIX 1: Landscape and Selection of Vaccine Epitopes in SARS-CoV-2

Please refer to https://doi.org/10.1186/s13073-021-00910-1 for full text.

APPENDIX 2: SARS-CoV-2 Peptide Vaccine Elicits T-cell Responses in Mice but Does Not Protect against Infection or Disease

Please refer to https://doi.org/10.1101/2022.02.22.481499 for full text.

APPENDIX 3: LENS: Landscape of Effective Neoantigens Software

Please refer to https://doi.org/10.1101/2022.04.01.486738 for full text.

APPENDIX 4: NeoSplice: a Bioinformatics Method for Prediction of Splice Variant Neoantigens

Please refer to https://doi.org/10.1093/bioadv/vbac032 for full text.

APPENDIX 5: Associations of Minor Histocompatibility Antigens with Clinical Outcomes Following Allogeneic Hematopoietic Cell Transplantation

Please refer to https://doi.org/10.1101/2022.08.31.506092 for full text.

REFERENCES

1. Liao S, von der Weid P-Y. Lymphatic System: An Active Pathway for Immune Protection. *Semin. Cell Dev. Biol.* 2015;38:83–89.

2. Chaplin DD. Overview of the Immune Response. *J. Allergy Clin. Immunol.* 2010;125(2 Suppl 2):S3-23.

3. Nicholson LB. The immune system. *Essays Biochem.* 2016;60(3):275–301.

4. Marshall JS, Warrington R, Watson W, Kim HL. An introduction to immunology and immunopathology. *Allergy Asthma Clin. Immunol.* 2018;14(2):49.

5. Alam R. A brief review of the immune system. *Prim. Care*. 1998;25(4):727–738.

6. Yokoyama WM. Natural killer cell immune responses. *Immunol. Res.* 2005;32(1–3):317–325.

7. Charles A Janeway J, Travers P, Walport M, Shlomchik MJ. The complement system and innate immunity. *Immunobiol. Immune Syst. Health Dis. 5th Ed.* 2001;

8. Henneke P, Golenbock DT. Phagocytosis, Innate Immunity, and Host–Pathogen Specificity. *J. Exp. Med.* 2004;199(1):1–4.

9. Bonilla FA, Oettgen HC. Adaptive immunity. *J. Allergy Clin. Immunol.* 2010;125(2):S33–S40.

10. Forthal DN. Functions of Antibodies. *Microbiol. Spectr.* 2014;2(4):1–17.

11. Charles A Janeway J, Travers P, Walport M, Shlomchik MJ. T cell-mediated cytotoxicity. *Immunobiol. Immune Syst. Health Dis. 5th Ed.* 2001;

12. Raskov H, Orhan A, Christensen JP, Gögenur I. Cytotoxic CD8+ T cells in cancer and cancer immunotherapy. *Br. J. Cancer*. 2021;124(2):359–367.

13. Tay RE, Richardson EK, Toh HC. Revisiting the role of CD4+ T cells in cancer immunotherapy—new insights into old paradigms. *Cancer Gene Ther.* 2021;28(1):5–17.

14. Kaiko GE, Horvat JC, Beagley KW, Hansbro PM. Immunological decision-making: how does the immune system decide to mount a helper T-cell response? *Immunology*. 2008;123(3):326–338.

15. Luckheeram RV, Zhou R, Verma AD, Xia B. CD4+T Cells: Differentiation and Functions. *Clin. Dev. Immunol.* 2012;2012:925135.

16. Duggleby R, Danby RD, Madrigal JA, Saudemont A. Clinical Grade Regulatory CD4+ T Cells (Tregs): Moving Toward Cellular-Based Immunomodulatory Therapies. *Front. Immunol.* 2018;9:.

17. Kondělková K, Vokurková D, Krejsek J, et al. Regulatory T cells (TREG) and their roles in immune system with respect to immunopathological disorders. *Acta Medica (Hradec Kralove)*. 2010;53(2):73–77.

18. Charles A Janeway J, Travers P, Walport M, Shlomchik MJ. Generation of lymphocytes in bone marrow and thymus. *Immunobiol. Immune Syst. Health Dis. 5th Ed.* 2001;

19. Parker ME, Ciofani M. Regulation of γδ T Cell Effector Diversification in the Thymus. *Front. Immunol.* 2020;11:.

20. Burtrum DB, Kim S, Dudley EC, Hayday AC, Petrie HT. TCR gene recombination and alpha beta-gamma delta lineage divergence: productive TCR-beta rearrangement is neither exclusive nor preclusive of gamma delta cell development. *J. Immunol.* 1996;157(10):4293–4296.

21. Charles A Janeway J, Travers P, Walport M, Shlomchik MJ. T-cell receptor gene rearrangement. *Immunobiol. Immune Syst. Health Dis. 5th Ed.* 2001;

22. Junctional diversity in signal joints from T cell receptor beta and delta loci via terminal deoxynucleotidyl transferase and exonucleolytic activity. *J. Exp. Med.* 1996;184(5):1919–1926.

23. Klein L, Kyewski B, Allen PM, Hogquist KA. Positive and negative selection of the T cell repertoire: what thymocytes see and don't see. *Nat. Rev. Immunol.* 2014;14(6):377–391.

24. Xiong Y, Bosselut R. CD4-CD8 differentiation in the thymus: connecting circuits and building memories. *Curr. Opin. Immunol.* 2012;24(2):139–145.

25. Takaba H, Takayanagi H. The Mechanisms of T Cell Selection in the Thymus. *Trends Immunol.* 2017;38(11):805–816.

26. Passos GA, Speck-Hernandez CA, Assis AF, Mendes-da-Cruz DA. Update on Aire and thymic negative selection. *Immunology*. 2018;153(1):10–20.

27. Gotter J, Brors B, Hergenhahn M, Kyewski B. Medullary Epithelial Cells of the Human Thymus Express a Highly Diverse Selection of Tissue-specific Genes Colocalized in Chromosomal Clusters. *J. Exp. Med.* 2004;199(2):155–166.

28. Alawam AS, Anderson G, Lucas B. Generation and Regeneration of Thymic Epithelial Cells. *Front. Immunol.* 2020;11:858.

29. Burger ML, Leung KK, Bennett MJ, Winoto A. T cell-specific inhibition of multiple apoptotic pathways blocks negative selection and causes autoimmunity. *eLife*. 2014;3:e03468.

30. Sckisel GD, Bouchlaka MN, Monjazeb AM, et al. Out-of-Sequence Signal 3 Paralyzes Primary CD4(+) T-Cell-Dependent Immunity. *Immunity*. 2015;43(2):240–250.

31. Tai Y, Wang Q, Korner H, Zhang L, Wei W. Molecular Mechanisms of T Cells Activation by Dendritic Cells in Autoimmune Diseases. *Front. Pharmacol.* 2018;9:.

32. Zhang S, Kohli K, Black RG, et al. Systemic Interferon-γ Increases MHC Class I Expression and T-cell Infiltration in Cold Tumors: Results of a Phase 0 Clinical Trial. *Cancer Immunol. Res.* 2019;7(8):1237–1243.

33. Shankaran V, Ikeda H, Bruce AT, et al. IFNγ and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature*. 2001;410(6832):1107–1111.

34. Zhou F. Molecular Mechanisms of IFN-γ to Up-Regulate MHC Class I Antigen Processing and Presentation. *Int. Rev. Immunol.* 2009;28(3–4):239–260.

35. Wieczorek M, Abualrous ET, Sticht J, et al. Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. *Front. Immunol.* 2017;8:.

36. Choo SY. The HLA System: Genetics, Immunology, Clinical Testing, and Clinical Implications. *Yonsei Med. J.* 2007;48(1):11–23.

37. Leone P, Shin E-C, Perosa F, et al. MHC Class I Antigen Processing and Presenting Machinery: Organization, Function, and Defects in Tumor Cells. *JNCI J. Natl. Cancer Inst.* 2013;105(16):1172–1187.

38. Rock KL, Reits E, Neefjes J. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends Immunol.* 2016;37(11):724–737.

39. Purcell AW, Ramarathinam SH, Ternette N. Mass spectrometry–based identification of MHC-bound peptides for immunopeptidomics. *Nat. Protoc.* 2019;14(6):1687–1707.

