THE HLA LIGAND ATLAS.
DEFINING ACTIONABLE TARGETS FOR T-CELL-BASED IMMUNOTHERAPY.

DISSERTATION

DER MATHEMATISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT

DER EBERHARD KARLS UNIVERSITÄT TÜBINGEN

ZUR ERLANGUNG DES GRADES EINES

DOKTORS DER NATURWISSENSCHAFTEN

(DR. RER. NAT.)

vorgelegt von

M. SC. ANA MARCU

aus Tîrgu Mureş, Rumänien

Tübingen

2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Prüfung:	29.09.2020
Stellvertretender Dekan:	Prof. Dr. József Fortágh
1. Berichterstatter:	Prof. Dr. Stefan Stevanović
2. Berichterstatter:	Prof. Dr. Oliver Planz

In accordance with the standard scientific protocol, the personal pronoun ‘we’ will be used to indicate the reader and the writer, or my scientific collaborators and myself.

SUMMARY

Over the past two decades, immunotherapy has become the fourth pillar in cancer treatment alongside surgery, radiotherapy, and chemotherapy. Immune checkpoint blockade shows objective clinical success but only a fraction of patients benefit from tumor regression in response to immunotherapy. The number of immune-oncology-related clinical trials initiated each year is increasing. And yet, NY-ESO-1, is the most targeted antigen type (31 clinical trials), and melanoma the most targeted tumor type (33 clinical trials)[1]. Thus, studying underrepresented cancers and characterizing novel antigens is essential for the development of new treatment options. T cells are the main mediators of anti-tumor immune responses and they recognize human leukocyte antigen (HLA)-presented peptides. The entirety of HLA-presented peptides is termed the immunopeptidome. This work addresses three aspects of antigen discovery for T-cell based immuno-oncology approaches.

The Results Part I section presents the HLA Ligand Atlas, a collection of benign HLA-I and -II immunopeptidomes covering 21 human subjects and 29 tissue types. We observe that immunopeptidomes are more similar within individuals as opposed to the expected tissue-based clustering reported for transcriptomes[2,3] and proteomes[4]. Furthermore, cryptic HLA ligands have been previously reported only in tumors and cell lines. In this dataset, we identified 1,407 cryptic peptides, showing that they are not *per se* tumor exclusive. Based on a case study in three glioblastoma patients, we illustrate the applicability of the multi-tissue HLA Ligand Atlas dataset to prioritize tumor-associated antigens for downstream experimental validation.

The Results Part II section describes the theoretical requirements for the use of mass spectrometers (LC-MS/MS) in immunopeptidomics experiments. In addition, a benchmark based on a serial dilution of a JY peptide eluate spiked with heavy-labeled peptides across five different LC-MS/MS systems reports performance differences and complementarity to the in house Orbitrap Fusion Lumos. These results showed that the timsTOF Pro from Bruker Daltonics and the q Exactive HF from Thermo Fisher Scientific could be suitable for immunopeptidomics experiments.

The results Part III section focuses on mapping the immunopeptidomic landscape of breast cancer, paving the way for generating a warehouse of shared tumor-associated antigens for frequent HLA allotypes. The selection of warehouse peptides will be the basis for a future phase I/II clinical study testing safety and toxicity of peptide vaccination. This project is still ongoing, as data acquisition has not been finalized. Data reanalysis and a stronger focus towards the triple negative breast cancer subtype will reshape the outcome of this project.

Thus, this work addresses three major steps related to the discovery of T cell-based antigens for cancer immunotherapy: i) an in-depth consideration of the mass spectrometric instrumentation required for immunopeptidomics studies, ii) the generation of a benign multi-tissue draft of the human immunopeptidome, which is required for the definition of tumor-association, and iii) mapping the breast cancer immunopeptidome and the definition of tumor-associated antigens.

ZUSAMMENFASSUNG

Die erfolgreiche klinische Anwendung der Immuncheckpoint-Inhibitoren stellt einen Meilenstein der Immunonkologie dar, der 2018 mit dem Nobelpreis für Medizin und Physiologie gekürt wurde. Die Anzahl an klinischen Studien im Bereich der Immunonkologie steigt stetig an, viele untersuchen bekannte Antigene wie NY-ESO1 (31 klinische Studien) in gut ansprechenden Tumoren wie das Melanom (33 klinische Studien)[1]. Dennoch, erfahren nicht alle Patienten eine Tumorregression als Antwort auf die Immuntherapie. Mittels Immuncheckpoint-Inhibitoren wird eine bereits vorhandene T-Zellantwort gestärkt, was die Untersuchung ihrer Antigene, der HLA-präsentierten Peptide, in den Fokus bringt. Die Gesamtheit aller HLA-präsentierter Peptide ist das Immunopeptidom. Diese Arbeit adressiert drei Bereiche, die essentiell für die Kartierung des Immunopeptidoms und somit für die Identifizierung neuer T-Zell-Antigene sind.

Der erste Ergebnisteil beschreibt den HLA Ligand Atlas, eine Sammlung an benignen HLA-I und -II Immunopeptidomen, welche 21 humane Spender und 29 Gewebetypen umfasst. Demnach wurde ersichtlich, dass das Immunopeptidom eines Spenders über unterschiedliche Gewebearten ähnlicher ist, als eine Gewebeart über mehrere Spender. Diese Erkenntnis steht im Gegensatz zur Gewebeähnlichkeit auf Transkriptom-[2,3] und Proteomebene[4]. Zusätzlich wurden 1,407 kryptische HLA-I Liganden identifiziert. Es wurde anhand von drei Glioblastompatienten gezeigt, dass der HLA Ligand Atlas Datensatz geeignet ist, um die Auswahl an tumorspezifischen Zielantigene für weitere Validierungsexperimente einzuschrenken.

Der zweite Ergebnisteil fokussiert sich auf die Analyse des Immunopeptidoms mittels Massenspektrometrie und den damit verbundenen nötigen technischen Parametern und Schwierigkeiten. Anhand einer seriellen Verdünnungsreihe eines JY-isolierten Peptidextraktes, konnte ein experimenteller Vergleich zwischen fünf Geräten durchgeführt werden. Die Ergebnisse einer komparative Auswertung von ca. 114 Messungen, zeigten dass die timsTOF Pro von Bruker Daltonics und die q Exactive HF von Thermo Fisher Scientific geeigent für Immunopeptidomanalysen sind. Insgesamt schnitt, die auf die Proben optimierte, bereits vorhandene Orbitrap Fusion Lumos am besten ab.

Der dritte Ergebnisteil beschreibt das Immunopeptidom im Mammakarzinom. Durch den Vergleich des Mammakarzinom-assoziierten und des benignen Immunopeptidoms, welches im HLA Ligand Atlas erfasst ist, wurden Tumor-assoziierte Antigene definiert. Die häufigsten, waren in den drei molekularen Subtypen repräsentiert. Diese Auswertung ist nicht vollständig, da die experimentelle Analyse der Proben nicht abgeschlossen ist.

Somit umfasst diese Arbeit drei essentielle Schritte für die Definition von T-Zell-Antigenen: i) eine detaillierte theoretische und praktische Betrachtung der massenspektrometrischen Voraussetzungen für vollständige Immunopeptidomanalysen, ii) die Erstellung einer sogenannten negativen Datenbank, welche benigne Immunopeptidome von 21 Spendern und 29 Gewebetypen charakterisiert, und iii) die Kartierung des Mammakarzinom-assoziierten Immunopeptidoms, welches durch den Vergleich mit dem benignen Datensatz, die Definition von T-Zell-Antigenen für neue Immuntherapien ermöglicht.

TABLE OF CONTENTS

1 PUBLICATIONS

1. **Marcu A**, Schurigt U, Müller K, Moll H, Krauth-Siegel RL, Prinz H. Inhibitory effect of phenothiazine- and phenoxazine-derived chloroacetamides on Leishmania major growth and Trypanosoma brucei trypanothione reductase. *Eur J Med Chem*. 2016;108:436-443.

2. Frank B, **Marcu A**, De Oliveira Almeida Petersen AL, Weber H, Stigloher C, Mottram JC, Scholz CJ, Schurigt U. Autophagic digestion of Leishmania major by host macrophages is associated with differential expression of BNIP3, CTSE, and the miRNAs miR-101c, miR-129, and miR-210. *Parasit Vectors*. 2015;8(1):404.

3. **Marcu A**, Eyrich M. Therapeutic vaccine strategies to induce tumor-specific T-cell responses. *Bone Marrow Transplant*. 2019;54(2):806-809.

4. Freudenmann LK, **Marcu A**, Stevanović S. Mapping the tumour human leukocyte antigen (HLA) ligandome by mass spectrometry. *Immunology*. 2018;154(3):331-345.

5. Lübke M, Spalt S, Kowalewski DJ, Zimmermann C, Bauersfeld L, Nelde A, Bichmann L, **Marcu A**, Peper JK, Kohlbacher O, Walz JS, Le-Trilling VTK, Hengel H, Rammensee H-G, Stevanović S, Halenius A. Identification of HCMV-derived T cell epitopes in seropositive individuals through viral deletion models. *J Exp Med*. 2020;217(3).

6. Bilich T, Nelde A, Bauer J, Walz S, Roerden M, Salih HR, Weisel K, Besemer B, **Marcu A**, Lübke M, Schuhmacher J, Neidert MC, Rammensee H-G, Stevanović S, Walz JS. Mass spectrometry-based identification of a B-cell maturation antigen-derived T-cell epitope for antigen-specific immunotherapy of multiple myeloma. *Blood Cancer J*. 2020;10(2):24.

7. Gierlich P, Lex V, Technau A, Keupp A, Morper L, Glunz A, Sennholz H, Rachor J, Sauer S, **Marcu A**, Grigoleit GU, Wölfl M, Schlegel PG, Eyrich M. Prostaglandin E2 in a TLR3- and 7/8-agonist-based DC maturation cocktail generates mature, cytokine-producing, migratory DCs but impairs antigen cross-presentation to CD8+ T cells. *Cancer Immunol Immunother*. 2020;69(6):1029-1042.

8. Ghosh M, Gauger M, **Marcu A**, Nelde A, Denk M, Schuster H, Rammensee HG, Stevanović S. Guidance document: validation of a high-performance liquid chromatography-tandem mass spectrometry immunopeptidomics assay for the identification of HLA class I ligands suitable for pharmaceutical therapies. *Mol Cell Proteomics*. 2020;19(3):432-443.

9. Bilich T, Nelde A, Bichmann L, Roerden M, Salih HR, Kowalewski DJ, Schuster H, Tsou C-CC, **Marcu A**, Neidert MC, Lübke M, Rieth J, Schemionek M, Brümmendorf TH, Vucinic V, Niederwieser D, Bauer J, Märklin M, Peper JK, Klein R, Kohlbacher O, Kanz L, Rammensee H-G, Stevanović S, Walz JS. The HLA ligandome landscape of chronic myeloid leukemia delineates novel T-cell epitopes for immunotherapy. *Blood Cancer J*. 2019;133(6):550-565.

10. Tudor Ilca F, Neerincx A, Hermann C, **Marcu A**, Stevanović S, Deane JE, Boyle LH. TAPBPR mediates peptide dissociation from MHC class I using a leucine lever. *Elife*. 2018;7.

11. Neidert MC, Kowalewski DJ, Silginer M, Kapolou K, Backert L, Freudenmann LK, Peper JK, **Marcu A**, Wang SSY, Walz JS, Wolpert F, Rammensee H-G, Henschler R, Lamszus K, Westphal M, Roth P, Regli L, Stevanović S, Weller M, Eisele G. The natural HLA ligandome of glioblastoma stem-like cells: antigen discovery for T cell-based immunotherapy. *Acta Neuropathol*. 2018;135(6):923-938.

2 INTRODUCTION

2.1 IMMUNOTHERAPY

The four pillars of cancer treatment are surgery, chemotherapy, radiotherapy, and immunotherapy. Immunotherapy is the most recent addition to cancer therapy, despite the observation of a connection between cancer and the immune system dating back to ancient Egypt and Greece. The first observations of tumors spontaneously going into remission after a simultaneous infection with high fever can be dated as early as ancient Egypt, while the similarity between cancer and inflammation has been attributed to the ancient Greek physician Galen[5,6]. However, during the same period, it was impossible to reproduce these results and decipher their mechanistic background due to experimental limitations, which allowed them to be forgotten.

William Bradley Coley is known today as the father of immunotherapy for making the similar observations of patients with various uncurable cancers that went into spontaneous remission after developing concomitant acute bacterial infections. In 1891 Coley developed different mixtures of live and heat-inactivated *Streptococcus pyogenes* and *Serratia marcescens* and gave them to patients with bone cancer and soft tissue sarcomas thereby achieving positive clinical results. Over the course of his career, he treated hundreds of patients and has thus developed the first immune-based treatment to cancer[7]. However, his results were met with skepticism due to the unknown underlying mechanisms of this therapeutic approach[7].

It was only after 1945 that interest in the immune system peaked. During the following years, interferon was discovered[8], the first cancer vaccine consisting of tumor lysate showed 22% remission in treated patients[9], the existence of T cells and their role in immunity had been discribed[10], and bone marrow transplantation as a treatment for hematological cancers had been pioneered[11]. Although Paul Ehrlich predicted in 1909 that tumor cells form continuously, and immune cells scan and eradicate them, Thomas and Burnet enunciated the theory of immunosurveillance in cancer in 1957, proposing that lymphocytes scavenge cancerous cells transformed by somatic mutations[12,13]. Again, this theory was not further elaborated due to the lack of experimental evidence and mechanistical understanding, until 1974 when Stutman showed that mice with impaired immune systems developed cancer at a faster rate than wild type controls[14]. Mechanistic clarification came from Schreiber, Dunn and Old, who in 1998 and 2001 showed that T cells can provide tumor-specific immune surveillance and antitumor responses[15,16]. The concept of immunosurveillance was broadened to encompass the observation that immunosurveillance promotes the formation of tumors with reduced immunogenicity. The dynamic interplay between tumor development and immunosurveillance is summarized in the three phases of cancer immunoediting: elimination (cancer immunosurveillance), equilibrium, and escape[17].

1) Elimination represents the classical theory of immunosurveillance in which adaptive and innate immunity act as an extrinsic tumor suppressor that protect immunocompetent individuals (and animals) from the development of cancers by eradicating them. When elimination of nascent tumor cells fails, the relationship with

the immune system shapes the tumor phenotype in different ways including but not limited to forming an immunodeficient tumor microenvironment.

2) Equilibrium represents a state of latency between the elimination and escape phase, in which the malignant disease can be clinically detected. While tumor growth is initially constrained by the immune system, it might reach a point of no return characterized by increased genomic instability that leads to an accumulation of mutations in tumor cells. The dynamic interplay between the two counterparts results in a heterogeneous tumor cell population that is selected to resist immune attack. However, the resulting tumor might progress to the next phase of immune escape, might be indefinitely maintained in the equilibrium phase, or might be eliminated by the immune system.
3) Escape: tumor cells surviving the immunoediting and equilibrium phase are unrestrained by immune pressure and can grow and metastasize.

Further key discoveries followed, such as the mechanism of immunoediting, evidence of tumor immune escape[16], and the first identification of a target antigen (melanoma antigen MAGE), that is recognized by cytotoxic T cells[18].

These milestones of the past century have consolidated our understanding of the interplay between the immune system and cancer. These discoveries enabled harnessing intrinsic mechanisms of the immune system for targeted cancer treatment. Immunotherapeutic options nowadays are multifaceted and act on mobilization or blockade of different cell types or cellular pathways. The Food and Drug Administration (FDA) and the European Medical Agency (EMA) approved a variety of novel immuno-oncological agents. Appraisal from the scientific community towards the field of immunology culminated in 2018 with the Nobel Prize in physiology and medicine awarded to James P. Allison and Tasuku Honjo for the discovery and application of immune checkpoint inhibitors. The recent success and popularity of immunotherapy is however overshadowed by our still limited understanding of the complex mechanism of action, as evidenced by clinical success in only a subset of cancer patients, and tumor types. In addition, the tremendous costs and efforts of production are an impediment towards their wide application in both developed and less developed countries.

Building on these discoveries, this work focuses on characterizing novel antigens for developing new T-cell based immunotherapies with an emphasis on peptide vaccination strategies. For this purpose, we have mapped the immunopeptidome in both health and disease.

To understand modern immunotherapy approaches, we will briefly introduce the HLA antigen presentation pathway and T cell activation. A further description of current cellular immunotherapies, antibodies, and viral vector vaccines will outline the state-of-the-art treatment options in immuno-oncology.

2.2 ANTIGEN PRESENTATION BY HUMAN LEUKOCYTE ANTIGEN (HLA) MOLECULES

The HLA is a large genetic region on chromosome 6 coding for class I (HLA-I) and class II (HLA-II) molecules as well as for other proteins related to the antigen presentation pathway. HLA genes are the most polymorphic in the human genome with a large number of allelic variants at each locus[19]. HLA-I molecules are composed of an α chain in a noncovalent complex with a nonpolymorphic invariant $\beta 2$ microglobulin chain and the presented peptide. HLA-II molecules contain two polymorphic chains, the α and the β chain. The peptide binding cleft has α helical sides and an eight-stranded antiparallel β-pleated sheet floor. The peptide binding cleft of the HLA-I molecule is comprised of the $\alpha 1$ and $\alpha 2$ segments and that of HLA-II molecules of the $\alpha 1$ and $\beta 1$ segments of the two chains. The Ig-like domains of HLA-I and -II molecules contain the binding sites for the T cell coreceptors CD8 and CD4, respectively. $CD8^+$ T cells recognize peptides presented on HLA-I, while $CD4^+$ T cells recognize peptides presented on HLA-II molecules. T cells can distinguish between self and foreign peptides. Figure 1 illustrates the crystal structure of the binding cleft of HLA-I and -II molecules while accommodating a peptide[20].

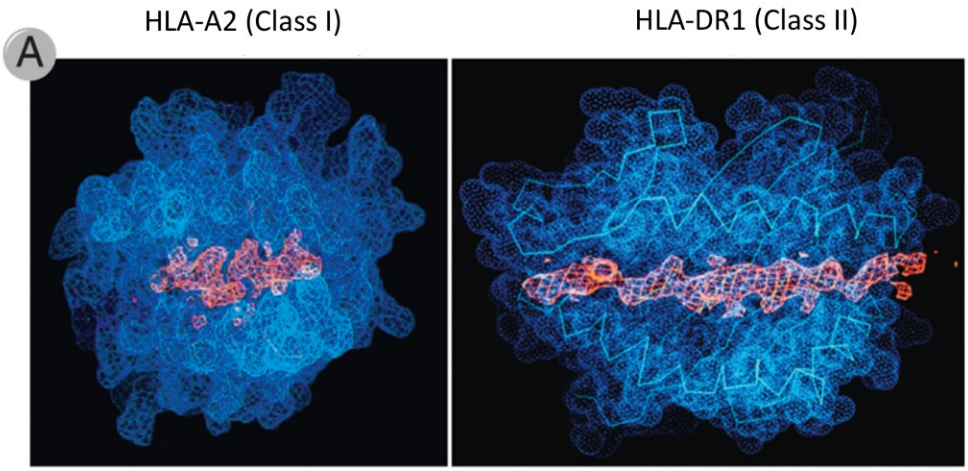


Figure 1: Peptide presentation on HLA-I and -II molecules.
The crystal structure illustrates the peptide-binding groove[20].

HLA-I allotypes can be divided into -A, -B, and -C allotypes, while HLA-II molecules can be divided into -DR, DP, and -DQ molecules. Each HLA allotype has different binding preferences for certain peptide sequences that can be condensed into anchor residues at specific positions. These binding preferences were characterized and resulted in allotype-specific binding motifs. The HLA molecule has further polymorphic residues that strongly interact with the T-cell receptor.

Briefly, HLA-I molecules present peptides of 8-12 aa originating from intracellular proteins[21], while HLA-II molecules present 8-25 aa long peptides[22], primarily originating from extracellular proteins. HLA-I molecules are presented on all nucleated cells, while HLA-II molecules are predominantly presented on professional antigen presenting cells (APCs), although some exceptions have been described[23]. The antigen processing pathways differ considerably between HLA-I and -II ligands.

As part of the regular protein turnover, HLA-I ligands are produced by proteolytic degradation of cytosolic proteins by the proteasome, or the IFN-γ-induced immunoproteasome. These proteins

might be self-proteins, or they might be of microbial or viral origin. Resulting peptides are transported to the endoplasmic reticulum (ER) by the transporter associated with antigen processing (TAP). At the ER-membrane, TAP mediates the ATP-dependent transfer of 8-16 amino acid long peptides with basic or hydrophobic C termini from the cytosol into the ER lumen. HLA-I molecules are synthesized and assembled in the ER with the help of two chaperones, calnexin and calreticulin. Within the ER, TAP associates with tapasin, which has a high affinity to newly synthesized HLA-I molecules. The aminopeptidase ERAAP trims the imported peptides to be able to bind to the HLA binding cleft. The complex of HLA-I molecules and the peptide has no affinity to tapasin anymore and can be transferred to the Golgi, from where it is transported via exocytic vesicles to the cell surface for presentation to $CD8^+$ T cells. Furthermore, defective ribosomal products (DRIPs) resulting directly from ribosomal translation can also be presented on HLA-I molecules[24].

HLA-II molecules present peptides originating from extracellular proteins captured by APCs via i.e. phagocytosis. These antigens are proteolytically degraded by lysosomal proteases such as cathepsins in phagolysosomes. HLA-II molecules are synthesized and assembled in the ER, with the invariant chain occupying the peptide binding cleft. The HLA-II complex is subsequently transported to the endosome, where endosomal proteases degrade the invariant chain, leaving a remnant peptide the same length as a presented peptide, the class II-associated invariant chain peptide (CLIP). HLA-DM, which is a nonpolymorphic HLA molecule facilitates the exchange of CLIP with an endosomal peptide. Due to the open peptide-binding cleft, peptides of up to 25 amino acids can bind to the HLA molecule. These are trimmed by proteases to the appropriate size for T-cell recognition. HLA-II molecules are stabilized by the bound peptide and the peptide-HLA-complex is delivered to the surface of the APC, where it is displayed for recognition by $CD4^+$ T cells.

Some APCs are able to take up, process, and present extracellular antigens on HLA-I molecules to $CD8^+$ T cells[25]. This process, termed cross presentation, permits the presentation of exogenously sampled antigens from *e.g.* apoptotic cells, on HLA-I molecules rather than HLA-II molecules, along the classical antigen presentation pathway, resulting in the stimulation of naïve cytotoxic $CD8^+$ T cells into activated $CD8^+$ T cells. Cross presentation is essential in developing an immunity against viruses and perhaps tumors, that do not readily infect APCs but are rather intracellular processes resulting in HLA-I presentation of abnormal antigens (*i.e.* viral or TAAs).

2.2.1 T CELL ACTIVATION

T CELLS ARE DEFINED BY THEIR TCR

T cells are lymphocytes that develop in the thymus and have the distinctive feature of a T-cell receptor (TCR) on their surface. The TCR is responsible for recognizing peptides presented by HLA molecules on other cells. The TCR is composed of two glycoproteins, the TCR α and β chain linked by disulfide bonds. Both chains have a variable and a membrane proximal constant domain, followed by a transmembrane region and a cytoplasmic tail. The variable regions of both α and β chains form three hypervariable complementarity determining regions (CDRs)[26]. The CDR mediates recognition of antigenic peptides. The VJ and V(D)J somatic recombination during T cell development generate the high variability of the CDR α and β chains, respectively[27]. The high CDR

diversity translates into a TCR repertoire with high antigenic diversity, estimated at around 10^6 different clonotypes[28]. The TCR is expressed on T cells in a non-covalently bound octameric complex with CD3εγ, a CD3εδ, and a TCRζζ homodimer. Both CD3εγ and CD3εδ have an intracellular immunoreceptor tyrosine-based activation motif (ITAM), while the ζ chains each contain three ITAM signaling domains[29]. Tyrosine phosphorylation within the ITAM motifs results in T cell activation after TCR binding to a peptide HLA complex[30].

Two major groups of αβ T cells can be differentiated, based on the expression of the CD4 or the CD8 coreceptor. Cytotoxic T cells are $CD8^+$ and recognize peptide antigens presented by HLA-I molecules, while T helper cells express the CD4 coreceptor and recognize peptides presented by HLA-II molecules.

THYMIC SELECTION

T cells originate from pluripotent hematopoietic stem cells in the bone marrow. Double positive thymocytes ($CD4^+/CD8^+$) migrate into the thymic cortex, where thymic cortical epithelial cells present self-antigens on HLA molecules. Only thymocytes that interact well with either HLA-I or HLA-II molecules receive a survival signal through this interaction. T cells then become single positive by downregulating the expression of the unstimulated CD4 or CD8 surface receptor. Thus, T cells that can interact with both the HLA molecule and the peptide are selected for further development, while the other thymocytes "die by neglect"[31].

During negative selection, T cells are confronted with self-peptides presented by medullary thymic epithelial cells (mTECs). The transcription factor AIRE enables the expression of all self-antigens in the human body by mTEC cells. A strong interaction between T cells and HLA presented self-peptides results in an apoptotic signal induced in the T cells. Thus, potentially autoreactive T cells are eliminated. The remaining thymocytes exit the thymus as mature naïve T cells (reviewed here[31]).

Negative selection dictates that only low affinity T cell responses can be elicited against non-mutated tumor-associated peptides, as high-affinity TCRs had already been eliminated.

T CELL ACTIVATION

Mature naïve T cells exit the thymus and circulate through the blood towards lymphoid organs. In lymphoid organs, they might encounter antigen presenting cells (APCs) that present a foreign antigen. Three signals are required for the activation of naïve T cells. The first signal is the interaction between the TCR and the respective HLA-peptide complex. Only this signal leads to T cell anergy, as it indicates an autoreactive T cell[32]. The second signal is mediated by costimulatory molecules on APCs. Multiple activating and inhibitory signals have been described, with CD28 being the most prominent[33] on T cells and interacting with CD80 and CD86 on APCs[34]. The co-stimulatory interaction induces IL-2 which promotes T cell proliferation[35]. Signal three is provided by DC-secreted cytokines that induce T cell polarization towards a certain effector function[36]. All three signals are required for successful T cell activation.

2.3 Cellular Therapies

The key to successful immunotherapy is the targeted activation of an anti-tumor T cell response. In this section, we will introduce different avenues in which this was achieved. Starting with hematopoietic stem cell transplantation, different modalities of modifying a patient's T cells in vitro and reinfusing them have been developed. Their success and limitations are described in this chapter.

2.3.1 Hematopoietic Stem Cell Transplantation (HSCT)

HSCT has pioneered the concepts of stem cell therapy and immunotherapy as a tool against cancer. During the last 50 years, HSCT has become the standard of care for multiple malignant and nonmalignant hematologic diseases[37]. First instances of HSCT were performed in 1957 by Thomas and colleagues, with all patients dying within 100 days after the procedure. The first reports of HLA-matched sibling donor transplants were published only in 1971[38]. Further research regarding HLA-matched donors, the conditioning regimen prior to HSCT, and the origin and management of adverse events optimized the procedure to the current standard. Milestones on this path were awarded with the Nobel Prize, in 1980 for the discovery of MHC and in 1990 for HSCT and organ transplant. Overall, HSCT requires full or partial bone marrow ablation with chemotherapy (*e.g.* cyclophosphamid) to allow a stable engraftment of donor hematopoietic stem cells (HSCs). The donor and recipient must be matched in as many HLA loci as possible, as every HLA mismatch negatively affects the success of HSCT[39]. To overcome the HLA-induced barrier, cord blood and partially HLA-matched HSCTs have emerged as a viable option, as cord blood stem cells require less strict HLA matching due to an immunologically naïve donor T cell repertoire[40]. However, engraftment and immune reconstitution are both slower than after HSCT from adult donors. The major side effects of HSCT are acute or chronic graft versus host disease (GvHD) and the accompanying graft versus leukemia (GvL) effect, as well as susceptibility to opportunistic infections and graft rejection. GvHD and GvL frequently occur concomitantly. Initially considered to be mediated by donor T cells, GvL is currently considered multifactorial, with an interplay of cytotoxic T cells, natural killer (NK) cells, dendritic cells (DCs) and minor histocompatibility antigens (MiHAs) playing a role[41]. Related therapeutic strategies have evolved from HSCT and are summarized by the umbrella term T-cell receptor (TCR)-T-cell therapy. These approaches include adoptive cell therapy (ACT), ACT with genetically modified TCRs, and ACT with engineered chimeric antigen receptors (CAR).

2.3.2 Chimeric Antigen Receptor (CAR)-T Cells

Autologous T cells are transduced with CARs and reinfused into the patient. The CAR is incorporated into autologous T cells via lentiviral-, mRNA- or transposon-based transduction[42]. CARs typically encode an extracellular domain that mediates tumor recognition linked to an intracellular domain that mediates signaling for T-cell activation[43]. The extracellular domain consists of a tumor antigen-specific single chain fragment variable (scFv) domain of an antibody linked to an intracellular TCR CD3-ζ signaling chain and optional costimulatory molecules[43]. First

generation CARs contain CD3-ζ alone, second generation CARs contain a costimulatory domain (CD3-ζ/4-1BB), while third generation CARs contain multiple costimulatory domains (reviewed in[44]). Antibody scFvs can recognize a wide array of intact antigens independently of the HLA processing and presentation pathway.

In vivo immune responses mounted against tumor self-antigens are limited by central or peripheral tolerance mechanisms. During thymic selection, high-affinity TCRs with specificities towards self-antigens are eliminated, and only low affinity TCRs are maintained that might be insufficient for inducing a strong anti-tumor immune response. Breaking self-tolerance is therefore possible with CARs due to engineering of high affinity scFv receptors.

Indeed, clinical success was obtained in the form of durable remission in a pediatric patient with treatment-refractory chronic lymphocytic leukemia (CLL) in 2011[45]. The patient received lentiviral transduced autologous T cells with a CAR, consisting of the B-cell antigen CD19 coupled to CD137 (costimulatory molecule 4-1BB) and CD3-ζ[45]. Since then, the CAR T-cell technology has been rapidly evolving, with two FDA-approved anti-CD-19 CARs (KYMRIAH™ and YESCARTA™) for the treatment of relapsed/refractory B-cell acute lymphoblastic leukemia and diffuse large B-cell lymphoma.

Unfortunately, CAR T cells elicit serious, sometimes life-threatening side effects, such as cytokine release syndrome or tumor lysis syndrome[46]. These side effects originate from expression of the tumor-associated antigen CD19 on healthy B cells, leading to their elimination as a side effect. A strategy to limit on-target/off-tumor toxicities is the inclusion of a suicide safety switch[47]. Nevertheless, further limitations of this technology include the time- and cost-intensive production of CAR T cells, acquired tumor resistance mechanisms such as antigen loss, and in some cases a high proportion of dysfunctional/exhausted T cells obtained from the patient which reduce the yield of functional CAR T cells for infusion.

CAR T cell therapy and TCR-engineered T cell therapy are two sides of the same coin, both having advantages and disadvantages. While CAR T cells target tumor-associated antigens via an HLA-independent mechanism, TCR-engineered T cells are dependent on HLA restriction and the HLA presentation pathway. Most proteins (~72%)[48] are localized intracellularly and they are only targetable by TCR-engineered T cells. As a result, the repertoire of antigens recognizable by CARs is far more limited. Still, CARs can recognize a variety of antigens, including glycosylated targets, irrespective of the HLA restriction, and HLA copy number on target cells. Despite the higher sensitivity of TCRs compared to CARs[49], TCRs mediate the release of less cytokines than CARs[50], thereby having a potentially reduced risk of cytokine release syndrome. Overall, CAR-T cell therapy has shown significant clinical benefit in hematologic diseases, while TCR-engineered T cells have shown clinical efficacy in both solid tumors (metastatic melanoma) and hematologic malignancies. Overall, both are expensive and difficult to manufacture individually for each patient and carry serious side effects and manufacturing errors.

On a similar note, dendritic cell (DC) vaccination with a tumor-associated antigen has also been implemented clinically. In 2010, Sipuleucel-T (Provenge®), the first DC vaccine was FDA approved for metastatic castration-resistant prostate cancer. Sipuleucel-T is an active cellular therapy

consisting of autologous peripheral blood mononuclear cells (PBMCs) that include DCs which have been activated with a recombinant fusion protein consisting of prostatic acid phosphatase (PAP) fused to granulocyte–macrophage colony-stimulating factor (GM-CSF). The clinical study showed a relative reduction of 22% in the risk of death as compared to the placebo group, which represented an increase of 4.1 months in the median overall survival[51].

2.3.3 Adoptive Cell Therapy (ACT)

ACT was pioneered by Steven Rosenberg in 1988[52] and is a procedure in which a patient's own T cells with anti-tumor activity are isolated, expanded *in vitro* and reinfused into the same patient. ACT-based strategies can be divided into two categories: i) isolation of naturally occurring tumor-specific T cells from existing tumor mass *i.e.* tumor-infiltrating T lymphocytes (TILs) and ii) genetic modification of blood-derived T cells with an engineered TCR. The goal of both strategies is to allow specific recognition of the tumor and mounting a robust anti-tumor immune response (reviewed in[53]). This was improved with the implementation of a nonmyeloablative lymphodepletion regimen followed by subsequent high-dose IL-2 treatment[54]. The immunodepleting preparative regimen given before the adoptive transfer facilitates the clonal repopulation of patients with anti-tumor T cells[55] and eliminates regulatory T cells and other lymphocytes that would compete for cytokines such as IL-7 and IL-15[56]. Infusion with autologous TILs grown from the resected tumor nodules of patients with metastatic melanoma has proved efficacy of ACT for solid tumors, mediating tumor regression in 50% of patients, and complete tumor regression in 10 – 25% of patients[1,56]. However, the high numbers of infused T cells (10^{11}) *in vivo* are still problematic to manufacture.

Recent developments focus on generating autologous T cells transduced with a tumor specific TCR. In the first proof of principle study, T cells from metastatic melanoma patients were transduced with a TCR directed against A*02:01/MART-1 peptide. Infused TCR-modified T cells were persistent for more than a year and a sustained response was observed in a minor subset of treated patients[57]. Other trials have subsequently tested TCR-modified T cells with specificities against NY-ESO-1 (in melanoma, synovial sarcoma, multiple myeloma), MAGE-A3 (myeloma, melanoma), MAGE-A4 (esophageal cancer and colorectal carcinoma) (summarized in table 1 from[53]).

2.4 Immune Checkpoint Inhibition

A further clinically effective immuno-oncology product was provided through the blockade of the cytotoxic T-lymphocyte antigen 4 (CTLA-4). CTLA-4 was initially identified in 1987[58], but its function as a critical immune checkpoint molecule that down-regulates pathways of T-cell activation was described by Jim Allison's group in 1995[59]. This discovery translated into the clinically applicable drug Ipilimumab (Yervoy®) which is a fully humanized monoclonal IgG1 antibody that blocks CTLA-4 to promote antitumor immunity[60]. In contrast to other approved therapeutic antibodies, Ipilimumab does not directly target tumor cells, but binds to inhibitory receptors on lymphocytes, thereby enhancing a pre-existing anti-tumor immune response[61]. The phase 3 clinical trial showing an overall survival advantage for patients treated with Ipilimumab over patients treated with a gp100 peptide vaccine was performed in 2010[62], and led to the FDA approval of Ipilimumab in 2011 in stage IV metastatic melanoma.

CTLA-4 is exclusively expressed on T lymphocytes and counteracts the costimulatory effects of CD28. Both CD28 and CTLA-4 receptors on T cells share the same ligand CD80 (B7.1) and CD86 (B7.2) on antigen presenting cells (APCs). Nevertheless, CTLA-4 has a higher affinity towards CD80 and CD86 and can thus outcompete CD28 for its ligand. Under physiological circumstances, CTLA-4 is expressed on T cells only after antigen recognition and T cell activation. Thus, CTLA-4 can avoid overactivation of T cells under abnormally high concentrations of antigens. CTLA-4 is also expressed on regulatory T cells (Tregs) and supports their immunosuppressive role. Thus, inhibitory receptors are essential for modulating the magnitude of the immune response to avoid autoimmunity and maintain self-tolerance[61]. During tumor development, mechanisms evolve that dysregulate the balance of inhibitory and stimulatory TCRs, in favor of higher abundance of inhibitory molecules. Thus, an immune-suppressive microenvironment is established that does not hinder tumor growth.

Despite severe immune-related toxicities, Ipilimumab is the standard therapy for late stage, heavily pretreated melanoma patients. Clinical trials are evaluating the efficacy of several additional checkpoint molecules such as PD1, PDL1, LAG3, and TIM3. The anti PD1 antibodies are the most advanced, with Pembrolizumab (Keytruda®) and Nivolumab (Opdivo®) having been approved as a second-line treatment in melanoma in 2014.

While effective in a wide range of solid tumors, the main mechanism of action of immune checkpoint inhibitors is boosting of a pre-existing immune response largely mediated by TILs. Thus, tumors with a poor TIL infiltration and low immunogenicity are less responsive to immune checkpoint inhibition and could be potentially targeted with ACT.

2.5 Immunotherapeutic Approaches in Development

Multiple vaccination strategies that aim to induce an anti-tumor immune response are being tested in clinical trials. Promising delivery platforms for tumor-associated antigens include RNA-based technologies frequently packaged in lipid nanoparticles (Curevac, BioNTech, Moderna), viral vector vaccines (Orf virus-based vaccine vector D1701-V[63], KISIMA[64], other platforms reviewed here[65]), or peptides administered together with adjuvants[66,67]. A prominent example of vaccine delivery via viral vectors is the second cancer vaccine that received FDA approval: talimogene laherparepvec (T-VEC) for the treatment of advanced melanoma. The oncolytic immunotherapy is administered intralesionally and comprises a genetically engineered attenuated herpes simplex virus type 1 (HSV-1) encoding GM-CSF[68].

All these immunotherapeutic approaches share the challenge of finding actionable target antigens to direct an immune response against. Despite the wide repertoire of therapeutic agents and formulations, only a handful of tumor-associated antigens are targeted. Currently employed antigens are not ideal, as their expression profile is not restricted to the tumor, but oftentimes is abundant on further healthy tissues. Thus, severe side effects arise. New potential TAAs can be divided into two categories: HLA dependent and HLA independent. HLA independent targets are best suited to be recognized by antibodies and CAR-T cells and represent cell surface antigens (not further described here).

HLA-dependent targets can be natural HLA ligands stemming from self-proteins abnormally expressed in the tumor, or cancer-testis antigens (CTAs), that have an exclusive expression on gametogenic cells and tumors. Further classes of tumor-associated HLA-dependent peptide antigens are discussed in the following sections. Tumor-specific HLA ligands can be targeted by ACT, therapeutic peptide vaccination, RNA vaccines, viral vector vaccines, TCR mimic antibodies and many more. Our work mainly focuses on peptide vaccination with directly identified naturally presented TAAs, therefore, the remaining part of the introduction will focus on this subset of TAAs with emphasis and examples from peptide vaccination trials.

2.6 IMMUNOPEPTIDOMICS

The study and characterization of HLA presented peptides (e.g. HLA ligands) is termed immunopeptidomics or HLA ligandomics. These two terms will be used interchangeably throughout this work. The first reasonably successful isolation of HLA-associated peptides was performed by proteolytic digestion with papain from large volumes of starting material in the 1970s[69,70]. However, the large quantity of starting material dissuaded the wider implementation of this method. The subsequent successful method of directly identifying endogenously processed and presented HLA ligands by LC-MS/MS was pioneered in the 1990s by Hunt and colleagues[71]. At the same time, Rammensee and colleagues used HPLC and Edman degradation to sequence HLA ligands[72,73]. Three main techniques of biochemical extraction of HLA ligands were developed roughly concomitantly: (1) strong acid elution of HLA-I and -II ligands from whole cell lysate using trifluoroacetic acid[74,75], (2) mild acid elution (MAE) of HLA-I ligands from the cell surface of suspension cell lines[76], and (3) immunoaffinity purification (IP) of HLA-I and -II peptide complexes from solubilized cells followed by peptide isolation[77].

These early HLA ligand profiling studies identified only tens of endogenously processed and presented ligands, while current studies report thousands[78,79] if not tens of thousands of HLA ligands[80]. The increasing identification depth of immunopeptidomes can be attributed to major improvements in mass spectrometry and chromatography instrumentation but also partially to computational tools capable of automizing spectral identification.

Nevertheless, the experimental isolation of HLA ligands is still roughly the same as in the 1990s, meaning that the same experimental uncertainties are a reality today. A first unknown factor is the yield of the HLA ligand isolation procedure, particularly the amount of HLA ligands lost throughout the immunopurification procedure, elution, ultrafiltration, and C18 purification prior to LC-MS/MS analysis. Secondly, a qualitative bias originates from an improved sequence-dependent detection in LC-MS/MS analyses towards peptides with basic N and C terminal residues. Due to the nature of the HLA binding motif, allotypes such as HLA-A*11 and HLA-A*03 have improved detection rates, due to lysine and arginine residues at the C terminus. A third bias resides in the antibody selection used to precipitate HLA-peptide complexes. The specificity, selectivity, and cross-reactivity of the selected antibodies introduces oftentimes unknown biases towards the peptide repertoire identified from a certain tissue or cell line. Two possibilities are available: employing pan HLA-I or -II antibodies or employing HLA-allotype-specific antibodies. The most frequent pan HLA-I antibody used in most immunopeptidomics experiments is W6/32[81]. Alternatives encompass HLA

allotype-specific antibodies such as BB7.2 against A*02[71,82]; GAP-A3 against A*03[83], ME1-1.2 against B*07[83,84] etc. After MS identification, peptides need to be associated with their respective HLA allotype. Allotype-specific antibodies have the advantage of rendering this step obsolete, provided they are specific to the HLA allotype of interest. However, only a hand-full of HLA allotype-specific antibodies have well documented specificities and clean HLA allotype cross-reactivity profiles.

In the meantime, the performance of HLA-I binding prediction algorithms has constantly improved, particularly since elution data has been included in training datasets. This allows for an increasingly confident deconvolution of multi-allelic peptide species to the HLA allotypes of the respective donors. Overall, for HLA-I immunopeptidomics experiments, binding prediction algorithms have a high accuracy and sensitivity that allow the convenient usage of the pan HLA-I antibody W6/32.

The opposite is the case for HLA-II immunopeptidomics studies. Firstly, HLA-II molecules have an open peptide binding groove molded by both the α and β chains. As HLA alleles are inherited co-dominantly, all permutations of α and β chains need to be computed and the resulting peptide-binding specificities accounted for, when deconvoluting multi-allelic immunopeptidomes. To experimentally compensate for the increased complexity, the HLA-DR-specific antibody L243[85] is being preferentially used to capture HLA-II peptide complexes. HLA-DR has a monomorphic α chain that facilitates a confident binding prediction as well. Pan HLA-II antibodies such as Tü39[86] are used by our group, but generally HLA-II ligand characterization has been mostly stagnant so far. Recently, multiple groups have been concomitantly working on improving HLA-II binding prediction algorithms resulting in four new tools published in October 2019: NetMHCIIpan4.0[87,88], MARIA[89], neoMHC2[90], and MixMHC2pred[91]. With substantial improvements in HLA-II binding prediction algorithms, it appears easy to fathom a future in which both HLA-I and -II immunopeptidomes are equally assessed for precision medicine, particularly as anti-cancer immune responses require the interplay of $CD4^+$ and $CD8^+$ T cells[92].

2.7 The Endless Search For Targets for Immunotherapy

Two common strategies of identifying tumor-specific HLA-ligands are employed, one based on reverse immunology, the other on direct identification of naturally presented HLA ligands. In reverse immunology, somatic mutations are identified by exome sequencing of tumor cells and PBMCs from blood. Resulting somatic mutations are translated to potential peptide sequences *in silico* and subjected to HLA-binding prediction tools to prioritize candidates that are most likely to be presented by the HLA allotypes of the respective subject. The major shortcoming of this approach is the enrichment of false positives, with reports estimating over 90% of neoantigens not being presented by HLA molecules, let alone eliciting T-cell dependent immune responses[93]. The underlying assumption that HLA-binding prediction algorithms account for all intermediate steps of the antigen processing pathway has proven faulty.

To overcome this limitation, immunopeptidomics analyses can deliver direct evidence of HLA-peptide presentation. Here, HLA-presented peptides are directly isolated and sequenced by LC-

MS/MS from primary tumor samples. However, the Human Immuno-Peptidome Project Consortium (HIPP) has concisely summarized the challenges impairing the wide implementation of immunopeptidomics in research and clinical settings. Beside the non-trivial experimental procedure that requires a high amount of sample input, frequent challenges regarding MS technologies include the cost, sensitivity, reproducibility, and accessibility to large volumes of generated MS data[94]. Consequently, the field has been expanding slowly giving way to the faster and cost-effective method of reverse-immunology albeit with only limited clinical success. Proteogenomic approaches combining exome and RNA sequencing with HLA immunopeptidomics-based validation of predicted neoantigens have gained in popularity[80,95]. Several lines of evidence suggest that actionable T-cell targets against cancer are not limited to mutated neoantigens[67], but can be expanded to non-mutated[78,96,97], post-translationally modified[84,98], cryptic[99–101], and proteasomally spliced[102–105] HLA ligands. Characteristic features for all these types of tumor-associated or tumor-specific antigens are listed below.

2.7.1 Non-mutated Self-antigens

Non-mutated natural HLA ligands have been isolated and characterized for the past three decades, mainly in mice[106,107], human tumors[97,108–110], or tumor cell lines[79,90,111,112]. Early studies that describe binding motifs and anchor residues mediating specific binding of peptides to HLA allotypes foresaw the potential of HLA ligands in immuno-oncology[75]. Non-mutated natural HLA ligands are derived from canonical proteins, in their canonical reading frames, and can therefore be identified by performing database search of LC-MS/MS data either against the standard reference human proteome or against an individualized canonical proteome translated *in silico* from whole exome sequencing (WES) data.

When defining non-mutated TAAs it is necessary to compare tumor HLA ligandomes with a benign counterpart to predict possible on-target/off tumor adverse events. On-target/off-tumor adverse events can lead to dramatic outcomes, even death, as has been observed after administration of an affinity-enhanced TCR directed to a HLA–A*01–restricted MAGE A3 antigen (EVDPIGHLY) for use in adoptive T cell therapy[113,114]. As benign human tissue is quite scarce, most studies rely on multi-tissue transcriptomics data from public repositories such as GTEx[2,3] for comparative analysis. The scientific community has widely accepted and repeatedly proven that evidence on transcript level correlates poorly with protein evidence[4,115], let alone HLA ligand presentation[106,116,117]. Alternatively, whenever possible, histologically normal tissue adjacent to the tumor is being employed as a benign counterpart[96,118,119]. In these approaches, comparative ligandome profiling between tumor and benign within one tissue type is possible. Furthermore, label-free quantitative LC-MS/MS analysis between adjacent benign tissue and the tumor ligandome is possible, allowing the definition of over-expressed ligands in tumor as candidates for immunotherapy. However, adjacent benign tissue has been in proximity with the tumor, thus potentially sharing the same tumor microenvironment[118]. Overall, in an ideal world, multi-tissue reference libraries of benign immunopeptidomics data, covering as many HLA allotypes as possible enable a more reliable definition of non-mutated TAAs. The first step toward such a library has been achieved with the HLA Ligand Atlas project, described in the results section 4.1[120].

The main advantage of defining non-mutated HLA ligands as TAAs is their potential to be shared among different individuals. Shared TAAs offer a major advantage in a clinical and pharmacological setting, as they allow manufacturing and quality control of the peptide vaccine prior to patient treatment, as described by the warehouse concept[121]. The warehouse concept encompasses semi-individualized treatment options that can be readily administered after *e.g.* HLA typing of the patient. Warehousing is common in peptide vaccination strategies[66] as a first line treatment, until the patient's tumor can be analyzed and a fully individualized peptide vaccine can be designed[66].

In contradiction with the warehouse concept, we have shown that the immunopeptidome varies considerably across individuals, leading to clustering of donors rather than tissue types[120]. This highlights the main caveat of employing shared non-mutated TAAs for immunotherapy that were discovered in one patient cohort onto another one: a high possibility that the peptide in question is not presented and can therefore not elicit an immune response. T-cell responses against non-mutated TAAs can be induced *in vitro*[78,97,108,110], although *in vivo* T cell responses have not yet been translated into clinical advantage over non-treated patient groups. By being derived from self-proteins, non-mutated TAAs must overcome central tolerance, potentially by being administered with the correct adjuvant[122].

2.7.2 NEOANTIGENS

Neoantigens are generated by somatic mutations occurring during tumor development and are per definition tumor exclusive. Neoantigens might have an increased immunogenicity based on the lack of previously induced central tolerance. Several lines of evidence seem to point towards an essential role of neoantigens in antitumor immune responses. First, meta-analyses based on RNA-sequencing data from thousands of tumor samples across 18 solid tumor types form The Cancer Genome Atlas (TCGA) show that the number of neoantigens per tumor type correlated positively with gene expression profiles of cytolytic T cells[123]. Consistently, a systematic analysis of whole-exome sequencing of 619 colorectal cancers showed that a high neoantigen load is correlated with high TIL infiltration and improved survival[124]. Second, preexisting TILs with specificities against neoantigens mediate anti-tumor immune responses after immune checkpoint inhibition[125] or adoptive T cell transfer[126,127]. Third, neoantigen-specific T cells can lyse autologous tumor cells that present mutated peptides[109]. First evidence of in-human tumor regression after adoptive transfer of neoepitope-specific $CD4^+$ T cells was shown in a patient with cholangiocarcinoma[128].

Based on these findings and the ease of access to WES and RNA-sequencing, multiple clinical studies were conducted to test efficacy of neoantigen vaccination in eliciting an anti-tumor immune response[66,67,129,130]. Multi-epitope vaccination either as peptide pools[66,67,130] or RNA formulation[129] were administered in a prime-boost approach in parallel to the standard of care and delivered reproducible results. All studies reported that patients immunized with neoantigen-based vaccines displayed $CD4^+$ T-cell responses more frequently than $CD8^+$ T cell responses. Mostly polyfunctional central or effector memory T cells with low PD1 expression were induced, indicating a possibility for combination therapy with immune checkpoint inhibitors. Furthermore, both $CD4^+$ and $CD8^+$ T cells could differentiate between the neoepitope and the wild type peptide. Overall, patients

immunized with neoantigen-based vaccines displayed expanded neoantigen-specific T-cells, and these studies confirmed safety, feasibility, and low toxicity.

Nevertheless, tumors with low mutational burden lack the advantage of having a large repertoire of somatic mutations. Given that the identification of actionable neoantigens is fraught with a high false discovery rate, a large proportion might not be naturally presented on HLA molecules, thereby unable to elicit T cell responses. Thus, direct identification of naturally presented neoantigens can be achieved via proteogenomic approaches[80]. As extensively reviewed[93], even if tens of thousands of mutations are called in published data, LC-MS/MS evidence is available only for a few of them, and even less are immunogenic. Furthermore, large quantities of tumor sample are required for immunopeptidomic analyses, and the mass spectrometric limit of detection might lead to missing values due to low copy numbers.

2.7.3 MiHAs

Minor histocompatibility antigens (MiHAs) are peptides presented on HLA-I and -II molecules, that contain germline-encoded non-synonymous single nucleotide polymorphisms (SNPs) that occur only in a fraction of the population[74]. To date, MiHAs are most reliably identified through proteogenomic approaches combining next-generation sequencing to identify patient-individual SNPs and LC-MS/MS based immunopeptidomics, to validate their presentation on HLA molecules[131]. Genomic data are still insufficient to discover MiHAs since only 0.5% of non-synonymous SNPs result in MiHAs[132].

MiHAs can lead to both GvHD and GvL effects after allogenic HSCT despite matched HLA allotypes between donor and recipient. Single amino acid differences in presented peptides are detectable by T cells and can become immunoreactive. Indeed, HSCT is an efficient treatment option for various hematologic diseases, but GvHD toxicity occurs, since unselected allogeneic T cells react against a multitude of MiHAs found in all tissues. Targeting tumor-specific MiHAs directly has been shown to elicit GvL effect, without causing GvHD[133].

Both MiHAs and tumor-specific neoantigens are seen as non-self-epitopes by infused therapeutic T cells. Nevertheless, MiHAs have a series of advantages over neoantigen targets. First, germline polymorphisms that produce MiHAs are present in all cells and have not been subjected to negative selection, because only transplanted allogeneic T cells recognize MiHAs as non-self[131]. Second, in contrast to neoantigens, MiHAs are present in all cancer cells. Thus, intratumoral heterogeneity does not affect the uniform presentation of MiHAs, as opposed to sporadically presented neoantigens that originate from passenger mutations. Third, MiHAs are shared between many subjects, whereas neoantigens are currently patient-individual (reviewed in[133,134]).

2.7.4 Post-translationally modified

Proteins undergo post-translational modifications (PTMs) that modulate their function and cellular localization. Such PTMs include glycosylation, phosphorylation and a few hundred more, resulting in a large number of proteoforms that originate from the ~20,000 protein-coding genes[135]. Phosphorylation is a key modification that regulates protein function in almost all cellular

processes. The aggregated number of phosphorylation sites is steadily increasing, with ~200,000 currently described[136]. Given that dysregulation of signaling pathways and metabolic dysfunction are hallmarks of cancer[137], it stands to reason, that peptides harboring tumor-associated PTMs are presented on HLA molecules. Indeed, both HLA-I and -II presentation of phosphopeptides has been repeatedly described[83,138,139].

HLA-presented phosphopeptides have been described to be tumor-specific and immunogenic[98]. Furthermore, the same study showed that healthy individuals display immune responses with memory characteristics against phosphopeptides. Specific T cells produced IFN$_\gamma$ when stimulated with the phosphopeptide but not when stimulated with the unphosphorylated peptide or other phosphopeptides, indicating that T cell recognition was phosphate-dependent and peptide sequence-specific. Also, after HSCT, some AML patients developed responses against phosphopeptides of the same magnitude as against immunodominant CMV epitopes indicating that beside MiHAs, phosphopeptides might be targets for GvL[98].

In agreement with these observations, the same group recently published a clinical study that showed safety and immunogenicity of peptide vaccination with two phosphorylated HLA-I peptides pBCAR3$_{126\text{-}134}$ (IMDR(pT)PEKL) and pIRS2$_{1097\text{-}1105}$ (RVA(pS)PTSGV) in 15 melanoma patients[139]. Thus, phosphorylated peptides represent a new source of tumor-associated antigens that can be added to the arsenal of antigens to target.

2.7.5 CRYPTIC

HLA ligands stemming from non-canonical transcripts have been termed differently in recent years: cryptic[101], non-canonical[100], nuORFs (novel unannotated open reading frames)[99], or aberrantly expressed[99]. Generally, cryptic HLA ligands refer to a subset of peptides stemming from RNA transcripts from presumed non-coding regions that cover the 5'UTR, 3'UTR, off-frame translation events, intronic regions, lncRNAs, pseudogenes, and transposable elements [24,99,100]. Their presentation on HLA molecules has been anecdotal until recently, but through ribosome profiling (RiboSeq), an orthogonal level of evidence has been achieved. Ribosome profiling assays mRNA translation by capturing and sequencing ribosome-protected mRNA fragments. Through this method a plethora of translated novel unannotated open reading frames were discovered[99].

The large-scale, confident identification of cryptic HLA ligands has been possible only recently, with the advent of proteogenomic approaches. Accounting for all possible translated genomic regions from which cryptic peptides can arise inflates the database search space enormously, leading to a higher than reasonable number of false annotations. Therefore, reducing the database size by employing personalized databases generated from RNA sequencing data[100,101] or RiboSeq data[140] of the same sample have been used as an alternative strategy. Alternatively, *de novo* peptide sequencing of tandem mass spectra has been adopted to circumvent the database size limitation. Thus, different approaches and computational workflows have been recently proposed to confidently identify cryptic HLA ligands. Two recent algorithms include NewAnce[100] and Peptide-PRISM[24]. Other approaches focus on reducing the size of the database.

For example, Laumont and collegues[101] focused their efforts on identifying actionable cryptic TAAs by performing RNA sequencing of both human tumors and HLA-II^{hi} mTECs as "normal control". mTEC cells mediate T-cell selection to induce central tolerance to HLA peptides, mainly through expression of the autoimmune regulator protein (AIRE). They identified tumor-specific RNA reads that were three-frame translated *in silico*. They generated a canonical protein database that was concatenated with the cancer-specific proteome to create a global database individual for each sample.

Ousepnskaia and colleagues[140] generated RiboSeq profiles for 29 malignant and healthy samples and developed an analytical approach for hierarchical ORF prediction. Over all samples, they generated a high-confidence database containing 86,421 annotated and 237,427 novel unannotated ORFs. According to the authors, nuORFdb has 1.46-fold more candidate HLA-I compatible 9mers than the reference proteome, making it an adequate size for routine use in immunopeptidomics studies.

In general cryptic peptides are thought to contribute between 3.3%[99] and 10%[95] to the HLA ligand pool. Cryptic peptides share the same physicochemical properties as canonical peptides[100], and seem to be HLA-I allotype specific, with certain allotypes having a greater predisposition of presenting them[24]. Cryptic peptides can contain somatic mutations, expanding the potential neoantigenic pool[99]. However, one study screened over 500 cryptic antigens for T-cell responses in autologous TILs/PBMCs, an endeavor that lead to only one positive response[100]. It remains to be seen if cryptic peptides can elicit sustained immune responses *in vitro* and *in vivo*, apart from proof of principle studies[95].

2.7.6 PROTEASOMALLY SPLICED

Splicing of peptides in the proteasome seems to be a byproduct of protein degradation[141]. Splicing involves a transpeptidation reaction based on the formation of an acyl-intermediate of a peptide fragment with the proteasome. The ester bond is not hydrolyzed, but the free amino group of a second peptide fragment drives a nucleophilic attack onto the acyl-enzyme intermediate resulting in a non-templated peptide sequence[142]. This model supports the observation that proteasomally spliced peptides can have different intervening sequences and fragments can originate from the same protein (cis splicing) or different proteins (trans splicing). The first described proteasomally-spliced peptides have been identified by characterizing epitopes recognized by cytotoxic CD8$^+$ T-cell clones in melanoma[143] and renal cell carcinoma[144].

In 2016, a study proposed a computational method aimed to identify proteasomally spliced peptides from LC-MS/MS-based immunopeptidomics data. The results led to the extraordinary claim that proteasomally-spliced peptides account for ~30% of HLA presented peptides[102]. These proteasomally-spliced HLA-eluted peptides had the same length distribution as canonical HLA ligands, but considerably lower binding affinities towards HLA-A and -B allotypes. At the time, the authors argued that the HLA binding algorithms were trained on canonical peptides and had a low predictive power for spliced peptides. It turned out that the identification of these proteasomally-spliced peptides was fraught with an exceedingly high false discovery rate and spectra annotated to proteasomally-spliced peptides could be oftentimes explained by a PTM (data not shown) or

non-canonical reading frame[24,104]. Overall, the existence of proteasomally-spliced peptides has not been contested, but their abundance has been corrected to cover < 5% of the epitope space[24,104,145]. Additionally, proteasomally-spliced peptides have to share the same physicochemical properties as canonical peptides, and their identification must be supported by a robust FDR control[103,104].

3 AIM OF THIS STUDY

The advent of immunotherapy has marked a new avenue of efficient cancer treatment options. However, immune checkpoint inhibition, adoptive T cell therapies, and antibody-based therapies are clinically effective only in a subset of patients. Recent advances show that clinical efficacy is dependent on various factors such as tumor mutational burden (TMB)[146], abundance of preferentially $CD8^+$ TILs, and HLA allele divergence[147]. As $CD8^+$ T cells and their interplay with $CD4^+$ T cells are currently considered to be the main mediators of immune-induced responses to cancer[92], the search for their HLA-peptide targets has captured the interest of the scientific community.

In this context, the present work aims to characterize the immunopeptidomic landscape of both benign and malignant tissues. While numerous efforts have mapped the tumor, adjacent benign, and PBMC-associated immunopeptidome, there are so far no reference immunopeptidomes of benign tissues. This work presents the HLA Ligand Atlas, a first draft of a comprehensive multi-tissue resource of benign human immunopeptidomes. Evaluating the presentation of TAAs on other tissues than the original tumor source enables prevention of potential on-target/off-tumor adverse events. On-target/off-tumor-mediated adverse events in immunotherapy settings can have lethal outcomes[114,148]. With the data encompassed in the HLA Ligand Atlas, potential TAAs with a broad HLA representation on benign tissues can be filtered out at an early stage of target discovery, a quality control step that can substantially improve future TAA safety profiles. Furthermore, shared cryptic HLA-presented peptides were identified from benign tissues, and the RAW LC-MS/MS data can be searched for post-translationally modified peptides, and other classes of HLA presented antigens.

A further objective this work is to address a series of technical aspects with respect to the mass spectrometric instrumentation, an essential prerequisite for in-depth characterization of immunopeptidomes from low amounts of sample input.

A third aim is to characterize the immunopeptidomic landscape of breast cancer. Breast cancer is the most frequent cancer in women and is a tumor type with a low mutational burden. Particularly the triple negative molecular subtype has a poor clinical prognosis and would benefit immensely from further therapeutic options. Therefore, we comparatively profiled the immunopeptidomes of breast cancer patients with the multi-tissue dataset encompassed in the HLA Ligand Atlas. We were able to define shared non-mutated TAAs, eligible for peptide vaccination. Following completion of the LC-MS/MS data acquisition and RNA and exome sequencing in 10 breast cancer patients, potential TAAs originating from other sources such as cryptic peptides or neoantigens, will be evaluated. All these HLA-based TAAs have a broad applicability for cellular therapies such as adoptive cell transfer, but more importantly for peptide vaccination purposes.

4 RESULTS

The results section of this thesis is divided into three parts.

Part I describes **"The HLA Ligand Atlas – A resource of natural HLA ligands presented on benign tissues."** The HLA Ligand Atlas is the first collection of benign ligandomes isolated from 29 tissues across 21 human subjects. The data is easily searchable via the online resource hla-ligand-atlas.org. This project is finalized and in the process of publication.

Part II describes a benchmark study between five different LC-MS/MS setups from three vendors with the purpose of evaluating complementary technologies suitable for immunopeptidomics experiments. For this purpose, samples comprising a serial dilution were distributed for LC-MS/MS analysis. Based on these results, two mass spectrometers were selected to complement the available Orbitrap Fusion Lumos, the q Exactive HF and the TIMS TOF Pro. This project is finalized, the instruments chosen, have been acquired.

Part III focuses on defining tumor-associated antigens in breast cancer. The basis for this project is the benign immunopeptidome defined in part I. Comparative immunopeptidomic profiling reveals a set of antigens shared between three molecular subtypes and multiple subjects. This project is not finalized yet, LC-MS/MS measurements, as well as RNA- and exome sequencing data for 10 patients are being acquired. The final data analysis will be complemented with immunogenicity experiments.

The HLA Ligand Atlas - A resource of natural HLA ligands presented on benign tissues

Ana Marcu§[1,2], Leon Bichmann§[1,3], Leon Kuchenbecker§[3], Daniel Johannes Kowalewski[1], Lena Katharina Freudenmann[1,2,4], Linus Backert[1,3], Lena Mühlenbruch[1,2,4], András Szolek[3], Maren Lübke[1,2], Philipp Wagner[1,5], Tobias Engler[5], Sabine Matovina[5], Jian Wang[6], Mathias Hauri-Hohl[7], Roland Martin[6], Konstantina Kapolou[8], Juliane Sarah Walz[2,9,10], Julia Velz[8], Holger Moch[11], Luca Regli[8], Manuela Silginer[12], Michael Weller[12], Markus W. Löffler[1,2,4,13,14], Florian Erhard[15], Andreas Schlosser[16], Oliver Kohlbacher[2,3,17,18,19,20,21], Stefan Stevanović[1,2,4], Hans-Georg Rammensee[1,2,4], Marian Christoph Neidert[12,22,23]

§ Ana Marcu, Leon Bichmann, and Leon Kuchenbecker contributed equally to this work.

Available at *bioRxiv*, https://doi.org/10.1101/778944

AM, LBi and LK contributed equally to this manuscript. AM performed the experimental work that encompasses the isolation of HLA ligands from the primary tissue samples; performed the LC-MS/MS data acquisition; the first data analysis using the vendor-specific software Proteome Discoverer 1.4; and further in-house available software tools (not described in this work). Furthermore, AM managed the sample and metadata documentation and project organization in Tübingen. AM drafted the manuscript and was involved in all steps of data analysis and figure generation and created Figure 1A and B, Figure 5A and F, Figure 7, and Figure S6. LBi systematically reprocessed all LC-MS/MS files with the open-source database search tool MHCquant. LK created the online resource available at hla-ligand-atlas.org and the entire database architecture behind. Data analysis and all remaining figures were collectively performed by AM, LBi and LK. LBi created Figure 3, Figure 4, Figure 5D and E, Figure S1, Figure S2, Figure S4A and B, and Figure S5. LK created Figure 1C and D, Figure 2, Figure 5B and C, Figure 7, Figure S3, Figure S4C and D.
ASz performed the high resolution typing from whole exome sequencing data. FE and ASc used Peptide-PRISM to search the LC-MS/MS files for cryptic peptides. LKF and LM performed experimental isolation of

Results

HLA ligands from 8 autopsy samples and 5 thymus samples, respectively. LKF performed the experimental isolation of HLA ligands and the LC-MS/MS acquisition from the three glioblastoma patients, performed the population frequency analysis, and helped with the experimental handling of the time series experiment. MWL organized the ethical approval in Tübingen and edited the manuscript. ML and JSW edited the manuscript. DJK, LBa, MCN initiated the project and edited the manuscript. MCN and HM provided all autopsy samples. KK, JV, LR, MS, MW organized the sample management in Zürich. PW, TE and SM provided the two benign ovary samples for the time series experiment. JW, MH-H, and RM provided the five thymus samples. OK edited the manuscript. OK, SS, HGR, MCN supervised the project and contributed resources.

1 Department of Immunology, Interfaculty Institute for Cell Biology, University of Tübingen, Tübingen, Baden-Württemberg, 72076, Germany.
2 Cluster of Excellence iFIT (EXC 2180) "Image-Guided and Functionally Instructed Tumor Therapies", University of Tübingen, Tübingen, Baden-Württemberg, 72076, Germany.
3 Applied Bioinformatics, Dept. of Computer Science, University of Tübingen, Tübingen, Baden-Württemberg, 72074, Germany.
4 DKFZ Partner Site Tübingen, German Cancer Consortium (DKTK), Tübingen, Baden-Württemberg, 72076, Germany.
5 Department of Obstetrics and Gynecology, University Hospital of Tübingen, Tübingen, Baden-Württemberg, 72076, Germany.
6 Neuroimmunology and MS Research Section (NIMS), Department of Neurology, University Hospital Zurich, University of Zurich, 8091 Zurich, Switzerland
7 Pediatric Stem Cell Transplantation, University Children's Hospital Zurich, 8032 Zurich, Switzerland
8 Clinical Neuroscience Center and Department of Neurosurgery, University Hospital and University of Zürich, Zürich, Zürich, 8091, Switzerland.
9 Clinical Collaboration Unit Translational Immunology, German Cancer Consortium (DKTK), University Hospital of Tübingen, Tübingen, Baden-Württemberg, 72076, Germany.
10 Department of Hematology and Oncology, University of Tübingen, Tübingen, Baden-Württemberg, 72076, Germany.
11 Department of Pathology, University Hospital and University of Zurich, Zürich, Zürich, 8091 Switzerland.
12 Clinical Neuroscience Center and Department of Neurology, University Hospital and University of Zurich, Zürich, Zürich, 8091 Switzerland.
13 Department of General, Visceral and Transplant Surgery, University Hospital of Tübingen, Tübingen, Baden-Württemberg, 72076, Germany.
14 Department of Clinical Pharmacology, University of Hospital Tübingen, Tübingen, Baden-Württemberg, 72074, Germany.
15 Institute for Virology and Immunobiology, Julius-Maximilians-University Würzburg, Würzburg, Freistaat Bayern, 97070, Germany.
16 Rudolf Virchow Center for Experimental Biomedicine, Julius-Maximilians-University Würzburg, Würzburg, Freistaat Bayern, 97080, Germany.
17 Quantitative Biology Center (QBiC), University of Tübingen, Tübingen, Baden-Württemberg, 72076, Germany.
18 Biomolecular Interactions, Max Planck Institute for Developmental Biology, Tübingen, Baden-Württemberg, 72076 Germany.
19 Cluster of Excellence Machine Learning in the Sciences (EXC 2064), University of Tübingen, Tübingen, Baden-Württemberg, 72076, Germany.
20 Institute for Translational Bioinformatics, University Hospital Tübingen, Tübingen, Baden-Württemberg, 72076 Germany
21 Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen, Baden-Württemberg, 72076 Germany
22 Neuroscience Center Zurich (ZNZ), University of Zurich and ETH Zurich, Zurich, Zürich, 8057, Switzerland
23 Department of Neurosurgery, Cantonal Hospital St. Gallen, St. Gallen, St. Gallen, 9007, Switzerland.

4.1.1 ABSTRACT

The human leukocyte antigen (HLA) complex controls adaptive immunity by presenting defined fractions of the intracellular and extracellular protein content to immune cells. Here, we describe the HLA Ligand Atlas, an extensive collection of mostly matched HLA-I and -II ligandomes from 225 benign samples (29 tissues, 21 subjects). The initial release covers 51 HLA-I and 86 HLA-II allotypes presenting 89,853 HLA-I- and 140,861 HLA-II ligands. We observe that the immunopeptidomes differ considerably between tissues and individuals on both source protein and HLA-ligand level. 1,407 HLA-I ligands stem from non-canonical genomic regions. We highlight the importance of comparatively analyzing both benign and malignant tissues to inform tumor association, based on a case study in three glioblastoma patients. The resource provides insights into applied and basic immune-associated questions in the context of cancer immunotherapy, infection, transplantation, allergy, and autoimmunity. It is publicly available at www.hla-ligand-atlas.org.

4.1.2 INTRODUCTION

In the past two decades, sequencing the human genome (genomics)[149,150], transcriptome (transcriptomics)[2,3], and proteome (proteomics)[48,151,152] have been milestones that enable a multi-dimensional understanding of biological processes. In the context of the immune system, a subsequent omics layer can be defined as the HLA ligandome or the immunopeptidome, comprising the entirety of HLA presented peptides. HLA molecules present peptides on the cell surface for recognition by T cells. These T cells can distinguish self from foreign[73,153] peptides, a crucial mechanism in adaptive immunity. Despite HLA-I ligands originating primarily from intracellular proteins, the correlation with their precursors (mRNA transcripts and proteins) is poor[97,116,117], limiting approaches based on in silico HLA-binding predictions in combination with transcriptomics and proteomics data alone[154,155].

The importance of investigating HLA ligandomes from human healthy and diseased tissues has been well recognized[94,156,157] to improve HLA-binding prediction algorithms[87,90,91,158], and immunogenicity prediction analysis[159,160], but also, to inform precision medicine[93,108,161]. Direct evidence of naturally presented HLA ligands is required to prove visibility of target peptides to T cells. This is a challenge, for example, in the context of cancer immunotherapy approaches that aim to identify optimal tumor-specific HLA-presented antigens[66,80,108]. While their discovery has been made possible by proteogenomics approaches, a major impediment still resides in the lack of benign tissues as a reference for the definition of tumor specificity of target peptides[97,100,101]. Due to the scarce availability of benign human tissue ligandomes, common alternative strategies are based on transcriptomic datasets either from the same patient, or from multiple tissues extracted from publicly available repositories[2,3]. Frequently, morphologically normal tissue adjacent to the tumor (NATs, normal tissues adjacent to tumor) is used as a control in cancer research. However, NATs have been shown to pose unique challenges, since they may be affected by disease and have been suggested to represent a unique intermediate state between healthy and malignant tissues, with a pan-cancer-induced inflammatory response[118]. Additionally, for some malignancies e.g. of the brain, surgical resection of NATs is inadmissible. Even in cancers that allow for the extraction

of NATs, it is still necessary to investigate the presence of potential tumor-associated targets (TAAs) on other tissues to anticipate on-target/off-tumor, systemic adverse events when administering immunotherapies to patients[114,148].

In this study we thus employed tissues originating from research autopsies of subjects that have not been diagnosed with any malignancy and have deceased for other reasons, an approach previously described as a surrogate source of benign tissue[118,162]. Although these tissues were affected by a range of non-malignant diseases, we designate their tissues as benign to emphasize morphological normality and absence of malignancy. This definition of benign is in agreement with the definition used by the Genotype-Tissue Expression Consortium[2,3], which provides RNA sequencing data of benign tissues originating from autopsy specimens.

We performed a large-scale mass spectrometry (LC-MS/MS)-based characterization of both HLA-I and -II ligands providing data from benign human tissues obtained at autopsy. The HLA Ligand Atlas is a first draft of a pan-tissue immunopeptidomics reference library from benign tissues comprising for the first time 225 mostly paired HLA-I (198) and -II (217) ligandomes from 29 different benign tissue types obtained from 21 human subjects. For the data analysis, we employed MHCquant[163], the first open-source customized computational tool for immunopeptidomics assays that provides database search, false discovery rate (FDR) scoring, label-free quantification and binding affinity predictions. In addition, we implemented a user-friendly, web-based interface to query and access the data at https://hla-ligand-atlas.org. Despite its unprecedented comprehensiveness, the HLA Ligand Atlas currently contains only a limited number of tissues and individuals. However, it has been designed as an open and extensible community resource and we strongly encourage the submission of additional data for inclusion. Consistent quality control and data processing will ensure a high quality of the data.

4.1.3 Materials and Methods

Experimental Model and Subject Details

Human tissue samples were obtained post-mortem during autopsy performed for medical reasons at the University Hospital Zürich. The study was approved by the Cantonal Ethics Committee Zürich (KEK) (BASEC-Nr. Req-2016-00604). For none of the included patients a refusal of post-mortem contribution to medical research was documented and study procedures are in accordance with applicable Swiss law for research on humans (Bundesgesetz über die Forschung am Menschen, Art. 38). In addition, the study protocol was reviewed by the ethics committee at the University of Tübingen and received a favorable assessment without any objections to the study conduct (Project Nr. 364/2017BO2).

None of the subjects included in this study was diagnosed with any malignant disease. Tissue samples were collected during autopsy, which was performed within 72 hours after death. Tissue annotation was performed by a board-certified pathologist. Tissue samples were immediately snap-frozen in liquid nitrogen.

Thymus samples were obtained from the University Children' s Hospital Zürich/ Switzerland. Thymus tissue was removed during heart surgery or for other medical reasons. Tissue samples from residual material not required for diagnostic or other medical purposes were obtained after informed consent from the parents of the respective children, in accordance with the principles of the Declaration of Helsinki. The study was approved by the Cantonal Ethics Committee Zürich (KEK) (EC-Nr. 2014-0699, PB_2017-00631) on February 27th 2015.

Furthermore, two benign ovarian tissue samples were collected for the project (OVA-DN278 and OVA-DN281). Both patients were postmenopausal and had a bilateral ovarectomy for cystadenofibromas, which were diagnosed by histopathologic examination of the specimen. The samples were obtained from a normal part of the ovary. The study was approved by the ethical committee of the University of Tübingen (354/2011BO2).

Finally, we included three primary glioblastoma tumor samples to illustrate a selection strategy for tumor associated antigens. The primary glioblastoma tumor was removed for patients GBM616 and GBM654, whereas, a recurrent tumor was analyzed for GBM753. The study was approved by the Cantonal Ethics Committee Zürich (KEK) (EC-Nr. 2014-0699, PB_2017-00631).

HLA Typing

Multiple HLA typing approaches were performed for the different sources of patient material. Autopsy subject AUT-DN08, AUT-DN16, and two benign ovary samples (OVA-DN278 and OVA-DN281) were typed at the Department of Transfusion Medicine of the University Hospital of Tübingen. High-resolution HLA typing was performed by next-generation sequencing on a GS Junior Sequencer using the GS GType HLA Primer Sets (both Roche, Basel, Switzerland). HLA typing was successful for HLA-A, -B, and -C alleles. However, HLA-II typing was only reliable for the HLA-DR locus, and incomplete for the HLA-DP and -DQ loci.

Therefore, we performed exome sequencing of lung tissue for remaining autopsy subjects. Exome sequencing data was processed and OptiType[164] was employed to identify HLA-I and -II alleles. Finally, sequence-based typing was performed for the five thymus samples and the three glioblastoma samples, by sequencing exons 1-8 for HLA-I alleles and exons 2-6 for HLA-II alleles (Histogenetics, Ossining, NY). The subject characteristics are summarized in Supplemental Table S1 encompassing information on sex, age, the number of collected tissues and HLA-I and II alleles.

HLA Immunoaffinity Purification

HLA-I and -II molecules were isolated from snap-frozen tissue using standard immunoaffinity chromatography[165]. The antibodies employed were the pan-HLA-I-specific antibody W6/32[81], and the HLA-DR-specific antibody L243[85], produced in house (University of Tübingen, Department of Immunology) from HB-95, and HB-55 cells (ATCC, Manassas, VA) respectively. Furthermore, the pan-HLA-II-specific antibody Tü39 was employed and produced in house from a hybridoma clone as previously described[86]. The antibodies were cross-linked to CNBr-activated sepharose (Sigma-Aldrich, St. Louis, MO) at a ratio of 40 mg sepharose to 1 mg antibody for 1 g tissue with 0.5 M NaCl, 0.1 M NaHCO3 at pH 8.3. Free activated CNBr reaction sites were blocked with 0.2 M glycine.

For the purification of HLA-peptide complexes, tissue was minced with a scalpel and further homogenized with the Potter-Elvehjem instrument (VWR, Darmstadt, Germany). The amount of tissue employed for each purification is documented in Supplemental Table S1. This information is not available for seven tissues, annotated as n.d. in said table. Tissue homogenization was performed in lysis buffer consisting of CHAPS (Panreac AppliChem, Darmstadt, Germany), and one cOmplete™ protease inhibitor cocktail tablet (Roche) in PBS. Thereafter, the lysate was sonicated and cleared by centrifugation for 45 min at 4,000 rpm, interspaced by 1 h incubation periods on a shaker at 4°C. Lysates were further cleared by sterile filtration employing a 5 µm filter unit (Merck Millipore, Darmstadt, Germany). The first column contained 1 mg of W6/32 antibody coupled to sepharose, whereas the second column contained equal amounts of Tü39 and L243 antibody coupled to sepharose. Finally, the lysates were passed through two columns cyclically overnight at 4°C. Affinity columns were then washed for 30 minutes with PBS and for 1 h with water. Elution of peptides was achieved by incubating four times successively with 100 – 200 µl 0.2% TFA on a shaker. All eluted fractions were subsequently pooled. Peptides were separated from the HLA molecule remnants by ultrafiltration employing 3 kDa and 10 kDa Amicon filter units (Merck Millipore) for HLA-I and HLA-II, respectively. The eluate volume was then reduced to approximately 50 µl by lyophilization or vacuum centrifugation. Finally, the reduced peptide solution was purified five times using ZipTip Pipette Tips with C18 resin and 0.6 µl bed volume (Merck,) and eluted in 32.5% ACN/0.2% TFA. The purified peptide solution was concentrated by vacuum centrifugation and supplemented with 1% ACN/0.05% TFA and stored at -80°C until LC-MS/MS analysis.

Time Series Experiments

We performed time series experiments to assess the suitability of tissues obtained from autopsies as a source of human tissues for the characterization of the benign immunopeptidome. We evaluated the degradation profile of the immunopeptidome, when tissues were stored at 4°C for up to 72 h after tissue removal, to mimic the conditions at autopsy. The time series experiment was repeated in three benign tissues from different individuals: one benign liver obtained at autopsy (AUT-DN16 Liver), and two benign ovaries removed surgically (OVA-DN278 and OVA-DN281). The tissues were extracted and incubated at 4°C until a certain time point and flash-frozen in liquid nitrogen until HLA ligand extraction. As more tissue was available form AUT-DN16 Liver, tissue samples were frozen after 8 h, 16 h, 24 h, 48 h, and 72 h. Due to the limited sample amount obtained from OVA-DN278 and OVA-DN281, only three time points could be accounted for: 0 h, 24 h, and 72 h. The HLA immunoaffinity purification was performed as mentioned, with the exception that mass to volume ratio in ovary samples was adjusted to the lowest mass across all time points before loading onto sepharose columns.

Mass Spectrometric Data Acquisition

HLA ligand characterization was performed on an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with a Nanospray Flex™ Ion Source (Thermo Fisher Scientific) coupled to an Ultimate 3000 RSLC Nano UHPLC System (Thermo Fisher Scientific). Peptide samples were loaded with 1% ACN/ 0.05% TFA on a 75 µm x 2 cm Acclaim™ PepMap™ 100 C18 Nanotrap column (Thermo Fisher Scientific) at a flow rate of 4 µl/min for 10 minutes. Separation was performed on a 50 µm x 25 cm PepMap RSLC C18 (Thermo Fisher Scientific) column, with a particle size of 2 µm. Samples were eluted with a linear gradient from 3% to 40% solvent B (80% / 0.15% FA in water) at a flow rate of 0.3 µl/min over 90 minutes. The column was subsequently washed by increasing to 95% B within 1 minute, and maintaining the gradient for 5 minutes, followed by reduction to 3% B and equilibration for 23 minutes.

Data acquisition was performed as technical triplicates in data-dependent mode, with customized top speed (3 s) methods for HLA-I- and HLA-II-eluted peptides. HLA-I peptides have a length of 8 - 12 amino acids[21,22] therefore, the scan range was restricted to 400 - 650 m/z and charge states of 2 - 3. MS1 and MS2 spectra were detected in the Orbitrap with a resolution of 120,000 and 30,000 respectively. Furthermore, we set the automatic gain control (AGC) targets to $1.5*10^5$ and $7.0*10^4$ and the maximum injection time to 50 ms and 150 ms for MS1 and MS2, respectively. The dynamic exclusion was set to 7 s. Peptides were fragmented with collision-induced dissociation (CID) while the collision energy was set to 35%.

HLA-II peptides have a length of 8 - 25 amino acids[22,166], thus the scan range was set to 400 -1,000 m/z and the charge states were restricted to 2 - 5. Readout for both MS1 and MS2 were performed in the Orbitrap with the same resolution and maximum injection times as for HLA-I peptides. The dynamic exclusion was set to 10 s and AGC values employed were $5.0*10^5$ and $7.0*10^4$ for MS1 and MS2, respectively. Higher-energy collisional dissociation (HCD) fragmentation with 30% collision energy was employed for HLA-II peptides.

Database Search with MHCquant

MS data obtained from HLA ligand extracts was analyzed using the nf-core[167] containerized, computational pipeline MHCquant[163] (release 1.5.1 - https://www.openms.de/mhcquant/) with default settings. The workflow comprises tools to analyze LC-MS/MS data of the open-source software library OpenMS (2.5)[168]. Identification and post-scoring were performed using the OpenMS adapters to Comet 2016.01 rev. 3[169] and Percolator 3.4[170] at a local peptide-level false discovery rate (FDR) threshold of 1% among the technical replicates per sample. Subsequently, we estimated the global peptide-level FDR by dividing the sum of expected false positive identifications from each sample (1% peptide level FDR) by the total number of identified peptides in the entire dataset (HLA-I: 4.5% FDR, HLA-II: 3.9% FDR)[171,172]. The human reference proteome (Swiss-Prot, Proteome ID UP000005640, 20,365 protein sequences) was used as a database reference. Database search was performed without enzymatic restriction, with methionine oxidation as the only variable modification. MHCquant settings for high-resolution instruments involving a precursor mass tolerance of 5 ppm and a fragment bin tolerance of 0.02 Da were applied. The peptide length restriction, digest mass and charge state range were set to 8-12 amino acids, 800-2500 Da and 2-3 for HLA-I and 8-25 amino acids, 800-5000 Da and 2-5 for HLA-II, respectively.

HLA Binding Prediction

Peptide binding predictions were computed based on the subject's HLA alleles. For HLA-I ligand extracts, we employed SYFPEITHI[173] and NetMHCpan-4.0[174] in ligand mode (default). The SYFPEITHI score s_{SYF} was computed by dividing the sum of amino acid-specific values for each position in the tested peptide by the maximally attainable score for the respective HLA allotype[175]. HLA-II ligand extracts were annotated with NetMHCIIpan-4.0[87] and MixMHC2pred[91] using the default settings.

Peptides were categorized as strong binders against a given HLA allotype if either netMHCpan-4.0, netMHCIIpan-4.0 or MixMHC2pred reported a percentile rank score $s_{\mathrm{rank}} \leq 0.5$. Peptides were reported as weak binders if any of the tools reported $s_{\mathrm{rank}} \leq 2.0$ or in case of SYFPEITHI $s_{\mathrm{SYF}} \geq 0.5$. All peptide-HLA allotype associations within these limits were included in the dataset, i.e., a single peptide sequence can be reported as a binder against multiple allotypes of the same donor. Unless allele associations are specified, all peptides including classified non-binders against any subject's allotype were included in the analysis.

Binding Prediction and Length Distribution-Based Quality Control

We defined the fraction of predicted binders of a sample as the ratio of predicted binders divided by the total number of peptide identifications. Technical replicates with a fraction of predicted binders lower than 50% for HLA-I and lower than 10% for HLA-II ligand extracts were excluded from the dataset. Furthermore, individual replicates were removed from the dataset if the mode of the length distribution differed from 9 amino acids for HLA-I and was not in the interval [12, 18] for HLA-II (see Figure S1).

Quantitative Time Series Analysis

Database search of LC-MS/MS data from the three time series experiments was performed with MHCquant 1.5.1 as previously described[163]. Identifications were matched between runs[176] based on retention time alignment and targeted feature extraction[177] to integrate respective MS1 areas for all time points and technical replicates. MS1 areas x were normalized to z-scores (standard scores) z per MS run by subtracting the mean and dividing by the standard deviation:

$$z = \frac{(x - \mu)}{\sigma}$$

The trajectory of scaled MS1 areas was clustered by k-means unsupervised clustering with 6 seeds using the tslearn (v.0.3.1) python package. All trajectories are related to the first time point by subtracting its median z-score from all other timepoints in the respective analysis.

Comparison of the HLA-Ligand-Atlas data base with IEDB and SysteMHC

All peptides contained in the HLA Ligand Atlas database were compared with peptides listed in the IEDB and SysteMHC databases for HLA-I and HLA-II ligands separately. The list of peptides stored in the IEDB was obtained by downloading the file "epitope_full_v3.zip" from the "Database Export" page. The obtained table was subsequently filtered for positive MS assays, linear peptides and human origin. Peptides with modifications were removed. Peptides stored in the SysteMHC database were obtained by downloading the file "180409_master_final.tgz" from "Builds_for_download" page. The obtained table was subsequently filtered for human as organism.

Gene Ontology (GO)-Term Enrichment

GO term enrichment analyses were performed with the Panther 15.0 database (Released 2020-02-21) with the integrated "statistical overrepresentation test" (Release 2019-07-11). Gene identifiers of source proteins presented exclusively by either HLA-I or -II allotypes were queried against the "GO cellular component complete" database using the default "Homo sapiens genes" reference list. GO terms were sorted by Fisher's exact raw p-value, and top 10 scoring terms reported as overrepresented and their corresponding p-values were selected for illustration.

Tissue-specific source proteins were defined as HLA-I or -II source proteins identified exclusively in one tissue across all subjects (Table S5). Gene identifiers of tissue-specific HLA-I and -II source proteins were queried against the "GO biological process complete" database, with the only difference that only the top 5 scoring terms reported as overrepresented were selected for illustration.

Tissue-Specific Gene Set enrichment

Analogously to the GO-term enrichment, tissue-specific HLA-I and -II source proteins were separately queried against the GTEx database for gene set enrichment analysis. Gene sets with upregulated gene expression profiles per tissue "GTEx_Tissue_Sample_Gene_Expression_Profiles_up" were retrieved using the gseapy

implementation (v.0.9.15, 2019-08-07) through the enrichr API. All tissues covered in the HLA Ligand Atlas were matched and compared against all tissues in the GTEx database that co-occur in the HLA Ligand Atlas. Fisher's exact raw p-values for the enrichment were computed for each pairwise comparison.

HLA-I and II Peptide Yield Correlation To Expression of Immune-Related Genes

We computed a linear model to compare the median HLA-I peptide yields per tissue with gene expression values (RPKM) of the following genes involved in the HLA-I presentation pathway: HLA-A, HLA-B, HLA-C, immunoproteasome, constitutive proteasome, TAP1, and TAP2. Median HLA-II peptide yields per tissue were correlated to genes involved in the HLA-II presentation pathway: HLA-DRB1, HLA-DRA, HLA-DQB1, HLA-DQA1, HLA-DPB1, HLA-DPA1. The corresponding gene expression values were taken from a previously published study[178].

An ordinary least squares linear model correlating gene expression and $\log_{10}$ median HLA-I and -II peptide yields was computed using R (v.3.5) and the corresponding stats (v.3.5) package reporting R^2, F-statistic p-value, and spearman rho. The cross correlation between all immune related genes and their individual linear models (Figure 3, Figure S5) was computed using R (v.3.5) and the corresponding packages corrplot (v. 0.84) and ggplot2 (v.3.2.1). As the expression levels of the investigated genes are highly covariant (Figure S5A, S5C), the regression would be overfitting when correlating peptide yields to multiple genes involved in the antigen presentation pathway, thus the analysis was limited to a single gene at a time.

Computation of Jaccard coefficients between samples

We investigated the similarity of immunopeptidomes between tissues and subjects by pairwise comparison of all samples in the HLA Ligand Atlas. Comparisons were performed both on HLA-I and -II level as well as on peptide and source protein level. The Jaccard index was calculated by dividing the set intersection by the set union for all pairwise comparisons:

$$j = \frac{A \cap B}{A \cup B}$$

Identification of Cryptic Peptides with Peptide-PRISM

Identification of cryptic HLA-I peptides from HLA-I LC-MS/MS data was performed as recently described in detail[24]. Briefly, *de novo* peptide sequencing was performed with PEAKS Studio X [179,180] (Bioinformatics Solutions Inc., Canada). Top10 sequence candidates were exported for each fragment ion spectrum. Database matching of all sequence candidates and stratified FDR-filtering was performed with Peptide-PRISM using the 6-frame translation of the human genome (HG38) and the 3-frame translation of the human transcriptome (Ensembl 90). Matched peptides were filtered to 10% FDR and peptides were predicted as binder to the corresponding HLA alleles by NetMHCpan-4.0[174].

Retention Time Model for Cryptic Peptide Validation

Retention time predictions were carried out using the OpenMS (2.5.0) RTModel based on oligo-kernel ν-support vector regression (ν=0.5, p=0.1, c=1, degree=1, border_length=22, kmer_length=1, Σ=5)[181]. The model was trained on all peptide identifications of canonical peptides identified with MHCquant and applied to all cryptic peptide identifications resulting from Peptide-PRISM. Predictions were evaluated by applying a linear least square fit to compute the 99% prediction interval around the predicted versus measured retention times using the statsmodels (v.0.11) function wls_prediction_std.

Synthesis of Isotope-Labeled Peptides

Peptides were synthesized using the Liberty Blue Automated Peptide Synthesizer (CEM) following the standard 9-fluorenylmethyl-oxycarbonyl/tert-butyl strategy. After removal from the resin by treatment with trifluoroacetic acid/triisopropylsilane/water (95/2.5/2.5 by vol.) for 1 h, peptides were precipitated from diethyl ether, washed three times with diethyl ether and resuspended in water prior to lyophilization. Purity and identity of the synthesis products were determined by C18-HPLC (Thermo Fisher Scientific, Darmstadt, Germany) and LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific), respectively.

Spectrum Validation

We selected 36 cryptic peptides, identified with 1% FDR for spectral validation with isotope-labeled synthetic peptides. Selected peptides were strong binders to the corresponding HLA alleles of the respective subject, with a netMHCpan-4.0 binding rank <0.5.

Isotope-labeled synthetic peptides were spiked into a sample matrix of native HLA eluted peptides from a JY cell line at a concentration of 20 fmol/µl, with the purpose of showing spectrum identity between the native and synthetic peptides.

The spectral similarity λ was computed analogous to the normalized spectral contrast angle[182] between eluted peptide spectra and synthetic isotope labeled peptide spectra:

$$\lambda(S_1, S_2) = 1 - \frac{2\cos^{-1}(S_1 \cdot \mathrm{S}_2)}{\pi},$$

where the spectra were encoded as intensity vectors (S_1 and S_2) based on their theoretical *b* and *y* fragment ions by using the mzR (v2.16.2), msdata (v0.20.0) and protViz (v0.4) R packages. Intensities of matching y- and b-ion pairs as encoded in the intensity vectors were compared, thereby avoiding the necessity to correct for the mass shift caused by the isotope label. Peaks present in at least one of the spectra were considered for the cross product ($S_1 \cdot S_2$). Intensities of missing peaks in the one spectrum compared to the other were set to zero.

Data Storage Web Interface

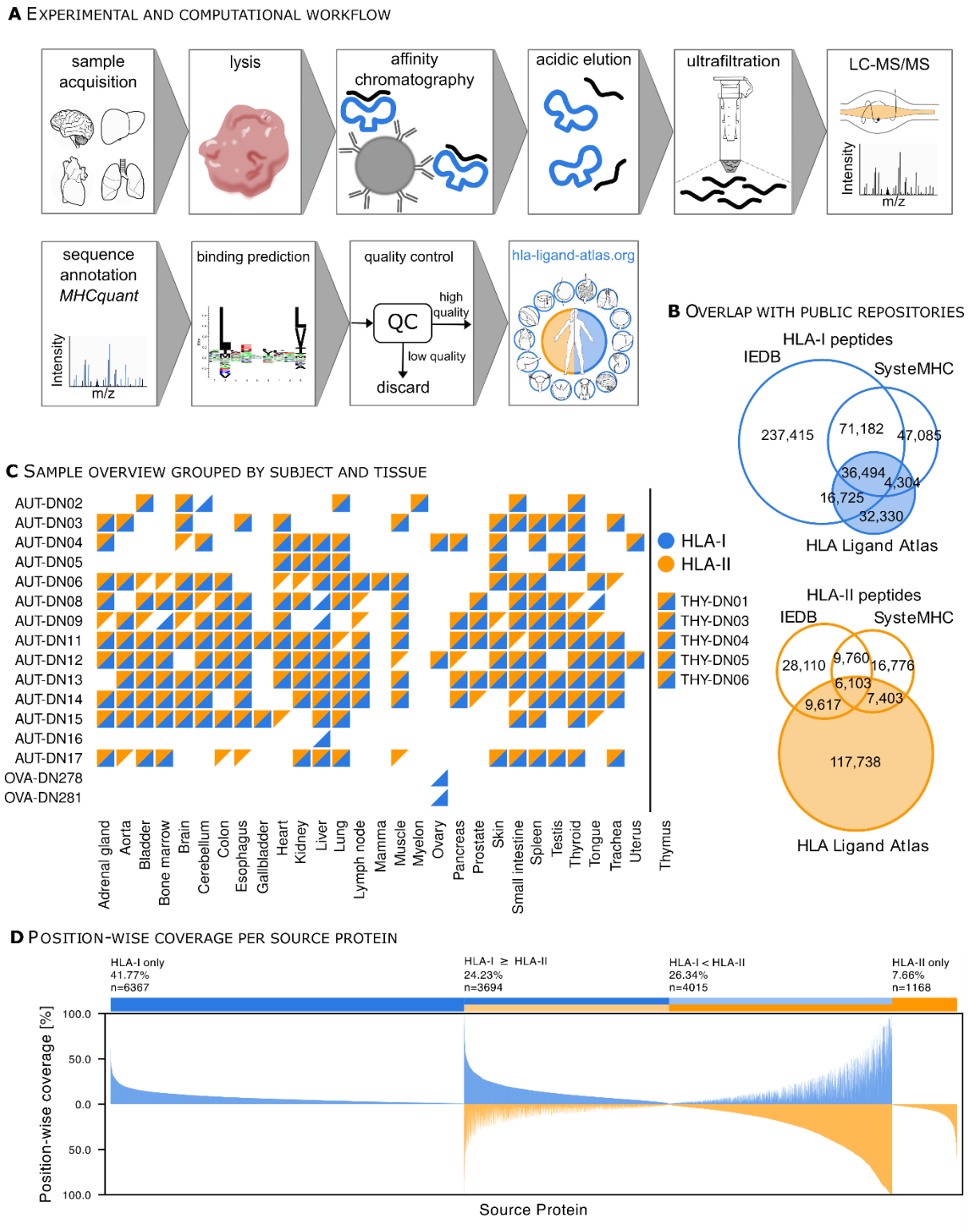
Data was stored and managed using the biomedical data-management platform qPortal[183]. HLA-I and -II peptides were complemented with their tissue and HLA allotype association and stored

in an SQL database. A public web server was implemented that allows users to formulate queries against the database, visualize results and allows data export for further analysis. The web front-end was implemented in HTML, CSS and JavaScript based on the front-end framework Bootstrap 4. The table plugin DataTables was used to provide rapid browsing and filtering for tabular data. Interactive plots were designed using Bokeh and ApexCharts.

Data Availability

The LC-MS/MS immunopeptidomics data comprised in the HLA Ligand Atlas has been deposited to the ProteomeXchange Consortium via the PRIDE [184] partner repository with the dataset identifier PXD019643 and the project DOI 10.6019/PXD019643. LC-MS/MS runs and sample not adhering to the implemented quality control thresholds are deposited as well.

The LC-MS/MS immunopeptiodmics data from the three glioblastoma patients can be accessed with the PXD020186, and the project DOI 10.6019/PXD020186.