40. Neerincx A, Castro W, Guarda G, Kufer T. NLRC5, at the Heart of Antigen Presentation. *Front. Immunol.* 2013;4:397.

41. Neefjes J, Jongsma MLM, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* 2011;11(12):823–836.

42. Reits E, Griekspoor A, Neijssen J, et al. Peptide Diffusion, Protection, and Degradation in Nuclear and Cytoplasmic Compartments before Antigen Presentation by MHC Class I. *Immunity*. 2003;18(1):97–108.

43. Blum JS, Wearsch PA, Cresswell P. Pathways of Antigen Processing. *Annu. Rev. Immunol.* 2013;31:443–473.

44. Wosen JE, Mukhopadhyay D, Macaubas C, Mellins ED. Epithelial MHC Class II Expression and Its Role in Antigen Presentation in the Gastrointestinal and Respiratory Tracts. *Front. Immunol.* 2018;9:.

45. Roche PA, Furuta K. The ins and outs of MHC class II-mediated antigen processing and presentation. *Nat. Rev. Immunol.* 2015;15(4):203–216.

46. Johansen TE, McCULLOUGH K, Catipovic B, et al. Peptide Binding to MHC Class I is Determined by Individual Pockets in the Binding Groove. *Scand. J. Immunol.* 1997;46(2):137–146.

47. Falk K, Rötzschke O, Stevanović S, Jung G, Rammensee H-G. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature*. 1991;351(6324):290–296.

48. Falk K, Rötzschke O. Consensus motifs and peptide ligands of MHC class I molecules. *Semin. Immunol.* 1993;5(2):81–94.

49. Zhang C, Anderson A, DeLisi C. Structural principles that govern the peptide-binding motifs of class I MHC molecules11Edited by I. Wilson. *J. Mol. Biol.* 1998;281(5):929–947.

50. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* 2020;48(W1):W449–W454.

51. O'Donnell TJ, Rubinsteyn A, Laserson U. MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing. *Cell Syst.* 2020;11(1):42-48.e7.

52. Wells DK, van Buuren MM, Dang KK, et al. Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction. *Cell*. 2020;183(3):818-834.e13.

53. Duan F, Duitama J, Al Seesi S, et al. Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J. Exp. Med.* 2014;211(11):2231–2248.

54. Ghorani E, Rosenthal R, McGranahan N, et al. Differential binding affinity of mutated peptides for MHC class I is a predictor of survival in advanced lung cancer and melanoma. *Ann. Oncol.* 2018;29(1):271–279.

55. Balachandran VP, Łuksza M, Zhao JN, et al. Identification of unique neoantigen qualities in long term pancreatic cancer survivors. *Nature*. 2017;551(7681):512–516.

56. Richman LP, Vonderheide RH, Rech AJ. Neoantigen Dissimilarity to the Self-Proteome Predicts Immunogenicity and Response to Immune Checkpoint Blockade. *Cell Syst.* 2019;9(4):375-382.e4.

57. Li G, Iyer B, Prasath VBS, Ni Y, Salomonis N. DeepImmuno: deep learning-empowered prediction and generation of immunogenic peptides for T-cell immunity. *Brief. Bioinform.* 2021;22(6):bbab160.

58. Lu T, Zhang Z, Zhu J, et al. Deep learning-based prediction of the T cell receptor–antigen binding specificity. *Nat. Mach. Intell.* 2021;3(10):864–875.

59. Springer I, Besser H, Tickotsky-Moskovitz N, Dvorkin S, Louzoun Y. Prediction of Specific TCR-Peptide Binding From Large Dictionaries of TCR-Peptide Pairs. *Front. Immunol.* 2020;11:.

60. Cai M, Bang S, Zhang P, Lee H. ATM-TCR: TCR-Epitope Binding Affinity Prediction Using a Multi-Head Self-Attention Model. *Front. Immunol.* 2022;13:893247.

61. Weber A, Born J, Rodriguez Martínez M. TITAN: T-cell receptor specificity prediction with bimodal attention networks. *Bioinformatics*. 2021;37(Supplement_1):i237–i244.

62. Lansford JL, Dharmasiri U, Chai S, et al. Computational modeling and confirmation of leukemia-associated minor histocompatibility antigens. *Blood Adv.* 2018;2(16):2052–2062.

63. Griffioen M, van Bergen CAM, Falkenburg JHF. Autosomal Minor Histocompatibility Antigens: How Genetic Variants Create Diversity in Immune Targets. *Front. Immunol.* 2016;7:100.

64. Mullally A, Ritz J. Beyond HLA: the significance of genomic variation for allogeneic hematopoietic stem cell transplantation. *Blood*. 2006;109(4):1355–1362.

65. Huang T, Shu Y, Cai Y-D. Genetic differences among ethnic groups. *BMC Genomics*. 2015;16(1):1093.

66. The MHC sequencing consortium. Complete sequence and gene map of a human major histocompatibility complex. *Nature*. 1999;401(6756):921–923.

67. den Haan JM, Meadows LM, Wang W, et al. The minor histocompatibility antigen HA-1: a diallelic gene with a single amino acid polymorphism. *Science*. 1998;279(5353):1054–1057.

68. Horowitz MM. Current Status of Blood and Marrow Transplantation. *Transplant. Hematol. Oncol.* 2000;11–21.

69. Gale RP, Horowitz MM, Ash RC, et al. Identical-Twin Bone Marrow Transplants for Leukemia. *Ann. Intern. Med.* 1994;120(8):646–652.

70. Martin PJ, Levine DM, Storer BE, et al. Genome-wide minor histocompatibility matching as related to the risk of graft-versus-host disease. *Blood*. 2017;129(6):791–798.

71. Bleakley M, Riddell SR. Exploiting T cells specific for human minor histocompatibility antigens for therapy of leukemia. *Immunol. Cell Biol.* 2011;89(3):396–407.

72. Loke J, Vyas H, Craddock C. Optimizing Transplant Approaches and Post-Transplant Strategies for Patients With Acute Myeloid Leukemia. *Front. Oncol.* 2021;11:666091.

73. Warren EH, Fujii N, Akatsuka Y, et al. Therapy of relapsed leukemia after allogeneic hematopoietic cell transplantation with T cells specific for minor histocompatibility antigens. *Blood*. 2010;115(19):3869–3878.

74. Lee SJ. Classification systems for chronic graft-versus-host disease. *Blood*. 2017;129(1):30–37.

75. Jagasia MH, Greinix HT, Arora M, et al. National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: I. The 2014 Diagnosis and Staging Working Group Report. *Biol. Blood Marrow Transplant.* 2015;21(3):389-401.e1.

76. Schoemans HM, Lee SJ, Ferrara JL, et al. EBMT−NIH−CIBMTR Task Force position statement on standardized terminology & guidance for graft-versus-host disease assessment. *Bone Marrow Transplant.* 2018;53(11):1401–1415.

77. Ferrara JLM, Reddy P. Pathophysiology of Graft-Versus-Host Disease. *Semin. Hematol.* 2006;43(1):3–10.

78. MacDonald KPA, Hill GR, Blazar BR. Chronic graft-versus-host disease: biological insights from preclinical and clinical studies. *Blood*. 2017;129(1):13–21.

79. Zeiser R, Blazar BR. Acute Graft-versus-Host Disease Biology, Prevention and Therapy. *N. Engl. J. Med.* 2017;377(22):2167–2179.

80. Filipovich AH. Diagnosis and manifestations of chronic graft-versus-host disease. *Best Pract. Res. Clin. Haematol.* 2008;21(2):251–257.

81. Sarantopoulos S, Blazar BR, Cutler C, Ritz J. B Cells in Chronic Graft-versus-Host Disease. *Biol. Blood Marrow Transplant.* 2015;21(1):16–23.

82. van Balen P, van Bergen CAM, van Luxemburg-Heijs SAP, et al. CD4 Donor Lymphocyte Infusion Can Cause Conversion of Chimerism Without GVHD by Inducing Immune Responses Targeting Minor Histocompatibility Antigens in HLA Class II. *Front. Immunol.* 2018;9:.

83. HOLLOWAY PA, KALDENHOVEN N, KOK-SCHOEMAKER HM, et al. Antigens shared by malignant plasma cells and normal B cells may be involved in graft versus myeloma. *Clin. Exp. Immunol.* 2003;131(2):340–346.

84. Singh AK, McGuirk JP. Allogeneic Stem Cell Transplantation: A Historical and Scientific Overview. *Cancer Res.* 2016;76(22):6445–6451.