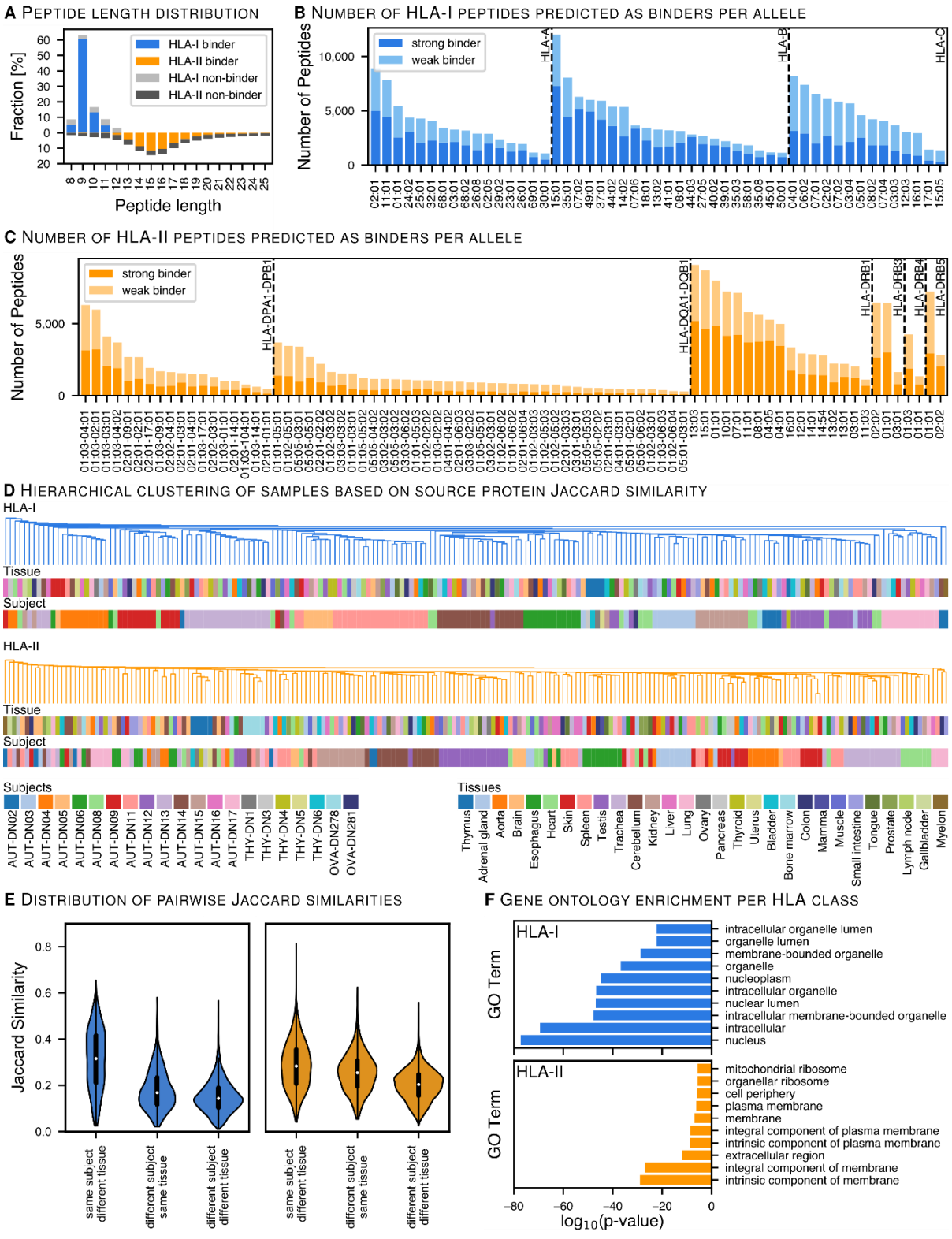

A EXPERIMENTAL AND COMPUTATIONAL WORKFLOW
sample acquisition
lysis
affinity chromatography
acidic elution
ultrafiltration
LC-MS/MS
Intensity
m/z
sequence annotation
MHCquant
binding prediction
quality control
QC
high quality
low quality
discard
hla-ligand-atlas.org
B OVERLAP WITH PUBLIC REPOSITORIES
HLA-I peptides
IEDB
SysteMHC
237,415
71,182
47,085
36,494
4,304
16,725
32,330
HLA Ligand Atlas
HLA-II peptides
IEDB
SysteMHC
28,110
9,760
16,776
6,103
9,617
7,403
117,738
HLA Ligand Atlas
C SAMPLE OVERVIEW GROUPED BY SUBJECT AND TISSUE
AUT-DN02
AUT-DN03
AUT-DN04
AUT-DN05
AUT-DN06
AUT-DN08
AUT-DN09
AUT-DN11
AUT-DN12
AUT-DN13
AUT-DN14
AUT-DN15
AUT-DN16
AUT-DN17
OVA-DN278
OVA-DN281
HLA-I
HLA-II
THY-DN01
THY-DN03
THY-DN04
THY-DN05
THY-DN06
Adrenal gland
Aorta
Bladder
Bone marrow
Brain
Cerebellum
Colon
Esophagus
Gallbladder
Heart
Kidney
Liver
Lung
Lymph node
Mamma
Muscle
Myelon
Ovary
Pancreas
Prostate
Skin
Small intestine
Spleen
Testis
Thyroid
Tongue
Trachea
Uterus
Thymus
D POSITION-WISE COVERAGE PER SOURCE PROTEIN
HLA-I only
41.77%
n=6367
HLA-I ≥ HLA-II
24.23%
n=3694
HLA-I < HLA-II
26.34%
n=4015
HLA-II only
7.66%
n=1168
Position-wise coverage [%]
100.0
50.0
0.0
50.0
100.0
Source Protein

4.1.4 RESULTS

THE HLA LIGAND ATLAS: CONTENT AND SCOPE OF THE DATA RESOURCE

We describe the HLA Ligand Atlas, a dataset of matched HLA-I and -II ligandomes of benign tissues. HLA-I and -II ligands were isolated via immunoaffinity purification and identified by LC-MS/MS. HLA-binding prediction algorithms and an assessment of peptide length distributions were used to identify high-quality samples and only these were integrated into the dataset (Figure S1 describes the QC steps employed). Our online resource https://hla-ligand-atlas.org provides access to the dataset comprising HLA-I and –II ligands (1% local peptide-level FDR), their source proteins, tissue and subject of origin, as well as all corresponding HLA allotypes classified as strong or weak binders through several user friendly views (Figure 1A, Figure S1). We have acquired HLA ligandome data from 29 distinct tissues obtained from 21 individuals, surmounting to 1,262 LC-MS/MS runs from 225 mostly paired HLA-I (198) and -II (217) samples (Figure 1C, Figure S1, Table S1). The majority of samples was obtained from 14 subjects after autopsy, while 7 additional subjects contributed 5 thymus and 2 ovary samples after surgery. We performed a time series experiment on three benign samples, two ovaries and one liver (Figure S2) and observed no qualitative or quantitative degradation of the immunopeptidome for up to 72 h after tissue removal, supporting the feasibility of employing autopsy tissue as input material for immunopeptidomics assays (Figure S2). Overall, we identified 89,853 HLA-I and 140,861 HLA-II peptides with a local peptide-level FDR of 1% and estimated global peptide-level FDRs of 4.5% and 3.9% for HLA-I and -II peptides, respectively. Identified peptides could be attributed to 51 HLA-I and 81 HLA-II allotypes. Ultimately, this dataset increases the total number of registered HLA ligands from 413,205 to 445,535 for HLA-I and from 77,769 to 195,507 for HLA-II, as currently encompassed in SysteMHC[185] and IEDB[186] (Figure 1B). Moreover, we sought to approximate the worldwide HLA allele frequency of subjects included in the HLA Ligand Atlas. For this purpose, we computed population coverages using the IEDB Analysis Resources (http://tools.iedb.org/population/42) (Table S2). When considering at least one HLA allele match per individual, we observe an allele frequency of 95.1%, 73.6%, 93.0%, for HLA-A (n=16), -B (n=21), and -C (n=14) alleles, respectively. Within the same constraints we observe allele frequencies of 78.8%, 99.5%, 98.2%, 92.3% for HLA-DPB1 (n=9), -DQA1 (n=11), -DQB1 (n=12), and DRB1 (n=19) alleles, respectively (Table S2).

Figure 1: The HLA Ligand Atlas: content and scope of the data resource.
(A) The high-throughput experimental and computational workflow steps used to analyze thousands of HLA-I and -II peptides isolated from benign tissues. The resulting HLA-I and -II immunopeptidomes are comprised in the searchable web resource: hla-ligand-atlas.org. See Figure S1 for details of the quality control workflow. See Figure S2 for proof of principle using autopsy tissues.
(B) HLA-I and -II peptides expand the know immunopeptidome as curated in the public repositories SysteMHC and IEDB.
(C) Sample matrix: HLA-I (blue triangles) and –II samples (orange triangles) included in the HLA Ligand Atlas cover 29 different tissues obtained from 21 human subjects. See Table S1 for patient characteristics.
(D) Position-wise coverage (%) of identified source proteins by HLA ligands binned into four groups: (1) exclusively covered by HLA-I peptides, (2) exclusively covered by HLA-II peptides and (3-4) covered by both and separated into higher position-wise coverage by either HLA-I or -II peptides.


A PEPTIDE LENGTH DISTRIBUTION
Fraction [%]
Peptide length
HLA-I binder
HLA-II binder
HLA-I non-binder
HLA-II non-binder
B NUMBER OF HLA-I PEPTIDES PREDICTED AS BINDERS PER ALLELE
Number of Peptides
strong binder
weak binder
HLA-A
HLA-B
HLA-C
C NUMBER OF HLA-II PEPTIDES PREDICTED AS BINDERS PER ALLELE
HLA-DPA1-DPB1
HLA-DQA1-DQB1
HLA-DRB1
HLA-DRB3
HLA-DRB4
HLA-DRB5
D HIERARCHICAL CLUSTERING OF SAMPLES BASED ON SOURCE PROTEIN JACCARD SIMILARITY
HLA-I
Tissue
Subject
HLA-II
Tissue
Subject
Subjects
Tissues
E DISTRIBUTION OF PAIRWISE JACCARD SIMILARITIES
Jaccard Similarity
same subject different tissue
different subject same tissue
different subject different tissue
F GENE ONTOLOGY ENRICHMENT PER HLA CLASS
HLA-I
GO Term
intracellular organelle lumen
organelle lumen
membrane-bounded organelle
organelle
nucleoplasm
intracellular organelle
nuclear lumen
intracellular membrane-bounded organelle
intracellular
nucleus
HLA-II
mitochondrial ribosome
organellar ribosome
cell periphery
plasma membrane
membrane
integral component of plasma membrane
intrinsic component of plasma membrane
extracellular region
integral component of membrane
intrinsic component of membrane
log10(p-value)

SOURCE PROTEINS AND HLA ALLOTYPE COVERAGE CHARACTERISTICS OF HLA LIGANDS

The HLA Ligands in the dataset were identified based on 15,244 of the 20,365 proteins in Swiss-Prot, hereinafter referred to as source proteins. About half of these source proteins yield both HLA-I and -II ligand identifications, 40% yield only HLA-I ligands and 8% only HLA-II ligands (Figure 1D). We performed a gene ontology enrichment analysis of HLA-I and -II exclusive source proteins, which corroborates the expected cellular compartments associated with the class-specific antigen presentation pathways, with HLA-I presenting primarily intracellular- and HLA-II extracellular proteins (Figure 2F).

When looking at single amino acid residues across all source proteins (position-wise), 10% of the single residue positions are covered by HLA ligands, a parameter that ranges from 0.02% to 1.9% for individual HLA allotypes (Figure S3). The mode of the overall peptide length distribution depicts the highest abundance of 9mers (60%) for HLA-I and of 15mers (18%) for HLA-II ligands (Figure 2A). While 81% of the HLA-I ligands are predicted to bind a subject's HLA allotype, this holds true for only 53% of the HLA-II ligands. A major shortcoming of HLA-II binding prediction models appears to be a negative bias towards the tails of the observed peptide length distribution, in particular towards short peptides (Figure 2A). The number of identified peptides that are predicted to bind against specific allotypes varies strongly between allotypes, with HLA-A*02:01, -B*15:01, -B*35:01, -C*04:01 and most HLA-DRB1 allotypes being among the highly represented ones (Figure 2B, C).

THE INTER-INDIVIDUAL HETEROGENEITY OUTWEIGHS SIMILARITIES BETWEEN TISSUE TYPES

An unaddressed question, relevant for the discovery and administration of shared TAAs, is if the similarity between tissue types outweighs that of individuals. We interrogated the HLA Ligand Atlas and assessed the similarity of the immunopeptidome on both source protein (Figure 2D, E) and HLA-ligand level (Figure S4C, D) between samples, as defined by subject-tissue combinations. For this purpose, we computed pairwise similarities between all samples as measured by the Jaccard similarity index and clustered the samples based on their similarity. We observe that the sample similarity, even on source protein-level, is dominated by the underlying HLA alleles governing peptide presentation in each subject, resulting in clusters largely reflecting the subjects rather than the tissues. Contrary to our expectations, the five thymus specimens show the same pattern of subject individuality without an increase in source protein overlap with

Figure 2: Source proteins and HLA allotype coverage characteristics of HLA ligands.
(A) Length distribution of identified HLA-I and -II peptides from all samples was analyzed. HLA-II peptide lengths are mirrored on the negative side of the x-axis.
(B, C) Global overview of HLA-I predicted binders distributed across HLA molecules. HLA binding prediction was performed with NetMHCpan 4.0 (% binding rank <2) and SYFPEITHI (Score >50%), while multiple HLA allotypes per peptide were allowed as long as their scores met the aforementioned thresholds. HLA binding prediction for HLA-II ligands was performed with NetMHCIIpan 4.0 and MixMHCPred (% binding rank 0.2 for both) See Figure S3
(D) Pairwise hierarchical clustering of samples based on the Jaccard similarity between HLA-I (blue) and HLA-II (orange) source proteins. The dendrogram illustrates the nearest neighbor based on the similarity between tissues and subjects. See Figure S4 C.
(E) Violin plots illustrate the distribution of the Jaccard similarity index for each pairwise comparison between the same subject - different tissues; different subjects - the same tissue, and different subject - different tissues.
(F) Gene ontology (GO) term enrichment of cellular components was performed for HLA-I and -II source proteins. Top10 enriched genes with respect to their log10 p-value (Fisher's exact test) differentiate between intracellular and extracellular antigen processing pathways.

other tissue types. The high subject individuality as indicated by the clustering of subjects rather than tissues holds true irrespective of the data level on which the analysis is based on.

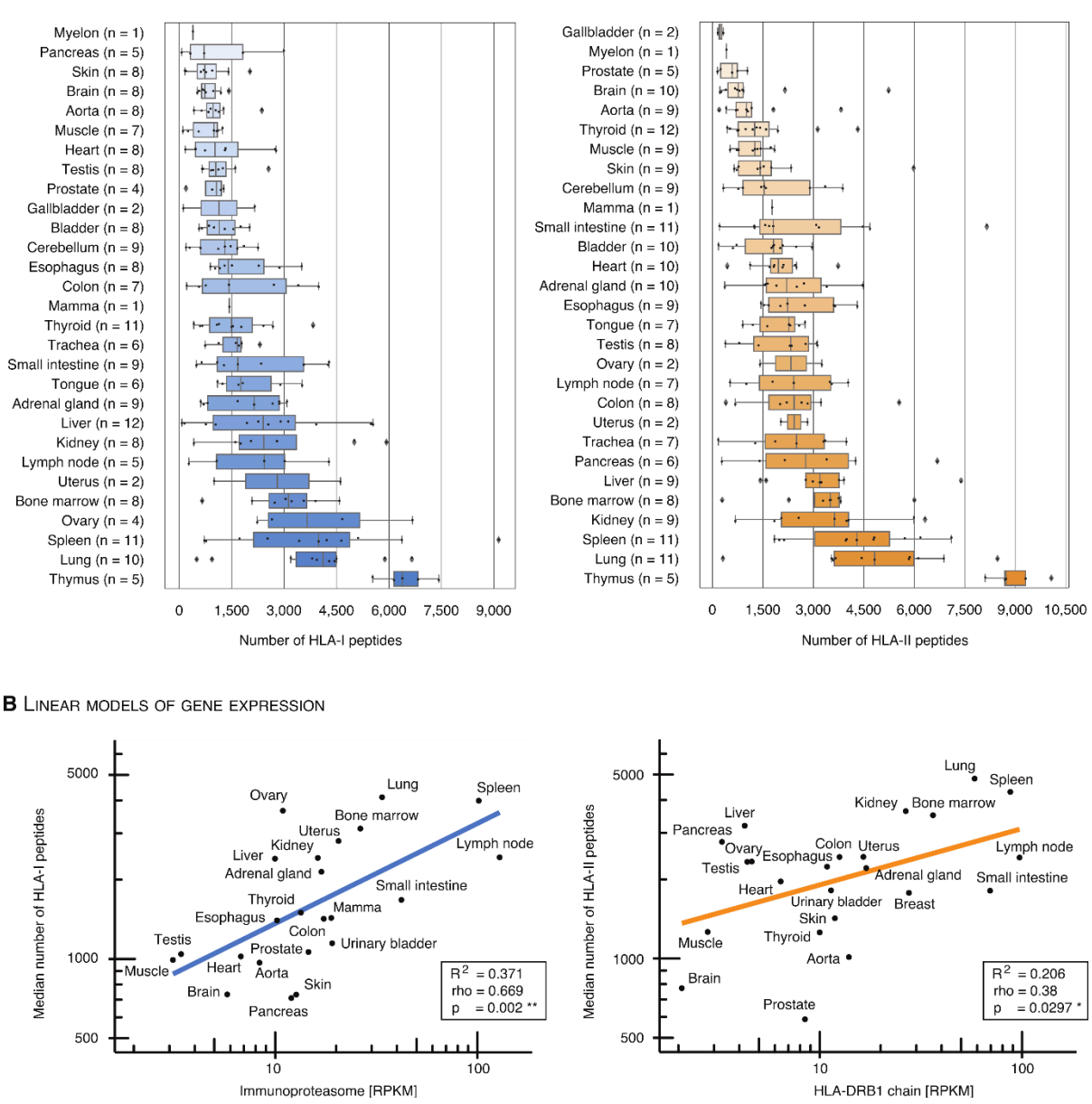


Figure 3: Tissues exhibit a gradual separation based on the immunopeptidome yield.
(A) The number of identified HLA-I and -II peptides per sample (subject and tissue combinations) was sorted and plotted by median immunopeptidome yield per tissue. Boxes span the inner two quantiles of the distribution and whiskers extend by the same length outside the box. Remaining outlier samples are indicated as black diamonds. The number of subjects contributing to each tissue is illustrated on the y-axis in parenthesis.
(B) A linear model was used to correlate the log transformed HLA-I and -II median peptide yields with log transformed median gene expression counts (RPKM) of the immunoproteasome and HLA-DRB1 per tissue. Corresponding R2, p-value (F-statistic) and spearman rho are indicated in the bottom right box. See Figure S5

THE IMMUNOPEPTIDOME YIELD VARIES CONSISTENTLY ACROSS TISSUES

We further investigated the immunopeptidome diversity and variance across all samples for both HLA-I and -II alleles. Overall, we observe a strong variance in the immunopeptidome yield, defined as the number of identified peptides per sample, across all tissues (Figure 3A) and subjects (Figure S4A, B). Despite the inter-individual (i.e., inter-allotype) variance, we can consistently differentiate between high-yielding and low-yielding tissues with respect to both HLA-I and -II peptides (Figure 3, Figure S4A, B). The separation of tissues based on the immunopeptidome yield is not abrupt, but gradual. Low-yielding tissues include skin, aorta, brain, and the gallbladder with a low number of both HLA-I and -II presented peptides across all subjects. On the other hand, high-yielding tissues include thymus, lung, spleen, bone marrow, and kidney (Figure 3A). These tissues have well-characterized immune-related functions or are central components of the lymphatic system.

We employed a linear model to systematically evaluate the correlation between the median HLA-I/-II immunopeptidome yield with RNA expression values (RPKM) of immune-related genes identified by targeted RNA sequencing from an external dataset[178] (Figure 3B and Figure S5). We observe a significant correlation between expression values of immune-related genes and HLA-I and -II immunopeptidome yields (Figure S 5A-D). Among these, genes of the immunoproteasome correlate well with the number of HLA-I ligand identifications per tissue (R^2=0.371, rho=0.669, p=0.002, Figure 3B). Independent studies mapping the healthy human proteome confirm expression of the immunoproteasome in a wide range of tissues, including tissues for which no primary immunological function would be expected[151,152].

HLA-II peptide yields correlate well with the expression of HLA-DRB1 genes (R2=0.206, rho=0.38, p=0.0297, Figure 3B). HLA-DR is well characterized due to the invariant α chain, and thus reduced complexity in the peptide binding groove. Through the high specificity of the L243 antibody for HLA-DR, and the presumably varying specificity of the second antibody Tü39 for different HLA-II allotypes, we cannot exclude a skewed identification in favor of HLA-DRB allotypes. However, higher expression values for HLA-DRB1 compared to other HLA-II allotypes have been described for example in earlier studies on gastric epithelium[187].

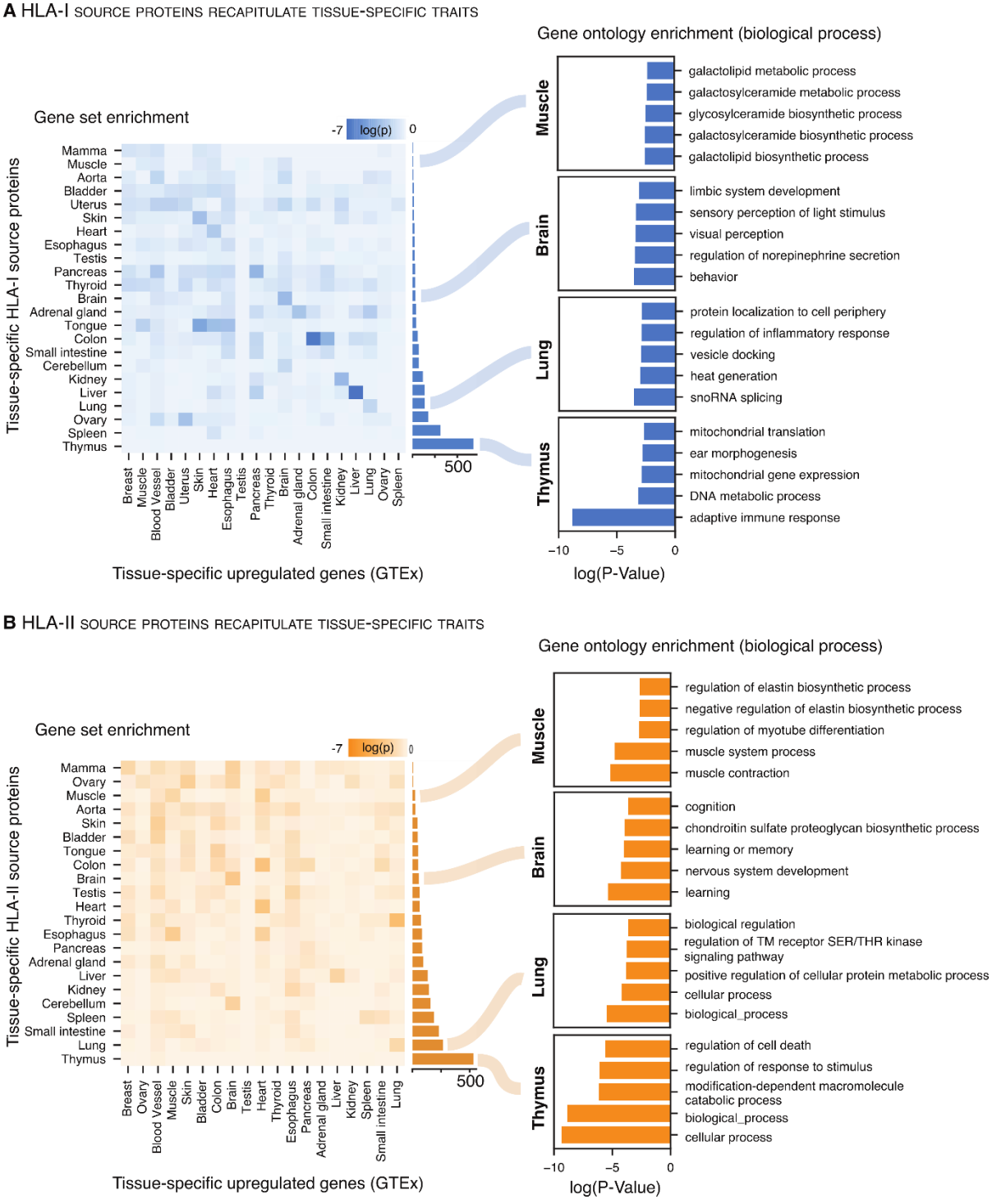


Figure 4: Small subsets of source proteins are tissue exclusive. See Table S5.
(A, B) Gene set enrichment (left) was tested for each tissue by correlating unique HLA-I and –II source proteins per tissue with upregulated genes as annotated in GTEx. Heatmaps depict log10 p-values (Fisher's exact test) for each pairwise comparison. The number of tissue-specific HLA-I and –II source proteins is depicted by the bar plot for each tissue at the right-hand side of the heatmaps. In addition, GO term enrichment (right) of biological processes was performed using the panther DB webservice for selected tissues with the same set of HLA-I and -II tissue-specific source proteins. Top 5 enriched terms with respect to their log10 p-value (Fisher's exact test) were selected.

SMALL SUBSETS OF SOURCE PROTEINS ARE TISSUE-EXCLUSIVE

Previous studies characterizing the human transcriptome and proteome across tissues have shown varying degrees of tissue-specificity for transcripts and proteins[4,115]. In this context, we analyzed source proteins of the benign immunopeptidome as a whole and grouped all samples by tissue of origin. We observe a particularly small number of HLA-I (ranging from 5 in mamma to 680 in thymus), and HLA-II (ranging from 8 in ovary to 567 in thymus) source proteins identified exclusively in one tissue (Figure 4A, B, Table S5). Concordantly, only small numbers of tissue-exclusive protein identifications have been observed in human tissue-wide proteomics studies[188]. Only recently, the systematic, quantitative analysis of the human proteome and transcriptome in multiple tissues has revealed that differences between tissues are rather quantitative than defined by the presence or absence of certain proteins[4,115].

Next, we sought to determine whether tissue-specific biology is conserved between the transcriptome and immunopeptidome. For this purpose, we compared tissue-enriched gene sets from the GTEx repository with tissue-exclusive HLA-I and -II source proteins (Figure 4A, B, left). We observe that tissue-specific biology is represented by HLA-I and -II source proteins through an enrichment with upregulated transcripts in the respective tissue. Gene set enrichment analysis further reflects functional proximity between tissues such as tongue, heart and muscle or brain and cerebellum.

We additionally observed that tissue-specific traits are recapitulated by gene ontology (GO) term enrichment of biological processes (Figure 4A and B, right). Enriched GO terms reveal tissue-specific biological functions such as ‘adaptive immune response’ in the thymus or ‘behavior’ in the brain. However, clear associations between enriched gene sets and HLA-I and -II source proteins are less evident in tissues such as spleen or testis, despite the disparity of tissue-exclusive HLA-I source protein identifications, accounting for only 23 in testis while spleen yields 309.

Overall, tissue-specific traits are more evident for HLA-I than for HLA-II source proteins, as supported by a higher significance, when assessing the correlation between tissue-exclusive source proteins with GTEx-enriched transcripts and function-specific GO terms. HLA-II source proteins are represented by more general GO terms, which still reflect distinct biological processes characteristic for the respective tissue.

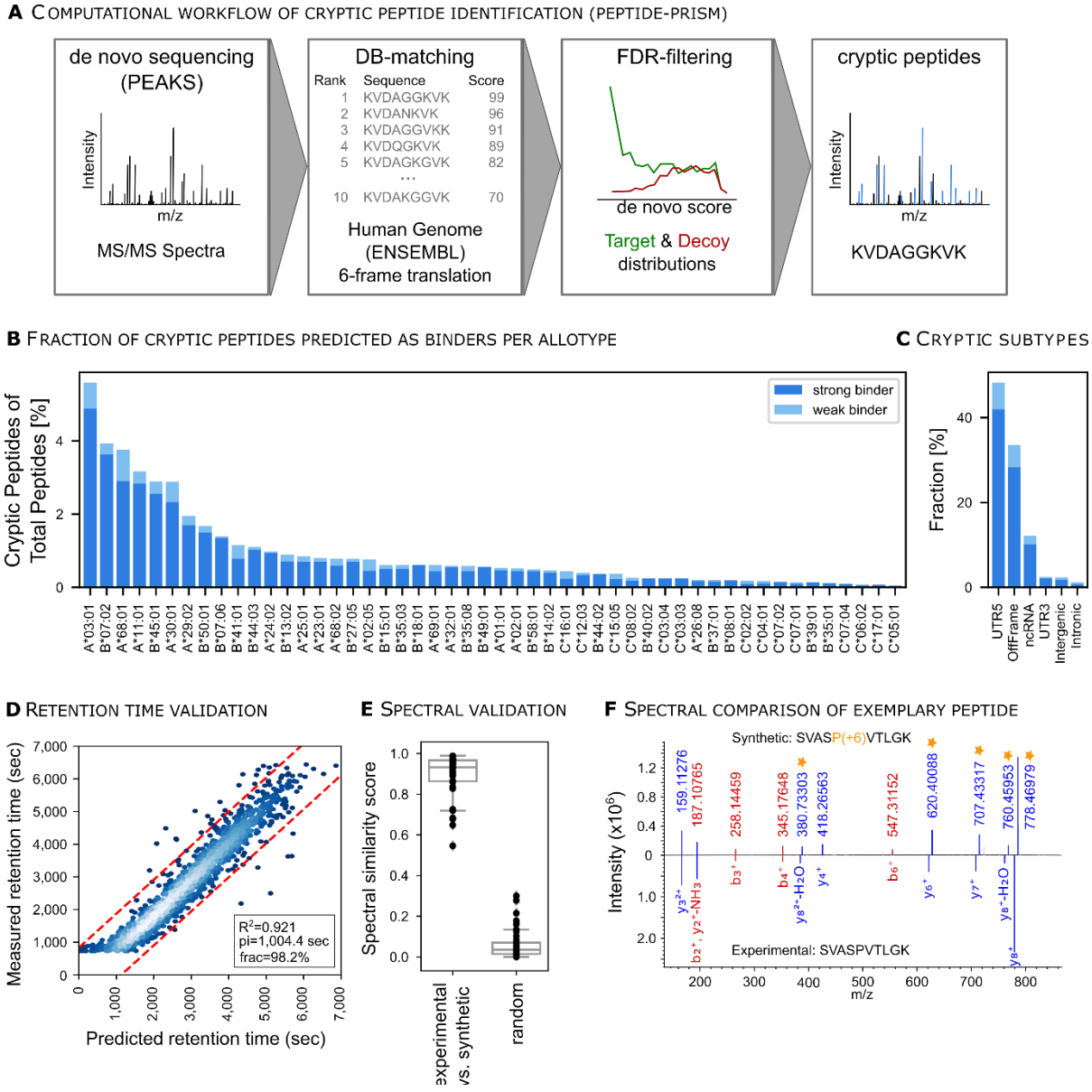


Figure 5: Cryptic peptides are part of the benign immunopeptidomes. See Table S3, Figure S6.

(A) Spectra were searched with Peptide-PRISM to identify peptides of cryptic origin. Briefly, de novo sequencing was performed, and top 10 sequences per spectra were queried against a database consisting of the 3-frame translated transcriptome (Ensemble 90). Target-Decoy search was performed per database stratum, separately for canonical and cryptic peptides.

(B) The HLA-allotype distribution of cryptic peptides was plotted in relation to cryptic and canonical peptides predicted to bind to the respective HLA allotype across all subjects and tissues.

(C) Distribution of identified cryptic peptides categorized into multiple non-coding genomic regions.

(D) Linear model correlating measured retention times (RT) of cryptic peptides with their predicted RTs trained on canonical peptide RTs. Corresponding R2, pi (width of the prediction interval – red dashed lines), and frac (the number of peptides falling into the prediction interval) are indicated in the bottom right.

(E) 36 cryptic peptides were selected for spectral validation with synthetic peptides. The similarity between the synthetic and experimental spectrum was computed by correlation scores. F) Exemplary spectral comparison of the cryptic peptide SVASPVTLGK and its synthesized heavy isotope-labeled counterpart (P+6). Matching b (red) and y ions (blue) are indicated as well as the isotope mass shifted ions (orange stars) of the synthesized peptide.

CRYPTIC PEPTIDES ARE PART OF THE BENIGN IMMUNOPEPTIDOME

Recently, cryptic HLA peptides come into focus as a new potential source of tumor-associated antigens (TAAs). Cryptic peptides originate from non-coding regions, i.e. 5'- and 3'-UTR, non-coding RNAs (ncRNA), intronic and intergenic regions, or from shifted reading frames in annotated protein coding regions (off-frame). Ribosome profiling and immunopeptidomics studies confirm their translation and presentation on HLA-I molecules[99–101]. So far, cryptic peptides have predominantly been characterized in tumors, while their presentation in benign tissues remains poorly charted. We analyzed the HLA-I-restricted LC-MS/MS data of the HLA Ligand Atlas with Peptide-PRISM[24] (Figure 5A) and identified 1,407 cryptic peptides, including the peptide SVASPVTLGK that was classified as a TAA in lung cancer tissue in a previously published study (Figure 5, Table S3)[101]. This peptide was identified in the HLA Ligand Atlas in five different subjects in lung and liver tissues. We find that 47% of cryptic peptides were identified in more than one subject (Table S3). Both cryptic and conventional peptides share similar physicochemical properties. Their predicted chromatographic retention time correlates with their experimentally observed retention time equally well as for conventional peptides (Figure 5D)[99,100,103,104]. The identified cryptic HLA-I ligands can be classified into following genomic categories with decreasing frequency: 5'-UTR (51%), followed by Off-Frame (33%), ncRNAs (13%), 3'-UTR (2%), intergenic (1%), and intronic regions (0.5%) (Figure 5C). The predominance of cryptic peptides from the 5'-UTR is in accordance with previous studies[24,99]. Overall, HLA allotypes show different presentation propensities of cryptic peptides, when related to cryptic and canonical peptides, with HLA-A*03:01 covering the largest fraction of all identified cryptic peptides, followed by -B*07:02 and -A*68:01, as previously observed (Figure 5B)[24].

We selected 36 top-ranking (1% FDR) cryptic peptides, shared among subjects for spectral validation by experimental comparison with the corresponding heavy isotopically labeled synthetic peptide (Table S3). We computed a similarity score between the spectra obtained from the experimental vs. synthetic peptides (Figure 5E, Table S3). A size-matched set of randomly selected comparisons was employed to create a reference negative distribution of the spectral similarity score. We were able to confirm the correct identification of selected cryptic HLA-I ligands, not only based on the computed similarity score, but also through individual inspection (Figure S6). Therefore, we can show that cryptic peptides are not per-se tumor-specific, albeit their frequency might be reduced in benign tissues[24].

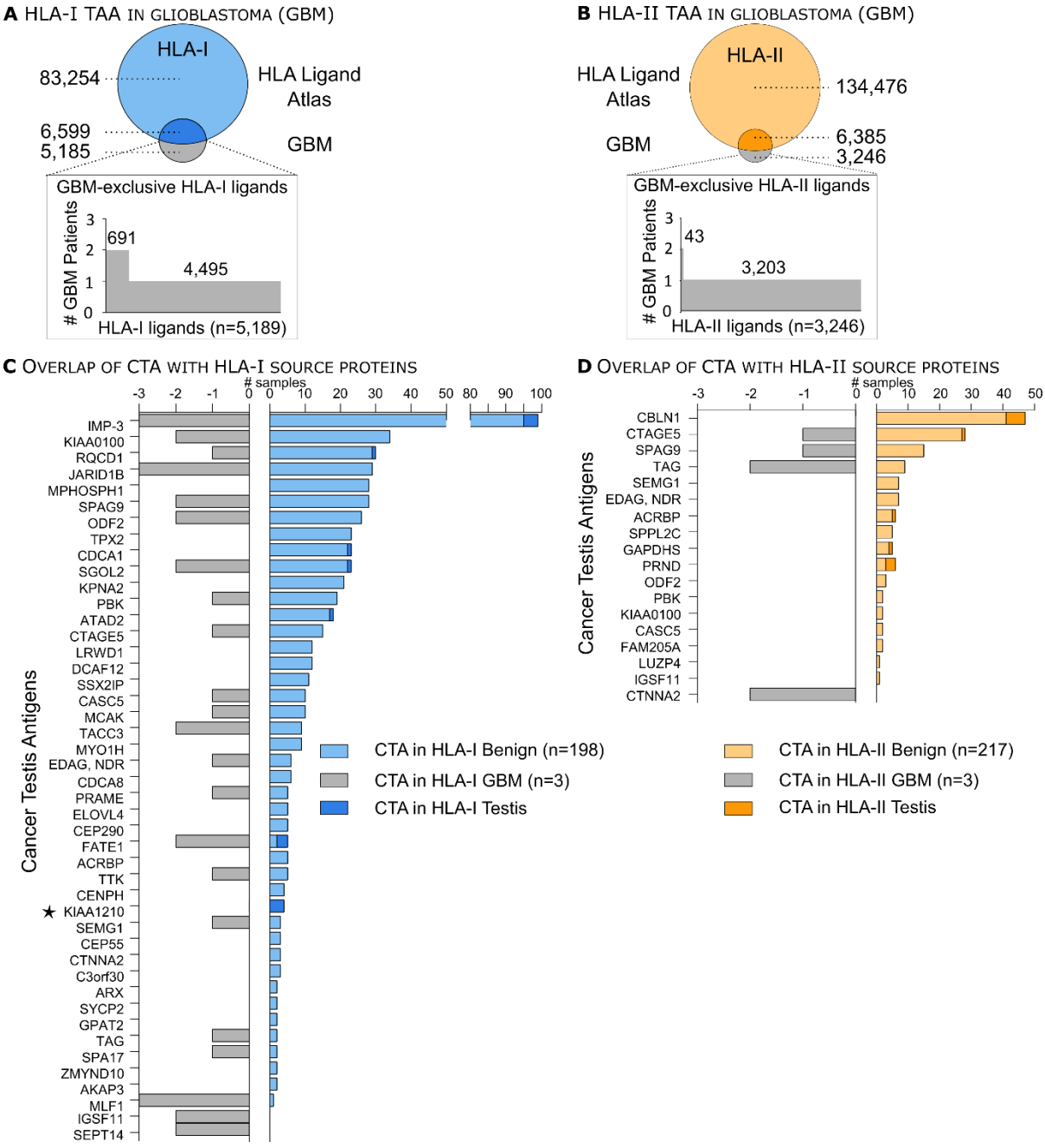


Figure 6: HLA Ligand Atlas data enables prioritization of tumor-associated antigens (TAAs).
(A, B) The size-proportional Venn diagram illustrates the overlap between the pooled glioblastoma (GBM) and benign HLA-I and -II immunopeptidomes, respectively. The waterfall plots show the number of glioblastoma-associated HLA-I ligands and their frequency among the three glioblastoma (GBM) patients analyzed. See Table S4.
(C, D) Published CTAs are presented as HLA-I or -II ligands on benign tissues, including testis but also in glioblastoma tumors. The number of identified samples either from the HLA Ligand Atlas or the glioblastoma dataset is depicted on the x-axis, provided that each CTA has been identified with at least two different HLA ligands. The CTA KIA1210 was identified exclusively on HLA-I source proteins in testis and is marked with an asterisk. See Table S4.

HLA LIGAND ATLAS DATA ENABLES PRIORITIZATION OF TUMOR-ASSOCIATED ANTIGENS (TAAS)

A general lack of multi-tissue immunopeptidomics reference libraries from benign tissues has been mentioned in previous studies aiming to identify TAAs[100,131]. Here, we propose the implementation of the HLA Ligand Atlas as a reference library of benign multi-tissue immunopeptidomes for comparative profiling with tumor immunopeptidomes for the discovery of actionable TAAs. As a case study, we selected three glioblastoma tumor samples from different individuals and analyzed their immunopeptidomes. We comparatively profiled the HLA-I and -II ligands of the glioblastoma samples against the benign dataset encompassed in the HLA Ligand Atlas (Figure 6A and B). The majority of HLA ligands is shared between both tumor and benign tissues, with 5,185 HLA-I TAAs and 3,246 HLA-II TAAs being unique to glioblastoma (Table S4). When assessing their presentation frequency, 691 HLA-I TAAs are found on two glioblastoma samples, while 4,495 are patient-individual. In the case of HLA-II TAAs, 43 are shared between two glioblastoma patients, and 3,203 are patient-individual. No identified HLA-I or –II ligands were common to all three glioblastoma patients.

Moreover, we investigated the presentation of cancer testis antigens (CTAs) by HLA-I and -II molecules on benign tissues. CTAs are immunogenic proteins preferentially expressed in normal gametogenic tissues and different types of tumors[189,190]. We compiled a list of 422 published CTAs from the curated CT database[189] and a recent publication aiming to identify CTAs from transcriptomics datasets[190]. Of 422 published CTAs, 40 CTAs were presented on either HLA-I or -II molecules and 10 CTAs on both HLA-I and -II molecules in the HLA Ligand Atlas, provided that respective source proteins were identified with at least two HLA ligands (Figure 6B, C, Table S3). CTAs, such as IMP-3, KIA0100, and CBLN1 were presented in numerous benign samples with HLA-I and II ligands (Figure 6C and D, Table S4). Furthermore, the CTA KIA1210 was only identified in the benign dataset on testis in accordance to its CTA status. Similarly, we queried all glioblastoma source proteins against the selected 422 CTAs and found three CTAs (two HLA-I and one HLA-II) exclusively presented in glioblastoma and not in our benign dataset, indicating promising targets against this tumor entity.

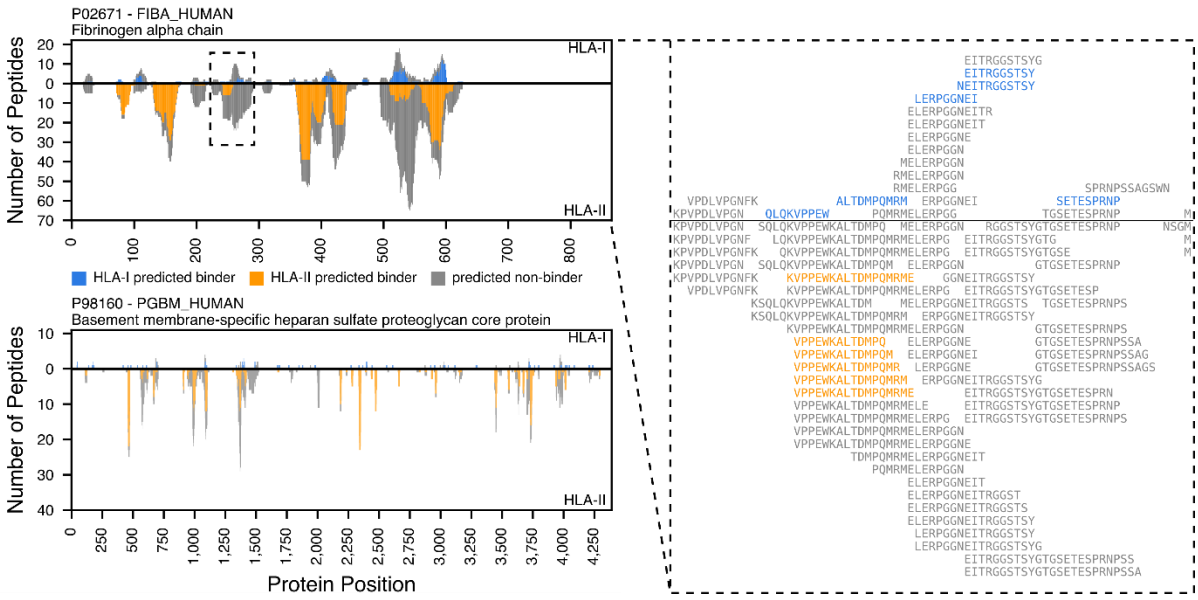


Figure 7: HLA ligands form hotspots in source proteins.
The position-wise HLA ligand coverage profiles as available in the HLA Ligand Atlas web interface for two exemplary proteins (left), the fibrinogen alpha chain (Uniprot ID P02671, length 866 aa, top) and the basement membrane-specific heparan sulfate proteoglycan core protein (Uniprot ID P98160, length 4,391 aa, bottom) are shown, illustrating the spatial clustering of HLA ligands into hotspots. For P02671 a close-up of such a cluster is shown in form of a multiple sequence alignment of the identified peptides (right).

HLA LIGANDS FORM HOTSPOTS IN SOURCE PROTEINS

When looking at the position-wise coverage profiles of individual source proteins across all HLA allotypes, we observe that HLA ligands seem to emerge from spatially clustered hotspot regions while other areas of the source protein do not contribute any HLA ligands at all (Figure 7, left). It has been shown previously that this clustering effect cannot be explained by the occurrence of HLA binding motifs as incorporated in epitope prediction tools[191]. The hotspot locations often coincide between HLA-I and -II ligands, however, we did not perform a large-scale statistical analysis to validate this class linkage effect. In the case of HLA-II, the clustering effect has to be distinguished from the co-occurrence of HLA-II ligand length variants, which leads to a large number of distinct peptides covering the same source protein position due to the nature of HLA-II antigen processing and binding[192]. Many of the observed clusters span ranges of distinct, non-overlapping HLA-II ligands (Figure 7, right). Position-wise coverage plots for all source proteins are available online at hla-ligand-atlas.org.

THE HLA LIGAND ATLAS WEB INTERFACE

The HLA Ligand Atlas web interface was designed to allow users to conveniently access the data we collected. Users can formulate queries in the form of filters based on peptide sequences, peptide sequence patterns, HLA allotypes, tissues and proteins of origin, or combinations thereof. Additionally, users can submit files with peptides or uniport IDs, either as plain lists or as a FASTA files. The peptide list is then queried against the database and the resulting hits can again be freely combined with the aforementioned filters. Query results are shown as a list of peptides with plots of the corresponding HLA allotype and tissue distributions. Additionally, detailed

views for single peptides and for coverage of proteins are available. Apart from the query interface, the web front-end also displays various aggregate views of the data stored in the database.

4.1.5 Discussion

In this study, we create a novel data resource termed the HLA Ligand Atlas which is publicly available and easily searchable at hla-ligand-atlas.org (Figure 1C). It provides for the first time a comprehensive collection of benign human HLA-I and -II immunopeptidomes. The large number of different HLA allotypes will help to considerably improve HLA-binding prediction algorithms, particularly for infrequent HLA alleles. HLA-II immunopeptidomes paired with high-resolution HLA-II typing are still scarcely available and therefore represent a valuable resource for improving HLA-II prediction models.

We find that HLA allotypes display varying presentation propensities towards certain peptide populations, with HLA-B*15:01 and HLA-DRB1*01:01 presenting the highest number of canonical self-peptides, and HLA-A*03:01 and B*07:02 presenting the highest proportion of cryptic peptides in our subject cohort and a previously published study[24]. The increased number of peptides presented on a subset of HLA alleles can be attributed to their frequency among the analyzed individuals or to their potentially high copy number on cells. Further technical biases can influence the immunopeptidome yield, such as antibody preferences towards certain HLA allotypes, ionization and fragmentation characteristics of eluted HLA ligands, but also binding prediction algorithms that perform better for frequent, well studied HLA allotypes. However, HLA allotypes have evolved to present different peptide subsets to T cells[193], examples ranging from HLA-B*40 being able to stabilize the negative charge of phosphorylated peptides[194], and HLA-B*57 conferring a survival advantage in HIV infections[157,195]. Moreover, we observed multiple HLA allele matches per peptide, which is indicative of binding similarities between HLA allotypes, or promiscuous HLA alleles that allow binding of a large repertoire of different peptides[193].

One fundamental and so far unanswered question concerns the similarity of immunopeptidomes across individuals. Our evaluation of the Jaccard similarity index across samples in the HLA Ligand Atlas provides evidence that differences between individuals exceed differences between tissue types in the same individual for both the immunopeptidome and their source proteins. On a proteome level, however, samples were previously separated by tissue type, rather than individuals[4]. Nonetheless, HLA-allotype-dependent selection and editing throughout the antigen presentation pathway shape the immunopeptidome, complicating its prediction from genomic, transcriptomic, and proteomic data sources. While we analyzed 21 human subjects in this study, a larger number would be required to answer this question unequivocally.

The high degree of individuality between immunopeptidomes, even when subjects share a subset of HLA allotypes, has major repercussions for clinical applications in emerging fields such as immuno-oncology. Our findings indicate that the immunopeptidome adds an additional layer of complexity to the well-described genomic and transcriptomic tumor-heterogeneity. Successful

induction of T cell responses after peptide vaccination with neoantigens[67,129] indicate that precision medicine will evolve to an increasingly individualized field, where treatment options will be tailored to the immunopeptidomic landscape of the tumor. Mapping the tumor HLA ligandome of an individual patient therefore needs to be paralleled by a broad and in-depth knowledge of its benign counterpart – the HLA Ligand Atlas is a first step in this direction. When dissecting the HLA Ligand Atlas tissue-wise, we observe a paired immunopeptidome yield between HLA-I and -II ligands that could be indicative of an increased infiltration of immune cells in high-yielding tissues. Alternatively, expression of HLA-II molecules on cells other than APCs could explain this observation. By analyzing bulk tissue, a definite statement whether peptide presentation occurred on tissue or tissue-infiltrating immune cells cannot be made. The immunopeptidome yields per tissue correlate positively with RNA expression profiles of genes related to antigen processing and presentation. Yet, the identified source proteins appear to be barely specific for the tissue of origin. The weak correlation between immunopeptidome yield and RNA expression values has been observed previously[97,117]. Although abundant HLA ligands stem from highly expressed transcripts, most HLA ligands span a wider dynamic range of gene expression[115]. Furthermore, it was recently shown that the immunopeptidome is better captured by the translatome as identified by ribosome profiling than by the transcriptome[99,100].

Low-yielding samples barely display any tissue-specific proteins. However, tissue-specific source proteins often reflect tissue-specific traits when correlated to enriched gene sets of GTEx transcriptomes from the respective tissues. Therefore, tissue-specific function is represented in the immunopeptidome, but differences between tissues cannot be imputed from the immunopeptidome alone. Studies mapping the whole proteome in multiple human tissues report few proteins with tissue-specific expression[115,152] and suggest that differences between tissues might be quantitative, and less dominated by the presence or absence of protein species[4,115].

Recent studies have focused on HLA-presented peptides derived from non-coding regions. Ribosome profiling, RNA sequencing, and immunopeptidomics studies have confirmed that cryptic HLA-I peptides expand the known HLA-I immunopeptidome by up to 15%[24], up to 3.3%[99], and about 10%[101]. These studies have mainly focused on tumors and tumor cell lines, PBMCs and mTEC cell lines, in most cases treated and expanded in vitro. We employed Peptide-PRISM and identified 1,407 cryptic HLA-I ligands from benign, primary, human samples. Corroborating other studies, we find that a large proportion (41%) are also shared between multiple subjects[99,100]. An essential application of the HLA Ligand Atlas is the selection of candidates for immunotherapy approaches. We propose a workflow to prioritize the large candidate pool of non-mutated tumor-associated targets by comparatively profiling immunopeptidomes of primary tumors and benign tissues, as comprised in the HLA Ligand Atlas. This approach would complement current strategies based on transcriptomes of benign tissues as comprised in GTEx. The HLA Ligand Atlas represents a first draft of a tissue-wide immunopeptidomics map covering both HLA-I and -II canonical peptides, but also HLA-I non-canonical peptides, that can be employed as an orthogonal level of quality control when defining TAAs. Furthermore, anecdotal observations of position-wise coverage of source proteins confirm a previously stated hypothesis, that immunopeptidomes cluster into hotspots of antigen presentation[78,191]. We envision that different

types of TAAs such as mutated, non-mutated, post-translationally modified, of cryptic or proteasomally spliced origin might cluster as well within these hotspots. Future studies that aim to enhance our understanding of such mechanistic patterns of peptide presentation will benefit greatly from the data the HLA Ligand Atlas comprises.

A series of systematic technical limitations in LC-MS/MS-based studies influences the identification depth in each sample. Such aspects include the still limited sensitivity and dynamic range of detection, the insufficient coverage of amino acids in peptide mass spectra, but also shortcomings in peptide identification algorithms. Advances in LC-MS/MS technology, data acquisition methods and computational tools are constantly improving the depth of investigation in immunopeptidomics experiments. Therefore, we encourage the reanalysis of the raw LC-MS/MS dataset with novel hypotheses and upcoming computational methods that will lead to additional insight. Overall, we anticipate that the number of charted human immunopeptidomes will increase, similarly as the human genome and transcriptome were mapped across multiple individuals. By generating larger datasets from many human individuals, population-wide conclusions can be drawn, and immunopeptidome-wide studies will provide insight into disease-associated HLA alleles and peptides[157]. The HLA Ligand Atlas strives to advance our understanding of fundamental aspects of immunology relating to autoimmunity, infection, transplantation, cancer immunotherapy and might provide a foundation for vaccine design. We hope that together with the scientific community we can expand the benign immunopeptidome to encompass more human subjects, tissues, and HLA alleles.

PART II:

BENCHMARKING MASS SPECTROMETERS

Ana Marcu[1,2], Lena Katharina Freudenmann[1,2,3], Jens Bauer[1,2], Stefan Stevanović[1,2,3], Hans-Georg Rammensee[1,2,3]

1 Department of Immunology, Interfaculty Institute for Cell Biology, University of Tübingen, Tübingen, Baden-Württemberg, 72076, Germany.

2 Cluster of Excellence iFIT (EXC 2180) "Image-Guided and Functionally Instructed Tumor Therapies", University of Tübingen, Tübingen, Baden-Württemberg, 72076, Germany.

3 DKFZ Partner Site Tübingen, German Cancer Consortium (DKTK), Tübingen, Baden-Württemberg, 72076, Germany.

Not for publication

AM performed the research, contacted the vendors, organized the measurements, analyzed the resulting data, and generated all plots. The demobox samples were collectively generated, AM grew the JY cell line and prepared the cell pellets for HLA ligandome analysis. JB experimentally isolated the HLA ligands and thereby generated the JY19#1 SST stock samples. LKF selected, weighed, and diluted the heavy-labeled RT-peptides to a stock solution of 10 fmol/µl. The serial dilution and subsequent LC-MS/MS analyses were performed by AM. SS and HGR supervised the project and contributed resources.

4.2.1 ABSTRACT

Success of immunopeptidomics experiments is highly dependent on the LC-MS/MS instrumentation. Technical limitations define the meaningfulness of our experiments, manifesting as missing values, low technical reproducibility, a limited dynamic range that cannot capture both highly and low-abundant peptides, and many more. Therefore, appropriate instrumentation that can be optimized to address challenges specific to immunopeptidomics must be chosen. The suitability for immunopeptidomics experiments was assessed both theoretically and experimentally between five different LC-MS/MS systems: the Orbitrap Fusion Lumos, the Orbitrap Fusion and the q Exactive HF from Thermo Fisher Scientific, the Triple TOF6600 from AB Sciex, and the timsTOF Pro from Bruker Daltonics. The main distinctive feature between the instruments is the detector. The Thermo Fisher Scientific instruments have an Orbitrap detector, while both the Triple TOF6600 and the timsTOF Pro have a time of flight (TOF) detector. Overall, 114 LC-MS/MS measurements were acquired across all systems. The analyzed samples consisted of a serial dilution series spiked with heavy isotope-labeled synthetic peptides. All instruments were compared to the in house Orbitrap Fusion Lumos with regard to sensitivity, dynamic range, mass accuracy, lower limit of detection, purity of identified HLA ligands, and complementarity. The resulting ranking revealed that the Orbitrap Fusion Lumos and the Orbitrap Fusion showed the highest sensitivity, purity of identified HLA ligands, reliably detected a constant number of RT peptides, and showed a concentration-dependent linear decrease in the cumulative area of RT peptides. Second ranked were the timsTOF Pro and the q Exactive HF. Furthermore, the timsTOF Pro revealed a lower detection limit than the Orbitrap instruments and revealed a complementary set of peptides. The Triple TOF 6600 performed poorly. In conclusion, the Tribrid instruments had the highest performance, owing to both intrinsic instrument parameters but also due to extensive optimization to immunopeptidomics samples. Thus, it stands to reason, that the timsTOF Pro and the q Exactive HF could be optimized similarly.

4.2.2 INTRODUCTION

LIQUID CHROMATOGRAPHY COUPLED TO TANDEM MASS SPECTROMETRY (LC-MS/MS) INSTRUMENTATION

Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) is an analytical method that combines the separation of analytes based on high-performance liquid chromatography (HPLC) and the mass analysis capabilities of mass spectrometry (MS). A complex peptide mixture can be separated based on the physicochemical properties of individual peptides, such as hydrophobicity, hydrophilicity, size, ion affinity, bioaffinity and many more. Once separated, peptides are ionized and transferred to the MS for mass detection of the unfragmented precursor ion (MS1) followed by fragmentation and detection of the fragments originating from the same precursor ion (MS2). The information obtained for each peptide is the chromatographic retention time (RT) and the mass spectrum containing a relative intensity and the mass to charge (m/z) ratio of the precursors or the fragmented ions. Ideally, each peptide could be chromatographically separated, ionized, and subsequently analyzed in the MS. However, this is not the case, as different biases and technical trade-offs must be considered throughout the analytical pathway.

Over the past years, significant improvements in HPLC systems with more reproducible nanoliter flows and increased column resolution help to considerably reduce sample complexity prior to MS analysis. Faster MS acquisition rates together with increased resolution and sensitivity have enabled recording of high accuracy mass spectra. A wider repertoire of fragmentation techniques such as CID, HCD, ETD, EThCD, UVPD and many more, allow adjustment of the fragmentation to the analyte at hand. Through increased detector acquisition speed, a multitude of peptides can be sequenced that roughly have the same RT. Furthermore, the development of the Orbitrap technology by Makarov[196] and colleagues represents a significant breakthrough in mass spectrometry, that has had a monumental impact on our understanding of proteomics today, but

ULTRA-HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY (UHPLC)

Chromatography is a technique in analytical chemistry used to separate, identify, and quantify components in a sample mixture. In principle, samples are loaded onto columns containing an adsorbent material, the stationary phase. Each component in the sample interacts differently with the stationary phase. A pressurized solvent mixture is passed through the column, called the mobile phase, allowing an interaction with analytes. If the interaction with the mobile phase predominates, analytes are eluted from the column at certain retention times (RT) and transported to the detector. By changing the composition of the mobile phase, ideally, all analytes can be transferred from the column to the mobile phase. Overall, analytes interact slightly differently with both the mobile and stationary phase, which results in a separation based on different RTs. Various physicochemical properties can be exploited for chromatographic separation, such as differences in analyte size (size exclusion chromatography), affinity towards a polar stationary surface (normal phase chromatography), hydrophobicity (reverse phase chromatography), and many more.

In this work, (U)HPLC was coupled to LC-MS/MS, and the separation method employed was reversed phase (RP) chromatography. Peptides have varying degrees of hydrophobicity, and by linearly increasing the concentration of the organic solvent, peptides are eluted with increasing hydrophobicity. The solid phase is generally based on silica or synthetic organic polymers and contain covalently bound alkyl chains of different lengths such as C4, C8, C12 or C18 residues[197]. C18 columns are frequently used for separation, desalting and concentration of peptides and proteins mainly because the sample is concentrated in a small volume of volatile solvent[198]. The volatile solvent is also advantageous to online coupling to a mass spectrometer.

Relationship between different HPLC parameters

The resolution, describes the distance between two chromatographic peaks at baseline, according to equation 1[199].

$$R = \frac{1}{4} \cdot \sqrt{N} \cdot \frac{k_2}{1 + k_2} \cdot \frac{\alpha - 1}{\alpha} \qquad 1$$

where R = resolution, N = plate number, α = separation or selectivity factor, k = retention factor[199].

The peak capacity is the number of peaks that can be separated in a given time window, with a sufficiently good resolution. The resolution can be improved by increasing efficiency, selectivity (α), and retention (k)[199].

The efficiency of a column is generally reported as the number of theoretical plates, N as reported in equation 2[199].

$$N = \frac{L}{H} = \frac{16 \cdot t_R^2}{w^2} \qquad 2$$

where N = plate number, L = length of the column, H = height of a theoretical plate, t_R = retention time measured from the injection, and w = width of the peak at its base. The concept of theoretical plates is a model that supposes that the column contains a large number of separate layers allowing interactions with the solid phase[200].The plate number is defined for isocratic separation and under the assumption that the peak follows a Gaussian curve. Excellent chromatograms have late appearing peaks (large t_R) that are narrow (small w) thereby producing large plate numbers (N). The column efficiency is dependent on the column dimensions, such as diameter, length, and particle size:

1. Reducing the diameter of a column results in improved sensitivity through increased peak height, lower flow rates required to achieve optimal linear velocity, and reduces solvent waste[201].
2. The plate number increases with longer HPLC columns, which translates into better resolution[198].
3. Columns containing smaller particle sizes have better resolution, because the diffusion paths are shorter allowing analytes to transfer in and out of the particle more quickly[202]. Columns in modern HPLC systems are packed with 2µm or > 3-µm particles[203].

Another option to increase the resolution is by increasing the gradient time, which results in slower flow rates and higher the peak capacity[204]. Optimal flow rates are inversely proportional to particle size, however, use of higher flow rates is possible if the system-pressure limit allows it[205].

DATABASE SEARCH

In proteomics and immunopeptidomics, LC-MS/MS-based studies generate a high number of fragment spectra that contain sequence information for peptides and other analytes from the sample. Due to the high acquisition rate of modern instruments, the number of MS2 spectra has increased to tens of thousands per LC-MS/MS run, rendering their manual interpretation impossible. As a result, three main strategies have been developed to computationally identify peptides from MS2 spectra[206]: i) database searching, where the experimental spectra are searched against a reference database of protein (or peptide) sequences, ii) de novo sequencing, where amino acid sequences are directly inferred from the spectra, and iii) spectral library searching, where experimental spectra are searched against a library of spectra from known peptides.

The most common computational approach is database search. Over the years, more than 30[207] different search algorithms were developed starting with SEQUEST[208,209] and Mascot[210], while Andromeda[211] and Comet[212] are more recent additions. Each implementation has different properties that have been reviewed here[207].

In principle, experimentally obtained MS2 spectra are queried against a target sequence database, such as the reviewed human proteome as encompassed in UniProt. The experimental MS2 spectra oftentimes undergo preprocessing, whereby low-intensity peaks are filtered out. The target database is digested *in silico* into peptides of a desired length, followed by fragmentation to generate theoretical spectra. The *in-silico* digestion mimics the experimental enzyme cleavage and fragmentation of sample proteins, and provides theoretical MS2 spectra that can be matched against experimental MS2 spectra[207]. A set of search parameters can tune the specificities of the comparison. Resulting peptide spectrum matches (PSMs) are then reported with a computed quality control score.

The choice of the sequence database impacts the results of the search. Ideally, the search database should cover only proteins that are likely to be present in the samples, without inflating the database unnecessarily. Expanding the search space negatively impacts sensitivity, leading to multiple peptide annotations per precursor, as well as increasing the false discovery rate (FDR). Verheggen et al. illustrate this effect in Figure 4, whereby enlarging the search space with additional PTMs, isoforms and many more parameters results in an increased number of peptides matched to a single precursor and inflates the false discovery rate (FDR)[207]. Therefore, all the search space should be tailored to the hypothesis underlying the experiment.

The key quality control step in database search is estimating the number of falsely annotated peptide sequences. For this purpose, the database search is performed against the target dataset

concatenated to a decoy database. Thus, it is possible to estimate the number of false positves[171,213,214].

Key Differences Between Proteomics and Immunopeptidomics Experiments

Key differences between proteomics and immunopeptidomics experiments relate to i) the search space, ii) peptide lengths, and iii) analyte concentration prior to LC-MS/MS analysis.

In proteomics experiments, proteins are isolated and digested with certain enzymes, whereby Trypsin, LysC, and many more have been described. Trypsin is a popular enzyme, as it cleaves proteins after an Arginine (R) or Lysine (K). This leads to peptide fragments with C-terminal R or K[215] and results in a substantially reduced theoretical search space yielding confident peptide and protein identifications. In immunopeptidomics, HLA molecules of the respective human subjects and donors determine which peptides are presented. Indeed, HLA molecules have preferred binding motifs, that can be frequently summarized by anchor amino acid residues at position 2 and the C-terminus. When accounting for all C-terminal anchor amino residues, no informed constraints can be imposed upon the search space.

Second, peptides eluted from HLA molecules have predetermined lengths, with a mode of 9 aa for HLA-I ligands and 15 aa for HLA-II ligands, as illustrated in Results Part I, Figure S1. The short HLA-I ligands are more difficult to discriminate between multiple theoretical sequences during database search. Longer peptides, such as tryptic peptides, or HLA-II ligands can be annotated to a theoretical spectrum with a higher confidence. Due to their predetermined length, all HLA-I ligands cluster in a narrow m/z window (400 - 650 m/z), necessitating a wide dynamic range to allow detection of both high- and low-abundance peptides.

Third, the analyte concentration is routinely determined in proteomics experiments, but seldomly in immunopeptidomics experiments. The limited sample input in biological experiments, together with the isolation of only HLA-presented peptides through immunoprecipitation results in very low concentrated peptide samples. In these situations, concentration estimation with *i.e.* the Bradford assay would require a large proportion of the sample. Furthermore, the final peptide concentration is dependent on the HLA expression in the sample of interest. Tissues or tumors with a high expression of HLA molecules yield higher numbers of peptide identifications. HLA expression can be analyzed in cell suspensions or hematologic malignancies[78], but is difficult to estimate in solid tissue samples.

All these particularities can be compensated by performing peptide sequencing on an instrument with a high resolving power, mass accuracy, sensitivity, and dynamic range. Typically, Orbitrap instruments are well suited to cover these requirements. These aspects also hamper the discovery of PTMs in immunopeptidomics. Considerable efforts were required to confidently identify PTMs. Nevertheless, novel technologies, such as IMS promise to facilitate the identification of PTMS and distinction between positional isomers.

Definitions in Mass Spectrometry

Mass analyzers record m/z-valued of molecules, such as peptides and their fragments, based on different physical methods. Widely employed mass analyzers are the Orbitrap, linear ion trap (IT), and time of flight (TOF) detectors. Relevant performance metrics that define efficacy of mass spectrometers are the resolving power, mass accuracy, sensitivity, dynamic range, and speed. These parameters differ with instrument and detector type, and unfortunately, no ideal instrument has been developed yet, that combines only favorable parameters for these performance criteria.

The mass resolving power, or resolution R has been defined as the ability to distinguish ions differing in the m/z ratio by a small increment[216] or 1 Da[217].

$$R = \frac{m}{\Delta m} \tag{3}$$

The resolution is calculated as illustrated in equation 3, where m designates the mass (m/z) and Δm the peak width necessary for separation defined by the full width at half maximum (FWHM). An exemplary calculation is illustrated in Figure 1.

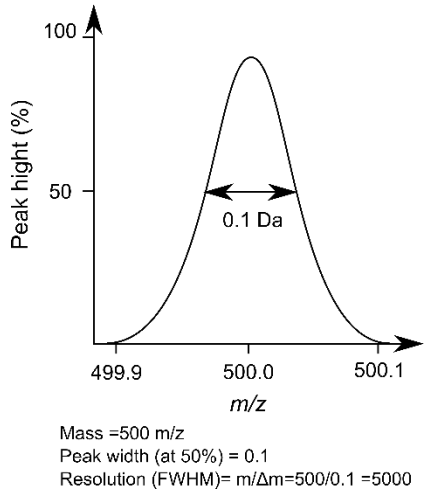


Figure 1: Exemplary calculation of the resolution

$$R = \frac{m}{\Delta m} = \frac{500\ m/z}{0.1\ FWHM} = 5000$$

The accuracy of a mass measurement refers to the degree of deviation from the true mass of the analyte. Mass accuracy is defined as the ratio of the m/z measurement error and the true m/z value of the analyte and is usually measured in ppm. High mass accuracy is required in particular when mixtures with similar m/z are analyzed. To separate similar peptide species, a high resolving power and chromatographic separation are necessary, while high mass accuracy alone is insufficient[218].

The sensitivity of a mass spectrometer is generally defined as “the slope of the calibration curve” relating signal intensity to absolute quantity or concentration[219]. The sensitivity is dependent on the resolving power of the instrument[220]. The relative detection limit depicts the smallest amount of material detectable in a matrix relative to the amount of material analyzed and is given in atomic, mole or weight fractions[220]. The terms “mass sensitivity” and “concentration sensitivity” are a source of misconceptions and need to be specified clearly[219]. NanoESI-MS operates at low flow rates and exhibits a high mass sensitivity, which is advantageous when working with low concentration, volume-limited samples[219].

The dynamic range refers to the ratio of a high intensity MS1 peak which is more intense than 95% of the peaks, to a low intensity MS1 peak which in turn is less intense than 95% of the peaks. The dynamic range only refers to MS1 spectra that have triggered an MS2 spectrum matched to a peptide, during the RT span where most of the peptides are eluting. Furthermore, the linear dynamic range is the range over which the ion signal increases linearly with analyte concentration.

The scan rate refers to the number of recorded mass spectra per second and is measured in Hz. It is an essential parameter, as it must be adjusted to the chromatographic elution profile of a certain peptide. The scan rate positively influences the resolving power in ion trap and Orbitrap instruments, meaning that higher resolution scans require a longer dwell time in the respective trap.

Orbitrap detectors have a high resolution, with a customary upper bound of 240,000, which is now extendable to 1,000,000. The high resolving power characteristic for Orbitrap instruments is dependent on the scan rate and m/z. However, only a limited number of ions can be analyzed in the Orbitrap in one scan due to space charge effects. Ion trap and Orbitrap instruments confine many ions in a small volume which results in mutual repulsion of particles with the same charge (ions and electrons) that influence the current in a charged particle beam. Ions deviate from the expected trajectories caused by external fields due to space charge effects[220].

The main advantage of TOF detectors is their immense scan rate, which can amount to 100 Hz (100 MS2 spectra per second) independently of the m/z and amount of analyzed ions. However, their resolution is significantly lower when compared to Orbitrap detectors ranging between 30,000 and 50,000 (Table 4-1).

Table 4-1: Instrumentation Comparison. The qExactive HF has been excluded, as it has been discontinued by Thermo Fisher. The Orbitral Eclipse and Orbitrap Exploris 480 have been inaugurated at ASMS 2019, which was too late for decision making. The Orbitrap Exploris

Company	Waters	Sciex	Bruker Daltonics	Thermo Fisher Scientific		
Instrument	Synapt G2-Si	Sciex 6600 triple TOF	TimsTOF Pro	qExactive HF	Orbitrap Fusion	Orbitrap Fusion Lumos
Detector	QuanTOF	QTrap/tripleTOF (accelarator)	TOF	Quad/IT	Quad/IT/OT	Quad/IT/OT
IMS	T-wave IMS – Drift time	Linear FAIMS, additional separation for PRM/MRM	TIMS		FAIMS	FAIMS
Fragmentation	CID; 2stepCID; ETD	CID	CID	CID, in source HCD	CID, HCD, ETD; hybrid fragmentations	CID; HCD; ETD; UVPD; hybrid fragmentaions
Collision Cell	Trap and Transfer T-wave region	Quad2/high Freq. curved LINAC collision cell	Quad	HCD collision cell	IRM, IT	IRM, IT
Resolving Power FWHM	40,000 @any m/z	40,000 – 42,000 @ any m/z	30,000 – 50,000 @ any m/z	240,000 @m/z 200	15,000 -450,000 @ m/z 200	240,000 @ m/z 200 up to 500,000
Mass Accuracy (ppm)	1 – 10 ppm	Ext. + Int. cal. 1 – 2 ppm	5 – 20 ppm improved s/n with PASEF	Int. >1 ppm RMS Ext. >3 ppm RMS	Int. >1 ppm RMS Ext. >3 ppm RMS	Int. >1 ppm RMS Ext. >3 ppm RMS
Dynamic Range	$>10^4$	10^4 - 10^5	10^{10}	OT 1:5,000	OT 1:5,000 FAIMS >105	OT 1:5,000 FAIMS $>10^5$
Acquisition Speed (Hz)	30 Hz	100 Hz 40 ms MS1 up to 1 ms MS2	100 Hz (100 ms cycle time)	40 Hz	20 Hz OT 20 HZ IT 15 Ht FAIMS	30 Hz OT 40 Hz IT 15 Hz FAIMS
Transmission/ Sensitivity	n.a.	50 -100 amol	n.a.			0.4 u - 1,200 u Quad: 50 -3,000 m/z
Ion Source	Nano source	NanoSpray III source, Ion Drive Qjet guide	Nano Source	NanoFlex	nanoFlex	nanoFlex
Comment	MSE: quant and ident. from same spectrum	SWATH-MS, IPr gas phase modifier	PASEF; DIA, IMEX			Operation Experience

Ion Mobility Separation in LC-MS/MS Setups

Recent advances have enabled the implementation of ion mobility separation (IMS) into mass spectrometers. The original limitation resided in the large size of drift tubes, which hindered their incorporation in mass spectrometers[221]. In IMS, ions are separated by their size, charge, and shape, more precisely summarized as the collisional cross section (CCS). In timsTOF data, the CCS is described through the Mason Schamp equation 4[222]:

$$CCS = \frac{3ze}{16}\frac{1}{k_0}\sqrt{\frac{2\pi}{\mu k_b T}} \quad (4)$$

Mobility values (k or k_0) can be measured and used akin to chromatographic RTs. Positioned between the HPLC and MS, IMS offers an orthogonal dimension of separation increasing the sensitivity and selectivity, particularly in highly complex sample mixtures. In immunopeptidomics this further separation step could ideally improve the signal to noise ratio, and deconvolute isobaric peptides prior to mass analysis. In general terms, ion mobility techniques can be separated into three concepts: temporally-dispersive, spatially-dispersive, and ion confinement (trapping) with selective release[221]. IMS applications from all three categories have been implemented in LC-IMS-MS settings by different vendors, as illustrated in Table 4-2 and reviewed[221].

Table 4-2: Applications of IMS and corresponding LC-IMS-MS instruments with high interest to immunopeptidomics[221].

IMS Technology		Instrument	Vendor
Temporally Dispersive	Drift tube IMS		
	Travelling wave IMS	Synapt G2Si	Waters
Spatially Dispersive	Field asymmetric (FAIMS)	Linear: SelexION® Cyclical: FAIMS Pro™	AB Sciex Thermo Fisher Scientific
	Differential mobility IMS		
Confinement and selective release	Trapped IMS	timsTOF Pro	Bruker

Temporally dispersive ion mobility methods lead ions along the same path, with different arrival times at the mass analyzer. Spatially dispersive ion mobility methods separate ions along different drift paths according to their ion mobility but does not influence the arriving time. Ion confinement and selective release methods trap ions at a certain voltage and sequentially eject them with modulation of the electric field[221].

Mechanism of Action of the timsTOF Pro in PASEF Modus

IMS was at the time of this benchmark, January 2019, a technology that had promise to be equally appropriate for immunopeptidomics samples, as the Orbitrap technology turned out to be. The timsTOF Pro was the only instrument with trapped ion mobility spectrometry (TIMS) capabilities included in this study and is therefore briefly explained.

Trapped ion mobility spectrometry (TIMS) implemented into Bruker Daltonic's timsTOF Pro belongs to the confinement and selective release category that traps ions within a pressurized region and selectively releases them based on different ion mobilities[221]. In the TIMS device, ions

are brought to rest at different positions in the ion tunnel by being balanced in an electrical field against a constant counteracting nitrogen stream. After a sufficient number of ions have been trapped and separated, the electrical potential is lowered and time-resolved ions are ejected into the downstream quadrupole and TOF mass analyzer[223,224].

As described by Meier et al.[222], an ingenious modality of operation termed parallel accumulation serial fragmentation (PASEF) was implemented. In PASEF mode, the precursor selection is synchronized with the TIMS separation. The TIMS funnel is "electronically" divided into two segments: i) the accumulation trap, that stores all ions entering the mass spectrometer ii) the second part performs the TIMS separation. As soon as the second part has finished, all ions from the first part are transferred into the second where they are separated. Thus, ions are not lost during the TIMS separation if the accumulation and separation have equal durations. In both TIMS regions ions experience a drag from incoming gas flow and a repulsion from the electrical field. Depending on their CCS, ions with high mobility accumulate at the entrance of the tunnel and ions with low ion mobility accumulate at the exit. The TIMS precursor selection algorithm considers the quadrupole switching and measuring time and estimates how many precursors can be selected and fragmented in one PASEF frame. The PASEF acquisition scheme is illustrated in Figure 2[225].

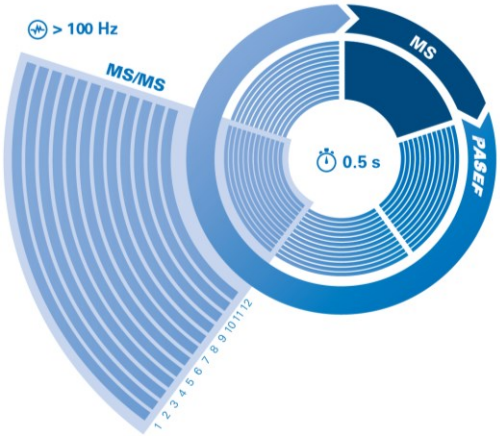


Figure 2: PASEF acquisition scheme

1. Acquisition of an MS1 heatmap 100 ms
2. Precursor isolation algorithm selects appropriate precursors for fragmentation considering their intensity and the quadrupole switching and acquisition time
3. 12 different precursors are fragmented and scanned => TIMS elution voltage is synchronized with the quadrupole selection.
4. The cycle is repeated.

TOF analysis time: 10 µs.
(https://www.bruker.com/products/mass-spectrometry-and-separations/lc-ms/o-tof/timstof-pro.html)

Briefly, an MS1 heatmap generated in PASEF mode has a duration of 100 ms and describes the m/z and 1/k0 of the ions contained. Based on this heatmap, the precursor selection algorithm can define a polygon that describes the most efficient order of precursor selection for fragmentation taking into consideration the ion intensities and quadrupole switching and measurement time. If an intense ion is not selected for fragmentation due to the quadrupole switching time, it will be prioritized in the next PASEF frame. If certain ions have a low intensity (below a set threshold), they can be selected again for fragmentation. In a postprocessing setting all these MS2 spectra can be added to obtain a better signal to noise ratio and increased spectral quality.

AIM OF THE ANALYSIS

With these considerations in mind, we set out to determine performance differences regarding immunopeptidomic samples between five different instruments. Therefore, we contacted all vendors mentioned in Table 4-2 for a closer description of their technology, as well as for measurements of a serial dilution series, the Demobox samples, described in the material and methods section. The Synpat G2Si instrument was disqualified due to the low-resolution mass accuracy, slow response times and long waiting times (> 3 months, and additionally 4-8 weeks to obtain the results). AB Sciex recommended the TripleTOF 6600 instead of the SelexION technology, as the ion mobility functionality could only be used for separating previously known isobaric compounds and was not suitable for constant usage. Thermo Fisher Scientific had difficulties providing the necessary EU-licenses, for FAIMS sale, and failed to provide the results to the Demobox measurements in the communicated time frame. As a result, the timsTOF pro was the only instrument with IMS capabilities included in our benchmark.

4.2.3 Materials and Methods

JY culturing conditions

The JY cell line (ECACC 94022533, IHW9287) is an Epstein–Barr virus-immortalized B cell lymphoblastoid line, with the advantageous property of being homozygous for all HLA-I loci (HLA-A*02:01, HLA-B*07:02, and HLA-C*07:02). JY has been established from a homozygous male individual in the Indiana Amish population[226]. Furthermore, it expresses between 200,000 – 1,000,000 HLA copies on its cell surface (personal communication Daniel Kowalewski), thereby allowing the isolation of large numbers of peptides from a limited number of cells. Due to the unpretentious culturing conditions and well-studied HLA allotypes, we employ the immunopeptidome of the JY cell line for standard suitability testing (SST) to regularly monitor and evaluate the performance of the mass spectrometers.

A frozen 1 ml JY cell suspension containing 3×10^6 cells in freezing medium consisting of 10% dimethyl sulfoxide (DMSO, WAK Chemie) in fetal calf serum (FCS, Capricorn Scientific) was thawed by dilution with RPMI 1640 medium (Thermo Fisher Scientific) supplemented with 1% sodium pryruvate (100 mM) (Thermo Fisher Scientific), 1% MEM non-essential amino acids (100x) (Thermo Fisher Scientific), 1% Penicillin (103 x U/ml)/Streptomycin (10 g/ml) (Sigma-Aldrich), and 10% FCS. Cells were cultured in humidified incubators at 37°C and 7.5% CO_2. Cells were split every two to three days at a ratio of 1:2, 1:3 or 1:5, depending on cell proliferation and medium discoloration. For harvesting, culture medium was removed by centrifugation (1500 rpm, 15 min, room temperature, with break), and cells were washed twice with Dulbecco's Phosphate-Buffered Saline (Thermo Fisher Scientific) to remove residual medium. Pellets were pooled to obtain an approximate 5 ml pellet containing about $3 - 4 \times 10^9$ cells and were frozen and stored at -80°C until immunopurification of HLA ligands was performed.

Isolation of HLA ligands from the JY cell line

JY19#1 HLA-I ligandome was isolated by Jens Bauer, using the protocol described in Results I under the Methods section “HLA immunoaffinity purification”. Deviations from the standard procedure were motivated by the large number of cells used. Briefly, a 7 ml cell pellet, approximately containing 4×10^9 cells was employed as input material. The cell pellet was lysed in 30 ml 2x lysis buffer, and 10 ml 1x lysis buffer. Subsequently, the cell lysate was cleared by two consecutive centrifugation steps at maximum speed, and 4°C for 45 minutes. The cleared lysate was loaded onto BioRad Econo Columns with 5 cm diameter, containing 20 ml antibody-bead conjugates (1 mg/ml) and was run cyclically over-night. Peptides were eluted into 5 ml low-bind Eppendorf tubes, the first elution being performed with 100 µl 10% TFA and 600 µl 0.2% TFA (followed by three subsequent elution steps with 600 µl 0.2% TFA). Both HLA-I and -II ligands were further purified by ultrafiltration employing a 10 kD filter unit suitable for a 15 ml Falcon tube. Peptide solutions from the JY19#1 experiment were lyophilized to complete dryness and resuspended in 3 ml Aload.

DEMOBOX SAMPLE PREPARATION: SERIAL DILUTION AND RT-PEPTIDE SPIKE-IN

The purified peptide solution was spiked with a cocktail of 10 heavy labeled synthetic peptides selected, weighed, and diluted by Lena Katharina Freudenmann to a stock solution of 10 fmol/µl. The RT-peptides spiked into JY19#1 No.1 sample are described in detail in Table 4-3. Ten different tryptic peptides covering the retention time range of HLA ligands were selected and synthesized with an L+7 heavy isotopic modification to allow easy distinction between native and spike-in peptides. The RT-peptide mix, having a stock concentration of 10 fmol/µl, was diluted in JY 19#1 to 0.2 fmol for each RT-peptide.

Table 4-3: Heavy isotope-labeled peptides that were spiked into sample No.1 consisting of the JY19#1 HLA ligand solution. The final concentration per RT-peptide in sample No.1 was 0.2 fmol/µl.

Peptide	Modification (L+7)	Spike-in concentration (No.1)
ISLGEHEGGGK	IS[L(^{13}C6; ^{15}N)]GEHEGGGK	0.2 fmol/µl
VGASTGYSGLK	VGASTGYSG[L(^{13}C6; ^{15}N)]K	0.2 fmol/µl
TLIAYDDSSTK	T[L(^{13}C6; ^{15}N)]IAYDDSSTK	0.2 fmol/µl
FLASSEGGFTK	F[L(^{13}C6; ^{15}N)]ASSEGGFTK	0.2 fmol/µl
GFLDYESTGAK	GF[L(^{13}C6; ^{15}N)]DYESTGAK	0.2 fmol/µl
ALFSSITDSEK	A[L(^{13}C6; ^{15}N)]FSSITDSEK	0.2 fmol/µl
HFALFSTDVTK	HFA[L(^{13}C6; ^{15}N)]FSTDVTK	0.2 fmol/µl
VYAETLSGFIK	VYAET[L(^{13}C6; ^{15}N)]SGFIK	0.2 fmol/µl
GASDFLSFAVK	GASDF[L(^{13}C6; ^{15}N)]SFAVK	0.2 fmol/µl
FFLTGTSIFVK	FF[L(^{13}C6; ^{15}N)]TGTSIFVK	0.2 fmol/µl

Table 4-4 describes the dilution of the JY19#1 HLA-I ligand sample with the isotopically labeled RT-peptide solution. Sample No.1 was spiked with RT-peptides. Subsequently, sample No.1 was diluted 1:2 with Aload resulting in sample No.2, which in turn was diluted 1:2 with Aload to obtain sample No.3, until sample No.5. The demobox samples thus contain a serial dilution series from the stock solution, that allow monitoring of MS parameters across different LC-MS/MS instruments.

Table 4-4: The five samples used to examine the comparability with other LC-MS/MS setups with the Orbitrap Fusion Lumos. The HLA ligand solution JY19#1 isolated from the JY cell line was spiked with RT-peptides resulting in sample No.1, which was subsequently serially diluted up to sample No.5.

Dilution factor	Sample	Dilution	Final concentration per RT-peptide
1	No.1	1:50 dilution of HLA-I JY19#1 ligands with 10 fmol/µl RT-peptide	0.2 fmol/µl
1:2	No.2	500 µl Aload + 500 µl Dilution No.1	0.1 fmol/µl
1:4	No.3	500 µl Aload + 500 µl Dilution No.2	0.05 fmol/µl
1:8	No.4	500 µl Aload + 500 µl Dilution No.3	0.025 fmol/µl
1:16	No.5	500 µl Aload + 500 µl Dilution No.4	0.0125 fmol/µl

The resulting five dilutions were aliquoted and stored at -80°C until LC-MS/MS analysis. We subsequently distributed the serial dilution series across five locations with instructions regarding the settings for both the HPLC and mass spectrometry system, as listed in the Results Part I, Section 4.1.3 Materials and Methods, subheading: Mass Spectrometric Data Acquisition. Furthermore, we strongly recommended the implementation of the same chromatographic conditions with respect to the separating column, solution composition, gradient and pick-up volume from the chromatographic vial. Thus, we aimed to transfer the extensive parameter optimization of our instrumentation to the other Demo-labs, to allow an informed decision regarding which instruments have a comparable performance and might be implemented either as a complementary or capacity-expanding instrument.

4.2.4 Results And Discussion

Generation of the Demobox Samples

We analyzed the serial dilution series on the in house Orbitrap Fusion Lumos to evaluate the baseline parameters we intended to use for the benchmark. The first measurements show a linear dilution profile in regard to both the immunopeptidome yield and the cumulative MS1 peak area of identified peptides (Figure 2A), but also considering the cumulative area of spiked-in heavy-isotope-labeled RT peptides (Figure 2B). In addition, all ten RT-peptides could be identified confidently at a lower-bound concentration of 25 amol/µl, while peptides ISLGEHEGGGK and GFLDYESTGAK could not be fragmented and identified at 12.5 amol/µl. Thus, the demobox samples show a linear serial dilution profile and highlight the lower limit of detection of the Orbitrap Fusion Lumos, making the samples ideal for LC-MS/MS system comparisons.

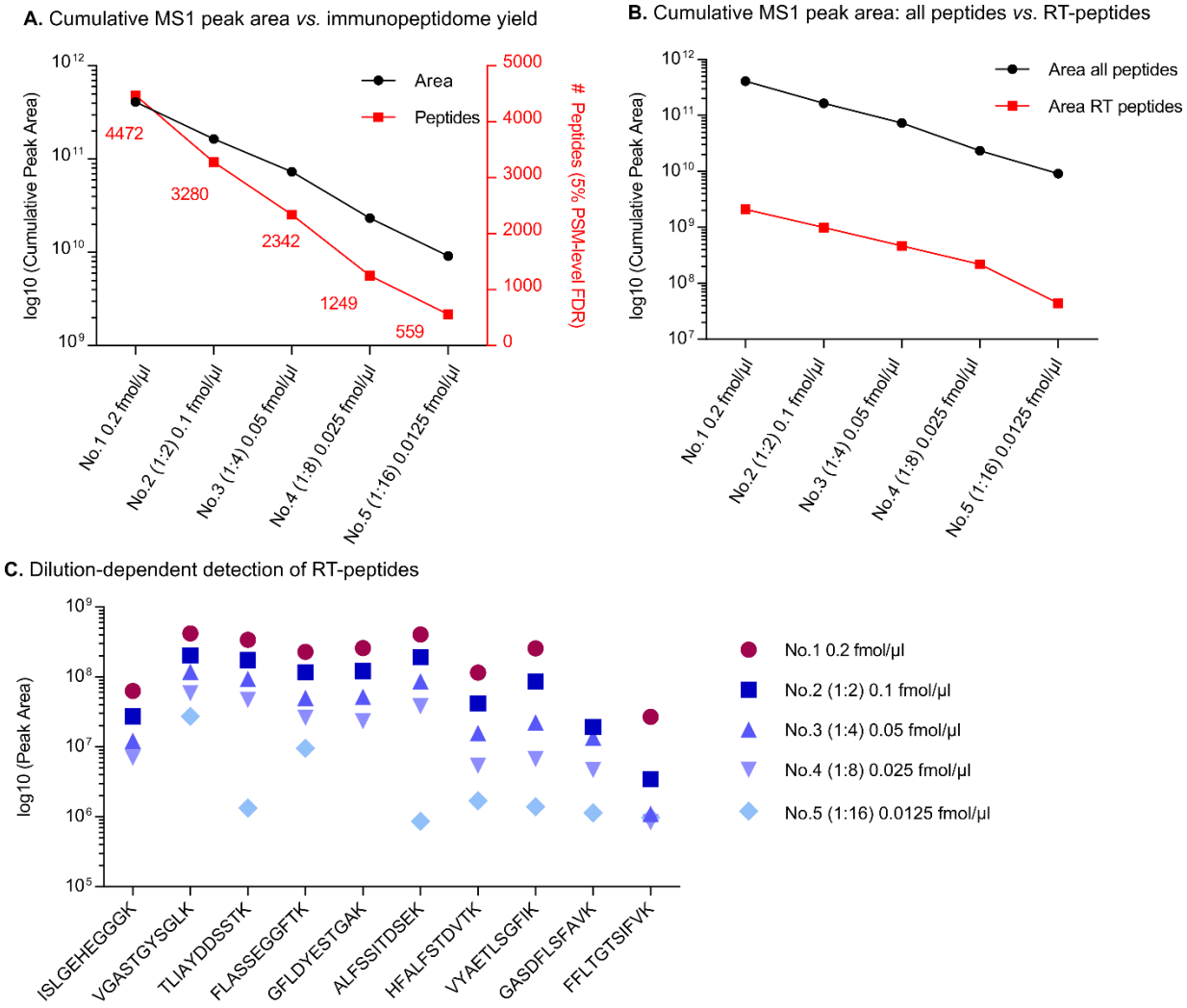


Figure 2: Serial dilution series of JY 19#1 Demobox.
(A) HLA ligands isolated from the JY cell line were spiked with 0.2 fmol/µl for each of the 10 isotope-labeled synthetic peptides (Sample No.1). A serial 1:2 dilution was performed. The number of identified HLA ligands per dilution is depicted on the right y-axis in red. The cumulative area of all MS1 precursors identified in the respective dilution is visible in black and associated to the left y-axis.
(B) The cumulative area of all MS1 precursors identified per dilution is depicted in black. The cumulative area of the spiked-in RT-peptides is illustrated in red.
(C) Concentration-dependent detection of isotype-labeled spiked-in RT-peptides. Not all peptides could be identified in all concentrations with the standard quality thresholds (5% PSM-level FDR).

INITIATING A SYSTEM-WIDE BENCHMARK

For the acquisition of new mass spectrometers, we evaluated the performance of six LC-MS/MS systems from three different vendors by comparatively evaluating LC-MS/MS runs of the five Demobox samples. The instrument Synpat G2Si the LC-MS/MS accessoir FAIMS could not be included in this comparison, as response times from Waters and Thermo Fisher Scientific exceeded the available time (> 3 months). Measurements on the Orbitrap Fusion were performed at Immatics Biotechnologies GmbH with the help of Daniel Kowalewski, on two different equivalent instruments. Furthermore, Marisa Schmidt from the Proteome Center Tübingen performed measurements on the q Exactive HF and q Exactive HFx (data not shown, as instrument was discontinued). Across all systems, 114 LC-MS/MS runs were acquired, as summarized in Table 4-5. Labs were instructed to implement the optimized solutions and instrument settings of our LC-MS/MS setup. However, further instrument-specific parameters were varied to adjust different technologies to the specificities of HLA ligandomics samples. Parameters that were varied frequently included the chromatographic gradient length and MS1 and MS2 resolution. As timsTOF data could at the time only be analyzed with PEAKS DB, we processed all runs with PEAKS DB to allow for a reproducible data analysis.

Table 4-5: The five Demobox JY19#1 samples were analyzed on six different LC-MS/MS instruments from three vendors. Details regarding the chromatographic gradient in minutes, the date of LC-MS/MS analysis, and optimization parameters (resolution: 20 K, 60 K) are listed.

	Orbitrap Fusion Lumos	**Orbitrap Fusion**	**q Exactive HF**	**LTQ Orbitrap XL**	**timsTOF Pro mit *PASEF***	**TripleTOF 6600**
Vendor	Thermo Fisher	Thermo Fisher	Thermo Fisher	Thermo Fisher	Bruker Daltonics	AB Sciex
MS Analysis	*Immunopepti-domics Core Facility*	immatics	Proteom Center Tübingen	*Immunopepti-domics Core Facility*	Bruker	AB Sciex
No.1	90 min /190303	70 min /N190321 70 min /N190315	90 min /190409	90 min /190329	60 K 30 min 60 K 60 min 20 K 100 min	90 min
No.2	90 min /190303	70 min /N190321 70 min /N190315	90 min /190409	90 min /190329	60 K 30 min 60 K 60 min 20 K 100 min	90 min
No.3	90 min /190303	70 min /N190321 70 min /N190315	90 min /190409	90 min /190329	60 K 30 min 60 K 60 min 20 K 100 min	90 min
No.4	90 min /190303	70 min /N190321 70 min /N190315	90 min /190409	90 min /190329	60 K 30 min 60 K 60 min 20 K 100 min	90 min
No.5	90 min /190303	70 min/N190321 70 min /N190315	90 min /190409	90 min /190329	60 K 30 min 60 K 60 min 20 K 100 min	90 min

Detection Limit Estimation and Quality of Peptide Identifications

A prerequisite for a new instrument is a high sensitivity, allowing for the detection of a high number of peptides from limited sample amounts. The Thermo Fisher Scientific instruments Orbitrap Fusion Lumos and Orbitrap Fusion together with their associated chromatography are highly similar technologies and have been optimized for immunopeptidomics samples in the respective laboratories. Therefore, it is not surprising that they show very similar performance patterns when looking at the immunopeptidome yield (Figure 3A). The q Exactive HF, distributed by Thermo Fisher Scientific, has a similar architecture and the same Orbitrap as the Tribrid instruments. Thus, similar numbers of identified peptides can be detected at high concentrations, but the performance between the two classes of instruments diverge at lower concentrations, owing in part to measurements performed on a non-optimized system. Three measurements, with different gradient times were performed on the timsTOF Pro. All three measurements clustered together, highlighting the possibility of drastically reducing the acquisition time to 30 minutes per run, without loss in sensitivity. The TripleTOF 6600 showed a reduced performance compared to the aforementioned instruments. Furthermore, the MS1 area could not be determined, with the software distributed by AB Sciex, Protein Pilot, but also using other algorithms such as SEQUEST embedded in the Proteome Discoverer 1.4, or PEAKS. Lastly, the LTQ Orbitrap XL, showed a similar performance profile as Sciex's TripleTOF 6600, despite being an outdated instrument. The linear dilution pattern of the Demobox samples was evident irrespective of the LC-MS/MS instruments tested (Figure 3).

Furthermore, the immunopeptidome purity was similarly high for all instruments. The 30-minute gradient run on Bruker's timsTOF Pro represents an outlier at ~75% purity irrespective of the concentration analyzed (Figure 3B). The number of identified RT-peptides was relatively constant for the reference instrument, the Orbitrap Fusion Lumos, with a steep decline at the lowest concentration (Figure 3C). Surprisingly, all 10 RT-peptides could be detected and identified only with the 60-minute gradient run on the timsTOF Pro. Even at the lowest concentration, it surpassed the lower bound detection limit of the Tribrid instruments. The worst performance was achieved with the q Exactive HF.

Taken together, the timsTOF Pro shows a comparable performance to the Tribrid instruments, with the advantage of a low detection limit, which can account for sensitivity, and faster fragmentation without loss of fragmentation information or intensity. The reduced gradient length is a desirable attribute as well.

Dynamic Range

The cumulative MS1 area under the curve (AUC) for RT-peptides and all identified peptides in the respective runs shows a linear gradient and a reliable quantification of low-abundant peptides (Figure 3D). This result indicates that the Thermo instruments and the timsTOF Pro show a high dynamic range which is highly advantageous for immunopeptidomic experiments.

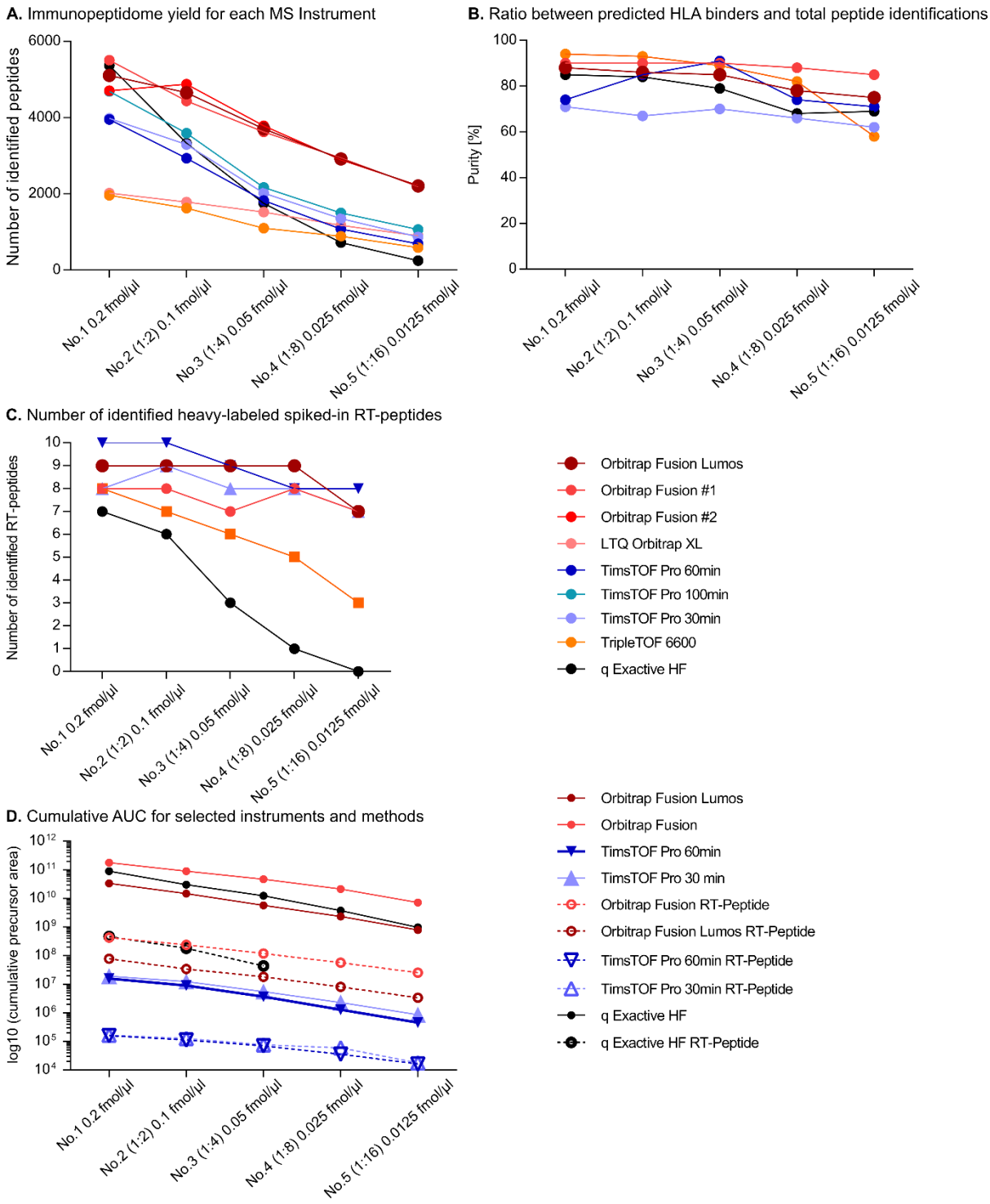


Figure 3: Qualitative benchmark between five LC-MS/MS systems.
(A) Instrument-dependent immunopeptidome yield varies for different concentrations. (B) The immunopeptidome purity determined as the ratio of predicted binders (netMHCpan4.0) and total number of identifications of each sample. (C) The number of detected RT-peptides varies with each instrument. Figures in A, B and C share the same legend. TimsTOF runs have been abbreviated by their gradient length, but according resolution settings are depicted in Table 3-4.
(D) The cumulative area of all MS1 precursors identified per dilution is illustrated by filled symbols connected by straight lines, while spiked-in RT-peptides are symbolized by hollow symbols connected by dotted lines.

To mimic high-yielding and low-yielding immunopeptidomic experiments, we investigated dilutions No.1 and No.4 due to their resemblance to standard samples (Figure 4A and B). The $\log_{10}$ -transformed AUC for each RT-peptide is depicted in Figure 4. As briefly mentioned previously, it was not possible to determine the MS1 intensities from measurements performed on the Sciex' TripleTOF 6600. Additionally, at 25 amol/µl, the area could not be determined for multiple peptides and LC-MS/MS instrumentations, despite an identification obtained from MS2 spectra. Overall, the Orbitrap instruments show the same order of magnitude of AUCs in sample No.1. The TimsTOF Pro shows reduced albeit highly reproducible areas at both concentrations. Areas obtained from the q Exactive HF measurements are on par with the Tribrid instruments Orbitrap Fusion and Orbitrap Fusion Lumos for sample No.1. For sample No.4, only one RT peptide was identified with the q Exactive HF, and the MS1 area could not be picked.

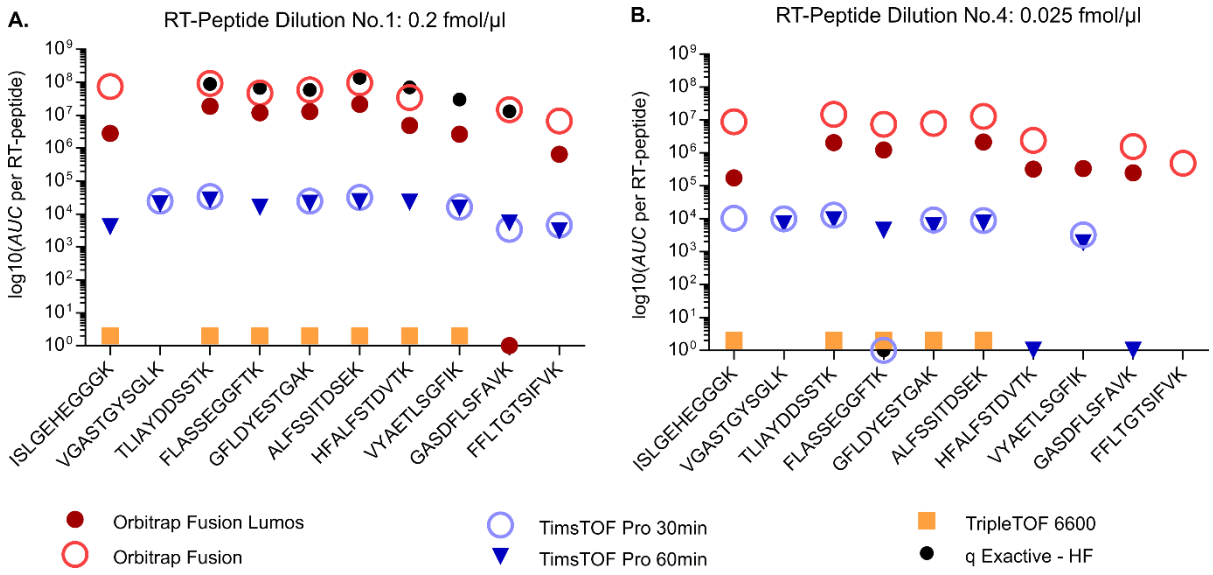


Figure 4: Concentration-dependent detection limit of different LC-MS/MS setups.
(A) $\log_{10}$-transformed AUCs of RT-peptides at the highest tested concentration: 0.2 fmol/µl per peptide. For peptides, that were identified via an MS2 spectrum, but no area could be picked from the corresponding MS1 spectrum, a default intensity of 1 was set.
(B) log10-transformed AUCs of RT-peptides at 25 amol/µl.