85. E. OSGOOD E, C. RIDDLE M, J. MATHEWS T. APLASTIC ANEMIA TREATED WITH DAILY TRANSFUSIONS AND INTRAVENOUS MARROW; CASE REPORT*. *Ann. Intern. Med.* 2008;

86. Storb R, Epstein RB, Rudolph RH, Thomas ED. Allogeneic canine bone marrow transplantation following cyclophosphamide. *Transplantation*. 1969;7(5):378–386.

87. Davies DA, Colombani J, Viza DC, Dausset J. Human HL-A transplantation antigens: separation of molecules carrying different immunological specificities determined by a single genotype. *Biochem. Biophys. Res. Commun.* 1968;33(1):88–93.

88. Dausset J, Colombani J, Feingold N, Rapaport F. [A LEUKOCYTE GROUP SYSTEM AND ITS RELATIONS WITH HISTOCOMPATIBILITY]. *Nouv. Rev. Fr. Hematol.* 1965;5:17–22.

89. Dausset J, Rapaport FT, Legrand L, et al. [Studies on transplantation antigens (HL-A) by means of skin grafts from 90 children onto their fathers]. *Nouv. Rev. Fr. Hematol.* 1969;9(2):215–229.

90. Dausset J, Hors J, Bigot J. [Genotypic study of HL-A histocompatibility in 91 kidney transplants]. *Presse Med.* 1969;77(47):1699–1704.

91. Dausset J, Colombani J, Legrand L, Feingold N. [The sub-loci of the HL-A system. The main system of histocompatibility in man]. *Presse Med.* 1969;77(23):849–852.

92. Rapaport FT, Dausset J, Lawrence HS, Converse JM. An approach to the facilitation of skin allograft survival in man. *Surg. Forum*. 1966;17:198–199.

93. Thomas ED, Buckner CD, Banaji M, et al. One hundred patients with acute leukemia treated by chemotherapy, total body irradiation, and allogeneic marrow transplantation. *Blood*. 1977;49(4):511–533.

94. Horowitz MM, Gale RP, Sondel PM, et al. Graft-versus-leukemia reactions after bone marrow transplantation. *Blood*. 1990;75(3):555–562.

95. Hamilton BK. Current approaches to prevent and treat GVHD after allogeneic stem cell transplantation. *Hematology*. 2018;2018(1):228–235.

96. Malard F, Huang X-J, Sim JPY. Treatment and unmet needs in steroid-refractory acute graft-versus-host disease. *Leukemia*. 2020;34(5):1229–1240.

97. Story CM, Wang T, Bhatt VR, et al. Genetics of HLA Peptide Presentation and Impact on Outcomes in HLA-Matched Allogeneic Hematopoietic Cell Transplantation. *Transplant. Cell. Ther.* 2021;27(7):591–599.

98. Loke J, Buka R, Craddock C. Allogeneic Stem Cell Transplantation for Acute Myeloid Leukemia: Who, When, and How? *Front. Immunol.* 2021;12:.

99. Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*. 2017;129(4):424–447.

100. Rowe JM. Optimal induction and post-remission therapy for AML in first remission. *Hematology*. 2009;2009(1):396–405.

101. Rowe JM, Tallman MS. How I treat acute myeloid leukemia. *Blood*. 2010;116(17):3147–3156.

102. Vasu S, Kohlschmidt J, Mrózek K, et al. Ten-year outcome of patients with acute myeloid leukemia not treated with allogeneic transplantation in first complete remission. *Blood Adv.* 2018;2(13):1645–1650.

103. Oran B, Weisdorf DJ. Survival for older patients with acute myeloid leukemia: a population-based study. *Haematologica*. 2012;97(12):1916–1924.

104. Zulu S, Kenyon M. Principles of Conditioning Therapy and Cell Infusion. *Eur. Blood Marrow Transplant. Textb. Nurses Auspices EBMT*. 2018;

105. Sengsayadeth S, Savani BN, Blaise D, et al. Reduced intensity conditioning allogeneic hematopoietic cell transplantation for adult acute myeloid leukemia in complete remission - a review from the Acute Leukemia Working Party of the EBMT. *Haematologica*. 2015;100(7):859–869.

106. BeTheMatch.org. How does a patient’s ethnic background affect matching? 2022;

107. Khaddour K, Hana CK, Mewawalla P. Hematopoietic Stem Cell Transplantation. *StatPearls*. 2022;

108. Anasetti C, Logan BR, Lee SJ, et al. Peripheral-blood stem cells versus bone marrow from unrelated donors. *N. Engl. J. Med.* 2012;367(16):1487–1496.

109. Dehn J, Spellman S, Hurley CK, et al. Selection of unrelated donors and cord blood units for hematopoietic cell transplantation: guidelines from the NMDP/CIBMTR. *Blood*. 2019;134(12):924–934.

110. Tiercy J-M. How to select the best available related or unrelated donor of hematopoietic stem cells? *Haematologica*. 2016;101(6):680–687.

111. Rutten CE, Luxemburg-Heijs SAP van, Meijden ED van der, et al. HLA-DPB1 Mismatching Results in the Generation of a Full Repertoire of HLA-DPB1-Specific CD4+ T Cell Responses Showing Immunogenicity of all HLA-DPB1 Alleles. *Biol. Blood Marrow Transplant.* 2010;16(9):1282–1292.

112. Sizzano F, Zito L, Crivello P, et al. Significantly higher frequencies of alloreactive CD4+ T cells responding to nonpermissive than to permissive HLA-DPB1 T-cell epitope disparities. *Blood*. 2010;116(11):1991–1992.

113. Zino E, Frumento G, Marktel S, et al. A T-cell epitope encoded by a subset of HLA-DPB1 alleles determines nonpermissive mismatches for hematologic stem cell transplantation. *Blood*. 2004;103(4):1417–1424.

114. Fleischhauer K, Shaw BE. HLA-DP in unrelated hematopoietic cell transplantation revisited: challenges and opportunities. *Blood*. 2017;130(9):1089–1096.

115. Kawase T, Morishima Y, Matsuo K, et al. High-risk HLA allele mismatch combinations responsible for severe acute graft-versus-host disease and implication for its molecular mechanism. *Blood*. 2007;110(7):2235–2241.

116. Morishima Y, Sasazuki T, Inoko H, et al. The clinical significance of human leukocyte antigen (HLA) allele compatibility in patients receiving a marrow transplant from serologically HLA-A, HLA-B, and HLA-DR matched unrelated donors. *Blood*. 2002;99(11):4200–4206.

117. Petersdorf EW, Anasetti C, Martin PJ, et al. Limits of HLA mismatching in unrelated hematopoietic cell transplantation. *Blood*. 2004;104(9):2976–2980.

118. Gooptu M, Romee R, St. Martin A, et al. HLA-haploidentical vs matched unrelated donor transplants with posttransplant cyclophosphamide-based prophylaxis. *Blood*. 2021;138(3):273–282.

119. Fuchs EJ. Haploidentical transplantation for hematologic malignancies: where do we stand? *Hematol. Educ. Program Am. Soc. Hematol. Am. Soc. Hematol. Educ. Program*. 2012;2012:230–236.

120. Saber W, Opie S, Rizzo JD, et al. Outcomes after matched unrelated donor versus identical sibling hematopoietic cell transplantation in adults with acute myelogenous leukemia. *Blood*. 2012;119(17):3908–3916.

121. Olsen KS, Jadi O, Dexheimer S, et al. Shared graft-vs-leukemia minor histocompatibility antigens in DISCOVeRY-BMT. *Blood Adv.* 2022;bloodadvances.2022008863.

122. Gibson T, Medawar PB. The fate of skin homografts in man. *J. Anat.* 1943;77(Pt 4):299-310.4.

123. Kohlhauser M, Luze H, Nischwitz SP, Kamolz LP. Historical Evolution of Skin Grafting—A Journey through Time. *Medicina (Mex.)*. 2021;57(4):348.

124. Owen RD. IMMUNOGENETIC CONSEQUENCES OF VASCULAR ANASTOMOSES BETWEEN BOVINE TWINS. *Science*. 1945;102(2651):400–401.