A further quality criterium is represented by the allotype distribution of identified HLA ligands and the accumulation of predicted non-binders. The purity of the sample and the allotype distribution correlate with each other (Figure 5). The Tribrid instruments show a reproducible purity of the sample, although with decreasing sample concentration less peptides match the binding motif of the respective allotype. Interestingly, a drastically reduced purity is visible with the 30-minute gradient acquired on the TimsTOF Pro. The 60-minute gradient shows varying percentages of non-binders, that do not necessarily correlate with the concentration of the sample. The TripleTOF6600 however, shows an increasing fraction of non-binders with decreasing concentration. Aside from varying proportions of non-binders, peptides were distributed across HLA-A, -B and -C allotypes as expected, with most peptides being presented by HLA-A, followed by -B and barely any peptides being presented by HLA-C.

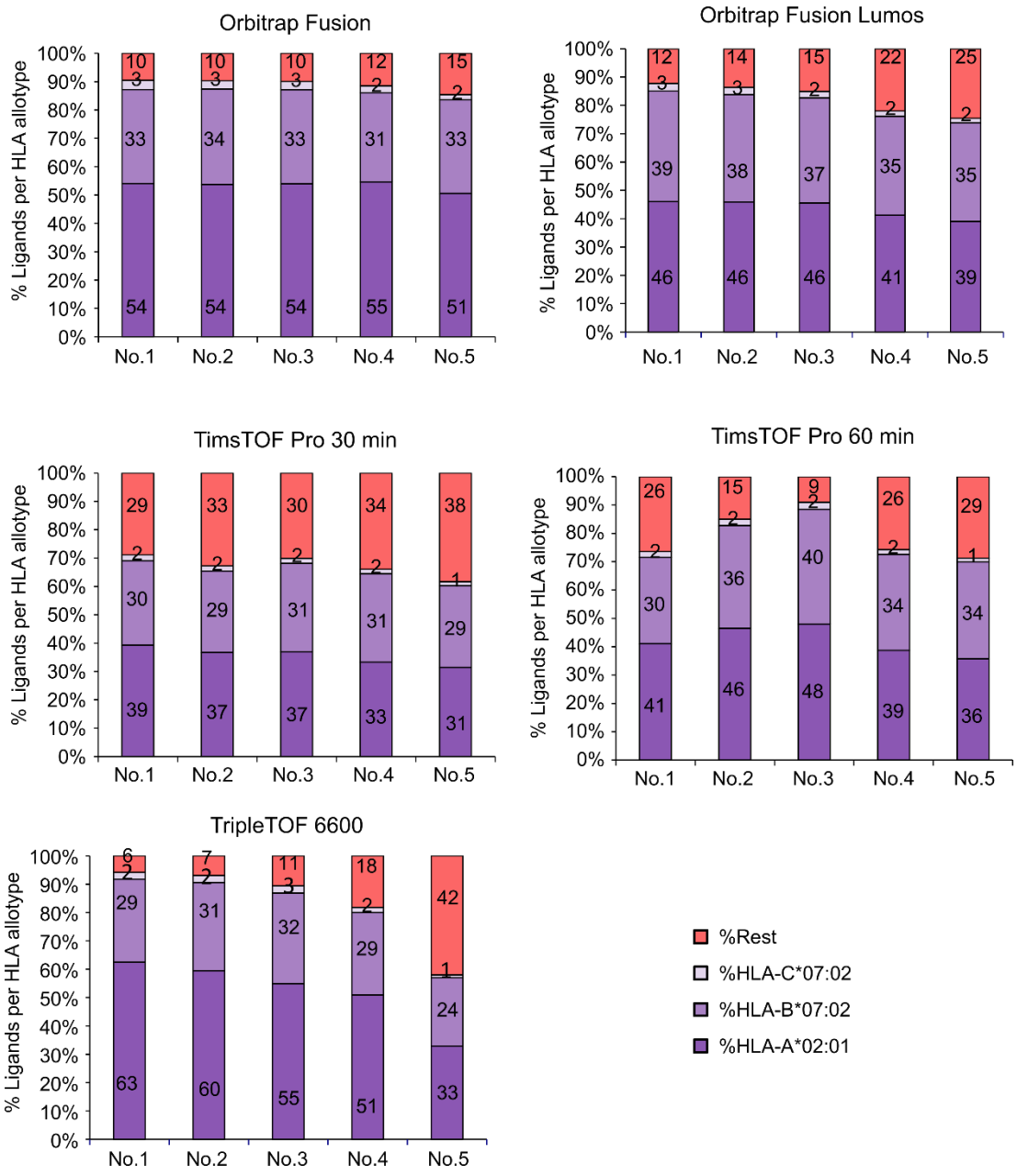


Figure 5: HLA Allotype Distribution.

Identified peptides were separated to be binders of A*02:01, B*07:02, C*07:02 or not bind any HLA allotype of the JY cell line. The allotype distribution was calculated in percent, but the absolute numbers of each category are added as data labels. HLA binding prediction was performed with netMHCpan 4.0[174], employing a binding rank threshold < 2%.

174

COMPLEMENTARITY

Our work focuses on expanding and mapping the known immunopeptidome in various primary tissue types. Therefore, an ideal instrument would be complementary to the in-house available Orbitrap Fusion Lumos and thereby unveil a different repertoire of HLA-presented peptides. To qualitatively assess complementarity of the benchmarked LC-MS/MS systems with the in-house available Orbitrap Fusion Lumos we performed proportional Venn Diagrams (Figure 6A-D) for the representative dilutions No.1 and No.4. In all pairwise comparisons, the number of unique HLA ligands is highest when analyzed with the Orbitrap Fusion Lumos in all comparisons. Yet, approximately 50% of peptides are uniquely identified with the timsTOF Pro. The Sciex TripleTOF 6600 only marginally increased the identifiable immunopeptidome. Taken together, the Bruker timsTOF Pro with PASEF represents a new type of technology that is complementary to Orbitrap instruments. The overall performance is still lagging behind the in house Orbitrap Fusion Lumos, an instrument which has been previously optimized. It may be expected that the performance of the remaining instruments can be massively improved by adjusting and optimizing the data acquisition parameters to best capture immunopeptidomes.

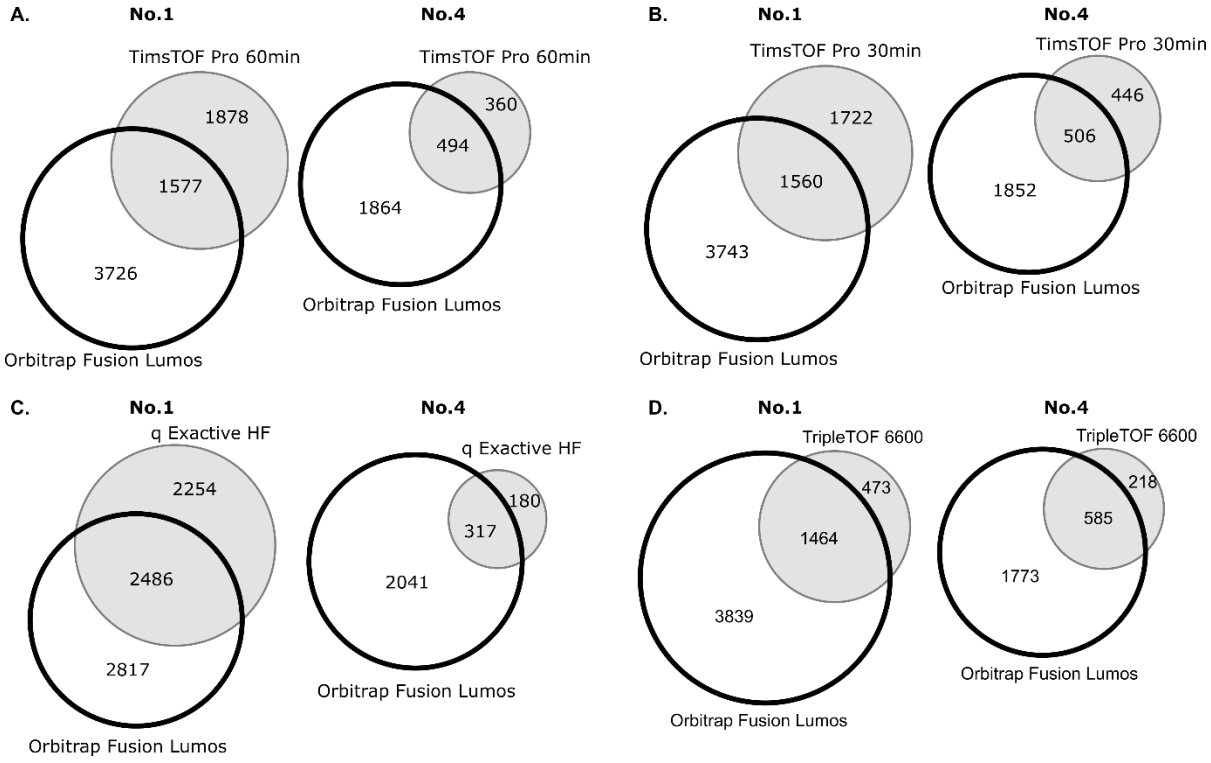


Figure 6: Size-proportional Venn diagrams illustrate complementarity between the available instrumentation (Orbitrap Fusion Lumos) and benchmarked LC-MS/MS systems. Samples No.1 and 4 resemble the range of naturally presented HLA ligands and have therefore been chosen to represent standard samples.

PART III:

THE IMMUNOPEPTIDOMIC LANDSCAPE OF BREAST CANCER

Ana Marcu[1,2], Tobias Engler[4], Sabine Matovina[4], Michael Ghosh[1,2], Stefan Stevanović[1,2,3], Hans-Georg Rammensee[1,2,3], Philipp Wagner[1,4]

1 Department of Immunology, Interfaculty Institute for Cell Biology, University of Tübingen, Tübingen, Baden-Württemberg, 72076, Germany.

2 Cluster of Excellence iFIT (EXC 2180) "Image-Guided and Functionally Instructed Tumor Therapies", University of Tübingen, Tübingen, Baden-Württemberg, 72076, Germany.

3 DKFZ Partner Site Tübingen, German Cancer Consortium (DKTK), Tübingen, Baden-Württemberg, 72076, Germany.

4 Department of Obstetrics and Gynecology, University Hospital of Tübingen, Tübingen, Baden-Württemberg, 72076, Germany.

AM performed the experimental isolation and LC-MS/MS-based identification of HLA ligands from tumor, adjacent benign, and benign mamma samples. AM performed the comparative analysis of HLA ligands identified in this results section and in Results Part I section and generated all plots. MG generated the prediction algorithm to define the HLA allotype of one sample based on its immunopeptidome. TE, SM and PW provided the samples, obtained patient informed consent, and managed the clinical patient metadata. SS, HGR and PW supervised this study.

4.3.1 ABSTRACT

Significant recent advances in cancer immunotherapies have led to long-term remission in patients with melanoma and non-small cell lung carcinoma. These therapies are currently tested in other cancer entities and it becomes evident that cancer immunotherapy is not beneficial for every patient. By knowing the antigenic repertoire presented by HLA molecules on breast cancer cells, patient-individualized immunotherapies can be tailored to target antigens exclusively presented on cancer cells.

In this study we employed LC-MS/MS-based characterization of the immunopeptidome of 31 primary breast cancer tissues (MaCa), 23 adjacent benign tissues (MaCa adj. ben.), and 6 healthy mamma (MaN) tissues. By comparatively profiling tumor and benign immunopeptidomes we were able to define a set of TAAs in breast cancer patients, which can be targeted by therapeutic peptide vaccination. Thereby, we identified 42,376 naturally eluted HLA-I ligands originating from 10,977 source proteins and 46,939 HLA-II ligands isolated from 5,289 source proteins. Overall, we have identified 44 HLA-I ligands (> 5/31 MaCa samples), 62 HLA-I source proteins (> 3/31 MaCa samples,), 46 HLA-II ligands (> 6/30), and 75 HLA-II source proteins (> 2/30 MaCa samples) identified exclusively on MaCa tissue. These frequently shared TAAs are shared by multiple molecular subtypes. Furthermore, they have been identified on other tumor types.

To further refine the set of naturally presented TAAs, we will expand the number of breast cancer specimens analyzed, perform exome and RNA sequencing from blood, tumor, and adjacent benign tissues for ten patients and generate a warehouse of immunogenic peptides for a potential clinical study. These warehouse peptides can then be confirmed in a validation dataset comprising tumor samples of the Triple Negative molecular subtype, the molecular subtype with the poorest prognosis and potentially the highest benefit in a clinical setting.

4.3.2 INTRODUCTION

BREAST CANCER PREVALENCE

Breast cancer is the most frequent type of cancer in women worldwide and is curable in ~70 - 80% of patients with early-stage, non-metastatic disease[227]. According to the GLOBOCAN report, approximately 2.1 million women were newly diagnosed with breast cancer and 626,679 women have died of breast cancer in 2018 worldwide[228]. Approximately 10% of breast cancers are inherited and are associated with a family history[229]. However, a family history of breast cancer is only loosely associated with an individual risk. Mutations in two high-penetrance tumor suppressor genes *BRCA1* (17q21) and *BRCA2* (13q13) are associated with an average cumulative risk of developing breast cancer in women by the age of 80 of 72% and 69%, respectively[230], while a relative risk of *BRCA2* mutations surmounts to 6% in men. *BRCA1* and *BRCA2* mutations are inherited in an autosomal-dominant manner and are tumor suppressor genes involved in double-stranded DNA break repair.

BREAST CANCER CLASSIFICATION BASED ON HISTOLOGY

Breast cancers can be divided into two main overarching groups: the carcinomas and the sarcomas. Carcinomas are cancers that arise from epithelial cells, while sarcomas arise from the stromal component of the tissue, mainly blood vessels and myofibroblasts. Sarcomas make up to 1% of breast cancers. Furthermore, breast cancers can be broadly categorized into *in situ* and invasive (infiltrating) carcinomas. The histopathological classification of breast cancer tumors is updated regularly by the WHO, incorporating molecular and genetic data to the originally morphological classification[231,232]. The most common type of breast cancer was previously known as invasive ductal carcinoma and is currently known as invasive breast carcinoma of no special type (NST). The most common premalignant lesions include the ductal carcinoma *in situ* (DCIS) and lobular carcinoma *in situ* (LCIS), but intraductal proliferative lesions and papillary lesions can also lead to malignant development. Table 4-6 illustrates the classification of frequent invasive breast cancer types.

Table 4-6: WHO classification of invasive breast carcinomas with the exception of the rare types[231,232].

Type
Invasive carcinoma of no special type (NST)
Pleomorphic carcinoma
Carcinoma with osteoclast-like stromal giant cells
Carcinoma with choriocarcinomatous features
Carcinoma with melanotic features
Invasive lobular carcinoma
Classic lobular carcinoma
Solid lobular carcinoma
Alveolar lobular carcinoma
Pleomorphic lobular carcinoma
Tubulolobular carcinoma
Mixed lobular carcinoma
Tubular carcinoma
Cribriform carcinoma
Mucinous carcinoma
Carcinoma with medullary features
Medullary carcinoma
Atypical medullary carcinoma
Invasive carcinoma NST with medullary features
Carcinoma with apocrine differentiation
Carcinoma with signet-ring-cell differentiation
Invasive micropapillary carcinoma
Metaplastic carcinoma of no special type
Low-grade adenosquamous carcinoma
Fibromatosis-like metaplastic carcinoma
Squamous cell carcinoma
Spindle cell carcinoma
Metaplastic carcinoma with mesenchymal differentiation
Chondroid differentiation
Osseous differentiation
Other types of mesenchymal differentiation
Mixed metaplastic carcinoma
Myoepithelial carcinoma
Epithelial-myoepithelial tumors
Adenomyoepithelioma with carcinoma
Adenoid cystic carcinoma

BREAST CANCER TUMOR GRADING

The American Joint Committee on Cancer (AJCC) TNM staging system is the classification universally used for a variety of cancer types, including breast cancer[233]. The pathologic prognostic stage groups are applicable to all patients for which the first treatment option is complete surgical excision, whereby a pathological report, grade, and the biomarkers ER, PR; and Her2 are included. The anatomic stage groups incorporate the tumor size (T), proximal lymph node metastasis (N), and distant metastasis (M) and are listed in table 4-7[234]. The updated and detailed classification table, that takes into account all biomarkers is available at[233].

Table 4-7: AJCC definition of primary Tumor (T), clinical Lymph Node (cN) infiltration, and Metastasis (M) anatomical staging (adapted from[234]).

T category	T criteria
T0	No evidence of primary tumor
Tis (DCIS)	Ductal carcinoma in situ (DCIS)
T1	Tumor ≤ 20 mm in greatest dimension
T2	Tumor > 20 mm but ≤ 50 mm in greatest dimension
T3	Tumor > 50 mm in greatest dimension
T4	Tumor of any size with direct extension to the chest wall and/or to the skin
cN0	No regional lymph node metastases
cN1	Metastases to movable ipsilateral level I and II axillary lymph node(s)
cN2	Metastases in ipsilateral level I and II axillary lymph nodes that are clinically fixed or matted
cN3	Metastases in ipsilateral infraclavicular (level III axillary) lymph node(s) with or without level I and II axillary lymph node involvement; or in ipsilateral internal mammary lymph node(s) with level I and II axillary lymph node metastases; or metastases in ipsilateral supraclavicular lymph node(s) with or without axillary or internal mammary lymph node involvement
M0	No clinical or radiographic evidence of distant metastases
M1	Distant metastases detected by clinical and radiographic means (cM) and/or histologically proven metastases larger than 0.2 mm (pM)

Based on TNM staging, tumor grading from 1 – 4 can be determined as listed in Table 4-8.

Table 4-8: Anatomical Breast cancer staging as defined by the AJCC (N1mi: micrometastases, approximately 200 cells, > 0.2 mm, but ≤ 2.0 mm). Adapted from[234]

Grading	Stage	T	N	M
G1	1A/ 1B/1B	T1/ T0/T1	N0/ N1mi/N1mi	M0/ M0/M0
G2	2A/2A/2A/ 2B/2B	T0/T1/T2/ T2/T3	N1/N1/N0/ N1/N0	M0/M0/M0/ M0/M0
G3	3A/3A/3A/3A/ 3B/3B/3B/3C	T1/T2/T3/T3/ T4/T4/T4/Any T	N2/N2/N1/N2/ N0/N1/N2/N3	M0/M0/M0/M0/ M0/M0/M0/M0
G4	4	Any T	Any N	M1

Breast Cancer Classification Based On Molecular Features

Human breast cancer is diverse in its physiology and responsiveness to treatment. Transcriptional variation among breast cancer tumors was originally assessed by complementary DNA microarrays[235] and showed that gene expression patterns in two tumors of the same patient almost always were more similar than tumors between patients. However, variation in mRNA could be related to features of physiological variation. As sequencing-based approaches were at the time and continue to be impractical in routine clinical settings, a classification of breast cancer subtypes based on pervasive differences in gene expression patterns was proposed[235]. Based on the same cDNA microarray data of 65 tissue samples, a

separation between two subtypes clinically described as hormone receptor positive (HR+) and hormone receptor negative (HR-) was dominant. The dendrogram further showed a separation of the HR- cluster into basal-like and Her2+ subtypes, while the HR+ branch separated into "normal breast-like" and "epithelial/ER+" subtypes. These four subtypes have been termed intrinsic subtypes, while surrogate intrinsic subtypes represent the clinically relevant, histologically determined breast cancer types. The four surrogate intrinsic subtypes have been increasingly characterized over the past 20 years[236,237] and their definition is condensed into four biomarkers: estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor (Her2), and by the proliferative grading, commonly assessed by Ki-67. The four tumor biological subtypes are Luminal A-like, Luminal B-like, Her2-like, and Basal-like or Triple Negative breast cancer (TNBC). These four subtypes cover > 90% of breast cancer cases, with Luminal A-like being the most frequent, followed by Luminal B-like, Her2-like, and TNBC[238]. This classification is summarized in Table 4-9 and is essential for therapy decisions as optimal therapies and responses vary across these subtypes.

Table 4-9: Classification of surrogate intrinsic breast cancer subtypes[236] and their frequency, common histological type and prognosis[227].

	Luminal A-like	Luminal B-like Her2-	Luminal B-Like Her2+	Her2-like	TNBC
Hormone receptor status	ER+PR+ ER+PR− ER−PR+	ER+PR+ ER+PR− ER−PR+	ER+PR+ ER+PR− ER−PR+	ER−PR−	ER−PR−
Grading	G1 G2	G1 G2 G3	G1 G2 G3	G1 G2 G3	G1 G2 G3
Her2 Proliferation	Her2− Ki-67 ≤15 %	Her2− Ki-67: >15 %	Her2− Ki-67: >15 %	Her2+ high Ki-67	Her2− high Ki-67
	Further Parameters				
Frequency	60-70%	10-20%	13-15%		10-15%
Histology	NST, tubular cribriform, and classic lobular	NST, micropapillary and lobular pleiomorphic	NST and pleiomorphic	NST	NST, special type (metaplastic, adenoid cystic, medullary-like, and secretory)
Prognosis	Good	Intermediate	Intermediate	Intermediate	Poor

Clinically, immune-histochemistry (IHC) is used to determine ER and PR status, Her2 expression is reported by IHC or *in situ* hybridization assays, while Ki-67 is assessed by immunostaining[237]. The Luminal A-like subtype is characterized by the presence of hormone receptors, either ER+ or PR+ or both. Furthermore, it is Her2 negative, has a low proliferation status and only includes tumors of G1 and G2 grading. The distinguishing parameter between the Luminal A-like and Her2 negative Luminal B-like subtypes is an increased Ki-67 cut-off > 15% and the inclusion of G3 grade tumors. However, this differentiation is problematic, due to the lack of a clinical standard operating procedure for the assessment of Ki-67, resulting in a low interlaboratory and interstudy comparability[239]. However, the clinical differentiation persists, but needs to be considered cautiously. The Her2 like subtype is characterized by the expression of Her2, while no further cut-off is applied to all remaining parameters. The TNBC subtype is characterized by the absence of hormone receptors and Her2, as well as any grading and any Ki-67.

THERAPY OPTIONS

Therapy options for breast cancer are a combination of surgical resection, radiation therapy, systemic therapy, and targeted therapy. Surgery of the primary tumor and affected lymph nodes is the main pillar of curative breast cancer treatment. Based on the tumor burden and its molecular expression profile, patients are given postoperative radiation therapy and/or systemic therapies comprising endocrine therapy, chemotherapy, targeted therapy, and bone-modifying agents. Tumor down-sizing with systemic therapies prior to surgery is encouraged for large tumors, if it is the same systemic therapy that is recommended post-surgery[227].

Radiation therapy following surgery has been shown to improve disease-free overall survival for patients with early breast cancer with lymph node involvement, but is also given to patients in breast-conserving therapy as it eliminates residual tumor cells[240] or induces an abscopal effect[241]. The proportional reduction of ipsilateral tumor recurrence with radiation therapy after surgery is for most indications ~75%[242]. As breast cancer is an 'immunologically cold' tumor, the microenvironment might be primed to stimulate an immune response through radiation therapy[243].

Chemotherapy can be given before (neoadjuvant) or after (adjuvant) surgery with equivalent outcomes[244]. Adjuvant endocrine therapy and adjuvant chemotherapy can decrease breast cancer mortality by approximately one third, independently of each other[245,246].

In luminal early breast cancer (Luminal A-like and Luminal B-like), adjuvant endocrine therapy is standard for at least 5 years after surgery. Tamoxifen is a common endocrine agent that binds to and inhibits ER and is standard therapy in premenopausal patients. High-risk premenopausal patients receive adjuvant chemotherapy and tamoxifen, while the addition of an ovarian suppressor such as a gonadotropin-releasing (GnRH) hormone analogue that inhibits estradiol production, has been shown to improve disease free survival (DFS) and overall survival compared to tamoxifen alone[247]. In postmenopausal women with early breast cancer, endocrine therapy and an aromatase inhibitor are standard, either as monotherapies or in sequence. Aromatase inhibitors block the aromatase enzyme that catalyzes the conversion of the enone ring of androgen precursors to a phenol, thereby completing the synthesis of estrogen. The choice of the endocrine agent depends on the relapse risk, tolerability, bone health, and patient preference.

CDK4/6 inhibitors together with endocrine therapy, have become the preferred treatment option in Luminal B-like (hormone receptor positive, Her2-negative) metastatic breast cancer[227]. Currently, four large clinical trials are evaluating the addition of a CDK4/6 inhibitor to endocrine therapy for 2-3 years in patients with Luminal B-like high-risk early breast cancer: PALLAS (NCT02513394), monarchE (NCT03155997), NATALEE (NCT03701334) and ADApPTcycle (EudraCT 2018-003749-40). Adjuvant chemotherapy in addition to endocrine therapy is added in Luminal B-like cancers depending on the individual risk of recurrence. Standard chemotherapy is based on anthracycline-taxane regimens, given preferentially in sequence to reduce excessive toxicity[248].

In Her2+ early breast cancer (Luminal B-like and Her2-like), neoadjuvant chemotherapy together with dual Her2 blockade with trastuzumab and pertuzumab improves rates of pathological complete response (pCR) and is considered standard[249]. Trastuzumab and pertuzumab are humanized monoclonal antibodies that bind to different domains of the Her2 molecule, inhibiting its dimerization. Both antibodies induce antibody-dependent cell-mediated toxicity. Therapy in Luminal B-like and Her2-like breast cancer tumors consists of chemotherapy either an anthracycline-taxane sequence or a combination of docetaxel and carboplatin together with anti-Her2 therapy for 1 year. Addition of neratinib, a tyrosine kinase inhibitor of Her2 and EGFR, leads to improved DFS. The standard for patients without pCR has shifted towards T-DM1, a trastuzumab–emtansine conjugate that combines HER2-blockade with the cytotoxic agent DM1[249].

In Triple-Negative breast cancer, neoadjuvant chemotherapy typically consisting of anthracycline-taxane or docetaxel-cyclophosphamide regimens.

Additionally, bone-modifying agents such as bisphosphonates or the RANK-L antibody Denosumab have been shown to have multiple positive effects: they improve bone mineral density and decrease treatment-related bone loss. Denosumab has been shown to be effective in lowering fracture rates in postmenopausal patients receiving aromatase inhibitors[250], and its use may also improve DFS in these patients, but not in a more general breast cancer population[250].

IMMUNOTHERAPY IN BREAST CANCER

Antibody-based therapies targeting antigens overexpressed in breast cancer such as Her2 or TROP-2 have been FDA approved and are already being used in a routine clinical setting. A brief overview of these is listed in Table 4-10.

Table 4-10: FDA-approved immunotherapeutic drugs that are part of the current standard of care, accessible at https://www.fda.gov/.

Immunotherapeutic drug	Indication
Pertuzumab Perjeta®	a monoclonal antibody that targets the Her2 pathway; approved for patients with Her2-positive breast cancer
Sacituzumab govitecan-hziy Trodelvy™	an antibody-drug conjugate that targets the TROP-2 and is conjugated to SN-38 via a cleavable linker; approved for patients with refractory TNBC, who received at least two prior therapies for metastatic disease
Trastuzumab Herceptin®	a monoclonal antibody that targets the Her2 pathway; approved for Her2-positive breast cancer
Trastuzumab deruxtecan Enhertu™	an antibody-drug conjugate composed of an anti-Her2 antibody, a cleavable linker, and a cytotoxic topoisomerase I inhibitor; approved for unresectable or metastatic Her2-positive breast cancer after two or more prior anti-Her2-based regimens in the metastatic setting.
Trastuzumab emtansine Kadcyla®	an antibody-drug conjugate, consisting of an anti-Her2 antibody covalently linked to the cytotoxic agent DM1; approved for subsets of patients with Her2-positive breast cancer
Atezolizumab Tecentriq®	a checkpoint inhibitor that targets the PD-1/PD-L1 pathway; approved in combination with the chemotherapy Abraxane® (nab-paclitaxel) for subsets of patients with advanced TNBC

Immunogenicity and T cell infiltration in breast cancer varies across molecular subtypes but is highest in TNBC and Her2+ tumors[251]. TILs from TNBC and Her2+ tumors are largely composed of $CD8^+$ T cells and $CD4^+$ T cells and to a smaller portion of B- and NK cells[252]. Also, TIL infiltration of the tumor or detection in the peripheral blood is a prognostic marker for improved survival[253]. Therefore, during the 2019 St. Gallen/Vienna consensus meeting, 66% of the panelists recommended routine characterization and reporting TILs[254]. Considerable efforts are being currently invested into expanding the anti-tumor T cell repertoire in a therapeutic setting, through adoptive T cell transfer strategies. A few ongoing phase I clinical studies have been initiated and are ongoing:

1. Phase I clinical study (NCT02792114) evaluating safety and tolerability of mesothelin-specific CAR T therapy in metastatic breast cancer patients.

2. Phase I clinical study evaluating feasibility, safety, and preliminary efficacy of CAR-T therapy with autologous T cells that express MET chimeric antigen receptors in TNBC and melanoma patients (NCT03060356).
3. Phase I clinical study (NCT02070406) evaluating toxicity, feasibility, and antitumor activity of ipilimumab + NY-ESO-1-specific T cells.

The involvement of T cells in the tumor microenvironment suggests that breast cancer could be an immunologically responsive tumor type. Beside adoptive T cell therapies, therapeutic peptide vaccination could boost preexisting T cell responses as observed in the GAPVAC study[66]. The difficulty still resides in defining actionable T cell targets. Therefore, we performed HLA ligandome analysis on a selection of 31 primary breast cancer tumor samples spanning different molecular subtypes, 23 adjacent benign samples from the same patients, and 6 healthy mamma tissue samples from primarily breast reduction surgeries. Comparative profiling of breast cancer-related immunopeptidomes with benign mamma tissue and the multi-tissue dataset detailed in Results Part I will reveal a set of TAAs that will be further validated.

4.3.3 Materials and Methods

Deviations from the HLA ligand isolation and data acquisition protocols

HLA ligands were isolated via HLA immunoaffinity purification and sequenced by LC-MS/MS. Details of the experimental procedure and data acquisition are described in the Results Part I, Materials and Methods, subheading: "HLA immunoaffinity purification" and "Mass Spectrometric Data Acquisition". Further patient characteristics, such as age, tumor grading, subtype and sample mass used as input material for the isolation of HLA ligands are described in Table 10-1. Overall, three technical replicates in DDA mode were acquired for each sample. The resulting technical LC-MS/MS runs were bundled per sample and database search was performed sample-wise using SEQUEST HT embedded in the Thermo Proteome Discoverer Version 1.4.0.288. A sample-wise PSM-level FDR of 5% was employed for HLA-I immunopeptidomes for both peptides and source proteins. We employed a sample-wise PSM-level FDR of 1% for HLA-II immunopeptidomes. As HLA-I ligands undergo further downstream quality control through their annotation to a corresponding HLA allotype, we estimate that false positive annotations would accumulate in peptides predicted as non-binders, as these peptides were not included in downstream analyses, allowing for the relaxed FDR threshold.

HLA Typing and Binding Prediction

HLA typing has not been performed by the time of this analysis for 21 out of 37 samples (31 MaCa, 6 MaN), we predicted the HLA-I allotypes of the respective samples using an in-house tool. This tool has been designed by Michael Ghosh and is based on HLA-I allotype-specific peptides, that have been identified only on samples with a certain HLA allotype, and never on samples negative for the respective HLA allotype. A threshold of at least two allotype-specific peptides is indicative of a positive *in silico* typing. The output was a two-digit typing, and the four-digit typing was derived from the highest-frequent HLA allele in the German population (according to the German pop 8 (39,689) cohort at allelefrequencies.net downloaded on 21.07.2020).

HLA binding prediction was performed with netMHCpan4.0 and SYFPEITHI for HLA-I ligands. As the HLA-II typing has not been performed at the time of analysis, no HLA-II binding prediction was performed.

4.3.4 Results and Discussion

Patient cohort characterization

In this study, we analyzed 38 mamma carcinoma (MaCa) samples, 9 (24%) patients were premenopausal with a mean age of 47 years (range: 43 years – 50 years), 29 (76%) patients were postmenopausal with a mean age of 71 years (range: 54 years - 88 years). The most common type of grading was G2 in premenopausal (n = 7, 78%) and G3 in postmenopausal (n = 11, 38%) women. The histological type for all tumors in this patient cohort is invasive ductal carcinoma of non-special type. Furthermore, 15 (48%) of tumors were Luminal A-like, 11 (35%) Luminal B-

like Her2 negative, 1 (3%) Her2-like, and 4 (13%) TNBC. Further 7 MaCa patients and 4 benign mamma samples were collected, but not included in this work, due to the time limitation for LC-MS/MS analysis. This resulted in 26 MaCa samples for comparative immunopeptidome profiling. The comparative ligandome profiling was performed for all subtypes together, surmounting to 31 different tumor samples. A further separation and immunohistochemical (IHC) classification of the tumor types and their frequency in this patient cohort is summarized in Table 4-11.

Table 4-11: IHC classification of breast cancer subtypes in the analyzed patient cohort.

Parameter	Premenopausal (n=9, 24%)	Postmenopausal (n=29, 76%)	Total (n= 38)
Estrogen receptor, n (%)			
Positive	8 (89%)	24 (83%)	32 (84%)
Negative	1 (11%)	5 (11%)	6 (16%)
Progesterone receptor, n (%)			
Positive	7 (77%)	23 (79%)	30 (79%)
Negative	2 (23%)	6 (20%)	8 (21%)
Progesterone receptor, n (%)			
ER+PR+	7 (78%)	22 (76%)	29 (76%)
ER+PR-	1 (11%)	2 (7%)	3 (8%)
ER-PR+	0 (0)	1 (3%)	1 (3%)
ER-PR-	1 (11%)	4 (14%)	5 (13%)
Grading, n (%)			
G1	0 (0%)	0 (0%)	0 (0%)
G2	7 (78%)	11 (38)	18 (47%)
G3	2 (22)	18 (62)	20 (52%)
Her2 status, n (%)			
Positive	1 (11%)	2 (7%)	3 (8%)
Negative	8 (89%)	27 (93%)	35 (92%)
Ki-67 categories (%), n (%)			
0-15	5 (56%)	11 (38%)	16 (42%)
16-25	1 (11%)	3 (10%)	4 (11%)
26-35	1 (11%)	4 (14%)	5 (13%)
36-45	0 (0%)	2 (7%)	2 (5%)
>45	2 (22%)	9 (31%)	11 (29%)

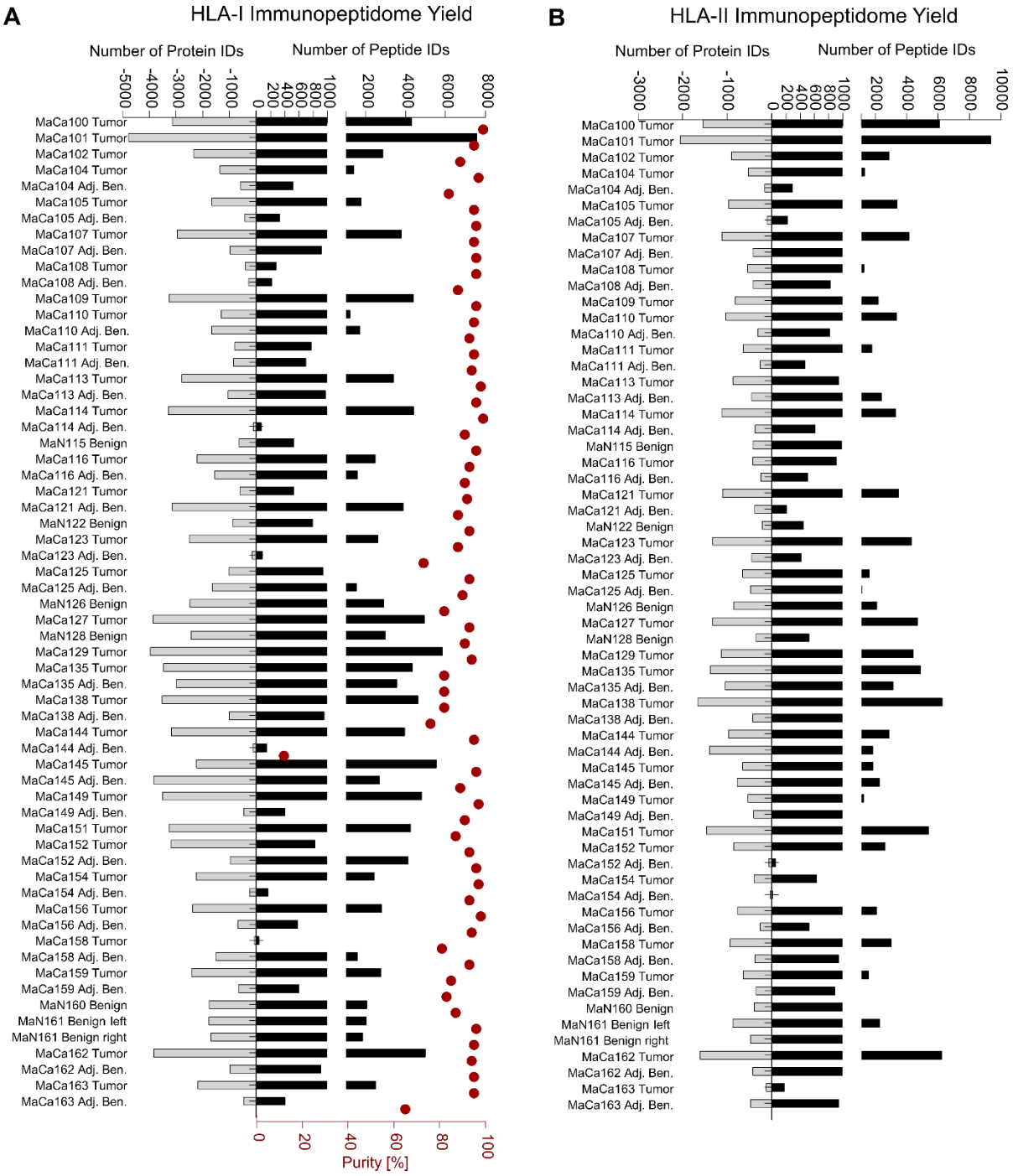


Figure 1: HLA-I and -II peptide and source protein yields.
(A) Bar plots indicate the number of identified HLA-I ligands, and the number of source proteins they can be mapped back to. The red dots represent the purity of each sample and are associated with the bottom x-axis. The purity is described by the ratio of predicted HLA binders divided by the total number of identified peptides. Binding prediction was performed with netMHCpan 4.0 integrated into the in-house tool "Ligandomat".
(B) Bar plots indicate the number of identified HLA-II peptides (black) and their corresponding source proteins (grey).

CHARACTERISTICS OF THE MACA HLA-I AND -II IMMUNOPEPTIDOME

In this work, we identified 42,376 HLA-I ligands which could be ascribed to 10,977 source proteins. The median HLA-I immunopeptidome yield per sample was 1811 (range: 37 - 7567). The median purity of each MaCa sample, as defined by the ratio between assigned HLA binders and all identified peptides amounted to 93% (range 18% - 99%). Overall, 46,939 HLA-II eluted peptides were identified that can be attributed to 5,289 source proteins, with a median HLA-II immunopeptidome yield per sample of 1285 (range 13 - 9363). Figure 1A and B illustrates the distribution of HLA-I and -II immunopeptidomes and the corresponding source proteins across all samples. Our patient cohort covers 39 HLA-I and 18 HLA-II allotypes. Nevertheless, typing had not been performed for all samples at the time of this analysis. The HLA-I allotypes were predicted with allotypic peptides as defined by Michael Ghosh for 20 samples. For these 20 samples, no information regarding the HLA-II allotypes is available, typing prediction is not possible, but allotypes must be experimentally determined.

HLA-I AND -II IMMUNOPEPTIDOME YIELDS DO NOT CORRELATE WITH SAMPLE MASS

HLA ligands are characterized by a class-specific length distribution that is slightly overlapping. The majority of HLA-I ligands are 9mers (Figure 2A), while most HLA-II ligands are 15mers and 16mers (Figure 2B). Frequently, the amount of starting material that can be used to isolate HLA ligands is limited, and it is oftentimes unclear, if increasing the input material would result in an increased immunopeptidome yield. In this study, most MaCa samples had a mass between 1 and 2 g (mean: 1.4 g, range: 0.15 g - 7.5 g) (Figure 2C and D, Table S1). However, the number of identified HLA-I and -II ligands are dispersed across different sample amounts used for their isolation. The largest input mass in this study is 7.5 g, yielding 2,905 HLA-I and 2,093 HLA-II ligands, both being in the upper part of the distribution but far from the highest number of identified ligands. Considering the tissue type, we observe that tumor samples yield higher numbers of isolated HLA ligands, while benign and adjacent benign samples resulted in moderate to low immunopeptidome yields. Indeed, the reduced numbers of identified HLA-II ligands per sample strongly correlate with tissue type, with lowest numbers for benign and adjacent benign samples. This effect might be ascribed to the cellular composition of the different tissue types. While tumor samples contain a variety of different cells, benign mamma samples were largely obtained from breast reduction surgery, and therefore consist primarily of fatty tissue and adipocytes. Furthermore, tumor tissue was less fatty in nature and might have an increased immune cell infiltration, leading to higher numbers of identified HLA-II ligands. The lower immunopeptidome yield from benign and adjacent benign tissues might be explained by the presumably lower HLA expression on adipocytes. Furthermore, fatty tissue reduces the efficacy

of tissue homogenization lysis and subsequent immunoprecipitation, and might lead to an increased loss of HLA ligands throughout the experimental workflow.

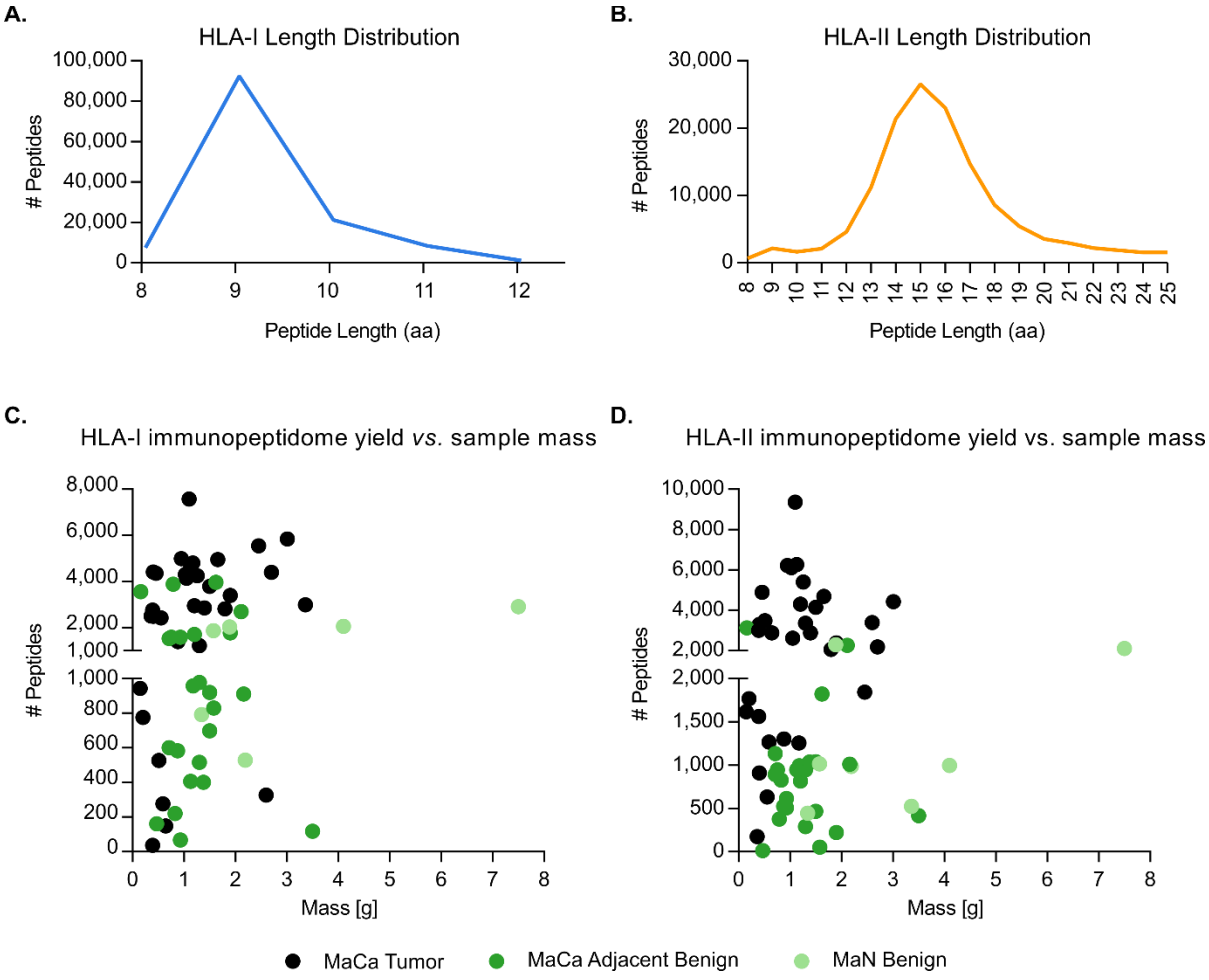


Figure 2: HLA-I and -II immunopeptidomics characteristics.
(A) – (B) Line plot illustrating the length distribution of HLA-I (blue) and -II (orange) eluted ligands, respectively. HLA-I ligands are binders according to SYFPEITHI and netMHCpan 4.0. HLA-II ligands have not been associated to their respective HLA allotypes, indicating the raw length distribution, as identified *via* LC-MS/MS.
(C) – (D) The scatter plot illustrates the relationship between the sample mass used for the HLA immunoaffinity chromatography and the number of identified HLA-I and -II ligands, respectively. The dots illustrate different samples, while the color coding highlights the dignity of each sample: tumor, adjacent benign and benign.

Definition of MaCa-associated HLA-I Ligands and Source proteins

We set out to determine TAAs in mamma carcinoma by comparatively profiling the immunopeptidomes of 31 MaCa samples with the immunopeptidomes of their corresponding adjacent benign tissues and a carefully curated benign dataset of HLA-I ligands. The benign sample cohort comprises the multi-tissue ligandome from the HLA Ligand Atlas (Results Part I) without testis, the set of 6 benign MaN samples, and the in house hematologic benign dataset (n = 98 samples, 54,613 HLA-I ligands) curated by Annika Nelde, version 190425. We found 11,622 tumor-associated HLA-I ligands that are presented on at least one donor (Figure 3 A and B). In fact, 80% of TAAs were uniquely presented on one MaCa sample, while 44 HLA-I TAAs showed representation frequencies > 15% (5 out of 31 MaCa samples). These 44 HLA-I TAAs are listed in Table 10-2, together with the samples they were found on, the corresponding gene name, HLA allotype, the number of further positive malignant samples, and the corresponding molecular subtypes of the primary tumor.

Furthermore, the absolute numbers that overlap between the three cohorts MaCa, adjacent benign and benign on HLA-I ligand level are illustrated by the size proportional Venn diagram (Figure 3B). HLA-I ligands eluted from adjacent benign tissues are shared to a large degree with the MaCa cohort, but to an even higher extent (when related to the immunopeptidome of adjacent benign tissue) with the benign dataset. Overall, these similarities can be based on the proximity of the tissues and HLA allotypes *in vivo* (tumor & adjacent benign), as well as the possibility of an overarching tumor microenvironment in histologically benign tissues.

To further prioritize the 11,622 TAAs, we queried them against our in-house database of malignant immunopeptidomes (version 190425). The malignant database encompasses HLA-I immunopeptidomes from 38 different solid and hematologic malignancies, that are covered by 950 samples and 287,978 HLA-I ligands. We observe that about 1,000 TAAs are unique to MaCa, while others are prevalent in other malignancies.

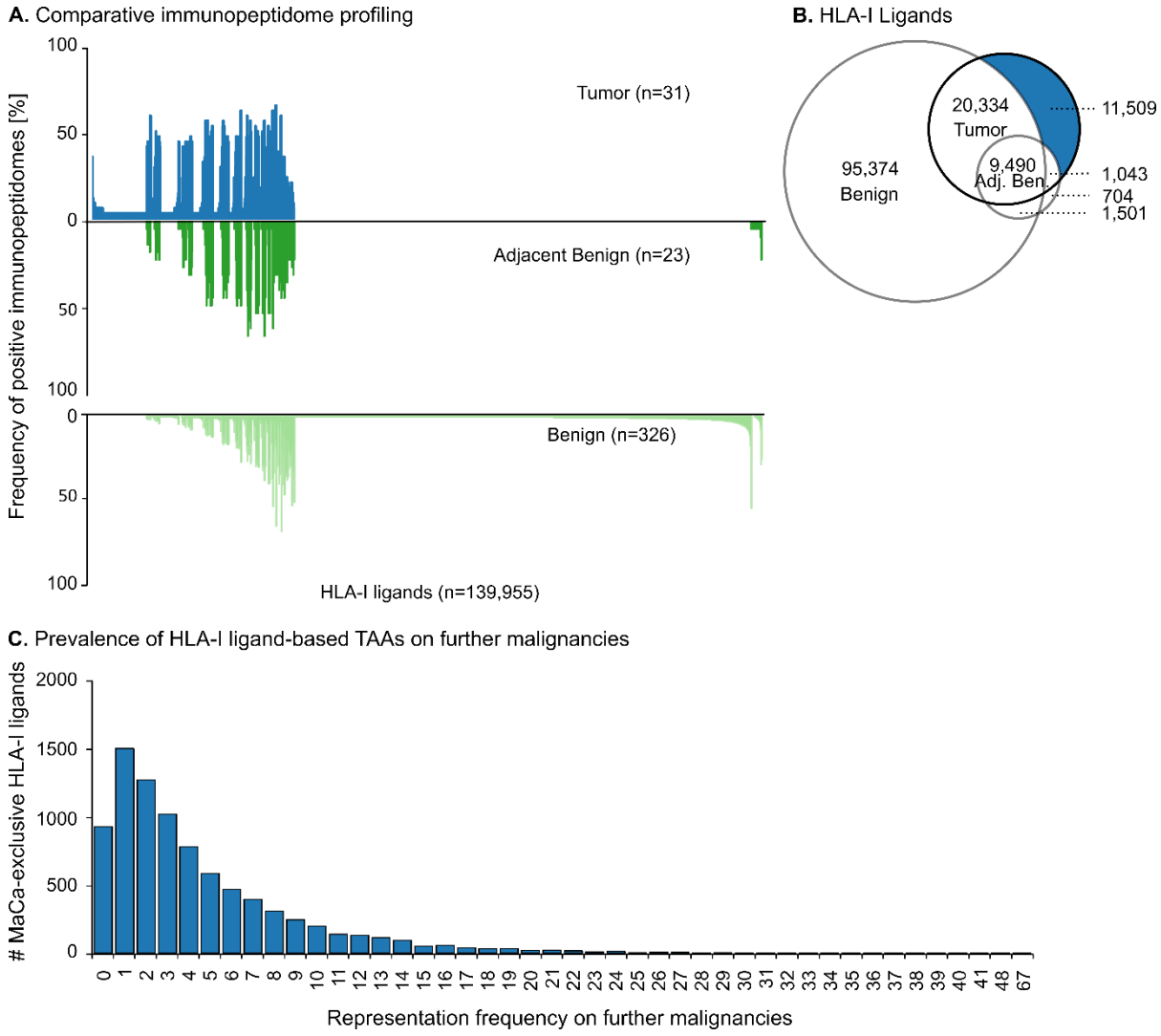


Figure 3: Comparative immunopeptidome profiling based on HLA-I ligands.
(A) Waterfall plot depicting the frequency of individual HLA-I ligands in their respective patient cohorts. A total of 137,957 HLA-I ligands are plotted on the x-axis, while their frequency in the respective population is illustrated on the y-axis. The frequency of these ligands in the 31 tumor samples is illustrated in blue, their frequency in the adjacent benign cohort (n=23) in dark green, while their frequency on benign tissues (n=326) is illustrated in light green. The data was sorted according to the decreasing frequency in tumor, adjacent benign and benign cohorts.
(B) The size-proportional Venn diagram offers an overview of the sample overlap in the three patient cohorts. Overall, 11,622 HLA-I ligands (blue color-coding) have been identified solely on MaCa.
(C) TAAs described in panel A and B have been queried against the in-house dataset of malignant immunopeptidomes, consisting of 950 samples, from 38 malignant entities and 287,978 HLA-I ligands. The y-axis shows the number of MaCa-TAAs, while the x-axis shows their prevalence on further malignant samples. For example, 1500 HLA-I TAAs have been found on one other malignant sample from the in-house dataset.

Similarly, comparative immunopeptidome profiling can be performed based on the source proteins of HLA-I ligands, to circumvent the high heterogeneity between samples based on HLA allotype mismatches (Figure 4). Multi-accession annotations may originate from conserved regions, such as certain protein domains, or could be false annotations. In this approach, we

allowed multi-accession annotations and performed a manual curation of the top scoring MaCa-associated source proteins to verify if they originate from the same protein groups. Overall, 747 HLA-I source proteins have been identified exclusively on MaCa (Figure 4A and B). Of these, 288 peptide sequences could be attributed to only one gene name, the majority, were multi-accession annotations.

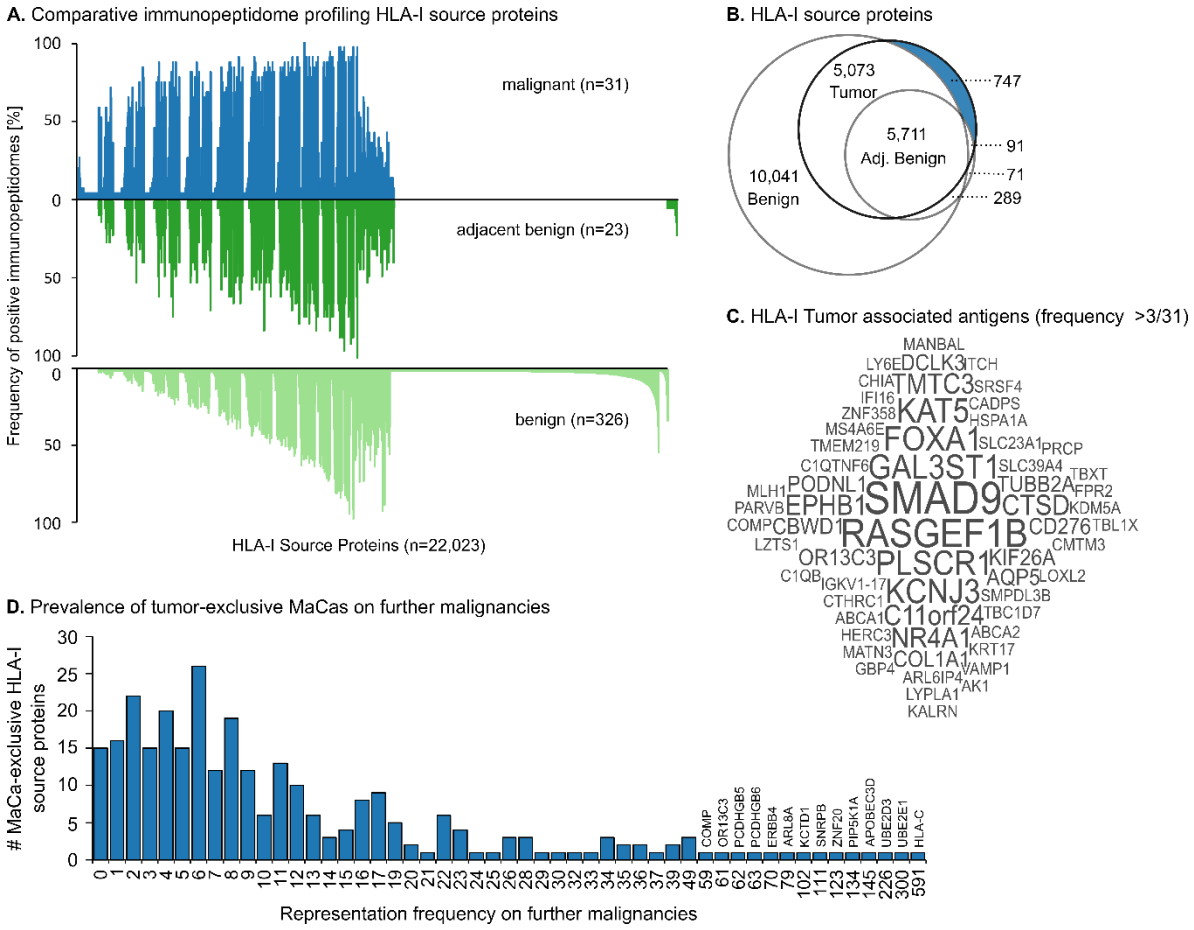


Figure 4: Comparative immunopeptidome profiling based on HLA-I source proteins

(A) Waterfall plot depicting the frequency of individual HLA-I source proteins in their respective patient cohorts. A total of 19,131 HLA-I source proteins are plotted on the x-axis, while their frequency in the respective population is illustrated on the y-axis. The frequency of these source proteins in the 31 tumor samples is illustrated in blue, their frequency in the adjacent benign cohort (n=23) in dark green, while their frequency on benign tissues (n=326) is illustrated in light green. The data was sorted according to the decreasing frequency in tumor, adjacent benign and benign cohorts.

(B) The Venn diagram offers an overview of the sample overlap in the three patient cohorts. Overall, 747 HLA-I source proteins (blue color-coding) have been identified solely on MaCa.

(C) The word cloud illustrates the MaCa source proteins identified on more than 3 samples. Their size is proportional to the number of positive MaCa samples.

(D) TAAs described in panel A and B have been queried against the in-house dataset of malignant immunopeptidomes, consisting of 950 samples, from 38 malignant entities. The y-axis shows the number of MaCa-TAAs, while the x-axis shows their prevalence on further malignant samples. For example, 16 source proteins have been found on one other malignant sample from the in-house dataset.

In line with the immunopeptide profiling performed on HLA-I ligand level, source proteins show a high similarity between MaCa adjacent benign and tumor samples (Figure 4B). Of the 747 MaCa-associated source proteins, 62 were shared between more than 3 MaCa samples, with SMAD9, SMAD9L being presented by 8 samples, followed by RASGEF1B, RASGEF1C (Figure 1C). Further details, such as the associated peptide/peptides, the corresponding HLA allotype, as well as their frequency on other malignancies are listed in Table 10-3. These 62 frequent TAAs have been identified on a multitude of further malignant samples (Figure 4D). Interestingly, these 62 top MaCa-associated source proteins and their ligands were not exclusively presented on one molecular breast cancer subgroup, but rather by multiple, if not all subgroups (Luminal A-like, Luminal B-like, TNBC).

Definition of MaCa-Associated HLA-II Ligands and Source Proteins

Equivalently, a systemic immune response against any antigen requires both a $CD4^+$ and a $CD8^+$ T cell response[92]. First hints indicate that after peptide vaccination $CD4^+$ T cell responses were frequently elicited first, while vaccine-induced $CD8^+$ T cell responses appeared later[255]. Other peptide vaccination studies similarly showed predominantly $CD4^+$ T cell responses in tumor[66,67,130]. Nevertheless, HLA-II ligands are less well characterized, as they pose a series of technical challenges: 1) the HLA-II binding groove is open and can accommodate different length variants, with the same core sequence; 2) the α and β chains of HLA-DP and HLA-DQ are inherited codominantly and their gene products both shape the antigen binding groove; 3) pan HLA-II antibody specificities used for IP might be skewed in favor of certain HLA-II allotypes. The main consequence of these complicating features is a poor HLA-II binding prediction. Only recently, HLA-II ligandomes identified by LC-MS/MS from monoallelic cells covering 40 HLA-II allotypes have been published[90]. Indeed this dataset considerably improved the pan HLA-II binding prediction algorithm NetMHCIIpan-4.0[87].

Due to the indubitable relevance of HLA-II ligands in a coordinated immune response, we comparatively analyzed HLA-II immunopeptidomes in MaCa, adjacent benign, and benign tissues on both HLA-II ligand (Figure 5) and source protein (Figure 6) level following the same workflow as for HLA-I. By the time of this analysis 20 MaCa and MaN samples have not been typed yet and the remaining samples have a poor-resolution serological typing with only HLA-DRB and HLA-DQB having been experimentally determined. Therefore, we looked at all eluted peptides and their source proteins together, without an additional filtering for HLA binding, as performed for HLA-I ligands and their source proteins.

In result, we observe 17,460 MaCa-associated HLA-II ligands, with 14,814 being unique to one individual. The number of HLA-II TAAs that is patient-individual is high when related to the presumably lower stringency of HLA-II allotypes towards peptide motifs Nevertheless, a large fraction of HLA-II ligands form clusters of length variants (Results Part I Figure7 and[191,192]). In accordance with these observations, the 14,814 patient-individual HLA-II TAAs originate from 3960 source proteins, with a median of 3.8 peptides per protein (range 1- 196). Of these 14,814 TAAs, 1944 were associated to only one protein identification.

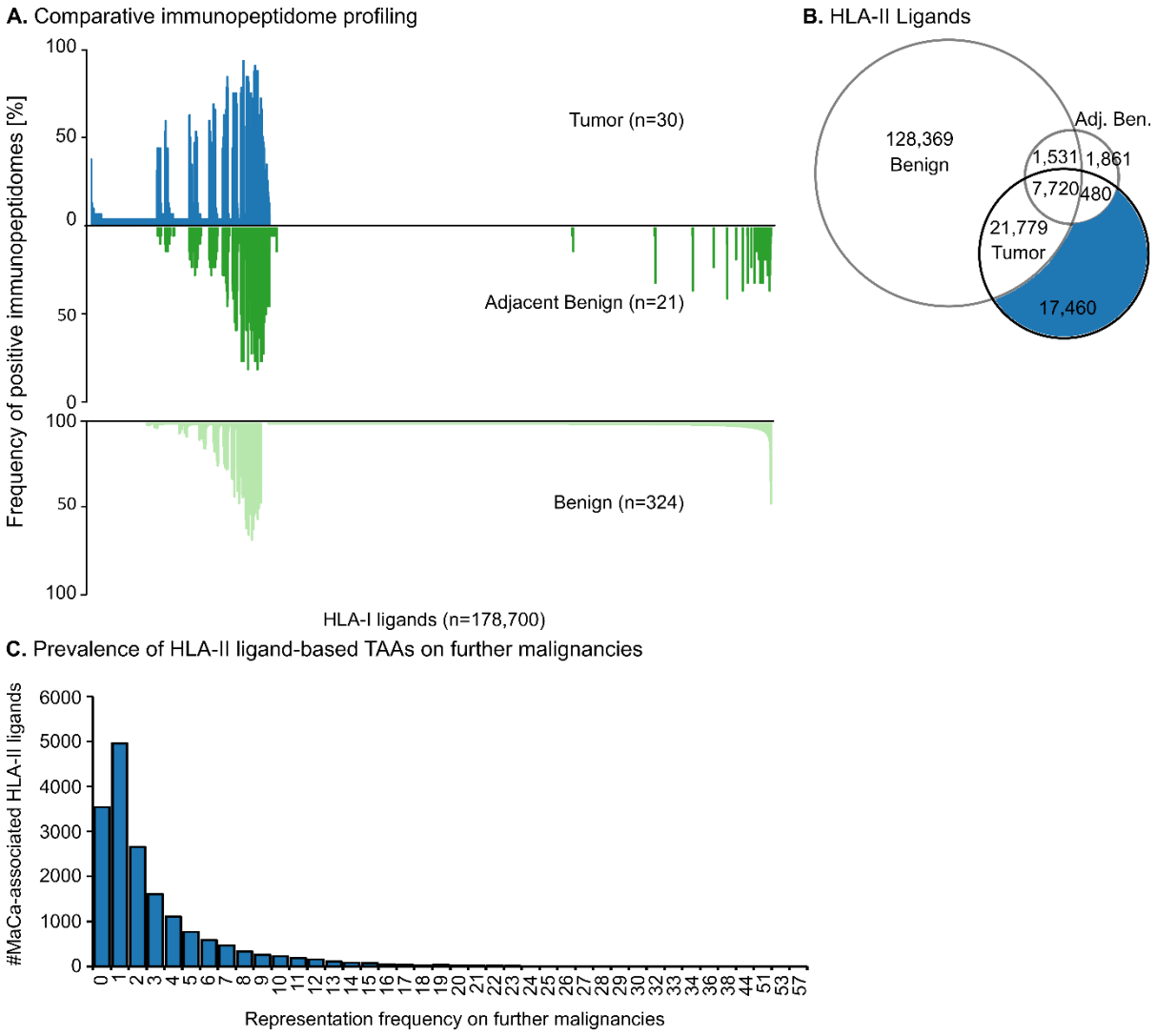


Figure 5: Comparative immunopeptidome profiling based on HLA-II eluted peptides.
(A) Waterfall plot depicting the frequency of individual HLA-II peptides in their respective patient cohorts. A total of 178,700 HLA-II peptides are plotted on the x-axis, while their frequency in the respective population is illustrated on the y-axis. The frequency of these ligands in the 30 tumor samples is illustrated in blue, their frequency in the adjacent benign cohort (n=21) in dark green, while their frequency on benign tissues (n=324) is illustrated in light green. The data was sorted according to the decreasing frequency in tumor, adjacent benign and benign cohorts.

(B) The Venn diagram offers an overview of the sample overlap in the three patient cohorts. Overall, 17,460 HLA-II peptides (blue color-coding) have been identified solely on MaCa.

(C) The prevalence of the TAAs described in panel A and B has been queried against the in-house dataset of malignant immunopeptidomes, consisting of 663 samples, from 29 malignant entities and 273,765 HLA-II ligands.

The three tissue cohorts overlap to a smaller degree than HLA-I ligands (Figure 5A and B). In addition, the representation frequency between MaCa-associated HLA-II peptides on other tumor samples is similar to that of HLA-I ligands with most peptides being unique to the MaCa samples or shared with a few other malignant samples. Table 10-4 illustrates the top 46 frequently shared (> 6/30 MaCa samples) tumor-specific HLA-II ligands.

Subsequently, comparative immunopeptidome profiling on HLA-II source protein level (without multi-accession annotations) revealed 572 MaCa-associated antigens. Of these, 75 were shared between at least 2 samples. These are listed and described in Table 10-5.

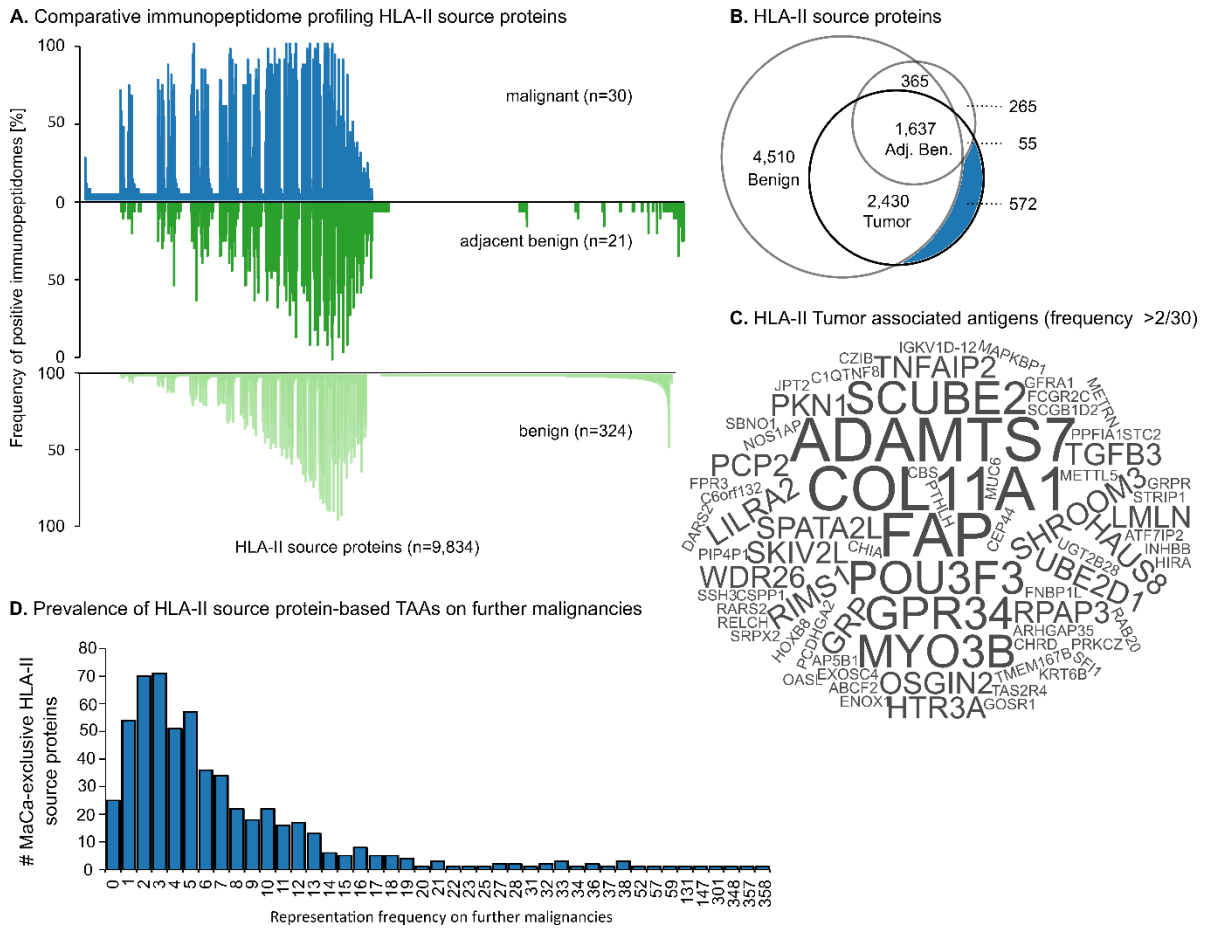


Figure 6: Comparative immunopeptidome profiling based on HLA-II source proteins.
(A) Waterfall plot depicting the frequency of individual HLA-II source proteins in their respective patient cohorts. A total of 9,834 HLA-II source proteins are plotted on the x-axis, while their frequency in the respective population is illustrated on the y-axis.

(B) The Venn diagram offers an overview of the sample overlap in the three patient cohorts. Overall, 575 HLA-II source proteins (blue color-coding) have been identified solely on MaCa.

(C) Word cloud illustrating the frequency of the MaCa-associated HLA-II source proteins, when identified in more than 2 samples.

(D) The prevalence of the TAAs described in panel A and B has been queried against the in-house dataset of malignant immunopeptidomes, consisting of 663 samples, from 29 malignant entities.

Their prevalence on further tumors is similar to that of HLA-I source proteins.

4.3.5 CONCLUSION

We comparatively profiled the HLA-I and -II immunopeptidomes from MaCa, adjacent benign and benign samples, with the purpose of defining targets for peptide vaccination in breast cancer. These MaCa associated targets could be tested in a follow-up clinical study. In this case, the type of clinical study should be defined before extensively validating potential vaccine candidates. Peptide vaccination has been tested based on the 1) one for all concept with the peptide cocktail IMA901 against renal clear cell carcinoma tailored for HLA-A*02 positive patients[256,257], 2) the warehouse concept via the APVAC1 arm in the GAPVAC clinical study[66], and 3) based on a fully individualized selection of candidate peptides, as in the APVAC2 arm in the GAPVAC study[66], or other peptide vaccination studies[67,130]. Designing a clinical study based on the warehouse concept is highly advantageous, as it enables premanufacturing of the drug product, which in this case corresponds to a set of target peptide antigens for frequent HLA allotypes[121]. Relevant premanufactured peptides are selected for each patient based on the HLA allotype, presentation of warehouse peptides on patient tumors, and perhaps immunogenic responses against warehouse peptides. For each patient, the choice of warehouse peptides and the resulting multi-peptide vaccine is individual. Nevertheless, a fully individualized peptide cocktail tailored to the immunopeptidomic landscape of the tumor, should elicit a specific T-cell response with the best safety profile. Particularly, when considering the highly individual immunopeptidomes across tissues described in Results Part I, it stands to reason that a better selection of actionable targets can be made individually.

As the ultimate proof of concept, a clinical study would be suitable to test whether peptide vaccination is beneficial for breast cancer patients. For this purpose, the target peptide selection must be adjusted to the specificities of the clinical study design that would ensue. The most interesting clinical subtype with currently the poorest prognosis and highest immunogenicity is TNBC. In this study, four patients who had not received neoadjuvant chemotherapy were recruited. Neoadjuvant chemotherapy is the standard of care in TNBC, and it leads to shrinkage and necrosis of the tumor. Unfortunately, necrotic tissue might impair the quality of the identified HLA ligandome. Following up on this study, we intend to adapt our analytic pipeline to test the similarity of the HLA ligandome prior and post chemotherapy, to evaluate if detection of warehouse peptides is possible, and perhaps if HLA loss has occurred. Oftentimes, only biopsies can be obtained from TNBC patients after neoadjuvant chemotherapy, and the lower bound of detection limit and sample mass will be determined with these experiments.

Interestingly, the preliminary results presented in this thesis show that top shared MaCa-associated source proteins and peptides are not restricted to one molecular subtype. If indeed, MaCa antigens discovered in the Luminal A-like subtype are confirmed in a validation dataset consisting of predominantly TNBC tumors, the warehouse peptide selection must not be filtered by tumor molecular subtype.

The completion of the warehouse is dependent on performing the remaining LC-MS/MS analyses, obtaining a high-resolution typing from all patient samples for reliable binding prediction, and selection of a set of frequent HLA-I and -II peptides for the warehouse. In addition, exome sequencing of blood and tumor tissue as well as RNA sequencing of tumor and adjacent benign

tissues is being performed for 10 patients. The proteogenomic investigation of breast cancer can enable the identification of neoantigens, cryptic peptides, and differentially expressed genes, thereby potentially expanding the available repertoire of target peptides for vaccination. Finally, carefully selected warehouse peptides must be tested for immunogenicity. Unfortunately, blood was not collected from the MaCa patients to allow isolation of patient-PBMCs. Thus, we will not be able to verify if patients had previously mounted an immune response against antigens presented on their tumor. Nevertheless, we can test if PBMCs from healthy volunteers can be activated to mount an immune response against relevant peptides.

5 DISCUSSION AND OUTLOOK

Immunotherapy has benefited from considerable advances over the past three decades which paved the way for numerous implementations into clinical routine. Immune checkpoint inhibitors, antibody therapy, and cellular therapies such as HSCT, CAR-T cell therapy, ACT, and peptide vaccines have each shown remarkable clinical success in a subset of patients. Nevertheless, all these therapeutic options depend on the choice of the target antigen. These must be carefully selected to be as specific to the tumor as possible, with a low expression profile on healthy tissues. Thus, on-target/off-tumor adverse events can be minimized.

Generally, HLA-dependent and HLA-independent T cell targets can be defined. Given that most proteins are expressed intracellularly (72%[48]), HLA-dependent tumor-associated or tumor-specific antigens are a highly sought after resource. For this reason, this work focuses on the discovery of tumor-associated and tumor-exclusive HLA-presented antigens. For the first time, we characterized the first draft of the ground-state, benign immunopeptidome in multiple tissues and human subjects, the HLA Ligand Atlas.

Immunopeptidomes of tumor and adjacent benign samples from the same subject have a high degree of similarity. Beside the HLA match in all 6 HLA-I loci, tissues are surgically removed from the same organ, with tumor and adjacent benign tissues being in proximity *in vivo* as well. The influence of the tumor microenvironment on the histologically benign tissue has been described on transcriptome level[118] which motivated the generation of the HLA Ligand Atlas.

Based on the HLA Ligand Atlas data, we observed that immunopeptidomes are largely patient-individual across tissues. Even when filtering for HLA-matched comparisons, a hierarchical clustering based on the Jaccard similarity index revealed that different tissue types within one subject clustered together, rather than the same tissue type across different human subjects. Nevertheless, only 13 human subjects were included in this study, for which immunopeptidomes covering multiple organs were available, and not all organs were represented in all subjects. Therefore, a higher number of human subjects would help to address the question with a higher statistical power. However, this observation poses relevant questions whether shared HLA-dependent TAAs are frequent and effective.

The Results Part III section, covering the immunopeptidome in breast cancer, adjacent benign tissues and benign mamma tissues from independent donors paints the same picture. Here most peptides identified in the adjacent benign patient cohort are shared with the mamma carcinoma cohort. A previous hypothesis stated that the high similarity between tumor and adjacent benign tissues is based on a tissue-specific effect, which has been described on the transcriptome[2,3], and on the proteome[4]. However, other proteomic studies did not see tissue-specific proteins[48,151] and observed that differences between organs are not dictated by the presence or absence of individual proteins, but rather by their abundance[115]. The scarcity of tissue-specific HLA ligands and source proteins was confirmed by a thorough search in the HLA Ligand Atlas data as well. Thus, the similarity between tumor and adjacent benign tissues on HLA ligand and source

protein-level can be rather attributed to an HLA-, and probably subject-based effect, rather than a tissue-dominated phenomenon.

Interestingly, the definition of TAAs in breast cancer patients revealed that top scoring peptides and source proteins which were highly abundant across multiple subjects were not restricted to a certain molecular subtype. Peptide YVYQNNIYL for example was identified in tumor samples from all three molecular subtypes covered in this dataset (Luminal-A like, Luminal-B like and TNBC). The focus regarding the molecular subtype that should be targeted in a putative clinical trial shifted at the end of this study from the Luminal A-like and Luminal B-like Her2 negative subtypes towards the TNBC subtype. Initially, the definition of TAAs was motivated by the high incidence of Luminal A-like and Luminal B-like cancers, and the long-term administration of the treatment. Nevertheless, TNBC is poorly treatable as no targeted therapies can be applied[236] and it is the most immunogenic subtype[251]. In this study four patients without neoadjuvant chemotherapy were included, all tumor samples resulting in a high immunopeptidome yield. Furthermore, a previous study by Ternette et al.[119] characterized the immunopeptidomic landscape of 15 HLA-A*02:01 TNBC tumors. They were able to characterize a subset of shared non-mutated TAAs but performed only a comparison to adjacent benign tissues.

Unfortunately, the data acquisition could not be finalized during this work. LC-MS/MS data acquisition for both benign mamma tissues and breast cancer – adjacent benign tissue pairs is still ongoing. Exome and RNA sequencing will enable the description of other classes of HLA-dependent TAAs, such as neoantigens, cryptic peptides, MiHAs and differentially expressed antigens. Thus, a thorough and in-depth characterization of the breast cancer immunopeptidome will ensue. Furthermore, the presence of these TAAs will be evaluated in a test dataset consisting of additional TNBCs. Further immunogenicity testing in PBMCs of healthy volunteers or breast cancer patients will confirm their suitability for inclusion in a warehouse of shared TAAs for frequent HLA allotypes.

All cellular immunotherapies and immune checkpoint inhibitors rely on a preexisting or *de novo*-induced (hopefully) tumor-specific T-cell mediated cytotoxic effect. Therefore, these warehouse peptides can be clinically tested in different formulations, such as therapeutic peptide vaccination, viral vector vaccination, engineered TCR-transduced T cells and many more.

Our efforts concentrate on designing therapeutic peptide vaccinations that can be administered together with the standard of care in the form of a subcutaneous administration of peptides in Montanide™ ISA51 and XS-15 as adjuvant[255]. The in-house GMP production facility has the approbation for mixing up to 10 different peptides per cocktail. This approach is currently being tested in several clinical studies (*e.g.* NCT02802943). Previous studies targeting non-mutated self-antigens showed a good induction of immunogenicity after peptide vaccination in glioblastoma, an immunologically cold tumor[66]. Generating a peptide cocktail for vaccination has the particular advantage, that tumor heterogeneity might rather be addressed with multi-epitope vaccines than with a single TAA. However, when using non-mutated self-peptides, the self-tolerance barrier needs to be overcome to induce a T cell response. Self-tolerance is theoretically not a problem when immunizing with neoantigens or further tumor-specific antigens. Studies

employing HLA ligands that cover somatic mutations have shown a neoantigen-specific T cell induction in clinical studies employing peptide vaccination[67,130]. Peptide-specific immune responses have been induced *in vivo* against non-mutated self-antigens[66], neoantigens[67,130], and phosphopeptides[139] as measured by antigen-specific T cell responses (ELISpot and tetramer staining) and have resulted in prolonged survival.

However, vaccine formulations can be improved regarding four key components: tumor antigens, formulations, immune adjuvants and delivery vehicles as (reviewed in[258]). Despite the increasing number of tumor antigens, their dose, route of delivery, number of administrations, interaction with other drugs concomitantly given must be optimized in humans. Comparative data on different approaches and different adjuvants are scarce. Recent vaccine strategies have shown that they can successfully increase the frequency and activity of tumor-specific T cells, but it is uncertain if these T cells infiltrate the tumor and exert their function within the tumor. Inadvertently, vaccination strategies have shed a light on the immune-suppressive and immune escape mechanisms that are active in cancer. Thus, a synergistic effect between cancer vaccination studies and immune checkpoint inhibitors might be expected to reverse the immune suppressive tumor microenvironment. Furthermore, therapeutic must address tumor immune evasion mechanisms (reviewed here[259]).

Overall, we strongly believe that peptide vaccination with personalized target antigens could be optimized to induce a long-lasting anti-tumor immune response. In this context, it is essential to expand our knowledge about the immunopeptidome in different cancers and benign tissues. Given that tumor classification evolved from a histologic perspective to a transcriptomic and molecular point of view, we envision a more immunologic-based distinction between tumor types. Thus, immuno-oncology could be optimized to address the particularities regarding tumor microenvironment, T cell infiltration but also the HLA immunopeptidomic landscape.

It is a challenging task to better capture the immunopeptidomic landscape on a population level, given the tremendous allelic polymorphism of HLA molecules, and the time and cost intensive analytical endeavor. Nevertheless, LC-MS/MS instrumentation is constantly improving, with an emphasis on speed and sensitivity, and new methods, both computational and experimental enable a constant improvement of our knowledge in each tumor.

6 ABBREVIATIONS

ACT	Adoptive cell transfer
APC	Professional antigen presenting cells
ATC	Adoptive T cell transfer
AUC	Area under the curve
CAR	Chimeric antigen receptor
CD	Cluster of differentiation
CID	Collision-induced dissociation
CLIP	Class II-associated invariant chain peptide
cN	Lymph node, clinical assessment
CTLA-4	Cytotoxic T-lymphocyte antigen 4
cT	Tumor, clinical assessment
DCs	Dendritic cell
DDA	Data dependent acquisition
DFS	Disease-free survival
DIA	Data independent acquisition
EGFR	Epidermal growth factor receptor
EMA	European Medical Agency
ER	Endoplasmic reticulum
ETD	Electron transfer dissociation
FAIMS	High-field asymmetric waveform ion mobility spectrometry
FCS	Fetal calf serum
FDR	False discovery rate
FDA	Food and Drug Administration
FWHM	Full width half maximum
GM-CSF	Granulocyte–macrophage colony-stimulating factor
GvHD	Graft vs. host disease
GvL	Graft vs. leukemia
HCD	Higher energy collisional dissociation
HLA	Human leukocyte antigen
HLA-I	Human leukocyte antigen class I
HLA-II	Human leukocyte antigen class II
HPLC	High-performance liquid chromatography
HSCT	Hematopoietic stem cell transplantation
Hz	Hertz
IFN-γ	Interferon gamma
Ig	Immunoglobulin
IL	Interleukin
IMS	Ion mobility separation
IP	Immunoaffinity purification
IRM	Ion routing multipole
IT	Ion trap
LC-MS/MS	Liquid chromatography coupled to tandem mass spectrometry
M	Metastasis
m/z	Mass to charge ratio
MiHAs	Minor histocompatibility antigens
mTEC	Medullary thymic epithelial cells
NK cells	Natural killer cells
OT	Orbitrap
PASEF	Parallel accumulation serial fragmentation
PBMC	Peripheral blood mononuclear cell
pCR	Pathological complete response
PD1	Programmed cell death protein 1
PDL1	Programmed death-ligand 1
ppm	Parts per million
PRM	Parallel reaction monitoring
Quad	Quadrupole
R	Resolving power, resolution
RiboSeq	Ribosome sequencing
RP	Reverse phase
RT	Retention time
scFv	Single chain fragment variable
SST	Standard suitability test
TAAs	Tumor-associated antigen
TAP	Transporter associated with antigen processing
TCR	T cell receptor
TFA	Trifluoroacetic acid
TILs	Tumor-infiltrating lymphocytes
TIMS	Trapped ion mobility spectrometry
TMB	Tumor mutational burden
TNBC	Triple negative breast cancer
TOF	Time of flight
UVPD	Ultraviolet photodissociation
VEGF	Vascular endothelial growth factor
WES	Whole exome sequencing

7 ACKNOWLEDGEMENTS

I want to thank Stefan and Hans-Georg for giving me the freedom to make mistakes. I learned that everything is doable even if it seems impossible at the beginning and even the toughest challenges can be fun when working together. Without their trust, I would have never engaged in such difficult endeavors. I learned that a can-do attitude and the willingness to learn by doing were all I needed to finish my PhD.

Leon Bichmann and Leon Kuchenbecker are not only collaborators but also friends. Working together has given me an insight into everything I don't know and may want to learn. They've given me asylum in their office and have showed me how effective and complementary teamwork can look like.

I want to thank everyone who keeps the lab afloat, both experimentally and bureaucratically: Beate Pömmerl, Claudia Falkenburger, Lynn Yakes, Gerhard Hörr, Carmen Höner, Franziska Löwenstein, and Georg Tiedemann who unfortunately passed away this year. I also want to thank the peptide synthesis lab for reliably supplying synthetic peptides: Uli, Patricia, Nici, Moni, Camille, Mirijam, and Marina.

I want to thank Daniel J. Kowalewski for always helping, despite not having to. In every crisis, he was reliably helpful.

Maren Lübke was my partner in crime, struggling with the same aspects of PhD life, imposter syndrome and procrastination. Together we worked late hours and had fun doing it. Lena Freudenmann grew very dear to me, and I learned to appreciate and value her unconditional help and friendship. Katja Fröhlich convinced me to do yoga, and through her humorous and smart ways shed a funny light on commonplace or annoying things. She became my dearest friend as we both, independently, went through similar difficulties and managed to get the best out of it.

I want to particularly thank Timo, Ferdi, Felix, Juliane, Patricia, Tati, Annika, Michi, Moni, Jens, Lena M, Léo, and Paul, who became my friends. With you, coming daily to work was fun!

I want to thank Paul S. Dobrotă, Linus Backert, Sergiu Sîmpălean, Thea Mihu, and Teodora Ică for supporting me particularly during the last year plagued by the Corona pandemic, home office, publication stress and lab construction work. During this last year, you kept checking on me to ensure, I was working and was reasonably amused.

Without constant help and support from my parents Cornelia-Lina & Tiberiu Marcu and my sister Catinca Varga, I couldn't have left home and studied abroad. Thank you for your help, love, and trust!

Most of all I want to thank Clemens for unconditionally believing in me, even when I didn't. Without you, this thesis wouldn't have been possible. Through your scientific and emotional support, I was able to work resiliently, without losing focus of what is most important to me. I am very grateful for having you by my side as a friend and partner!

8 REFERENCES

1. Zhang J, Wang L. The Emerging World of TCR-T Cell Trials Against Cancer: A Systematic Review. *Technol Cancer Res Treat*. 2019;18:1533-0346.
2. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ, Johnson R, Segrè A V., Djebali S, Niarchou A, Wright FA, Lappalainen T, Calvo M, Getz G, Dermitzakis ET, Ardlie KG, Guigó R. The human transcriptome across tissues and individuals. *Science (80-)*. 2015;348(6235):660-665.
3. Ardlie KG, DeLuca DS, Segrè A V., Sullivan TJ, Young TR, Gelfand ET, Trowbridge CA, Maller JB, Tukiainen T, Lek M, Ward LD, Kheradpour P, Iriarte B, Meng Y, Palmer CD, Esko T, Winckler W, Hirschhorn JN, Kellis M, MacArthur DG, Getz G, Shabalin AA, Li G, Zhou YH, Nobel AB, Rusyn I, Wright FA, Lappalainen T, Ferreira PG, Ongen H, Rivas MA, Battle A, Mostafavi S, Monlong J, Sammeth M, Melé M, Reverter F, Goldmann JM, Koller D, Guigó R, McCarthy MI, Dermitzakis ET, Gamazon ER, Im HK, Konkashbaev A, Nicolae DL, Cox NJ, Flutre T, Wen X, Stephens M, Pritchard JK, Tu Z, Zhang B, Huang T, Long Q, Lin L, Yang J, Zhu J, Liu J, Brown A, Mestichelli B, Tidwell D, Lo E, Salvatore M, Shad S, Thomas JA, Lonsdale JT, Moser MT, Gillard BM, Karasik E, Ramsey K, Choi C, Foster BA, Syron J, Fleming J, Magazine H, Hasz R, Walters GD, Bridge JP, Miklos M, Sullivan S, Barker LK, Traino HM, Mosavel M, Siminoff LA, Valley DR, Rohrer DC, Jewell SD, Branton PA, Sobin LH, Barcus M, Qi L, McLean J, Hariharan P, Um KS, Wu S, Tabor D, Shive C, et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science (80-)*. 2015;348(6235):648-660.
4. Jiang L, Wang M, Lin S, Jian R, Li X, Chan J, Fang H, Dong G, Tang H, Snyder MP, Consortium Gte, Tang H, Snyder MP. A Quantitative Proteome Map of the Human Body. *bioRxiv*. Published online 2019:797373.
5. Decker WK, da Silva RF, Sanabria MH, Angelo LS, Guimarães F, Burt BM, Kheradmand F, Paust S. Cancer immunotherapy: Historical perspective of a clinical revolution and emerging preclinical animal models. *Front Immunol*. 2017;8.
6. Dobosz P, Dzieciątkowski T. The Intriguing History of Cancer Immunotherapy. *Front Immunol*. 2019;10(December):2965.
7. McCarthy EF. The toxins of William B. Coley and the treatment of bone and soft-tissue sarcomas. *Iowa Orthop J*. 2006;26:154-158.
8. Isaacs A, Lindenmann J, Andrewes CH. Virus interference. I. The interferon. *Proc R Soc London Ser B - Biol Sci*. 1957;147(927):258-267.
9. Graham JB, Graham RM. The Effect of vaccine on cancer patients. *Plast Reconstr Surg*. 1959;24(5).
10. Miller JFA, Mitchell GF, Weiss NS. Cellular Basis of the Immunological Defects in Thymectomized Mice. *Nature*. 1967;214(5092):992-997.
11. Parkman R. The application of bone marrow transplantation to the treatment of genetic diseases. *Science (80-)*. 1986;232(4756):1373 LP - 1378.
12. Burnet M. Cancer: a biological approach. III. Viruses associated with neoplastic conditions. IV. Practical applications. *Br Med J*. 1957;1(5023):841-847.
13. Thomas L. On immunosurveillance in human cancer. *Yale J Biol Med*. 1982;55(3-4):329-333.
14. Stutman O. Tumor Development after 3-Methylcholanthrene in Immunologically Deficient Athymic-Nude Mice. *Science (80-)*. 1974;183(4124):534 LP - 536.
15. Dunn GP, Old LJ, Schreiber RD. The Three Es of Cancer Immunoediting. *Annu Rev Immunol*. 2004;22(1):329-360.
16. Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: from immunosurveillance to tumor escape. *Nat Immunol*. 2002;3(11):991-998.
17. Dunn GP, Old LJ, Schreiber RD. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity*. 2004;21(2):137-148.
18. van der Bruggen P, Traversari C, Chomez P, Lurquin C, De Plaen E, Van den Eynde B, Knuth A, Boon T. A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. *Science (80-)*. 1991;254(5038):1643-1647.
19. Howell WM, Carter V, Clark B. The HLA system: immunobiology, HLA typing, antibody screening and crossmatching techniques. *J Clin Pathol*. 2010;63(5):387—390.
20. Abbas AK, Lichtman AH, Pillai S. *Cellular and Molecular Immunology*. Saunders/Elsevier; 2010.
21. Rammensee H-G. Chemistry of peptides associated with MHC class I and class II molecules. *Curr Opin Immunol*. 1995;7(1):85-96.
22. Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, Wiley DC. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature*. 1994;368(6468):215-221.
23. Branchet MC, Boisnic S, Blétry O, Robert L, Charron D, Francés D. Expression of HLA class II antigens on skin fibroblasts in scleroderma. *Br J Dermatol*. 1992;126(5):431-435.

24. Erhard F, Dölken L, Schilling B, Schlosser A. Identification of the Cryptic HLA-I Immunopeptidome. *Cancer Immunol Res*. 2020;8(8):canimm.0886.2019.

25. Rock KL. A new foreign policy: MHC class I molecules monitor the outside world. *Immunol Today*. 1996;17(3):131-137.

26. Garcia KC, Degano M, Stanfield RL, Brunmark A, Jackson MR, Peterson PA, Teyton L, Wilson IA. An αβ T Cell Receptor Structure at 2.5 Å and Its Orientation in the TCR-MHC Complex. *Science (80-)*. 1996;274(5285):209 LP - 219.

27. Davis MM, Bjorkman PJ. T-cell antigen receptor genes and T-cell recognition. *Nature*. 1988;334(6181):395-402.

28. Laydon DJ, Bangham CRM, Asquith B. Estimating T-cell repertoire diversity: Limitations of classical estimators and a new approach. *Philos Trans R Soc B Biol Sci*. 2015;370(1675).

29. Birnbaum ME, Berry R, Hsiao Y-S, Chen Z, Shingu-Vazquez MA, Yu X, Waghray D, Fischer S, McCluskey J, Rossjohn J, Walz T, Garcia KC. Molecular architecture of the αβ T cell receptor–CD3 complex. *Proc Natl Acad Sci*. 2014;111(49):17576 LP - 17581.

30. Love PE, Hayes SM. ITAM-mediated signaling by the T-cell antigen receptor. *Cold Spring Harb Perspect Biol*. 2010;2(6):a002485.

31. 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.

32. Jenkins MK, Schwartz RH. Pillars Article: Antigen Presentation by Chemically Modified Splenocytes Induces Antigen-Specific T Cell Unresponsiveness In Vitro and In Vivo. J. Exp. Med. 1987. 165: 302–319. *J Exp Med*. 2009;183(7):4150 LP - 4167.

33. Vasaturo A, Di Blasio S, Peeters DGA, de Koning CCH, de Vries JM, Figdor CG, Hato S V. Clinical implications of co-inhibitory molecule expression in the tumor microenvironment for DC vaccination: A game of stop and go. *Front Immunol*. 2013;4(DEC):1-14.

34. Green JM, Noel PJ, Sperling AI, Walunas TL, Gray GS, Bluestone JA, Thompson CB. Absence of B7-dependent responses in CD28-deficient mice. *Immunity*. 1994;1(6):501-508.

35. Linsley PS, Brady W, Grosmaire L, Aruffo A, Damle NK, Ledbetter JA. Binding of the B cell activation antigen B7 to CD28 costimulates T cell proliferation and interleukin 2 mRNA accumulation. *J Exp Med*. 1991;173(3):721-730.

36. Curtsinger JM, Schmidt CS, Mondino A, Lins DC, Kedl RM, Jenkins MK, Mescher MF. Inflammatory Cytokines Provide a Third Signal for Activation of Naive CD4+ and CD8+ T Cells. *J Immunol*. 1999;162(6):3256 LP - 3262.

37. Singh AK, McGuirk JP. Allogeneic stem cell transplantation: A historical and scientific overview. *Cancer Res*. 2016;76(22):6445-6451.

38. Thomas ED, Buckner CD, Rudolph RH, Fefer A, Storb R, Neiman PE, Bryant JI, Chard RL, Clift RA, Epstein RB, Fialkow PJ, Funk DD, Giblett ER, Lerner KG, Reynolds FA, Slichter S. Allogeneic Marrow Grafting for Hematologic Malignancy Using HLA Matched Donor-Recipient Sibling Pairs. *Blood Cancer J*. 1971;38(3):267-287.

39. Lee SJ, Klein J, Haagenson M, Baxter-Lowe LA, Confer DL, Eapen M, Fernandez-Vina M, Flomenberg N, Horowitz M, Hurley CK, Noreen H, Oudshoorn M, Petersdorf E, Setterholm M, Spellman S, Weisdorf D, Williams TM, Anasetti C. High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. *Blood Cancer J*. 2007;110(13):4576-4583.

40. Garderet L, Dulphy N, Douay C, Chalumeau N, Schaeffer V, Zilber MT, Lim A, Even J, Mooney N, Gelin C, Gluckman E, Charron D, Toubert A. The umbilical cord blood αβ T-cell repertoire: Characteristics of a polyclonal and naive but completely formed repertoire. *Blood Cancer J*. 1998;91(1):340-346.

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

42. Zhang C, Liu J, Zhong JF, Zhang X. Engineering CAR-T cells. *Biomark Res*. 2017;5(1):22.

43. Sadelain M, Brentjens R, Rivière I. The basic principles of chimeric antigen receptor design. *Cancer Discov*. 2013;3(4):388-398.

44. Feins S, Kong W, Williams EF, Milone MC, Fraietta JA. An introduction to chimeric antigen receptor (CAR) T-cell immunotherapy for human cancer. *Am J Hematol*. 2019;94(S1):S3-S9.

45. Porter DL, Levine BL, Kalos M, Bagg A, June CH. Chimeric antigen receptor-modified T cells in chronic lymphoid leukemia. *N Engl J Med*. 2011;365(8):725-733.

46. Bonifant CL, Jackson HJ, Brentjens RJ, Curran KJ. Toxicity and management in CAR T-cell therapy. *Mol Ther - Oncolytics*. 2016;3:16011.

47. Arndt C, Bachmann M, Bergmann R, Berndt N, Feldmann A, Koristka S. Theranostic CAR T cell targeting: A brief review. *J Label Compd Radiopharm*. 2019;62(8):533-540.

48. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson IM, Edlund K, Lundberg E, Navani S, Szigyarto CA-KK, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist P-HH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, Von Feilitzen K, Forsberg M, Persson L,

Johansson F, Zwahlen M, von Heijne G, Nielsen J, Pontén F. Tissue-based map of the human proteome. *Science (80-)*. 2015;347(6220):1260419.

49. Stone JD, Harris DT, Kranz DM. TCR affinity for p/MHC formed by tumor antigens that are self-proteins: impact on efficacy and toxicity. *Curr Opin Immunol*. 2015;33:16-22.

50. Harris DT, Hager M V, Smith SN, Cai Q, Stone JD, Kruger P, Lever M, Dushek O, Schmitt TM, Greenberg PD, Kranz DM. Comparison of T Cell Activities Mediated by Human TCRs and CARs That Use the Same Recognition Domains. *J Immunol*. 2018;200(3):1088 LP - 1100.

51. Kantoff PW, Higano CS, Shore ND, Berger ER, Small EJ, Penson DF, Redfern CH, Ferrari AC, Dreicer R, Sims RB, Xu Y, Frohlich MW, Schellhammer PF. Sipuleucel-T Immunotherapy for Castration-Resistant Prostate Cancer. *N Engl J Med*. 2010;363(5):411-422.

52. Rosenberg SA, Packard BS, Aebersold PM, Solomon D, Topalian SL, Toy ST, Simon P, Lotze MT, Yang JC, Seipp CA, Simpson C, Carter C, Bock S, Schwartzentruber D, Wei JP, White DE. Use of Tumor-Infiltrating Lymphocytes and Interleukin-2 in the Immunotherapy of Patients with Metastatic Melanoma. *N Engl J Med*. 1988;319(25):1676-1680.

53. Met Ö, Jensen KM, Chamberlain CA, Donia M, Svane IM. Principles of adoptive T cell therapy in cancer. *Semin Immunopathol*. 2019;41(1):49-58.

54. Dudley ME, Wunderlich JR, Yang JC, Sherry RM, Topalian SL, Restifo NP, Royal RE, Kammula U, White DE, Mavroukakis SA, Rogers LJ, Gracia GJ, Jones SA, Mangiameli DP, Pelletier MM, Gea-Banacloche J, Robinson MR, Berman DM, Filie AC, Abati A, Rosenberg SA. Adoptive cell transfer therapy following non-myeloablative but lymphodepleting chemotherapy for the treatment of patients with refractory metastatic melanoma. *J Clin Oncol*. 2005;23(10):2346-2357.

55. Dudley ME, Wunderlich JR, Robbins PF, Yang JC, Hwu P, Schwartzentruber DJ, Topalian SL, Sherry R, Restifo NP, Hubicki AM, Robinson MR, Raffeld M, Duray P, Seipp CA, Rogers-Freezer L, Morton KE, Mavroukakis SA, White DE, Rosenberg SA. Cancer regression and autoimmunity in patients after clonal repopulation with antitumor lymphocytes. *Science (80-)*. 2002;298(5594):850-854.

56. Rosenberg SA, Restifo NP, Yang JC, Morgan RA, Dudley ME. Adoptive cell transfer: A clinical path to effective cancer immunotherapy. *Nat Rev Cancer*. 2008;8(4):299-308.

57. Morgan RA, Dudley ME, Wunderlich JR, Hughes MS, Yang JC, Sherry RM, Royal RE, Topalian SL, Kammula US, Restifo NP, Zheng Z, Nahvi A, de Vries CR, Rogers-Freezer LJ, Mavroukakis SA, Rosenberg SA. Cancer Regression in Patients After Transfer of Genetically Engineered Lymphocytes. *Science (80-)*. 2006;314(5796):126 LP - 129.