125. Silverstein AM. The curious case of the 1960 Nobel Prize to Burnet and Medawar. *Immunology*. 2016;147(3):269–274.

126. Roopenian D. A Methods Paper That Led to Much More. *J. Immunol. Baltim. Md 1950*. 2014;192(1):3–4.

127. Gorer PA, Lyman S, Snell GD. Studies on the Genetic and Antigenic Basis of Tumour Transplantation. Linkage between a Histocompatibility Gene and “Fused” in Mice. *Proc. R. Soc. Lond. B Biol. Sci.* 1948;135(881):499–505.

128. Dausset J. [Iso-leuko-antibodies]. *Acta Haematol.* 1958;20(1–4):156–166.

129. Carosella ED. From MAC to HLA: Professor Jean Dausset, the pioneer. *Hum. Immunol.* 2009;70(9):661–662.

130. Levine BB, Ojeda A, Benacerraf B. STUDIES ON ARTIFICIAL ANTIGENS. III. THE GENETIC CONTROL OF THE IMMUNE RESPONSE TO HAPTEN-POLY-L-LYSINE CONJUGATES IN GUINEA PIGS. *J. Exp. Med.* 1963;118(6):953–957.

131. Green I, Paul WE, Benacerraf B. Histocompatibility-linked genetic control of the immune response to hapten guinea pig albumin conjugates in inbred guinea pigs. *J. Immunol. Baltim. Md 1950*. 1972;109(3):457–463.

132. Goulmy E, Termijtelen A, Bradley BA, Van Rood JJ. Y-antigen killing by T cells of women is restricted by HLA. *Nature*. 1977;266(5602):544–545.

133. Goulmy E, Termijtelen A, Bradley BA, van Rood JJ. Alloimmunity to human H-Y. *Lancet Lond. Engl.* 1976;2(7996):1206.

134. Eichwald EJ, Silmser CR. Skin. *Transplant. Bull.* 1955;2:148–149.

135. Simpson E. Review Lecture: Immunology of H-Y Antigen and its Role in Sex Determination. *Proc. R. Soc. Lond. B Biol. Sci.* 1983;220(1218):31–46.

136. Popli R, Sahaf B, Nakasone H, Lee JYY, Miklos DB. Clinical impact of H-Y alloimmunity. *Immunol. Res.* 2014;58(0):249–258.

137. Falk K, Rötzschke O, Rammensee H-G. Cellular peptide composition governed by major histocompatibility complex class I molecules. *Nature*. 1990;348(6298):248–251.

138. Loveland B, Simpson E. The non-MHC transplantation antigens: neither weak nor minor. *Immunol. Today*. 1986;7(7):223–229.

139. Mendoza LM, Paz P, Zuberi A, et al. Minors held by majors: the H13 minor histocompatibility locus defined as a peptide/MHC class I complex. *Immunity*. 1997;7(4):461–472.

140. de Bueger M, Bakker A, Van Rood JJ, Van der Woude F, Goulmy E. Tissue distribution of human minor histocompatibility antigens. Ubiquitous versus restricted tissue distribution indicates heterogeneity among human cytotoxic T lymphocyte-defined non-MHC antigens. *J. Immunol.* 1992;149(5):1788–1794.

141. den Haan J, Sherman N, Blokland E, et al. Identification of a graft versus host disease-associated human minor histocompatibility antigen. *Science*. 1995;268(5216):1476–1480.

142. Goulmy E. Human minor histocompatibility antigens. *Curr. Opin. Immunol.* 1996;8(1):75–81.

143. van Els CACM, D'Amaro J, Pool J, et al. Immunogenetics of human minor histocompatibility antigens: their polymorphism and immunodominance. *Immunogenetics*. 1992;35(3):161–165.

144. Goulmy E. Minor histocompatibility antigens: allo target molecules for tumor-specific immunotherapy. *Cancer J. Sudbury Mass*. 2004;10(1):1–7.

145. Dolstra H, Fredrix H, Maas F, et al. A Human Minor Histocompatibility Antigen Specific for B Cell Acute Lymphoblastic Leukemia. *J. Exp. Med.* 1999;189(2):301–308.

146. Brickner AG, Warren EH, Caldwell JA, et al. The Immunogenicity of a New Human Minor Histocompatibility Antigen Results from Differential Antigen Processing. *J. Exp. Med.* 2001;193(2):195–206.

147. Akatsuka Y, Nishida T, Kondo E, et al. Identification of a Polymorphic Gene, BCL2A1, Encoding Two Novel Hematopoietic Lineage-specific Minor Histocompatibility Antigens. *J. Exp. Med.* 2003;197(11):1489–1500.

148. Spierings E, Brickner AG, Caldwell JA, et al. The minor histocompatibility antigen HA-3 arises from differential proteasome–mediated cleavage of the lymphoid blast crisis (Lbc) oncoprotein. *Blood*. 2003;102(2):621–629.

149. Murata M, Warren EH, Riddell SR. A Human Minor Histocompatibility Antigen Resulting from Differential Expression due to a Gene Deletion. *J. Exp. Med.* 2003;197(10):1279–1289.

150. Rijke B de, Horssen-Zoetbrood A van, Beekman JM, et al. A frameshift polymorphism in *P2X5* elicits an allogeneic cytotoxic T lymphocyte response

associated with remission of chronic myeloid leukemia. *J. Clin. Invest.* 2005;115(12):3506–3516.

151. Torikai H, Akatsuka Y, Miyazaki M, et al. The human cathepsin H gene encodes two novel minor histocompatibility antigen epitopes restricted by HLA-A*3101 and -A*3303. *Br. J. Haematol.* 2006;134(4):406–416.

152. Brickner AG, Evans AM, Mito JK, et al. The PANE1 gene encodes a novel human minor histocompatibility antigen that is selectively expressed in B-lymphoid cells and B-CLL. *Blood*. 2006;107(9):3779–3786.

153. Warren EH, Vigneron NJ, Gavin MA, et al. An Antigen Produced by Splicing of Noncontiguous Peptides in the Reverse Order. *Science*. 2006;313(5792):1444–1447.

154. van Bergen CAM, Kester MGD, Jedema I, et al. Multiple myeloma–reactive T cells recognize an activation-induced minor histocompatibility antigen encoded by the ATP-dependent interferon-responsive (ADIR) gene. *Blood*. 2007;109(9):4089–4096.

155. Wölfel C, Lennerz V, Lindemann E, et al. Dissection and molecular analysis of alloreactive CD8+ T cell responses in allogeneic haematopoietic stem cell transplantation. *Cancer Immunol. Immunother.* 2008;57(6):849–857.

156. Tykodi SS, Fujii N, Vigneron N, et al. C19orf48 Encodes a Minor Histocompatibility Antigen Recognized by CD8+ Cytotoxic T Cells from Renal Cell Carcinoma Patients. *Clin. Cancer Res.* 2008;14(16):5260–5269.

157. Spaapen RM, Lokhorst HM, van den Oudenalder K, et al. Toward targeting B cell cancers with CD4+ CTLs: identification of a CD19-encoded minor histocompatibility antigen using a novel genome-wide analysis. *J. Exp. Med.* 2008;205(12):2863–2872.

158. Griffioen M, van der Meijden ED, Slager EH, et al. Identification of phosphatidylinositol 4-kinase type II β as HLA class II-restricted target in graft versus leukemia reactivity. *Proc. Natl. Acad. Sci.* 2008;105(10):3837–3842.

159. Stumpf AN, van der Meijden ED, van Bergen CAM, et al. Identification of 4 new HLA-DR–restricted minor histocompatibility antigens as hematopoietic targets in antitumor immunity. *Blood*. 2009;114(17):3684–3692.

160. Spaapen RM, de Kort RAL, van den Oudenalder K, et al. Rapid Identification of Clinical Relevant Minor Histocompatibility Antigens via Genome-Wide Zygosity-Genotype Correlation Analysis. *Clin. Cancer Res.* 2009;15(23):7137–7143.

161. Kamei M, Nannya Y, Torikai H, et al. HapMap scanning of novel human minor histocompatibility antigens. *Blood*. 2009;113(21):5041–5048.

162. Bleakley M, Otterud BE, Richardt JL, et al. Leukemia-associated minor histocompatibility antigen discovery using T-cell clones isolated by in vitro stimulation of naive CD8+ T cells. *Blood*. 2010;115(23):4923–4933.

163. Van Bergen CAM, Rutten CE, Van Der Meijden ED, et al. High-Throughput Characterization of 10 New Minor Histocompatibility Antigens by Whole Genome Association Scanning. *Cancer Res.* 2010;70(22):9073–9083.

164. Oostvogels R, Lokhorst HM, Mutis T. Minor histocompatibility Ags: identification strategies, clinical results and translational perspectives. *Bone Marrow Transplant.* 2016;51(2):163–171.

165. Dharamsiri U, Hunsucker SA, Vincent BG, et al. UNC-GRK4-1: An Allele Specific Cancer Testis Antigen Identified Through Genomic Screening. *Blood*. 2013;122(21):3246.

166. Amado-Azevedo J, Reinhard NR, van Bezu J, et al. The minor histocompatibility antigen 1 (HMHA1)/ArhGAP45 is a RacGAP and a novel regulator of endothelial integrity. *Vascul. Pharmacol.* 2018;101:38–47.

167. Kremer AN, Bausenwein J, Lurvink E, et al. Discovery and Differential Processing of HLA Class II-Restricted Minor Histocompatibility Antigen LB-PIP4K2A-1S and Its Allelic Variant by Asparagine Endopeptidase. *Front. Immunol.* 2020;11:381.

168. Pont MJ, van der Lee DI, van der Meijden ED, et al. Integrated Whole Genome and Transcriptome Analysis Identified a Therapeutic Minor Histocompatibility Antigen in a Splice Variant of ITGB2. *Clin. Cancer Res.* 2016;22(16):4185–4196.

169. Armistead PM, Liang S, Li H, et al. Common Minor Histocompatibility Antigen Discovery Based upon Patient Clinical Outcomes and Genomic Data. *PLOS ONE*. 2011;6(8):e23217.

170. Hombrink P, Hadrup SR, Bakker A, et al. High-Throughput Identification of Potential Minor Histocompatibility Antigens by MHC Tetramer-Based Screening: Feasibility and Limitations. *PLOS ONE*. 2011;6(8):e22523.

171. Rodenko B, Toebes M, Hadrup SR, et al. Generation of peptide–MHC class I complexes through UV-mediated ligand exchange. *Nat. Protoc.* 2006;1(3):1120–1132.

172. Bakker AH, Hoppes R, Linnemann C, et al. Conditional MHC class I ligands and peptide exchange technology for the human MHC gene products HLA-A1, -A3, -A11, and -B7. *Proc. Natl. Acad. Sci.* 2008;105(10):3825–3830.

173. Hadrup SR, Bakker AH, Shu CJ, et al. Parallel detection of antigen-specific T-cell responses by multidimensional encoding of MHC multimers. *Nat. Methods*. 2009;6(7):520–526.

174. Bentzen AK, Marquard AM, Lyngaa R, et al. Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nat. Biotechnol.* 2016;34(10):1037–1045.

175. Rasmussen M, Fenoy E, Harndahl M, et al. Pan-Specific Prediction of Peptide–MHC Class I Complex Stability, a Correlate of T Cell Immunogenicity. *J. Immunol.* 2016;197(4):1517–1524.

176. Dolstra H, Rijke B de, Fredrix H, et al. Bi-directional allelic recognition of the human minor histocompatibility antigen HB-1 by cytotoxic T lymphocytes. *Eur. J. Immunol.* 2002;32(10):2748–2758.

177. Spaapen R, Mutis T. Targeting haematopoietic-specific minor histocompatibility antigens to distinguish graft-versus-tumour effects from graft-versus-host disease. *Best Pract. Res. Clin. Haematol.* 2008;21(3):543–557.

178. Broen K, Levenga H, Vos J, et al. A Polymorphism in the Splice Donor Site of ZNF419 Results in the Novel Renal Cell Carcinoma-Associated Minor Histocompatibility Antigen ZAPHIR. *PLOS ONE*. 2011;6(6):e21699.

179. Griffioen M, Honders MW, Meijden ED van der, et al. Identification of 4 novel HLA-B*40:01 restricted minor histocompatibility antigens and their potential as targets for graft-versus-leukemia reactivity. *Haematologica*. 2012;97(8):1196–1204.

180. Hombrink P, Hassan C, Kester MGD, et al. Discovery of T cell epitopes implementing HLA-peptidomics into a reverse immunology approach. *J. Immunol. Baltim. Md 1950*. 2013;190(8):3869–3877.

181. Oostvogels R, Lokhorst HM, Minnema MC, et al. Identification of minor histocompatibility antigens based on the 1000 Genomes Project. *Haematologica*. 2014;99(12):1854–1859.

182. van Bergen CAM, Verdegaal EME, Honders MW, et al. Durable Remission of Renal Cell Carcinoma in Conjuncture with Graft versus Host Disease following Allogeneic Stem Cell Transplantation and Donor Lymphocyte Infusion: Rule or Exception? *PLoS ONE*. 2014;9(1):e85198.

183. Oostvogels R, Minnema MC, van Elk M, et al. Towards effective and safe immunotherapy after allogeneic stem cell transplantation: identification of hematopoietic-specific minor histocompatibility antigen UTA2-1. *Leukemia*. 2013;27(3):642–649.

184. Hombrink P, Hassan C, Kester MGD, et al. Identification of Biological Relevant Minor Histocompatibility Antigens within the B-lymphocyte–Derived HLA-Ligandome Using a Reverse Immunology Approach. *Clin. Cancer Res.* 2015;21(9):2177–2186.

185. Fred Hutchinson Cancer Center. Phase I Study of Adoptive Immunotherapy With CD8 Minor Histocompatibility (H) Antigen-Specific CTL Clones for Patients With

Relapsed of AML or ALL After Allogeneic Hematopoietic Stem Cell Transplant. clinicaltrials.gov; 2010.

186. Janelle V, Rulleau C, Del Testa S, Carli C, Delisle J-S. T-Cell Immunotherapies Targeting Histocompatibility and Tumor Antigens in Hematological Malignancies. *Front. Immunol.* 2020;11:276.

187. Li N, Matte-Martone C, Zheng H, et al. Memory T cells from minor histocompatibility antigen–vaccinated and virus-immune donors improve GVL and immune reconstitution. *Blood*. 2011;118(22):5965–5976.

188. Loenen MM van, Boer R de, Hagedoorn RS, et al. Optimization of the HA-1-specific T-cell receptor for gene therapy of hematologic malignancies. *Haematologica*. 2011;96(3):477–481.

189. Loenen MM van, Boer R de, Liempt E van, et al. A Good Manufacturing Practice procedure to engineer donor virus-specific T cells into potent anti-leukemic effector cells. *Haematologica*. 2014;99(4):759–768.

190. Meij P, Jedema I, van der Hoorn MAWG, et al. Generation and administration of HA-1-specific T-cell lines for the treatment of patients with relapsed leukemia after allogeneic stem cell transplantation: a pilot study. *Haematologica*. 2012;97(8):1205–1208.

191. Medigene AG. A Dose-Escalation, Open Label Phase I Study to Assess the Safety, Feasibility and Preliminary Efficacy of HA-1H TCR Modified T Cells, MDG1021, in Patients With Relapsed or Persistent Hematologic Malignancies After Allogeneic HSCT With or Without Unmanipulated DLI. clinicaltrials.gov; 2021.

192. Inaguma Y, Akahori Y, Murayama Y, et al. Construction and molecular characterization of a T-cell receptor-like antibody and CAR-T cells specific for minor histocompatibility antigen HA-1H. *Gene Ther.* 2014;21(6):575–584.

193. University of Chicago. A Phase I/II Study of Vaccination Against Minor Histocompatibility Antigens HA1 or HA2 After Allogeneic Stem Cell Transplantation for Advanced Hematologic Malignancies. clinicaltrials.gov; 2014.

194. Oostvogels R, Kneppers E, Minnema MC, et al. Efficacy of host-dendritic cell vaccinations with or without minor histocompatibility antigen loading, combined with donor lymphocyte infusion in multiple myeloma patients. *Bone Marrow Transplant.* 2017;52(2):228–237.

195. Radboud University Medical Center. Vaccination With PD-L1/L2-silenced Minor Histocompatibility Antigen-loaded Donor DC Vaccines to Boost Graft-versus-tumor Immunity After Allogeneic Stem Cell Transplantation (a Phase I/II Study). clinicaltrials.gov; 2021.

196. Dossa RG, Cunningham T, Sommermeyer D, et al. Development of T-cell immunotherapy for hematopoietic stem cell transplantation recipients at risk of leukemia relapse. *Blood*. 2018;131(1):108–120.