58. Brunet J-F, Denizot F, Luciani M-F, Roux-Dosseto M, Suzan M, Mattei M-G, Golstein P. A new member of the immunoglobulin superfamily—CTLA-4. *Nature*. 1987;328(6127):267-270.

59. Krummel MF, Allison JP. CD28 and CTLA-4 have opposing effects on the response of T cells to stimulation. *J Exp Med*. 1995;182(2):459-465.

60. Fellner C. Ipilimumab (yervoy) prolongs survival in advanced melanoma: serious side effects and a hefty price tag may limit its use. *Pharm Ther*. 2012;37(9):503-530.

61. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer*. 2012;12(4):252-264.

62. Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, Gonzalez R, Robert C, Schadendorf D, Hassel JC, Akerley W, van den Eertwegh AJM, Lutzky J, Lorigan P, Vaubel JM, Linette GP, Hogg D, Ottensmeier CH, Lebbé C, Peschel C, Quirt I, Clark JI, Wolchok JD, Weber JS, Tian J, Yellin MJ, Nichol GM, Hoos A, Urba WJ. Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. *N Engl J Med*. 2010;363(8):711-723.

63. Reguzova A, Ghosh M, Müller M, Rziha H-J, Amann R. Orf Virus-Based Vaccine Vector D1701-V Induces Strong CD8+ T Cell Response against the Transgene but Not against ORFV-Derived Epitopes. *Vaccines*. 2020;8(2):295.

64. Belnoue E, Mayol J-F, Carboni S, Di Berardino Besson W, Dupuychaffray E, Nelde A, Stevanović S, Santiago-Raber M-L, Walker PR, Derouazi M. Targeting self and neo-epitopes with a modular self-adjuvanting cancer vaccine. *JCI insight*. 2019;5(11):e127305.

65. Lundstrom K. Viral Vectors in Gene Therapy. *Diseases*. 2018;6(2):42.

66. Hilf N, Kuttruff-Coqui S, Frenzel K, Bukur V, Stevanović S, Gouttefangeas C, Platten M, Tabatabai G, Dutoit V, van der Burg SH, thor Straten P, Martínez-Ricarte F, Ponsati B, Okada H, Lassen U, Admon A, Ottensmeier CH, Ulges A, Kreiter S, von Deimling A, Skardelly M, Migliorini D, Kroep JR, Idorn M, Rodon J, Piró J, Poulsen HS, Shraibman B, McCann K, Mendrzyk R, Löwer M, Stieglbauer M, Britten CM, Capper D, Welters MJP, Sahuquillo J, Kiesel K, Derhovanessian E, Rusch E, Bunse L, Song C, Heesch S, Wagner C, Kemmer-Brück A, Ludwig J, Castle JC, Schoor O, Tadmor AD, Green E, Fritsche J, Meyer M, Pawlowski N, Dorner S, Hoffgaard F, Rössler B, Maurer D, Weinschenk T, Reinhardt C, Huber C, Rammensee H-G, Singh-Jasuja H, Sahin U, Dietrich PY, Wick W. Actively personalized vaccination trial for newly diagnosed glioblastoma. *Nature*. 2019;565(7738):240-245.

67. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, Zhang W, Luoma A, Giobbie-Hurder A, Peter L, Chen C, Olive O, Carter TA, Li S, Lieb DJ, Eisenhaure T, Gjini E, Stevens J, Lane WJ, Javeri I, Nellaiappan K, Salazar AM, Daley H, Seaman M, Buchbinder EI, Yoon CH, Harden M, Lennon N, Gabriel S, Rodig SJ, Barouch DH, Aster JC, Getz G, Wucherpfennig K, Neuberg D, Ritz J,

Lander ES, Fritsch EF, Hacohen N, Wu CJ. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature*. 2017;547(7662):217-221.

68. Ott PA, Hodi FS. Talimogene Laherparepvec for the Treatment of Advanced Melanoma. *Clin Cancer Res*. 2016;22(13):3127 LP - 3131.

69. Cresswell P, Turner MJ, Strominger JL. Papain-solubilized HLA antigens from cultured human lymphocytes contain two peptide fragments. *Proc Natl Acad Sci U S A*. 1973;70(5):1603-1607.

70. Peterson PA, Rask L, Lindblom JB. Highly purified papain-solubilized HL-A antigens contain beta2-microglobulin. *Proc Natl Acad Sci U S A*. 1974;71(1):35—39.

71. Hunt DF, Henderson RA, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, Cox AL, Appella E, Engelhard VH. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science (80-)*. 1992;255(5049):1261-1263.

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

73. Rammensee H-G, Rötzschke O, Falk K. Self tolerance of natural MHC class I ligands. *Int Rev Immunol*. 1993;10(4):291-300.

74. Rötzschke O, Falk K, Wallny HJ, Faath S, Rammensee HG. Characterization of naturally occurring minor histocompatibility peptides including H-4 and H-Y. *Science (80-)*. 1990;249(4966):283-287.

75. Rötzschke O, Falk K, Deres K, Schild H, Norda M, Metzger J, Jung G, Rammensee H-G. Isolation and analysis of naturally processed viral peptides as recognized by cytotoxic T cells. *Nature*. 1990;348(6298):252-254.

76. Storkus WJ, Zen HJIII, Salter RD, Lotze MT, Zeh HJ, Salter RD, Lotze MT. Identification of T-Cell Epitopes: Rapid Isolation of Class I-Presented Peptides from Viable Cells by Mild Acid Elution. *J Immunother*. 1993;14(2):94-103.

77. Bleek GM Van, Nathenson SG. Isolation of an endogenously processed immunodominant viral peptide from the class I H–2Kb molecule. *Nature*. 1990;348(6298):213-216.

78. Bilich T, Nelde A, Bichmann L, Roerden M, Salih HR, Kowalewski DJ, Schuster H, Tsou C-CC, Marcu A, Neidert MC, Lübke M, Rieth J, Schemionek M, Brümmendorf TH, Vucinic V, Niederwieser D, Bauer J, Märklin M, Peper JK, Klein R, Kohlbacher O, Kanz L, Rammensee H-G, Stevanović S, Walz JS. The HLA ligandome landscape of chronic myeloid leukemia delineates novel T-cell epitopes for immunotherapy. *Blood Cancer J*. 2019;133(6):550-565.

79. Chong C, Marino F, Pak H, Racle J, Daniel RT, Müller M, Gfeller D, Coukos G, Bassani-Sternberg M. High-throughput and sensitive immunopeptidomics platform reveals profound interferon γ-mediated remodeling of the human leukocyte antigen (HLA) ligandome. *Mol Cell Proteomics*. 2018;17(3):533-548.

80. Bassani-Sternberg M, Bräunlein E, Klar R, Engleitner T, Sinitcyn P, Audehm S, Straub M, Weber J, Slotta-Huspenina J, Specht K, Martignoni ME, Werner A, Hein R, H. Busch D, Peschel C, Rad R, Cox J, Mann M, Krackhardt AM, Busch DH, Peschel C, Rad R, Cox J, Mann M, Krackhardt AM. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun*. 2016;7(1):13404.

81. Barnstable CJ, Bodmer WF, Brown G, Galfre G, Milstein C, Williams AF, Ziegler A. Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens-new tools for genetic analysis. *Cell*. 1978;14(1):9-20.

82. Zhang M, Fritsche J, Roszik J, Williams LJ, Peng X, Chiu Y, Tsou CC, Hoffgaard F, Goldfinger V, Schoor O, Talukder A, Forget MA, Haymaker C, Bernatchez C, Han L, Tsang YH, Kong K, Xu X, Scott KL, Singh-Jasuja H, Lizee G, Liang H, Weinschenk T, Mills GB, Hwu P. RNA editing derived epitopes function as cancer antigens to elicit immune responses. *Nat Commun*. 2018;9(1).

83. Zarling AL, Ficarro SB, White FM, Shabanowitz J, Hunt DF, Engelhard VH. Phosphorylated peptides are naturally processed and presented by major histocompatibility complex class I molecules in vivo. *J Exp Med*. 2000;192(12):1755-1762.

84. Malaker SA, Penny SA, Steadman LG, Myers PT, Loke JC, Raghavan M, Bai DL, Shabanowitz J, Hunt DF, Cobbold M. Identification of glycopeptides as posttranslationally modified neoantigens in Leukemia. *Cancer Immunol Res*. 2017;5(5):376-384.

85. Goldman JM, Hibbin J, Kearney L, Orchard K, Th'ng KH. HLA-DA monoclonal antibodies inhibit the proliferation of normal and chronic granulocytic leukaemia myeloid progenitor cell. *Br J Haematol*. 1982;52(3):411-420.

86. Pawelec G, Ziegler A, Wernet P. Dissection of human allostimulatory determinants with cloned T cells: Stimulation inhibition by monoclonal antibodies TÜ22, 34, 36, 37, 39, 43, and 58 against distinct human MHC class II molecules. *Hum Immunol*. 1985;12(3):165-176.

87. Reynisson B, Barra C, Kaabinejadian S, Hildebrand WH, Peters B, Nielsen M. Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass spectrometry MHC eluted ligand data. *J Proteome Res*. Published online April 18, 2020:1535-3893.

88. Alvarez B, Reynisson B, Barra C, Buus S, Ternette N, Connelley T, Andreatta M, Nielsen M. NNAlign-MA; MHC peptidome deconvolution for accurate MHC binding motif characterization and improved t-cell epitope predictions. *Mol Cell Proteomics*. 2019;18(12):2459-2477.

89. Chen B, Khodadoust MS, Olsson N, Wagar LE, Fast E, Liu CL, Muftuoglu Y, Sworder BJ, Diehn M, Levy R, Davis MM, Elias JE,

Altman RB, Alizadeh AA. Predicting HLA class II antigen presentation through integrated deep learning. *Nat Biotechnol*. 2019;37(11):1332-1343.

90. Abelin JG, Harjanto D, Malloy M, Suri P, Colson T, Goulding SP, Creech AL, Serrano LR, Nasir G, Nasrullah Y, McGann CD, Velez D, Ting YS, Poran A, Rothenberg DA, Chhangawala S, Rubinsteyn A, Hammerbacher J, Gaynor RB, Fritsch EF, Greshock J, Oslund RC, Barthelme D, Addona TA, Arieta CM, Rooney MS. Defining HLA-II Ligand Processing and Binding Rules with Mass Spectrometry Enhances Cancer Epitope Prediction. *Immunity*. 2019;51(51(4)):766-779.

91. Racle J, Michaux J, Rockinger GA, Arnaud M, Bobisse S, Chong C, Guillaume P, Coukos G, Harari A, Jandus C, Bassani-Sternberg M, Gfeller D. Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes. *Nat Biotechnol*. 2019;37(11):1283-1286.

92. Alspach E, Lussier DM, Miceli AP, Kizhvatov I, DuPage M, Luoma AM, Meng W, Lichti CF, Esaulova E, Vomund AN, Runci D, Ward JP, Gubin MM, Medrano RFV, Arthur CD, White JM, Sheehan KCF, Chen A, Wucherpfennig KW, Jacks T, Unanue ER, Artyomov MN, Schreiber RD. MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature*. 2019;574(7780):696-701.

93. Freudenmann LKLK, Marcu A, Stevanović S. Mapping the tumour human leukocyte antigen (HLA) ligandome by mass spectrometry. *Immunology*. 2018;154(3):331-345.

94. Caron E, Aebersold R, Banaei-Esfahani A, Chong C, Bassani-Sternberg M. A Case for a Human Immuno-Peptidome Project Consortium. In: *Immunity*. Vol 47. ; 2017:203-208.

95. Laumont CM, Daouda T, Laverdure J-P, Bonneil É, Caron-Lizotte O, Hardy M-P, Granados DP, Durette C, Lemieux S, Thibault P, Perreault C. Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames. *Nat Commun*. 2016;7(1):10238.

96. Reustle A, Di Marco M, Meyerhoff C, Nelde A, Walz JS, Winter S, Kandabarau S, Büttner F, Haag M, Backert L, Kowalewski DJ, Rausch S, Hennenlotter J, Stühler V, Scharpf M, Fend F, Stenzl A, Rammensee H-G, Bedke J, Stevanović S, Schwab M, Schaeffeler E. Integrative -omics and HLA-ligandomics analysis to identify novel drug targets for ccRCC immunotherapy. *Genome Med*. 2020;12(1):32.

97. Schuster H, Peper JK, Bösmüller HC, Röhle K, Backert L, Bilich T, Ney B, Löffler MW, Kowalewski DJ, Trautwein N, Rabsteyn A, Engler T, Braun S, Haen SP, Walz JS, Schmid-Horch B, Brucker SY, Wallwiener D, Kohlbacher O, Fend F, Rammensee HG, Stevanovic S, Staebler A, Wagner P. The immunopeptidomic landscape of ovarian carcinomas. *Proc Natl Acad Sci U S A*. 2017;114(46):E9942-E9951.

98. Cobbold M, De La Peña H, Norris A, Polefrone JM, Qian J, English AM, Cummings KL, Penny S, Turner JE, Cottine J, Abelin JG, Malaker SA, Zarling AL, Huang HW, Goodyear O, Freeman SD, Shabanowitz J, Pratt G, Craddock C, Williams ME, Hunt DF, Engelhard VH. MHC class I-associated phosphopeptides are the targets of memory-like immunity in leukemia. *Sci Transl Med*. 2013;5(203).

99. Ouspenskaia T, Law T, Clauser KR, Klaeger S, Sarkizova S, Aguet F, Li B, Christian E, Knisbacher BA, Le PM, Hartigan CR, Keshishian H, Apffel A, Oliveira G, Zhang W, Chow YT, Ji Z, Shukla SA, Bachireddy P, Getz G, Hacohen N, Keskin DB, Carr SA, Wu CJ, Regev A. Thousands of novel unannotated proteins expand the MHC I immunopeptidome in cancer. *bioRxiv*. Published online 2020:2020.02.12.945840.

100. Chong C, Müller M, Pak HS, Harnett D, Huber F, Grun D, Leleu M, Auger A, Arnaud M, Stevenson BJ, Michaux J, Bilic I, Hirsekorn A, Calviello L, Simó-Riudalbas L, Planet E, Lubiński J, Bryśkiewicz M, Wiznerowicz M, Xenarios I, Zhang L, Trono D, Harari A, Ohler U, Coukos G, Bassani-Sternberg M. Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nat Commun*. 2020;11(1):1293.

101. Laumont CM, Vincent K, Hesnard L, Audemard É, Bonneil É, Laverdure J-PP, Gendron P, Courcelles M, Hardy M-PP, Côté C, Durette C, St-Pierre C, Benhammadi M, Lanoix J, Vobecky S, Haddad E, Lemieux S, Thibault P, Perreault C. Noncoding regions are the main source of targetable tumor-specific antigens. *Sci Transl Med*. 2018;10(470):eaau5516.

102. Liepe J, Marino F, Sidney J, Jeko A, Bunting DE, Sette A, Kloetzel PM, Stumpf MPH, Heck AJR, Mishto M. A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science (80-)*. 2016;354(6310):354 LP - 358.

103. Rolfs Z, Solntsev SK, Shortreed MR, Frey BL, Smith LM. Global Identification of Post-Translationally Spliced Peptides with Neo-Fusion. *J Proteome Res*. 2018;18(1):349-358.

104. Mylonas R, Beer I, Iseli C, Chong C, Pak HH-S, Gfeller D, Coukos G, Xenarios I, Müller M, Bassani-Sternberg M. Estimating the Contribution of Proteasomal Spliced Peptides to the HLA-I Ligandome. *Mol Cell Proteomics*. 2018;17(12):2347 LP - 2357.

105. Faridi P, Li C, Ramarathinam SH, Vivian JP, Illing PT, Mifsud NA, Ayala R, Song J, Gearing LJ, Hertzog PJ, Ternette N, Rossjohn J, Croft NP, Purcell AW. A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands. *Sci Immunol*. 2018;3(28):eaar3947.

106. Schuster H, Shao W, Weiss T, Pedrioli PGAA, Roth P, Weller M, Campbell DS, Deutsch EW, Moritz RL, Planz O, Rammensee H-G, Aebersold R, Caron E. A tissue-based draft map of the murine MHC class I immunopeptidome. *Sci Data*. 2018;5(1):180157.

107. Wan X, Vomund AN, Peterson OJ, Chervonsky A V., Lichti CF, Unanue ER. The MHC-II peptidome of pancreatic islets identifies key features of autoimmune peptides. *Nat Immunol*. 2020;21(4):455-463.

108. Löffler MW, Mohr C, Bichmann L, Freudenmann LK, Walzer M, Schroeder CM, Trautwein N, Hilke FJ, Zinser RS, Mühlenbruch L, Kowalewski DJ, Schuster H, Sturm M, Matthes J, Riess O, Czemmel S, Nahnsen S, Königsrainer I, Thiel K, Nadalin S, Beckert S, Bösmüller H, Fend F, Velic A, Maček B, Haen SP, Buonaguro L, Kohlbacher O, Stevanović S, Königsrainer A, Rammensee H-G, Mayer-Mokler A, Weinschenk T, Flohr C, Reinhardt C, Singh-Jasuja H, Accolla RS, Tosi G, Forlani G, Ma YT, Adams D, Valmori D, Chaumette T, Heidenreich R, Gouttefangeas C, Sangro B, Francque S, Vonghia L, Tagliamonte M, Petrizzo A, Tornesello ML, Buonaguro FM, Consortium H. Multi-omics discovery of exome-derived neoantigens in hepatocellular carcinoma. *Genome Med*. 2019;11(1):28.

109. Bilich T, Nelde A, Bauer J, Walz S, Roerden M, Salih HR, Weisel K, Besemer B, Marcu A, Lübke M, Schuhmacher J, Neidert MC, Rammensee H-G, Stevanović S, Walz JS. Mass spectrometry-based identification of a B-cell maturation antigen-derived T-cell epitope for antigen-specific immunotherapy of multiple myeloma. *Blood Cancer J*. 2020;10(2):24.

110. Nelde A, Kowalewski DJ, Backert L, Schuster H, Werner JO, Klein R, Kohlbacher O, Kanz L, Salih HR, Rammensee H-G, Stevanović S, Walz JS. HLA ligandome analysis of primary chronic lymphocytic leukemia (CLL) cells under lenalidomide treatment confirms the suitability of lenalidomide for combination with T-cell-based immunotherapy. *Oncoimmunology*. 2018;7(4).

111. Ritz D, Gloger A, Weide B, Garbe C, Neri D, Fugmann T. High-sensitivity HLA class I peptidome analysis enables a precise definition of peptide motifs and the identification of peptides from cell lines and patients' sera. *Proteomics*. 2016;16(10):1570-1580.

112. Abelin JG, Keskin DB, Sarkizova S, Hartigan CR, Zhang W, Sidney J, Stevens J, Lane W, Zhang GL, Eisenhaure TM, Clauser KR, Hacohen N, Rooney MS, Carr SA, Wu CJ. Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction. *Immunity*. 2017;46(2):315-326.

113. Cameron BJ, Gerry AB, Dukes J, Harper J V., Kannan V, Bianchi FC, Grand F, Brewer JE, Gupta M, Plesa G, Bossi G, Vuidepot A, Powlesland AS, Legg A, Adams KJ, Bennett AD, Pumphrey NJ, Williams DD, Binder-Scholl G, Kulikovskaya I, Levine BL, Riley JL, Varela-Rohena A, Stadtmauer EA, Rapoport AP, Linette GP, June CH, Hassan NJ, Kalos M, Jakobsen BK. Identification of a titin-derived HLA-A1-presented peptide as a cross-reactive target for engineered MAGE A3-directed T cells. *Sci Transl Med*. 2013;5(197):197ra103 LP-197ra103.

114. Linette GP, Stadtmauer EA, Maus M V, Rapoport AP, Levine BL, Emery L, Litzky L, Bagg A, Carreno BM, Cimino PJ, Binder-Scholl GK, Smethurst DP, Gerry AB, Pumphrey NJ, Bennett AD, Brewer JE, Dukes J, Harper J, Tayton-Martin HK, Jakobsen BK, Hassan NJ, Kalos M, June CH. Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma. *Blood Cancer J*. 2013;122(6):863-871.

115. Wang D, Eraslan B, Wieland T, Hallström B, Hopf T, Zolg DP, Zecha J, Asplund A, Li L, Meng C, Frejno M, Schmidt T, Schnatbaum K, Wilhelm M, Ponten F, Uhlen M, Gagneur J, Hahne H, Kuster B. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol Syst Biol*. 2019;15(2):e8503.

116. Weinzierl AO, Lemmel C, Schoor O, Müller M, Krüger T, Wernet D, Hennenlotter J, Stenzl A, Klingel K, Rammensee H-G, Stevanović S. Distorted relation between mRNA copy number and corresponding major histocompatibility complex ligand density on the cell surface. *Mol Cell Proteomics*. 2007;6(1):102-113.

117. Fortier MH, Caron É, Hardy MP, Voisin G, Lemieux S, Perreault C, Thibault P. The MHC class I peptide repertoire is molded by the transcriptome. *J Exp Med*. 2008;205(3):595-610.

118. Aran D, Camarda R, Odegaard J, Paik H, Oskotsky B, Krings G, Goga A, Sirota M, Butte AJ. Comprehensive analysis of normal adjacent to tumor transcriptomes. *Nat Commun*. 2017;8(1):1077.

119. Ternette N, Olde Nordkamp MJM, Müller J, Anderson AP, Nicastri A, Hill AV, Kessler BM, Li D. Immunopeptidomic Profiling of HLA-A2-Positive Triple Negative Breast Cancer Identifies Potential Immunotherapy Target Antigens. *Proteomics*. 2018;18(12).

120. Marcu A, Bichmann L, Kuchenbecker L, Backert L, Kowalewski DJ, Freudenmann LK, Löffler MW, Lübke M, Walz JS, Velz J, Moch H, Regli L, Silginer M, Weller M, Schlosser A, Kohlbacher O, Stevanović S, Rammensee H-G, Neidert MC, Backert L, Mühlenbruch L, Szolek A, Lübke M, Wagner P, Engler T, Matovina S, Wang J, Hauri-Hohl M, Martin R, Kapolou K, Walz JS, Velz J, Moch H, Regli L, Silginer M, Weller M, Löffler MW, Erhard F, Schlosser A, Kohlbacher O, Stevanović S, Rammensee H-G, Neidert MC. The HLA Ligand Atlas - A resource of natural HLA ligands presented on benign tissues. *bioRxiv*. Published online 2020:778944.

121. Haen SP, Löffler MW, Rammensee H-G, Brossart P. Towards new horizons: characterization, classification and implications of the tumour antigenic repertoire. *Nat Rev Clin Oncol*. Published online 2020.

122. Rammensee H-G, Wiesmüller KH, Chandran PA, Zelba H, Rusch E, Gouttefangeas C, Kowalewski DJ, Di Marco M, Haen SP, Walz JS, Gloria YC, Bödder J, Schertel JM, Tunger A, Müller L, Kießler M, Wehner R, Schmitz M, Jakobi M, Schneiderhan-Marra N, Klein R, Laske K, Artzner K, Backert L, Schuster H, Schwenck J, Weber ANR, Pichler BJ, Kneilling M, La Fougère C, Forchhammer S, Metzler G, Bauer J, Weide B, Schippert W, Stevanović S, Löffler MW. A new synthetic toll-like receptor 1/2 ligand is an efficient adjuvant for peptide vaccination in a human volunteer. *J Immunother Cancer*. 2019;7(1):307.

123. Brown SD, Warren RL, Gibb EA, Martin SD, Spinelli JJ, Nelson BH, Holt RA. Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. *Genome Res*. 2014;24(5):743-750.

124. Giannakis M, Mu XJ, Shukla SA, Qian ZR, Cohen O, Nishihara R, Bahl S, Cao Y, Amin-Mansour A, Yamauchi M, Sukawa Y,

Stewart C, Rosenberg M, Mima K, Inamura K, Nosho K, Nowak JA, Lawrence MS, Giovannucci EL, Chan AT, Ng K, Meyerhardt JA, Van Allen EM, Getz G, Gabriel SB, Lander ES, Wu CJ, Fuchs CS, Ogino S, Garraway LA. Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Rep*. 2016;15(4):857-865.

125. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, Lee W, Yuan J, Wong P, Ho TS, Miller ML, Rekhtman N, Moreira AL, Ibrahim F, Bruggeman C, Gasmi B, Zappasodi R, Maeda Y, Sander C, Garon EB, Merghoub T, Wolchok JD, Schumacher TN, Chan TA. Mutational landscape determines sensitivity to PD-1 blockade in non–small cell lung cancer. *Science (80-)*. 2015;348(6230):124 LP - 128.

126. Tran E, Ahmadzadeh M, Lu Y-C, Gros A, Turcotte S, Robbins PF, Gartner JJ, Zheng Z, Li YF, Ray S, Wunderlich JR, Somerville RP, Rosenberg SA. Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science (80-)*. 2015;350(6266):1387 LP - 1390.

127. Linnemann C, van Buuren MM, Bies L, Verdegaal EME, Schotte R, Calis JJA, Behjati S, Velds A, Hilkmann H, Atmioui D el, Visser M, Stratton MR, Haanen JBAG, Spits H, van der Burg SH, Schumacher TNM. High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4+ T cells in human melanoma. *Nat Med*. 2015;21(1):81-85.

128. Tran E, Turcotte S, Gros A, Robbins PF, Lu Y-C, Dudley ME, Wunderlich JR, Somerville RP, Hogan K, Hinrichs CS, Parkhurst MR, Yang JC, Rosenberg SA. Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science (80-)*. 2014;344(6184):641-645.

129. Sahin U, Derhovanessian E, Miller M, Kloke B-P, Simon P, Löwer M, Bukur V, Tadmor AD, Luxemburger U, Schrörs B, Omokoko T, Vormehr M, Albrecht C, Paruzynski A, Kuhn AN, Buck J, Heesch S, Schreeb KH, Müller F, Ortseifer I, Vogler I, Godehardt E, Attig S, Rae R, Breitkreuz A, Tolliver C, Suchan M, Martic G, Hohberger A, Sorn P, Diekmann J, Ciesla J, Waksmann O, Brück A-K, Witt M, Zillgen M, Rothermel A, Kasemann B, Langer D, Bolte S, Diken M, Kreiter S, Nemecek R, Gebhardt C, Grabbe S, Höller C, Utikal J, Huber C, Loquai C, Türeci Ö. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature*. 2017;547(7662):222-226.

130. Keskin DB, Anandappa AJ, Sun J, Tirosh I, Mathewson ND, Li S, Oliveira G, Giobbie-Hurder A, Felt K, Gjini E, Shukla SA, Hu Z, Li L, Le PM, Allesøe RL, Richman AR, Kowalczyk MS, Abdelrahman S, Geduldig JE, Charbonneau S, Pelton K, Iorgulescu JB, Elagina L, Zhang W, Olive O, McCluskey C, Olsen LR, Stevens J, Lane WJ, Salazar AM, Daley H, Wen PY, Chiocca EA, Harden M, Lennon NJ, Gabriel S, Getz G, Lander ES, Regev A, Ritz J, Neuberg D, Rodig SJ, Ligon KL, Suvà ML, Wucherpfennig KW, Hacohen N, Fritsch EF, Livak KJ, Ott PA, Wu CJ, Reardon DA. Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature*. 2019;565(7738):234-239.

131. Granados DP, Rodenbrock A, Laverdure J-P, Côté C, Caron-Lizotte O, Carli C, Pearson H, Janelle V, Durette C, Bonneil E, Roy DC, Delisle J-S, Lemieux S, Thibault P, Perreault C. Proteogenomic-based discovery of minor histocompatibility antigens with suitable features for immunotherapy of hematologic cancers. *Leukemia*. 2016;30(6):1344-1354.

132. Granados DP, Sriranganadane D, Daouda T, Zieger A, Laumont CM, Caron-Lizotte O, Boucher G, Hardy M-P, Gendron P, Côté C, Lemieux S, Thibault P, Perreault C. Impact of genomic polymorphisms on the repertoire of human MHC class I-associated peptides. *Nat Commun*. 2014;5(1):3600.

133. Vincent K, Roy D-C, Perreault C. Next-generation leukemia immunotherapy. *Blood Cancer J*. 2011;118(11):2951-2959.

134. Summers C, Sheth VS, Bleakley M. Minor Histocompatibility Antigen-Specific T Cells. *Front Pediatr*. 2020;8(June):1-17.

135. Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, Boja ES, Costello CE, Cravatt BF, Fenselau C, Garcia BA, Ge Y, Gunawardena J, Hendrickson RC, Hergenrother PJ, Huber CG, Ivanov AR, Jensen ON, Jewett MC, Kelleher NL, Kiessling LL, Krogan NJ, Larsen MR, Loo JA, Ogorzalek Loo RR, Lundberg E, Maccoss MJ, Mallick P, Mootha VK, Mrksich M, Muir TW, Patrie SM, Pesavento JJ, Pitteri SJ, Rodriguez H, Saghatelian A, Sandoval W, Schlüter H, Sechi S, Slavoff SA, Smith LM, Snyder MP, Thomas PM, Uhlén M, Van Eyk JE, Vidal M, Walt DR, White FM, Williams ER, Wohlschlager T, Wysocki VH, Yates NA, Young NL, Zhang B. How many human proteoforms are there? *Nat Chem Biol*. 2018;14(3).

136. Ochoa D, Jarnuczak AF, Viéitez C, Gehre M, Soucheray M, Mateus A, Kleefeldt AA, Hill A, Garcia-Alonso L, Stein F, Krogan NJ, Savitski MM, Swaney DL, Vizcaíno JA, Noh K-M, Beltrao P. The functional landscape of the human phosphoproteome. *Nat Biotechnol*. 2020;38(3):365-373.

137. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. *Cell*. 2011;144(5):646-674.

138. Meyer VS, Drews O, Günder M, Hennenlotter J, Rammensee H-G, Stevanović S. Identification of natural MHC class II presented phosphopeptides and tumor-derived MHC class I phospholigands. *J Proteome Res*. 2009;8(7):3666-3674.

139. Engelhard VH, Obeng RC, Cummings KL, Petroni GR, Ambakhutwala AL, Chianese-Bullock KA, Smith KT, Lulu A, Varhegyi N, Smolkin ME, Myers P, Mahoney KE, Shabanowitz J, Büttner N, Hall EH, Haden K, Cobbold M, Hunt DF, Weiss G, Gaughan E, Slingluff Jr CL. MHC-restricted phosphopeptide antigens: preclinical validation and first-in-humans clinical trial in participants with high-risk melanoma. *J Immunother cancer*. 2020;8(1):e000262.

140. Sarkizova S, Klaeger S, Le PM, Li LW, Oliveira G, Keshishian H, Hartigan CR, Zhang W, Braun DA, Ligon KL, Bachireddy P, Zervantonakis IK, Rosenbluth JM, Ouspenskaia T, Law T, Justesen S, Stevens J, Lane WJ, Eisenhaure T, Lan Zhang G, Clauser KR, Hacohen N, Carr SA, Wu CJ, Keskin DB. A large peptidome dataset improves HLA class I epitope prediction across most of the human population. *Nat Biotechnol*. 2020;38(2):199-209.

141. Vigneron N, Stroobant V, Ferrari V, Abi Habib J, Van den Eynde BJ. Production of spliced peptides by the proteasome. *Mol*

Immunol. 2019;113:93-102.

142. Vigneron N, Ferrari V, Stroobant V, Habib JA, Van Den Eynde BJ. Peptide splicing by the proteasome. *J Biol Chem*. 2017;292(51):21170-21179.

143. Vigneron N, Stroobant V, Chapiro J, Ooms A, Degiovanni G, Morel S, van der Bruggen P, Boon T, Van den Eynde BJ. An Antigenic Peptide Produced by Peptide Splicing in the Proteasome. *Science (80-)*. 2004;304(5670):587 LP - 590.

144. Hanada K, Yewdell JW, Yang JC. Immune recognition of a human renal cancer antigen through post-translational protein splicing. *Nature*. 2004;427(6971):252-256.

145. Faridi P, Li C, Ramarathinam SH, Vivian JP, Illing PT, Mifsud NA, Ayala R, Song J, Gearing LJ, Hertzog PJ, Ternette N, Rossjohn J, Croft NP, Purcell AW. A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands. *Sci Immunol*. 2018;3(28):eaar3947.

146. Galuppini F, Dal Pozzo CA, Deckert J, Loupakis F, Fassan M, Baffa R. Tumor mutation burden: From comprehensive mutational screening to the clinic. *Cancer Cell Int*. 2019;19(1).

147. Chowell D, Krishna C, Pierini F, Makarov V, Rizvi NA, Kuo F, Morris LGT, Riaz N, Lenz TL, Chan TA. Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy. *Nat Med*. 2019;25(11):1715-1720.

148. Cameron BJ, Gerry AB, Dukes J, Harper J V., Kannan V, Bianchi FC, Grand F, Brewer JE, Gupta M, Plesa G, Bossi G, Vuidepot A, Powlesland AS, Legg A, Adams KJ, Bennett AD, Pumphrey NJ, Williams DD, Binder-Scholl G, Kulikovskaya I, Levine BL, Riley JL, Varela-Rohena A, Stadtmauer EA, Rapoport AP, Linette GP, June CH, Hassan NJ, Kalos M, Jakobsen BK. Identification of a Titin-Derived HLA-A1–Presented Peptide as a Cross-Reactive Target for Engineered MAGE A3–Directed T Cells. *Sci Transl Med*. 2013;5(197):197ra103 LP-197ra103.

149. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, Fitzhugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, Levine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860-921.

150. Venter CJ, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov G V, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Yuan Wang Z, Wang A, Wang X, Wang J, Wei MH, Wides R, et al. The sequence of the human genome. *Science (80-)*. 2001;291(5507):1304-1351.

151. Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, Thomas JK, Muthusamy B, Leal-Rojas P, Kumar P, Sahasrabuddhe NA, Balakrishnan L, Advani J, George B, Renuse S, Selvan LDN, Patil AH, Nanjappa V, Radhakrishnan A, Prasad S, Subbannayya T, Raju R, Kumar M, Sreenivasamurthy SK, Marimuthu A, Sathe GJ, Chavan S, Datta KK, Subbannayya Y, Sahu A, Yelamanchi SD, Jayaram S, Rajagopalan P, Sharma J, Murthy KR, Syed N, Goel R, Khan AA, Ahmad S, Dey G, Mudgal K, Chatterjee A, Huang TC, Zhong J, Wu X, Shaw PG, Freed D, Zahari MS, Mukherjee KK, Shankar S, Mahadevan A, Lam H, Mitchell CJ, Shankar SK, Satishchandra P, Schroeder JT, Sirdeshmukh R, Maitra A, Leach SD, Drake CG, Halushka MK, Prasad TSK, Hruban RH, Kerr CL, Bader GD, Iacobuzio-Donahue CA, Gowda H, Pandey A. A draft map of the human proteome. *Nature*. 2014;509(7502):575-581.

152. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, Mathieson T, Lemeer S, Schnatbaum K, Reimer U, Wenschuh H, Mollenhauer M, Slotta-Huspenina J, Boese JH, Bantscheff M, Gerstmair A, Faerber F, Kuster B. Mass-spectrometry-based draft of the human proteome. *Nature*. 2014;509(7502):582-587.

153. Rammensee H-G, Rötzschke O, Falk K. MHC Class I-Restricted Antigen Processing — Lessons from Natural Ligands. In: *Chemical Immunology and Allergy*. ; 1993:113-133.

154. Boegel S, Castle JC, Kodysh J, O'Donnell T, Rubinsteyn A. Bioinformatic methods for cancer neoantigen prediction. In: *Progress in Molecular Biology and Translational Science*. Vol 164. Academic Press; 2019:25-60.

155. Finotello F, Rieder D, Hackl H, Trajanoski Z. Next-generation computational tools for interrogating cancer immunity. *Nat Rev Genet*. 2019;20(12):724-746.

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

157. Vizcaíno JA, Kubiniok P, Kovalchik KA, Ma Q, Duquette JD, Mongrain I, Deutsch EW, Peters B, Sette A, Sirois I, Caron E. The Human Immunopeptidome Project: A Roadmap to Predict and Treat Immune Diseases. *Mol Cell Proteomics*. 2020;19(1):31 LP - 49.

158. Bassani-Sternberg M, Chong C, Guillaume P, Solleder M, Pak H, Gannon PO, Kandalaft LE, Coukos G, Gfeller D. Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allostery regulating HLA specificity. *PLoS Comput Biol*. 2017;13(8):e1005725-e1005725.

159. Brown SD, Holt RA. Neoantigen characteristics in the context of the complete predicted MHC class I self-immunopeptidome. *Oncoimmunology*. 2018;8(3):1556080.

160. Calis JJA, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, Keşmir C, Peters B. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput Biol*. 2013;9(10):e1003266-e1003266.

161. Fritsche J, Rakitsch B, Hoffgaard F, Römer M, Schuster H, Kowalewski DJ, Priemer M, Stos-Zweifel V, Hörzer H, Satelli A, Sonntag A, Goldfinger V, Song C, Mahr A, Ott M, Schoor O, Weinschenk T. Translating Immunopeptidomics to Immunotherapy-Decision-Making for Patient and Personalized Target Selection. *Proteomics*. 2018;18(12):e1700284-e1700284.

162. Iacobuzio-Donahue CA, Michael C, Baez P, Kappagantula R, Hooper JE, Hollman TJ. Cancer biology as revealed by the research autopsy. *Nat Rev Cancer*. 2019;19(12):686-697.

163. Bichmann L, Nelde A, Ghosh M, Heumos L, Mohr C, Peltzer A, Kuchenbecker L, Sachsenberg T, Walz JS, Stevanović S, Rammensee H-G, Kohlbacher O. MHCquant: Automated and Reproducible Data Analysis for Immunopeptidomics. *J Proteome Res*. 2019;18(11):3876-3884.

164. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics*. 2014;30(23):3310-3316.

165. Nelde A, Kowalewski DJ, Stevanović S. Purification and Identification of Naturally Presented MHC Class I and II Ligands BT - Antigen Processing: Methods and Protocols. In: van Endert P, ed. Springer New York; 2019:123-136.

166. Lippolis JD, White FM, Marto JA, Luckey CJ, Bullock TNJ, Shabanowitz J, Hunt DF, Engelhard VH. Analysis of MHC Class II Antigen Processing by Quantitation of Peptides that Constitute Nested Sets. *J Immunol*. 2002;169(9):5089 LP - 5097.

167. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Di Tommaso P, Nahnsen S. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol*. 2020;38(3):276-278.

168. Röst HL, Sachsenberg T, Aiche S, Bielow C, Weisser H, Aicheler F, Andreotti S, Ehrlich HC, Gutenbrunner P, Kenar E, Liang X, Nahnsen S, Nilse L, Pfeuffer J, Rosenberger G, Rurik M, Schmitt U, Veit J, Walzer M, Wojnar D, Wolski WE, Schilling O, Choudhary JS, Malmström L, Aebersold R, Reinert K, Kohlbacher O. OpenMS: A flexible open-source software platform for mass spectrometry data analysis. *Nat Methods*. 2016;13(9):741-748.

169. Eng JK, Hoopmann MR, Jahan TA, Egertson JD, Noble WS, MacCoss MJ. A Deeper Look into Comet - Implementation and Features. *J Am Soc Mass Spectrom*. 2015;26(11):1865-1874.

170. The M, MacCoss MJ, Noble WS, Käll L. Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. *J Am Soc Mass Spectrom*. 2016;27(11):1719-1727.

171. Savitski MM, Wilhelm M, Hahne H, Kuster B, Bantscheff M. A Scalable Approach for Protein False Discovery Rate Estimation in Large Proteomic Data Sets. *Mol Cell Proteomics*. 2015;14(9):2394-2404.

172. Reiter L, Claassen M, Schrimpf SP, Jovanovic M, Schmidt A, Buhmann JM, Hengartner MO, Aebersold R. Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. *Mol Cell Proteomics*. 2009;8(11):2405-2417.

173. Rammensee H-G, Bachmann J, Emmerich NPN, Bachor OA, Stevanović S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*. 1999;50(3):213-219.

174. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol*. 2017;199(9):3360-3368.

175. Di Marco M, Schuster H, Backert L, Ghosh M, Rammensee H-G, Stevanović S. Unveiling the Peptide Motifs of HLA-C and HLA-G from Naturally Presented Peptides and Generation of Binding Prediction Matrices. *J Immunol*. 2017;199(8):2639-2651.

176. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc*. 2016;11(12):2301-2319.

177. Weisser H, Choudhary JS. Targeted Feature Detection for Data-Dependent Shotgun Proteomics. *J Proteome Res*. 2017;16(8):2964-2974.

178. Boegel S, Löwer M, Bukur T, Sorn P, Castle JC, Sahin U. HLA and proteasome expression body map. *BMC Med Genomics*. 2018;11(1):36.

179. Tran NH, Qiao R, Xin L, Chen X, Liu C, Zhang X, Shan B, Ghodsi A, Li M. Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. *Nat Methods*. 2019;16(1):63-66.

180. Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, Zhang W, Zhang Z, Lajoie GA, Ma B. PEAKS DB: De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. *Mol Cell Proteomics*. 2012;11(4):M111.010587.

181. Pfeifer N, Leinenbach A, Huber CG, Kohlbacher O. Statistical learning of peptide retention behavior in chromatographic separations: a new kernel-based approach for computational proteomics. *BMC Bioinformatics*. 2007;8(1):468.

182. Toprak UH, Gillet LC, Maiolica A, Navarro P, Leitner A, Aebersold R. Conserved peptide fragmentation as a benchmarking tool for mass spectrometers and a discriminating feature for targeted proteomics. *Mol Cell Proteomics*. 2014;13(8):2056-2071.

183. Mohr C, Friedrich A, Wojnar D, Kenar E, Polatkan AC, Codrea MC, Czemmel S, Kohlbacher O, Nahnsen S. qPortal: A platform for data-driven biomedical research. *PLoS One*. 2018;13(1):e0191603.

184. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, Pérez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yilmaz S, Tiwary S, Cox J, Audain E, Walzer M, Jarnuczak AF, Ternent T, Brazma A, Vizcaíno JA. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res*. 2019;47(D1):D442-D450.

185. Shao W, Pedrioli PGA, Wolski W, Scurtescu C, Schmid E, Vizcaíno JA, Courcelles M, Schuster H, Kowalewski D, Marino F, Arlehamn CSL, Vaughan K, Peters B, Sette A, Ottenhoff THM, Meijgaarden KE, Nieuwenhuizen N, Kaufmann SHE, Schlapbach R, Castle JC, Nesvizhskii AI, Nielsen M, Deutsch EW, Campbell DS, Moritz RL, Zubarev RA, Ytterberg AJ, Purcell AW, Marcilla M, Paradela A, Wang Q, Costello CE, Ternette N, Van Veelen PA, Van Els CACM, Heck AJR, De Souza GA, Sollid LM, Admon A, Stevanovic S, Rammensee HG, Thibault P, Perreault C, Bassani-Sternberg M, Aebersold R, Caron E. The SysteMHC Atlas project. *Nucleic Acids Res*. 2018;46(D1):D1237-D1247.

186. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, Wheeler DK, Gabbard JL, Hix D, Sette A, Peters B. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res*. 2015;43(D1):D405-D412.

187. Ishii N, Chiba M, Iizuka M, Watanabe H, Ishioka T, Masamune O. Expression of MHC class II antigens (HLA-DR, -DP, and -DQ) on human gastric epithelium. *Gastroenterol Jpn*. 1992;27(1):23-28.

188. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, Mathieson T, Lemeer S, Schnatbaum K, Reimer U, Wenschuh H, Mollenhauer M, Slotta-Huspenina J, Boese JH, Bantscheff M, Gerstmair A, Faerber F, Kuster B. Mass-spectrometry-based draft of the human proteome. *Nature*. 2014;509(7502):582-587.

189. Almeida LG, Sakabe NJ, DeOliveira AR, Silva MCC, Mundstein AS, Cohen T, Chen Y-TT, Chua R, Gurung S, Gnjatic S, Jungbluth AA, Caballero OL, Bairoch A, Kiesler E, White SL, Simpson AJGG, Old LJ, Camargo AA, Vasconcelos ATR, de Oliveira AR, Silva MCC, Mundstein AS, Cohen T, Chen Y-TT, Chua R, Gurung S, Gnjatic S, Jungbluth AA, Caballero OL, Bairoch A, Kiesler E, White SL, Simpson AJGG, Old LJ, Camargo AA, Vasconcelos ATR. CTdatabase: A knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Res*. 2009;37(Database issue):D816-D819.

190. Wang C, Gu Y, Zhang K, Xie K, Zhu M, Dai N, Jiang Y, Guo X, Liu M, Dai J, Wu L, Jin G, Ma H, Jiang T, Yin R, Xia Y, Liu L, Wang S, Shen B, Huo R, Wang Q, Xu L, Yang L, Huang X, Shen H, Sha J, Hu Z. Systematic identification of genes with a cancer-testis expression pattern in 19 cancer types. *Nat Commun*. 2016;7:10499.

191. Müller M, Gfeller D, Coukos G, Bassani-Sternberg M. "Hotspots" of antigen presentation revealed by human leukocyte antigen ligandomics for neoantigen prioritization. *Front Immunol*. 2017;8:1367.

192. Álvaro-Benito M, Morrison E, Abualrous ET, Kuropka B, Freund C. Quantification of HLA-DM-Dependent Major Histocompatibility Complex of Class II Immunopeptidomes by the Peptide Landscape Antigenic Epitope Alignment Utility. *Front Immunol*. 2018;9:872.

193. Kaufman J. Generalists and Specialists: A New View of How MHC Class I Molecules Fight Infectious Pathogens. *Trends Immunol*. 2018;39(5):367-379.

194. Alpízar A, Marino F, Ramos-Fernández A, Lombardía M, Jeko A, Pazos F, Paradela A, Santiago C, Heck AJR, Marcilla M. A molecular basis for the presentation of phosphorylated peptides by HLA-B antigens. *Mol Cell Proteomics*. 2017;16(2):181-193.

195. Marino F. Gaining Insight Into Posttranslationally Modified HIV Antigens and Their Underlying Characteristics. *Proteomics*. 2018;18(12):1800041.

196. Makarov A. Orbitrap journey: taming the ion rings. *Nat Commun*. 2019;10(1):3743.

197. Otter D. PROTEIN | Determination and Characterization. In: Caballero BBT-E of FS and N (Second E, ed. Academic Press; 2003:4824-4830.

198. Evans DRH, Romero JK, Westoby M. Chapter 9 Concentration of Proteins and Removal of Solutes. In: Burgess RR, Deutscher MPBT-M in E, eds. *Guide to Protein Purification, 2nd Edition*. Vol 463. Academic Press; 2009:97-120.

199. Kromidas S. When Should I Use My UHPLC as a UHPLC? *HPLC Expert II*. Published online February 28, 2017:1-25.

200. Pakhomov VP. Chromatography in Pharmaceutical Chemistry (100 Years of the Discovery of Chromatography by M. S. Tswett). *Pharm Chem J*. 2003;37(8):451-452.

201. Teutenberg T, Hetzel T, Portner C, Wiese S, vom Eyser C, Tuerk J. Aspects of the Development of Methods in LC/MS Coupling.

HPLC-MS Handb Pract. Published online July 5, 2017:73-137.

202. McPolin O. *An Introduction to HPLC for Pharmaceutical Analysis*. Vol 44.; 2003.

203. Majors RE. Developments in HPLC column packing design. *LCGC North Am*. 2006;24(4 SUPPL.):8-15.

204. Fisher Scientific T. *Chromatogrpahy Acclaim PepMap and Acclaim PepMap RSLC Columns for High-Resolution Peptide Mapping*.; 2016.

205. Cielecka-Piontek J, Zalewski P, Jelińska A, Garbacki P. UHPLC: The Greening Face of Liquid Chromatography. *Chromatographia*. 2013;76(21):1429-1437.

206. McHugh L, Arthur JW. Computational methods for protein identification from mass spectrometry data. *PLoS Comput Biol*. 2008;4(2):e12-e12.

207. Verheggen K, Ræder H, Berven FS, Martens L, Barsnes H, Vaudel M. Anatomy and evolution of database search engines—a central component of mass spectrometry based proteomic workflows. *Mass Spectrom Rev*. 2020;39(3):292-306.

208. Eng JK, Fischer B, Grossmann J, MacCoss MJ. A fast SEQUEST cross correlation algorithm. *J Proteome Res*. 2008;7(10):4598-4602.

209. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom*. 1994;5(11):976-989.

210. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 1999;20(18):3551-3567.

211. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen J V, Mann M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J Proteome Res*. 2011;10(4):1794-1805.

212. Eng JK, Hoopmann MR, Jahan TA, Egertson JD, Noble WS, MacCoss MJ. A Deeper Look into Comet—Implementation and Features. *J Am Soc Mass Spectrom*. 2015;26(11):1865-1874.

213. Nesvizhskii AI. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J Proteomics*. 2010;73(11):2092-2123.

214. Rosenberger G, Bludau I, Schmitt U, Heusel M, Hunter CL, Liu Y, Maccoss MJ, Maclean BX, Nesvizhskii AI, Pedrioli PGA, Reiter L, Röst HL, Tate S, Ting YS, Collins BC, Aebersold R. Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses. *Nat Methods*. 2017;14(9):921-927.

215. Lenz C, Dihazi H. Introduction to Proteomics Technologies. In: Jung K, ed. *Statistical Analysis in Proteomics*. Springer New York; 2016:3-27.

216. *IUPAC, Compendium of Chemical Terminology, 2nd Ed. (the "Gold Book") (1997). Online Corrected Version: (2006–) "Resolving Power in Mass Spectrometry".*

217. Marshall AG, Hendrickson CL. High-Resolution Mass Spectrometers. *Annu Rev Anal Chem*. 2008;1(1):579-599.

218. Kind T, Fiehn O. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics*. 2006;7(1):234.

219. Urban PL. Clarifying Misconceptions about Mass and Concentration Sensitivity. *J Chem Educ*. 2016;93(6):984-987.

220. Price P. Standard definitions of terms relating to mass spectrometry. *J Am Soc Mass Spectrom*. 1991;2(4):336-348.

221. May JC, McLean JA. Ion mobility-mass spectrometry: Time-dispersive instrumentation. *Anal Chem*. 2015;87(3):1422-1436.

222. Meier F, Brunner AD, Koch S, Koch H, Lubeck M, Krause M, Goedecke N, Decker J, Kosinski T, Park MA, Bache N, Hoerning O, Cox J, Räther O, Mann M. Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol Cell Proteomics*. 2018;17(12):2534-2545.

223. Meier F, Beck S, Grassl N, Lubeck M, Park MA, Raether O, Mann M. Parallel accumulation-serial fragmentation (PASEF): Multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. *J Proteome Res*. 2015;14(12):5378-5387.