197. Fred Hutchinson Cancer Center. Phase I Study of Adoptive Immunotherapy With CD8+ and CD4+ Memory T Cells Transduced to Express an HA-1-Specific T Cell Receptor (TCR) for Children and Adults With Recurrent Acute Leukemia After Allogeneic Hematopoietic Stem Cell Transplantation (HCT). clinicaltrials.gov; 2022.

198. Roex MCJ, van Balen P, Germeroth L, et al. Generation and infusion of multi-antigen-specific T cells to prevent complications early after T-cell depleted allogeneic stem cell transplantation—a phase I/II study. *Leukemia*. 2020;34(3):831–844.

199. Ciusss de L'Est de l'Île de Montréal. An Exploratory, Open-label, Multicenter Study to Evaluate the Safety and Efficacy of Anti-minor Histocompatibility Complex (MiHA) Donor T-lymphocytes Expanded ex Vivo, in Patients With a Hematologic Malignancy, With Molecular or Clinical Relapse After Hematopoietic Stem Cell Transplantation From a Matched Donor. clinicaltrials.gov; 2017.

200. Staff N. The ALFA dataset: New aggregated allele frequency from dbGaP and dbSNP now available. *NCBI Insights*. 2020;

201. Maiers M, Gragert L, Klitz W. High-resolution HLA alleles and haplotypes in the United States population. *Hum. Immunol.* 2007;68(9):779–788.

202. Hahn T, Sucheston-Campbell LE, Preus L, et al. Establishment of Definitions and Review Process for Consistent Adjudication of Cause-specific Mortality after Allogeneic Unrelated-donor Hematopoietic Cell Transplantation. *Biol. Blood Marrow Transplant.* 2015;21(9):1679–1686.

203. Wang J, Clay-Gilmour AI, Karaesmen E, et al. Genome-Wide Association Analyses Identify Variants in IRF4 Associated With Acute Myeloid Leukemia and Myelodysplastic Syndrome Susceptibility. *Front. Genet.* 2021;12:.

204. Tang H, Hahn T, Karaesmen E, et al. Validation of genetic associations with acute GVHD and nonrelapse mortality in DISCOVeRY-BMT. *Blood Adv.* 2019;3(15):2337–2341.

205. Karaesmen E, Rizvi AA, Preus LM, et al. Replication and validation of genetic polymorphisms associated with survival after allogeneic blood or marrow transplant. *Blood*. 2017;130(13):1585–1596.

206. Hahn T, Wang J, Preus LM, et al. Novel genetic variants associated with mortality after unrelated donor allogeneic hematopoietic cell transplantation. *eClinicalMedicine*. 2021;40:.

207. Zhu Q, Yan L, Liu Q, et al. Exome chip analyses identify genes affecting mortality after HLA-matched unrelated-donor blood and marrow transplantation. *Blood*. 2018;131(22):2490–2499.

208. Baldwin RM, Owzar K, Zembutsu H, et al. A Genome-Wide Association Study Identifies Novel Loci for Paclitaxel-Induced Sensory Peripheral Neuropathy in CALGB 40101. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 2012;18(18):5099–5109.

209. Witherspoon DJ, Wooding S, Rogers AR, et al. Genetic Similarities Within and Between Human Populations. *Genetics*. 2007;176(1):351–359.

210. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164–e164.

211. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74.

212. Kent WJ, Sugnet CW, Furey TS, et al. The Human Genome Browser at UCSC. *Genome Res.* 2002;12(6):996–1006.

213. Jadi O, Tang H, Olsen K, et al. Associations of Minor Histocompatibility Antigens with Clinical Outcomes Following Allogeneic Hematopoietic Cell Transplantation. 2022;2022.08.31.506092.

214. The Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* 2013;45(10):1113–1120.

215. Golub TR, Slonim DK, Tamayo P, et al. Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. *Science*. 1999;286(5439):531–537.

216. CardioRenal Systems, Inc. TARGET: A Study to Evaluate the Treatment of Patients With Acute Decompensated Heart Failure (ADHF) Using an Automated Fluid Management System. clinicaltrials.gov; 2019.

217. Loh ML, Zhang J, Harvey RC, et al. Tyrosine kinome sequencing of pediatric acute lymphoblastic leukemia: a report from the Children's Oncology Group TARGET Project. *Blood*. 2013;121(3):485–488.

218. Liu H. Emerging agents and regimens for AML. *J. Hematol. Oncol.J Hematol Oncol*. 2021;14(1):49.

219. Pidala J, Kim J, Schell M, et al. Race/ethnicity affects the probability of finding an HLA-A, -B, -C and -DRB1 allele-matched unrelated donor and likelihood of subsequent transplant utilization. *Bone Marrow Transplant.* 2013;48(3):346–350.

220. Hoffmann TJ, Zhan Y, Kvale MN, et al. Design and coverage of high throughput genotyping arrays optimized for individuals of East Asian, African American, and Latino race/ethnicity using imputation and a novel hybrid SNP selection algorithm. *Genomics*. 2011;98(6):422–430.

221. Medina-Gomez C, Felix JF, Estrada K, et al. Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study. *Eur. J. Epidemiol.* 2015;30(4):317–330.

222. Knerr S, Wayman D, Bonham VL. Inclusion of Racial and Ethnic Minorities in Genetic Research: Advance the Spirit by Changing the Rules? *J. Law Med. Ethics J. Am. Soc. Law Med. Ethics*. 2011;39(3):502–512.

223. Mersha TB, Abebe T. Self-reported race/ethnicity in the age of genomic research: its potential impact on understanding health disparities. *Hum. Genomics*. 2015;9(1):1.

224. Lachance J, Tishkoff SA. SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. *BioEssays News Rev. Mol. Cell. Dev. Biol.* 2013;35(9):780–786.

225. Eller E. Effects of ascertainment bias on recovering human demographic history. *Hum. Biol.* 2001;73(3):411–427.

226. Chong C, Coukos G, Bassani-Sternberg M. Identification of tumor antigens with immunopeptidomics. *Nat. Biotechnol.* 2022;40(2):175–188.

227. Hu A, Noble WS, Wolf-Yadlin A. Technical advances in proteomics: new developments in data-independent acquisition. *F1000Research*. 2016;5:F1000 Faculty Rev-419.

228. Li KW, Gonzalez-Lozano MA, Koopmans F, Smit AB. Recent Developments in Data Independent Acquisition (DIA) Mass Spectrometry: Application of Quantitative Analysis of the Brain Proteome. *Front. Mol. Neurosci.* 2020;13:.

229. Faria SS, Morris CFM, Silva AR, et al. A Timely Shift from Shotgun to Targeted Proteomics and How It Can Be Groundbreaking for Cancer Research. *Front. Oncol.* 2017;7:13.

230. Olshina MA, Sharon M. Mass Spectrometry: A Technique of Many Faces. *Q. Rev. Biophys.* 2016;49:e18.

231. Davies V, Wandy J, Weidt S, et al. Rapid Development of Improved Data-Dependent Acquisition Strategies. *Anal. Chem.* 2021;93(14):5676–5683.

232. Faridi P, Purcell AW, Croft NP. In Immunopeptidomics We Need a Sniper Instead of a Shotgun. *PROTEOMICS*. 2018;18(12):1700464.

233. Ritz D, Kinzi J, Neri D, Fugmann T. Data-Independent Acquisition of HLA Class I Peptidomes on the Q Exactive Mass Spectrometer Platform. *PROTEOMICS*. 2017;17(19):1700177.

234. Mommen GPM, Frese CK, Meiring HD, et al. Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proc. Natl. Acad. Sci.* 2014;111(12):4507–4512.

235. Ronsein GE, Pamir N, von Haller PD, et al. Parallel reaction monitoring (PRM) and selected reaction monitoring (SRM) exhibit comparable linearity, dynamic range and precision for targeted quantitative HDL proteomics. *J. Proteomics*. 2015;0:388–399.

236. Vidova V, Spacil Z. A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition. *Anal. Chim. Acta*. 2017;964:7–23.

237. Rauniyar N. Parallel Reaction Monitoring: A Targeted Experiment Performed Using High Resolution and High Mass Accuracy Mass Spectrometry. *Int. J. Mol. Sci.* 2015;16(12):28566–28581.

238. Park J, Talukder AH, Lim SA, et al. SLC45A2: A Melanoma Antigen with High Tumor Selectivity and Reduced Potential for Autoimmune Toxicity. *Cancer Immunol. Res.* 2017;5(8):618–629.

239. Sarkizova S, Klaeger S, Le PM, et al. A large peptidome dataset improves HLA class I epitope prediction across most of the human population. *Nat. Biotechnol.* 2020;38(2):199–209.

240. Trentini DB, Pecoraro M, Tiwary S, et al. Role for ribosome-associated quality control in sampling proteins for MHC class I-mediated antigen presentation. *Proc. Natl. Acad. Sci.* 2020;117(8):4099–4108.

241. Lubben NB, Sahlender DA, Motley AM, et al. HIV-1 Nef-induced Down-Regulation of MHC Class I Requires AP-1 and Clathrin but Not PACS-1 and Is Impeded by AP-2. *Mol. Biol. Cell*. 2007;18(9):3351–3365.

242. Kalli A, Smith GT, Sweredoski MJ, Hess S. Evaluation and Optimization of Mass Spectrometric Settings during Data-Dependent Acquisition Mode: Focus on LTQ-Orbitrap Mass Analyzers. *J. Proteome Res.* 2013;12(7):3071–3086.

243. Zhang M, Sukhumalchandra P, Enyenihi AA, et al. A novel HLA-A*0201 restricted peptide derived from cathepsin G is an effective immunotherapeutic target in acute myeloid leukemia. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 2013;19(1):247–257.

244. Sarri CA, Papadopoulos GE, Papa A, et al. Amino acid signatures in the HLA class II peptide-binding region associated with protection/susceptibility to the severe West Nile Virus disease. *PLoS ONE*. 2018;13(10):e0205557.

245. Piehowski PD, Petyuk VA, Orton DJ, et al. Sources of Technical Variability in Quantitative LC-MS Proteomics: Human Brain Tissue Sample Analysis. *J. Proteome Res.* 2013;12(5):2128–2137.

246. Tabb DL, Vega-Montoto L, Rudnick PA, et al. Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography—Tandem Mass Spectrometry. *J. Proteome Res.* 2010;9(2):761.

247. Kote S, Pirog A, Bedran G, Alfaro J, Dapic I. Mass Spectrometry-Based Identification of MHC-Associated Peptides. *Cancers*. 2020;12(3):535.

248. Borland K, Diesend J, Ito-Kureha T, et al. Production and Application of Stable Isotope-Labeled Internal Standards for RNA Modification Analysis. *Genes*. 2019;10(1):26.

249. Stokvis E, Rosing H, Beijnen JH. Stable isotopically labeled internal standards in quantitative bioanalysis using liquid chromatography/mass spectrometry: necessity or not? *Rapid Commun. Mass Spectrom.* 2005;19(3):401–407.

250. Prasad B, Unadkat JD. Comparison of Heavy Labeled (SIL) Peptide versus SILAC Protein Internal Standards for LC-MS/MS Quantification of Hepatic Drug Transporters. *Int. J. Proteomics*. 2014;2014:451510.

251. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.

252. Scholtalbers J, Boegel S, Bukur T, et al. TCLP: an online cancer cell line catalogue integrating HLA type, predicted neo-epitopes, virus and gene expression. *Genome Med.* 2015;7:118.

253. Cornaby C, Montgomery MC, Liu C, Weimer ET. Unique Molecular Identifier-Based High-Resolution HLA Typing and Transcript Quantitation Using Long-Read Sequencing. *Front. Genet.* 2022;13:.

254. Terai YL, Huang C, Wang B, et al. Valid-NEO: A Multi-Omics Platform for Neoantigen Detection and Quantification from Limited Clinical Samples. *Cancers*. 2022;14(5):1243.

255. Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res.* 1999;9(8):677–679.

256. Roy DC, Perreault C. Major vs minor histocompatibility antigens. *Blood*. 2017;129(6):664–666.

257. Pearson H, Daouda T, Granados DP, et al. MHC class I–associated peptides derive from selective regions of the human genome. *J. Clin. Invest.* 2016;126(12):4690–4701.

258. Spellman S, Warden MB, Haagenson M, et al. Effects of Mismatching for Minor Histocompatibility Antigens on Clinical Outcomes in HLA-Matched, Unrelated Hematopoietic Stem Cell Transplants. *Biol. Blood Marrow Transplant.* 2009;15(7):856–863.

259. Spierings E, Kim Y-H, Hendriks M, et al. Multicenter analyses demonstrate significant clinical effects of minor histocompatibility antigens on GvHD and GvL after HLA-matched related and unrelated hematopoietic stem cell transplantation. *Biol. Blood Marrow Transplant. J. Am. Soc. Blood Marrow Transplant.* 2013;19(8):1244–1253.

260. Marijt WAE, Heemskerk MHM, Kloosterboer FM, et al. Hematopoiesis-restricted minor histocompatibility antigens HA-1- or HA-2-specific T cells can induce complete remissions of relapsed leukemia. *Proc. Natl. Acad. Sci.* 2003;100(5):2742–2747.

261. Gallardo D, Aróstegui JI, Balas A, et al. Disparity for the minor histocompatibility antigen HA-1 is associated with an increased risk of acute graft-versus-host disease (GvHD) but it does not affect chronic GvHD incidence, disease-free survival or overall survival after allogeneic human leucocyte antigen-identical sibling donor transplantation. *Br. J. Haematol.* 2001;114(4):931–936.

262. Katagiri T, Shiobara S, Nakao S, et al. Mismatch of minor histocompatibility antigen contributes to a graft-versus-leukemia effect rather than to acute GVHD, resulting in long-term survival after HLA-identical stem cell transplantation in Japan. *Bone Marrow Transplant.* 2006;38(10):681–686.

263. Pierce RA, Field ED, Mutis T, et al. The HA-2 Minor Histocompatibility Antigen Is Derived from a Diallelic Gene Encoding a Novel Human Class I Myosin Protein. *J. Immunol.* 2001;167(6):3223–3230.

264. Goulmy E, Schipper R, Pool J, et al. Mismatches of Minor Histocompatibility Antigens between HLA-Identical Donors and Recipients and the Development of Graft-Versus-Host Disease after Bone Marrow Transplantation. *N. Engl. J. Med.* 1996;334(5):281–285.

265. Akatsuka Y, Warren EH, Gooley TA, et al. Disparity for a newly identified minor histocompatibility antigen, HA-8, correlates with acute graft-versus-host disease after haematopoietic stem cell transplantation from an HLA-identical sibling. *Br. J. Haematol.* 2003;123(4):671–675.

266. McCarroll SA, Bradner JE, Turpeinen H, et al. Donor-recipient mismatch for common gene deletion polymorphisms in graft-versus-host disease. *Nat. Genet.* 2009;41(12):1341–1344.

267. Stephens AJ, Burgess-Brown NA, Jiang S. Beyond Just Peptide Antigens: The Complex World of Peptide-Based Cancer Vaccines. *Front. Immunol.* 2021;12:.

268. Kumai T, Kobayashi H, Harabuchi Y, Celis E. Peptide vaccines in cancer - old concept revisited. *Curr. Opin. Immunol.* 2017;45:1–7.

269. Liu J, Fu M, Wang M, et al. Cancer vaccines as promising immuno-therapeutics: platforms and current progress. *J. Hematol. Oncol*.*J Hematol Oncol*. 2022;15(1):28.

270. Mizukoshi E, Nakagawa H, Tamai T, et al. Peptide vaccine-treated, long-term surviving cancer patients harbor self-renewing tumor-specific CD8+ T cells. *Nat. Commun.* 2022;13(1):3123.

271. Nelde A, Rammensee H-G, Walz JS. The Peptide Vaccine of the Future. *Mol. Cell. Proteomics*. 2021;20:100022.

272. Tornesello AL, Tagliamonte M, Tornesello ML, Buonaguro FM, Buonaguro L. Nanoparticles to Improve the Efficacy of Peptide-Based Cancer Vaccines. *Cancers*. 2020;12(4):1049.

273. Shafer P, Kelly LM, Hoyos V. Cancer Therapy With TCR-Engineered T Cells: Current Strategies, Challenges, and Prospects. *Front. Immunol.* 2022;13:835762.

274. Wölfl M, Greenberg PD. Antigen-specific activation and cytokine-facilitated expansion of naive, human CD8+ T cells. *Nat. Protoc.* 2014;9(4):950–966.

275. Choo JAL, Liu J, Toh X, Grotenbreg GM, Ren EC. The Immunodominant Influenza A Virus M158–66 Cytotoxic T Lymphocyte Epitope Exhibits Degenerate Class I Major Histocompatibility Complex Restriction in Humans. *J. Virol.* 2014;88(18):10613–10623.

276. Campillo-Davo D, Versteven M, Roex G, et al. Rapid Assessment of Functional Avidity of Tumor-Specific T Cell Receptors Using an Antigen-Presenting Tumor Cell Line Electroporated with Full-Length Tumor Antigen mRNA. *Cancers*. 2020;12(2):256.

277. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data.

278. Bolotin DA, Poslavsky S, Davydov AN, et al. Antigen receptor repertoire profiling from RNA-seq data. *Nat. Biotechnol.* 2017;35(10):908–911.

279. Rizzuto GA, Merghoub T, Hirschhorn-Cymerman D, et al. Self-antigen–specific CD8+ T cell precursor frequency determines the quality of the antitumor immune response. *J. Exp. Med.* 2009;206(4):849–866.

280. Alanio C, Lemaitre F, Law HKW, Hasan M, Albert ML. Enumeration of human antigen–specific naive CD8+ T cells reveals conserved precursor frequencies. *Blood*. 2010;115(18):3718–3725.

281. Blattman JN, Antia R, Sourdive DJD, et al. Estimating the Precursor Frequency of Naive Antigen-specific CD8 T Cells. *J. Exp. Med.* 2002;195(5):657–664.

282. Scordo M, Shah GL, Kosuri S, et al. Effects of Late Toxicities on Outcomes in Long-Term Survivors of Ex-Vivo CD34+-Selected Allogeneic Hematopoietic Cell Transplantation. *Biol. Blood Marrow Transplant. J. Am. Soc. Blood Marrow Transplant.* 2018;24(1):133–141.

283. Ferrara JLM, Levine JE, Reddy P, Holler E. Graft-versus-host disease. *Lancet Lond. Engl.* 2009;373(9674):1550–1561.

284. Shlomchik WD. Graft-versus-host disease. *Nat. Rev. Immunol.* 2007;7(5):340–352.

285. Kosuri S, Herrera DA, Scordo M, et al. The Impact of Toxicities on First Year Outcomes after Ex-vivo CD34+ Selected Allogeneic Hematopoietic Cell Transplantation in Adults with Hematologic Malignancies. *Biol. Blood Marrow Transplant. J. Am. Soc. Blood Marrow Transplant.* 2017;23(11):2004–2011.

286. Urbano-Ispizua A, Carreras E, Marín P, et al. Allogeneic transplantation of CD34(+) selected cells from peripheral blood from human leukocyte antigen-identical siblings: detrimental effect of a high number of donor CD34(+) cells? *Blood*. 2001;98(8):2352–2357.

287. Lee SJ, Logan B, Westervelt P, et al. Comparison of Patient-Reported Outcomes in 5-Year Survivors Who Received Bone Marrow vs Peripheral Blood Unrelated Donor Transplantation: Long-term Follow-up of a Randomized Clinical Trial. *JAMA Oncol.* 2016;2(12):1583–1589.

288. Kröger N, Solano C, Wolschke C, et al. Antilymphocyte Globulin for Prevention of Chronic Graft-versus-Host Disease. *N. Engl. J. Med.* 2016;374(1):43–53.

289. Nash RA, Antin JH, Karanes C, et al. Phase 3 study comparing methotrexate and tacrolimus with methotrexate and cyclosporine for prophylaxis of acute graft-versus-host disease after marrow transplantation from unrelated donors. *Blood*. 2000;96(6):2062–2068.

290. Cutler C, Logan B, Nakamura R, et al. Tacrolimus/sirolimus vs tacrolimus/methotrexate as GVHD prophylaxis after matched, related donor allogeneic HCT. *Blood*. 2014;124(8):1372–1377.

291. Bejanyan N, Rogosheske J, DeFor TE, et al. Sirolimus and Mycophenolate Mofetil as Calcineurin Inhibitor-Free Graft-versus-Host Disease Prophylaxis for

Reduced-Intensity Conditioning Umbilical Cord Blood Transplantation. *Biol. Blood Marrow Transplant. J. Am. Soc. Blood Marrow Transplant.* 2016;22(11):2025–2030.

292. Povlsen HR, Bentzen AK, Kadivar M, et al. ATRAP - Accurate T cell Receptor Antigen Pairing through data-driven filtering of sequencing information from single-cells. 2022;2022.08.31.506001.

293. Kula T, Dezfulian MH, Wang CI, et al. T-Scan: A Genome-wide Method for the Systematic Discovery of T Cell Epitopes. *Cell*. 2019;178(4):1016-1028.e13.

294. Dobson CS, Reich AN, Gaglione S, et al. Antigen identification and high-throughput interaction mapping by reprogramming viral entry. *Nat. Methods*. 2022;19(4):449–460.

295. Ehx G, Larouche J-D, Durette C, et al. Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I-associated epitopes. *Immunity*. 2021;54(4):737-752.e10.

296. Laumont CM, Vincent K, Hesnard L, et al. Noncoding regions are the main source of targetable tumor-specific antigens. *Sci. Transl. Med.* 2018;10(470):eaau5516.

297. Perreault C, Roy DC, Fortin C. Immunodominant minor histocompatibility antigens: the major ones. *Immunol. Today*. 1998;19(2):69–74.

298. Nevala WK, Wettstein PJ. The preferential cytolytic T lymphocyte response to immunodominant minor histocompatibility antigen peptides. *Transplantation*. 1996;62(2):283–291.

299. Wettstein PJ, Bailey DW. Immunodominance in the immune response to “multiple” histocompatibility antigens. *Immunogenetics*. 1982;16(1):47–58.

300. Nevala WK, Paul C, Wettstein PJ. IMMUNODOMINANT MINOR HISTOCOMPATIBILITY ANTIGEN PEPTIDES RECOGNIZED BY CYTOLYTIC T LYMPHOCYTES PRIMED BY INDIRECT PRESENTATION1. *Transplantation*. 1998;65(4):559.

301. Cruz CRY, Bo N, Bakoyannis G, et al. Antigen-specific T cell responses correlate with decreased occurrence of acute GVHD in a multicenter contemporary cohort. *Bone Marrow Transplant.* 2022;57(2):279–281.

302. Zhang P, Tey S-K. Adoptive T Cell Therapy Following Haploidentical Hematopoietic Stem Cell Transplantation. *Front. Immunol.* 2019;10:1854.

303. Baumeister SHC, Rambaldi B, Shapiro RM, Romee R. Key Aspects of the Immunobiology of Haploidentical Hematopoietic Cell Transplantation. *Front. Immunol.* 2020;11:.

304. Sahillioglu AC, Schumacher TN. Safety switches for adoptive cell therapy. *Curr. Opin. Immunol.* 2022;74:190–198.

305. Hazini A, Fisher K, Seymour L. Deregulation of HLA-I in cancer and its central importance for immunotherapy. *J. Immunother. Cancer*. 2021;9(8):e002899.

306. Toffalori C, Zito L, Gambacorta V, et al. Immune signature drives leukemia escape and relapse after hematopoietic cell transplantation. *Nat. Med.* 2019;25(4):603–611.

307. Christopher MJ, Petti AA, Rettig MP, et al. Immune Escape of Relapsed AML Cells after Allogeneic Transplantation. *N. Engl. J. Med.* 2018;379(24):2330–2341.

308. Vago L, Perna SK, Zanussi M, et al. Loss of Mismatched HLA in Leukemia after Stem-Cell Transplantation. *N. Engl. J. Med.* 2009;361(5):478–488.

309. Zhou M, Sacirbegovic F, Zhao K, Rosenberger S, Shlomchik WD. T cell exhaustion and a failure in antigen presentation drive resistance to the graft-versus-leukemia effect. *Nat. Commun.* 2020;11(1):4227.

310. Janelle V, Delisle J-S. T-Cell Dysfunction as a Limitation of Adoptive Immunotherapy: Current Concepts and Mitigation Strategies. *Cancers*. 2021;13(4):598.

311. Lee J, Ahn E, Kissick HT, Ahmed R. Reinvigorating Exhausted T Cells by Blockade of the PD-1 Pathway. *Forum Immunopathol. Dis. Ther.* 2015;6(1–2):7–17.