224. Meier F, Brunner A-D, Frank M, Ha A, Bludau I, Voytik E, Kaspar-Schoenefeld S, Lubeck M, Raether O, Aebersold R, Collins BC, Röst HL, Mann M. Parallel accumulation – serial fragmentation combined with data-independent acquisition (diaPASEF): Bottom-up proteomics with near optimal ion usage. *bioRxiv*. Published online January 1, 2020:656207.

225. Bruker. No Title. https://www.bruker.com/products/mass-spectrometry-and-separations/lc-ms/o-tof/timstof-pro.html

226. Terhorst C, Parham P, Mann DL, Strominger JL. Structure of HLA antigens: amino acid and carbohydrate compositions and NH2 terminal sequences of four antigen preparations. *Proc Natl Acad Sci U S A*. 1976;73(3):910-914.

227. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, Ruddy K, Tsang J, Cardoso F. Breast cancer. *Nat Rev Dis Prim*. 2019;5(1):66.

228. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of

incidence and mortality worldwide for 36 cancers in 185 countries. *A Cancer J Clin*. 2018;68(6):394-424.

229. Shiovitz S, Korde LA. Genetics of breast cancer: a topic in evolution. *Ann Oncol Off J Eur Soc Med Oncol*. 2015;26(7):1291-1299.

230. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips K-A, Mooij TM, Roos-Blom M-J, Jervis S, van Leeuwen FE, Milne RL, Andrieu N, Goldgar DE, Terry MB, Rookus MA, Easton DF, Antoniou AC, Consortium B and BC. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA*. 2017;317(23):2402-2416.

231. Tan PH, Ellis I, Allison K, Brogi E, Fox SB, Lakhani S, Lazar AJ, Morris EA, Sahin A, Salgado R, Sapino A, Sasano H, Schnitt S, Sotiriou C, van Diest P, White VA, Lokuhetty D, Cree IA, For the W H O Classification of Tumours Editorial Board. The 2019 WHO classification of tumours of the breast. *Histopathology*. 2020;77(2):181-185.

232. Sinn HP, Kreipe H. A brief overview of the WHO classification of breast tumors, 4th edition, focusing on issues and updates from the 3rd edition. *Breast Care*. 2013;8(2):149-154.

233. Hortobagyi GN, Edge SB, Giuliano A. New and Important Changes in the TNM Staging System for Breast Cancer. *Am Soc Clin Oncol Educ B*. 2018;(38):457-467.

234. Giuliano AE, Connolly JL, Edge SB, Mittendorf EA, Rugo HS, Solin LJ, Weaver DL, Winchester DJ, Hortobagyi GN. Breast Cancer—Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin*. 2017;67(4):290-303.

235. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge Ø, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale A-L, Brown PO, Botstein D. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747-752.

236. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thürlimann B, Senn H-J. Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol*. 2011;22(8):1736-1747.

237. Curigliano G, Burstein HJ, Winer EP, Gnant M, Dubsky P, Loibl S, Colleoni M, Regan MM, Piccart-Gebhart M, Senn H-J, Thürlimann B, André F, Baselga J, Bergh J, Bonnefoi H, Brucker SY, Cardoso F, Carey L, Ciruelos E, Cuzick J, Denkert C, Di Leo A, Ejlertsen B, Francis P, Galimberti V, Garber J, Gulluoglu B, Goodwin P, Harbeck N, Hayes DF, Huang C-S, Huober J, Khaled H, Jassem J, Jiang Z, Karlsson P, Morrow M, Orecchia R, Osborne KC, Pagani O, Partridge AH, Pritchard K, Ro J, Rutgers EJT, Sedlmayer F, Semiglazov V, Shao Z, Smith I, Toi M, Tutt A, Viale G, Watanabe T, Whelan TJ, Xu B. De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. *Ann Oncol*. 2017;28(8):1700-1712.

238. Inwald EC, Koller M, Klinkhammer-Schalke M, Zeman F, Hofstädter F, Gerstenhauer M, Brockhoff G, Ortmann O. 4-IHC classification of breast cancer subtypes in a large cohort of a clinical cancer registry: use in clinical routine for therapeutic decisions and its effect on survival. *Breast Cancer Res Treat*. 2015;153(3):647-658.

239. Penault-Llorca F, Radosevic-Robin N. Ki67 assessment in breast cancer: an update. *Pathology*. 2017;49(2):166-171.

240. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer mortality: meta-analysis of individual patient data for 8135 women in 22 randomised trials. *Lancet*. 2014;383(9935):2127-2135.

241. Jatoi I, Benson JR, Kunkler I. Hypothesis: can the abscopal effect explain the impact of adjuvant radiotherapy on breast cancer mortality? *Breast Cancer*. 2018;4(1):8.

242. Bartelink H, Maingon P, Poortmans P, Weltens C, Fourquet A, Jager J, Schinagl D, Oei B, Rodenhuis C, Horiot J-C, Struikmans H, Van Limbergen E, Kirova Y, Elkhuizen P, Bongartz R, Miralbell R, Morgan D, Dubois J-B, Remouchamps V, Mirimanoff R-O, Collette S, Collette L. Whole-breast irradiation with or without a boost for patients treated with breast-conserving surgery for early breast cancer: 20-year follow-up of a randomised phase 3 trial. *Lancet Oncol*. 2015;16(1):47-56.

243. Bernier J. Immuno-oncology: Allying forces of radio- and immuno-therapy to enhance cancer cell killing. *Crit Rev Oncol Hematol*. 2016;108:97-108.

244. Rastogi P, Anderson SJ, Bear HD, Geyer CE, Kahlenberg MS, Robidoux A, Margolese RG, Hoehn JL, Vogel VG, Dakhil SR, Tamkus D, King KM, Pajon ER, Wright MJ, Robert J, Paik S, Mamounas EP, Wolmark N. Preoperative Chemotherapy: Updates of National Surgical Adjuvant Breast and Bowel Project Protocols B-18 and B-27. *J Clin Oncol*. 2008;26(5):778-785.

245. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet*. 2011;378(9793):771-784.

246. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100 000 women in 123 randomised trials. *Lancet*. 2012;379(9814):432-444.

247. Francis PA, Pagani O, Fleming GF, Walley BA, Colleoni M, Láng I, Gómez HL, Tondini C, Ciruelos E, Burstein HJ, Bonnefoi HR, Bellet M, Martino S, Geyer CE, Goetz MP, Stearns V, Pinotti G, Puglisi F, Spazzapan S, Climent MA, Pavesi L, Ruhstaller T, Davidson NE, Coleman R, Debled M, Buchholz S, Ingle JN, Winer EP, Maibach R, Rabaglio-Poretti M, Ruepp B, Di Leo A, Coates

AS, Gelber RD, Goldhirsch A, Regan MM. Tailoring Adjuvant Endocrine Therapy for Premenopausal Breast Cancer. *N Engl J Med*. 2018;379(2):122-137.

248. Mackey JR, Pieńkowski T, Crown J, Sadeghi S, Martin M, Chan A, Saleh M, Sehdev S, Provencher L, Semiglazov V, Press MF, Sauter G, Lindsay M, Houé V, Buyse M, Drevot P, Hitier S, Bensfia S, Eiermann W. Long-term outcomes after adjuvant treatment of sequential versus combination docetaxel with doxorubicin and cyclophosphamide in node-positive breast cancer: BCIRG-005 randomized trial. *Ann Oncol*. 2016;27(6):1041-1047.

249. von Minckwitz G, Procter M, de Azambuja E, Zardavas D, Benyunes M, Viale G, Suter T, Arahmani A, Rouchet N, Clark E, Knott A, Lang I, Levy C, Yardley DA, Bines J, Gelber RD, Piccart M, Baselga J. Adjuvant Pertuzumab and Trastuzumab in Early HER2-Positive Breast Cancer. *N Engl J Med*. 2017;377(2):122-131.

250. Gnant M, Pfeiler G, Dubsky PC, Hubalek M, Greil R, Jakesz R, Wette V, Balic M, Haslbauer F, Melbinger E, Bjelic-Radisic V, Artner-Matuschek S, Fitzal F, Marth C, Sevelda P, Mlineritsch B, Steger GG, Manfreda D, Exner R, Egle D, Bergh J, Kainberger F, Talbot S, Warner D, Fesl C, Singer CF. Adjuvant denosumab in breast cancer (ABCSG-18): a multicentre, randomised, double-blind, placebo-controlled trial. *Lancet*. 2015;386(9992):433-443.

251. Solinas C, Carbognin L, De Silva P, Criscitiello C, Lambertini M. Tumor-infiltrating lymphocytes in breast cancer according to tumor subtype: Current state of the art. *The Breast*. 2017;35:142-150.

252. Pruneri G, Vingiani A, Denkert C. Tumor infiltrating lymphocytes in early breast cancer. *The Breast*. 2018;37:207-214.

253. Meng S, Li L, Zhou M, Jiang W, Niu H, Yang K. Distribution and prognostic value of tumor-infiltrating T cells in breast cancer. *Mol Med Rep*. 2018;18(5):4247-4258.

254. Balic M, Thomssen C, Würstlein R, Gnant M, Harbeck N. St. Gallen/Vienna 2019: A Brief Summary of the Consensus Discussion on the Optimal Primary Breast Cancer Treatment. *Breast Care*. 2019;14(2):103-110.

255. Rammensee H-G, Stevanović S, Gouttefangeas C, Heidu S, Klein R, Preuß B, Walz JS, Nelde A, Haen S, Reth M, Yang J, Tabatabai G, Bösmüller H, Hoffmann H, Schindler M, Planz O, Wiesmüller K-H, Löffler M. Designing a therapeutic SARS-CoV-2 T-cell-inducing vaccine for high-risk patient groups. Published online 2020:1-21.

256. Walter S, Weinschenk T, Stenzl A, Zdrojowy R, Pluzanska A, Szczylik C, Staehler M, Brugger W, Dietrich PY, Mendrzyk R, Hilf N, Schoor O, Fritsche J, Mahr A, Maurer D, Vass V, Trautwein C, Lewandrowski P, Flohr C, Pohla H, Stanczak JJ, Bronte V, Mandruzzato S, Biedermann T, Pawelec G, Derhovanessian E, Yamagishi H, Miki T, Hongo F, Takaha N, Hirakawa K, Tanaka H, Stevanović S, Frisch J, Mayer-Mokler A, Kirner A, Rammensee H-G, Reinhardt C, Singh-Jasuja H. Multipeptide immune response to cancer vaccine IMA901 after single-dose cyclophosphamide associates with longer patient survival. *Nat Med*. 2012;18(8):1254-1261.

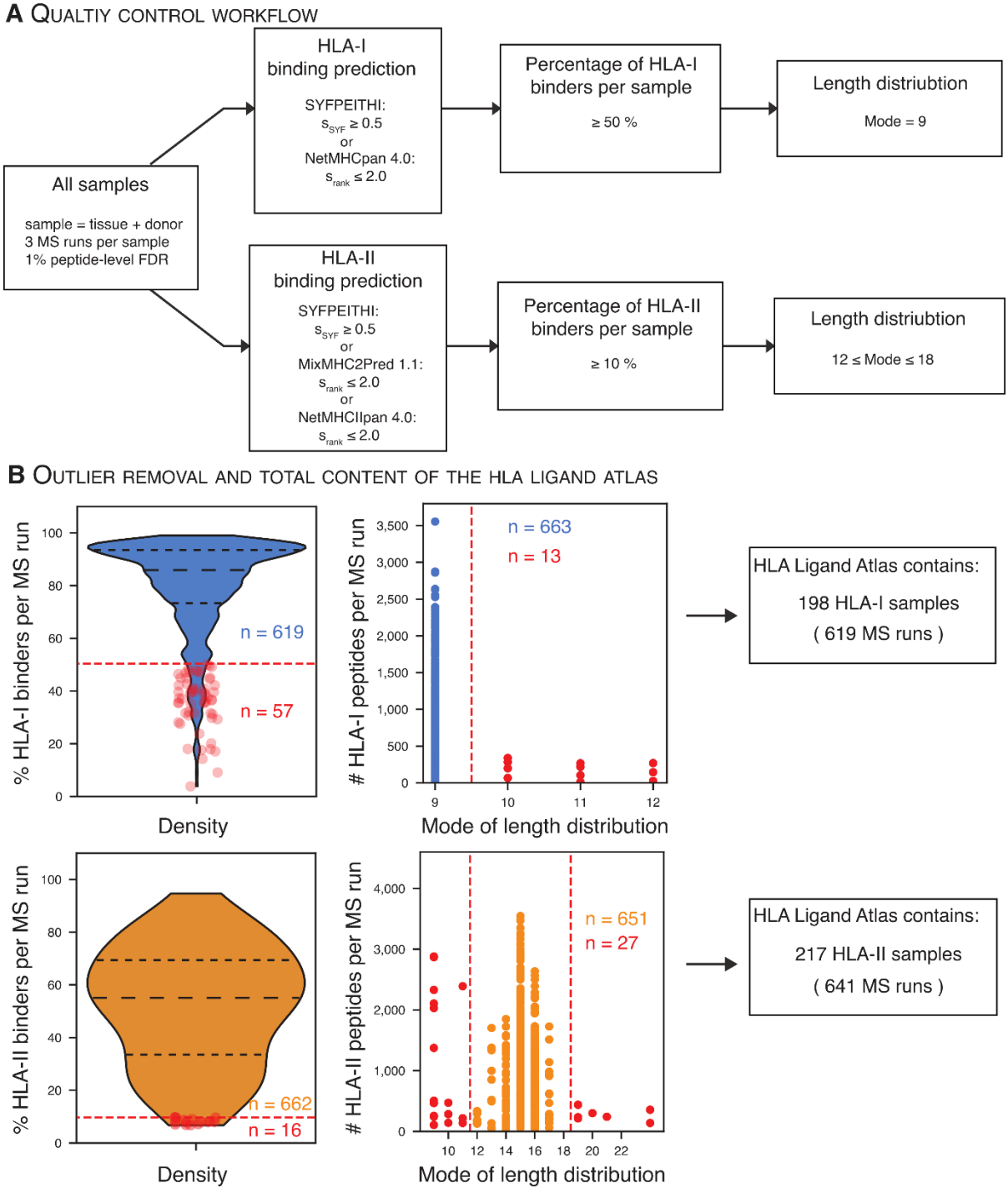
257. Rini BI, Stenzl A, Zdrojowy R, Kogan M, Shkolnik M, Oudard S, Weikert S, Bracarda S, Crabb SJ, Bedke J, Ludwig J, Maurer D, Mendrzyk R, Wagner C, Mahr A, Fritsche J, Weinschenk T, Walter S, Kirner A, Singh-Jasuja H, Reinhardt C, Eisen T. IMA901, a multipeptide cancer vaccine, plus sunitinib versus sunitinib alone, as first-line therapy for advanced or metastatic renal cell carcinoma (IMPRINT): a multicentre, open-label, randomised, controlled, phase 3 trial. *Lancet Oncol*. 2016;17(11):1599-1611.

258. Hu Z, Ott PA, Wu CJ. Towards personalized, tumour-specific, therapeutic vaccines for cancer. *Nat Rev Immunol*. 2018;18(3):168-182.

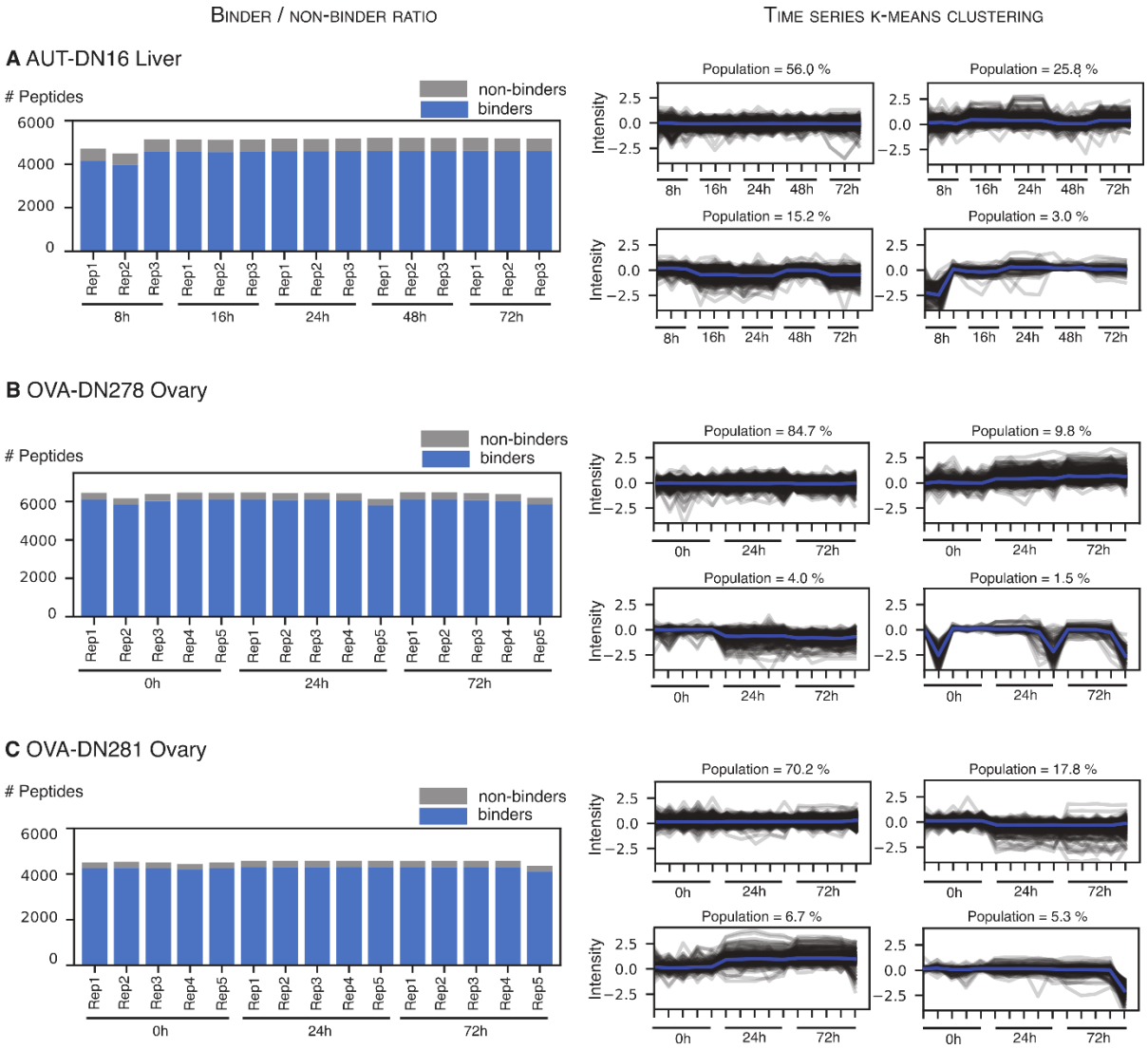
259. Van Der Burg SH, Arens R, Ossendorp F, Van Hall T, Melief CJM. Vaccines for established cancer: Overcoming the challenges posed by immune evasion. *Nat Rev Cancer*. 2016;16(4):219-233.

References

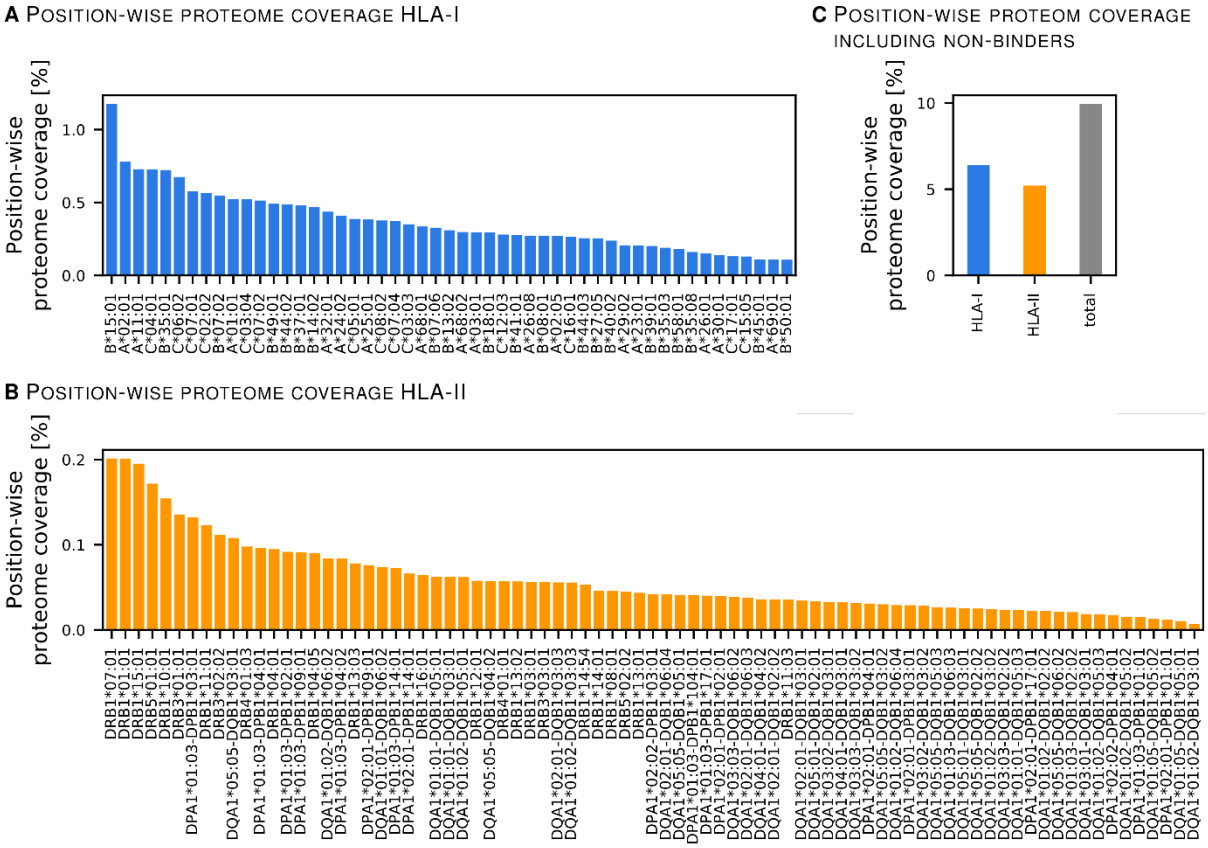
9 SUPPLEMENTAL DATA: HLA LIGAND ATLAS



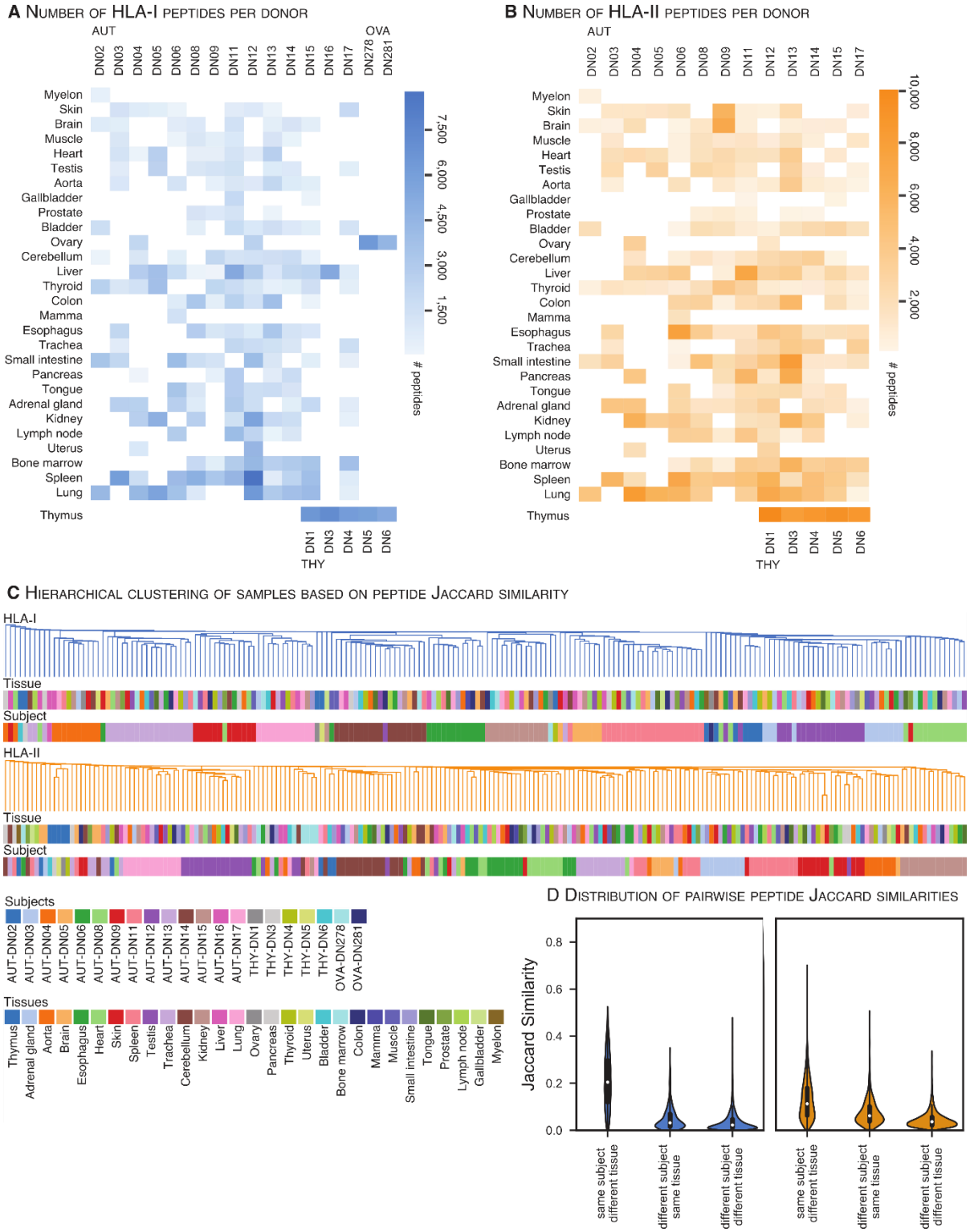
Supplemental Figure S1: Computational workflow and quality control thresholds (related to Figure 1).
(A) Data from three LC-MS/MS runs (technical replicates) per sample (one tissue from one subject) were processed with MHCquant with a local peptide-level FDR of 1%. Identified peptides were categorized into peptides predicted as strong, weak, or non-binders. HLA-I peptides were predicted by SYFPEITHI and NetMHCpan 4.0, employing a cut-off threshold of $s_{\mathrm{SYF}} \geq 0.5$ and $s_{\mathrm{rank}} \leq 2.0$, respectively. HLA-II peptides were predicted with MixMHCPred, NetMHCIIpan 4.0 and SYFPEITHI with equal thresholds. Samples with less than 50% and 10% predicted binders per technical replicate for HLA-I and HLA-II immunopeptidomes were excluded from further analysis and are not included in the HLA Ligand Atlas database release. Finally, we employed a stringent cut-off concerning the mode of the peptide length distribution for HLA-I (mode equal to 9) and -II (mode between 12 to18) immunopeptidomes.
(B) LC-MS/MS runs not pertaining to the aforementioned QC thresholds (dashed red lines) concerning the percentage of predicted HLA-binders and the peptide length distribution were not included in the HLA Ligand Atlas database release: Violin plots (left) depict the percentage of peptides predicted to be HLA-binders per LC-MS/MS run and the quality control cut-off for LC-MS/MS runs employed for HLA-I and -II immunopeptidomes. Dot plots (center) depict the mode of the peptide length distribution encountered per LC-MS/MS run and the quality control cut-off employed for HLA-I and -II. The number of LC-MS/MS runs (HLA-I - blue and HLA-II orange) failing the QC thresholds is indicated by red dots. The final number of LC-MS/MS runs and samples selected for inclusion into the HLA Ligand Atlas database release and data analysis are indicated on the right and differ for HLA-I and –II experiments.



Supplemental Figure S2: Time-Series Experiment (related to Figure 1).
The time series experiment was carried out on three biological samples from three subjects (A: AUT-DN16 Liver, B: OVA-DN278, C: OVA-DN281).
A-C: Bar plots (left hand side) indicate the number of identified HLA-I predicted binders (blue) and predicted non-binders (grey) across technical replicates and timepoints for each timeseries.
Time series (right hand side) indicate individual clusters (K-Means using four seeds) of trajectories of quantified MS1 intensity across technical replicates and timepoints. Intermediate trend lines for each cluster are indicated in blue and the percentage of trajectories populating a given cluster is annotated at the top edge of each plot. The trajectories of all timepoints were set in relation to the initial timepoint. The analysis reveals that the number of identified peptides and their percentage of predicted HLA-binders is constant and that most trajectories do not vary much across time points.



Supplemental Figure S3: Position-wise proteome coverage (related to Figure 2).
(A, B) Analysis of the different propensities of HLA-I (blue) and –II (orange), allotypes concerning the presentation of peptides. In turn different percentages of the source proteome are position-wise covered. For example, HLA-B*15:01 ligands cover over 1% of the HLA-I source proteome.
(C) Global analysis of the position-wise coverage of all proteins presented on either HLA-I (blue), HLA-II (orange), or both independent of specific allotypes.

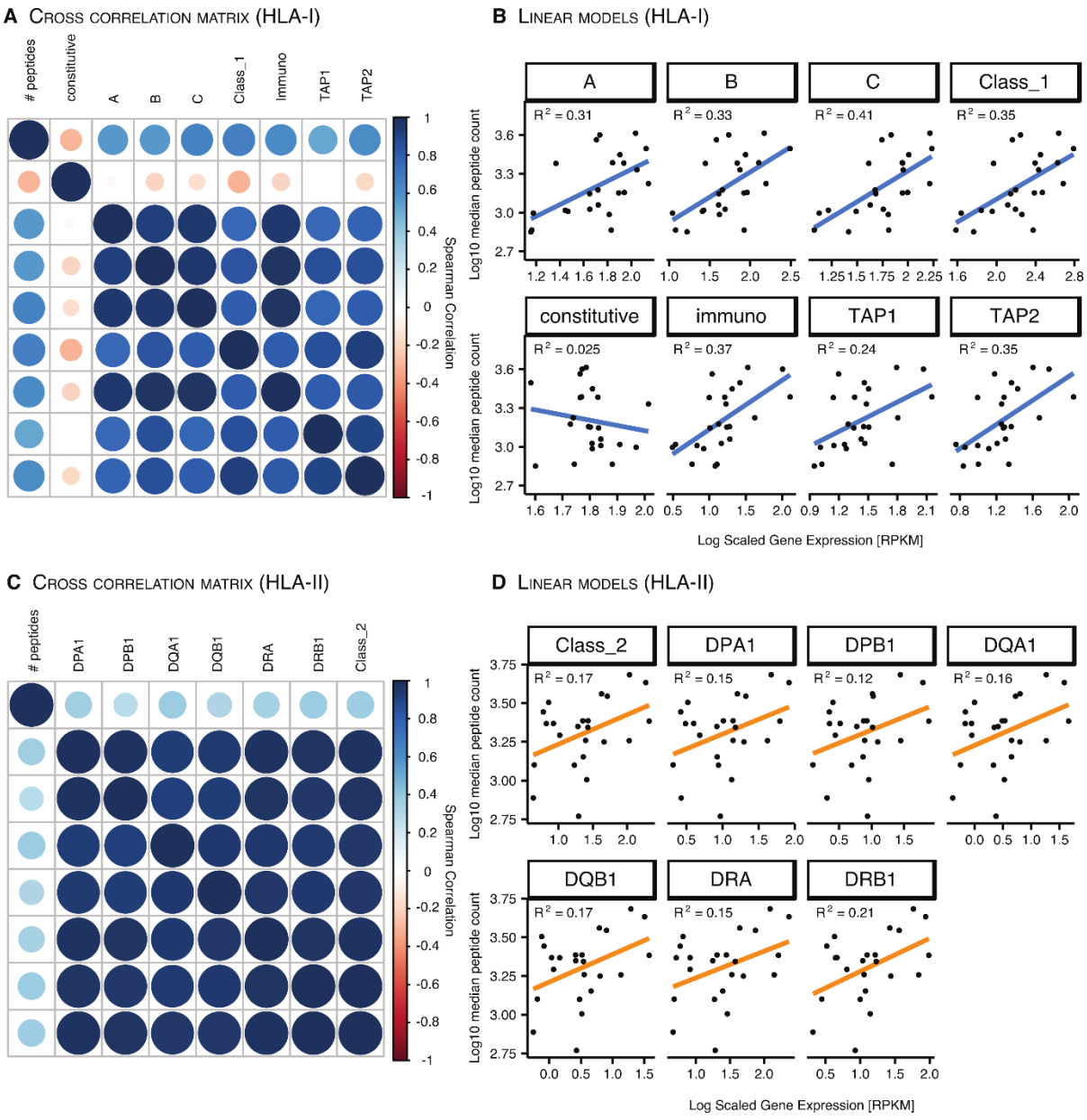

A Number of HLA-I peptides per donor
AUT
OVA
DN02 DN03 DN04 DN05 DN06 DN08 DN09 DN11 DN12 DN13 DN14 DN15 DN16 DN17 DN278 DN281
Myelon
Skin
Brain
Muscle
Heart
Testis
Aorta
Gallbladder
Prostate
Bladder
Ovary
Cerebellum
Liver
Thyroid
Colon
Mamma
Esophagus
Trachea
Small intestine
Pancreas
Tongue
Adrenal gland
Kidney
Lymph node
Uterus
Bone marrow
Spleen
Lung
Thymus
7,500
6,000
4,500
3,000
1,500
peptides
DN1 DN3 DN4 DN5 DN6
THY
B Number of HLA-II peptides per donor
AUT
DN02 DN03 DN04 DN05 DN06 DN08 DN09 DN11 DN12 DN13 DN14 DN15 DN17
10,000
8,000
6,000
4,000
2,000
peptides
DN1 DN3 DN4 DN5 DN6
THY
C Hierarchical clustering of samples based on peptide Jaccard similarity
HLA-I
Tissue
Subject
HLA-II
Tissue
Subject
Subjects
AUT-DN02 AUT-DN03 AUT-DN04 AUT-DN05 AUT-DN06 AUT-DN08 AUT-DN09 AUT-DN11 AUT-DN12 AUT-DN13 AUT-DN14 AUT-DN15 AUT-DN16 AUT-DN17 THY-DN1 THY-DN3 THY-DN4 THY-DN5 THY-DN6 OVA-DN278 OVA-DN281
Tissues
Thymus Adrenal gland Aorta Brain Esophagus Heart Skin Spleen Testis Trachea Cerebellum Kidney Liver Lung Ovary Pancreas Thyroid Uterus Bladder Bone marrow Colon Mamma Muscle Small intestine Tongue Prostate Lymph node Gallbladder Myelon
D Distribution of pairwise peptide Jaccard similarities
Jaccard Similarity
0.0
0.2
0.4
0.6
0.8
same subject different tissue
different subject same tissue
different subject different tissue
same subject different tissue
different subject same tissue
different subject different tissue

Supplemental Figure S4: Distribution of peptide identifications and yields across tissues and subjects and peptide-based hierarchical sample clustering (related to Figure 2D, E, and Figure 4A).

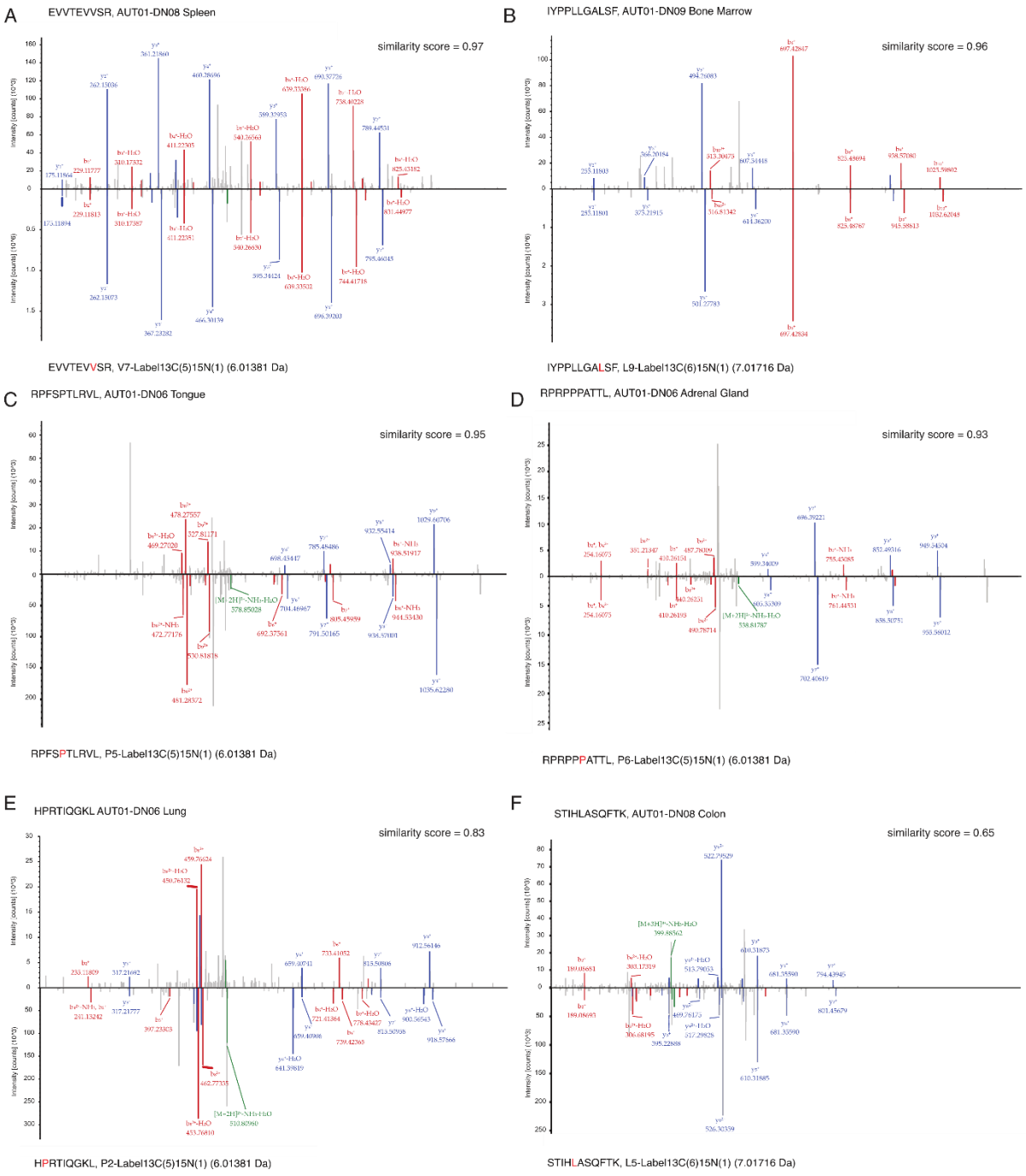
(A, B) HLA-I peptide yields per tissue and subject are illustrated in a heatmap for HLA-I and –II. The color range is in accordance with the number of peptides identified in each sample as indicated in the legend on the right (HLA-I – blue, HLA-II – orange).

(C) Pairwise hierarchical clustering of samples based on the Jaccard similarity between HLA-I (blue) and HLA-II (orange) peptides. The dendrogram illustrates the nearest neighbor based on the similarity between tissues and subjects. See Figure 2D.

(D) Violin plots illustrate the distribution of the peptide Jaccard similarity index for each pairwise comparison between the same subject - different tissues; different subjects - the same tissue, and different subject - different tissues. Results confirmed for HLA-I and -II source proteins Figure 2D and E, respectively.



Supplemental Figure 5: Linear models of gene expression across tissues (related to Figure 3).
The correlation of HLA-I and -II peptide yields across tissues was assessed with respect to the expression levels (RPKM) of multiple immune related genes of the HLA-I and -II antigen presentation pathway. (RNA expression values were taken from a different publication [178]
A, C) Symmetrical cross-correlation matrices were generated, illustrating the spearman correlation coefficients (rho) of the number of identified HLA-I and -II ligands with expression levels (RPKM) of relevant genes in the HLA-I and -II antigen presentation pathways and among each other. The color-coded dots and their size represent the degree (Spearman rho) of positive (blue) or negative correlation (red).
B, D) In addition, linear models illustrate HLA-I and -II immunopeptidome yields correlated to log scaled gene expression values (RPKM) of genes involved in the HLA-I and -II antigen processing pathway, respectively. The R^2 correlation coefficient is depicted in the top left corner for each comparison.



Supplemental Figure 6: Manual validation of cryptic peptide identifications with isotope labeled synthetic peptides (related to Figure 6).

A-F) Six exemplary spectral comparisons depict cryptic peptides identified in diverse tissues and subjects (upper spectrum) related to their synthetic isotope labeled counterpart (lower spectrum). The isotope labeled amino acid is highlighted in red and its corresponding label and mass are annotated. Matching b- (red), y-ions (blue) and neutral losses (green) and their corresponding fragment masses are annotated to each peak in the spectra. The spectral similarity score of the given comparison is annotated in the top right.

Supplemental Table S1: Patient Characteristics

Subject	Sex	Age (years)	# Tissues	Typing	HLA-A	HLA-B	HLA-C
AUT01-DN02	Female	56.7	9	WES + Optitype	11:01/68:01	15:01/35:03	03:03/04:01
AUT01-DN03	Male	80.6	18	WES + Optitype	01:01/11:01	15:01/35:01	03:03/04:01
AUT01-DN04	Female	47	13	WES + Optitype	02:01/23:01	27:05/50:01	02:02/06:02
AUT01-DN05	Male	89.8	7	WES + Optitype	01:01/11:01	07:02/49:01	07:01/07:02
AUT01-DN06	Female	60.1	22	WES + Optitype	03:01/68:02	07:02/14:02	07:02/08:02
AUT01-DN08	Male	59	22	SBT & Luminex	32:01/68:01	15:01/44:02	03:03/07:04
AUT01-DN09	Male	75.1	22	WES + Optitype	24:02/30:01	13:02/35:08	04:01/06:02
AUT01-DN11	Male	69.7	25	WES + Optitype	01:01/69:01	37:01/49:01	06:02/07:01
AUT01-DN12	Female	82.7	23	WES + Optitype	02:01/11:01	15:01/35:01	03:04/04:01
AUT01-DN13	Male	48.9	23	WES + Optitype	02:05/11:01	40:02/58:01	02:02/07:01
AUT01-DN14	Male	55.6	22	WES + Optitype	02:01/68:02	14:02/27:05	02:02/08:02
AUT01-DN15	Female	82.7	15	WES + Optitype	01:01/02:01	08:01/44:02	07:01/07:04
AUT01-DN16	Male	80.4	1	WES + Optitype	01:01/24:02	08:01/41:01	07:01/17:01
AUT01-DN17	Male	76.6	21	WES + Optitype	03:01/24:02	35:03/45:01	04:01/16:01
OvN278	Female	58	1	SBT & Luminex	02:01:01	44:02	05:01
OvN281	Female	62	1	SBT & Luminex	11:01/26:01	08:01/35:01	07:02/04:01
THY-DN1	Male	4 days	1	ngs-based typing*	03:01:01/29:02:01	07:02:01/44:03:01	07:02:01/16:01:01
THY-DN3	Female	4 months	1	ngs-based typing*	24:02:01/25:01:01	18:01:01/41:01:01	12:03:01/17:01:01
THY-DN4	Male	6 months	1	ngs-based typing*	02:01:01/26:08:01	15:01:01/44:02:01	03:04:01/05:01:01
THY-DN5	Male	2 months	1	ngs-based typing*	01:01:01/03:01:01	07:06:01/07:02:01	07:02:01/15:05:02
THY-DN6	Male	6 months	1	ngs-based typing*	01:01:01/25:01:01	13:02:01/39:01:01	06:02:01/12:03:01
GBM 16 Primary	Female	86.9	1	ngs-based typing*	02:01:01/29:02:01	07:02:01/44:02:01	05:01:01/07:02:01
GBM753 Recurrence	Male	72.4	1	ngs-based typing*	02:01:01/02:01:01	08:01:01/51:01:01	07:01:01/15:02:01
GBM 654Primary	Female	60.9	1	ngs-based typing*	01:01:01/31:01:02	51:01:01/57:01:01	06:02:01/15:06:01

*Histogenetics

Supplemental Table S1: Patient Characteristics (continued)

Subject	HLA-DRB1	HLA-DRB3	HLA-DRB4	HLA-DRB5	HLA-DQA1	HLA-DQB1	HLA-DPA1	HLA-DPB1
AUT01-DN02	04:01		01:03		03:02/03:01	03:01	01:03	04:02/04:01
AUT01-DN03	11:03/07:0	02:02	01:01		02:01/05:05	02:02/03:01	01:03	04:01/03:01
AUT01-DN04	13:01/07:0	01:01	01:01		02:01/01:03	02:02/06:03	02:02/01:03	04:01/03:01
AUT01-DN05	11:01/14:5	02:02			01:01/05:05	05:03/03:01	01:03	04:01
AUT01-DN06	13:03/08:0	01:01			04:01/05:05	04:02/03:01	01:03	04:02/04:01
AUT01-DN08	13:03/14:0				05:05	03:01		
AUT01-DN09	07:01		01:03		02:01	02:02	01:03	04:02/03:01
AUT01-DN11	13:02/07:0	03:01	01:03		02:01/01:02	06:04/03:03	01:03	04:01/03:01
AUT01-DN12	01:01/12:0	02:02			01:01/05:05	05:01/03:01	02:01/01:03	02:01/09:01
AUT01-DN13	15:01/10:0				01:01/01:02	05:01/06:02	01:03	02:01/04:02
AUT01-DN14	13:03/04:0	01:01	01:03		03:02/05:05	03:02/03:01	02:01/01:03	02:01
AUT01-DN15	11:01/03:0	02:02/01:01			05:01	02:01/03:01	02:01/01:03	01:01/03:01
AUT01-DN16	03:01/04:0				05:01/03:02	02:01/02:02		
AUT01-DN17	15:01/14:5	02:02			01:01/01:02	05:03/06:02	01:03	02:01
THY01-DN1	01:01/15:0			01:01	01:01/01:02	05:01/06:02	02:01/01:03	04:01/14:01
THY01-DN3	04:05/15:0		01:03	01:01	03:03/01:02	02:02/06:02	02:01/01:03	02:01/09:01
THY01-DN4	11:01/15:0	02:02		01:01	05:05/01:02	06:02/03:01	01:03	02:01/04:01
THY01-DN5	04:05/15:0		01:03	01:01	03:03/01:02	03:02/06:02	01:03	104:01/04:01
THY01-DN6	10:01/16:0			02:02	01:05/01:02	05:01/05:02	02:01/01:03	02:01/17:01
OVA01-DN278	11:01/13:0					03:01/06:03		
OVA01-DN281	04:04/11:0					03:02/03:01		
GBM161 Primary	12:01/15:0	02:02		01:01	01:02/05:01	03:01/06:02		
GBM753Recurrence	01:01/03:0	01:01			01:01/05:01	02:01/05:01		
GBM654Primary	04:04/07:0		01:03		02:01/03:01	03:03/03:02		

Supplemental Table S2: Sample Amount used for immunoprecipitation experiments in g.

Subject/Mass [g]	Aorta	Adrenal Gland	Bone Marrow	Brain	Cerebellum	Colon	Esophagus	Gallbladder	Heart	Kidney	Liver	Lung	Mamma	Lymph Node	Muscle	Ovary	Pancreas	Prostate	Skin	Small Intestine	Spinal Chord	Spleen	Stomach	Testis	Thymus	Thyroid	Tongue	Trachea	Urinary Bladder	Uterus
AUT01-DN02			n.a.	n.a.							4.5						2.3		2.3	4.1	3.1		2.6			0.8				
AUT01-DN03	0.7		n.a.	3.7		1.6		1.0	1.2	3.2	2.0		2.8				1.5		2.7	3.0		3.6	2.6	1		1.3		0.6		
AUT01-DN04			2.9	2.5				1.1	2.9	3.0	2.4					2.6	2.3		1.2			3.3				0.7				0.7
AUT01-DN05								2.6	1.0	2.4	2.8								0.4					1.0		1.0				
AUT01-DN06	2.1	n.a.	3.2	2.5	1.3	n.a.		2.3	2.8	0.9	2.3	0.4	1.9	1.9					0.7	2.4		1.2	1.9			n.a.	1.2	n.a.	0.9	
AUT01-DN08	0.3	1.0	1.0	1.1	1.1	0.9		0.7	0.6	0.7	1.0	1.1	0.8					1.0	0.4	0.8		1.9	1.0	1.1		0.8	0.8		0.4	
AUT01-DN09	1.1	1.4	1.4	1.1	1.1	1.1		1.1		4.7		1.0	1.1				1.1	1.1	0.6	1.2		4.1	1.8	0.9		0.8	0.8		1.1	
AUT01-DN11	1.1	1.0	1.0	0.7	1.1	1.1	1.0	0.7	1.1	1.2	1.3	0.5	1.0				1.1	1.1	0.3	1.0		1.1	1.2	1.1		0.8	1.0	1.1	1.1	
AUT01-DN12	1.0	0.5	1.2	1.0	1.1	1.1		0.9	1.0	1.1	1.2	0.2	0.7			0.7	1.2		0.8	0.9			1.2			0.9	0.9	0.7	1.1	1.0
AUT01-DN13	1.1	1.2	1.1	1.0	1.0	1.0		1.0	0.9	1.0	1.0	0.2	1.1				0.9	1.0	1.0	**0.9**		**1.1**	**1.1**	**0.9**		**1.1**	**0.9**	**1.1**	**1.0**	
AUT01-DN14	1.0	0.7	0.9	1.0		1.1	0.9		1.0	1.1	1.0	0.9	1.1				1.0	1.0		**1.0**		**1.1**	1.0	0.8		1.0	0.8	0.9	1.1	
AUT01-DN15	0.8	1.2	1.1	1.0	1.1	1.0	0.7	1.0		1.0	1.0											1.1	0.7			1.0			0.8	
AUT01-DN16	1.1	0.7	0.9		1.1	1.2	1.0	1.1	1.1	1.1	1.0		1.1					1.0	1.1	1.1		1.1		1.1		1.3	1.0	1.1	1.0	
AUT01-DN17			n.a.	n.a.							4.5						2.3		2.3	4.1	3.1		2.6			0.8				
THY-DN1																									1.0					
THY-DN3																									1.0					
THY-DN4																									1.1					
THY-DN5																									1.0					
THY-DN6																									1.1					
GBM 16 Primary				1.16																										
GBM753 Recurrence				2.48																										
GBM 654Primary				2.75																										

10 SUPPLEMENTAL DATA: MAMMA CARCINOMA

Table 10-1: Overview of available MaCa Samples and their preparation status.

Subject	Dignity	Mass IP [g]	Remainig mass [g]	MS	Tumor Subclass	Typing
MaCa 100	Tumor	1.03	0.53	Yes	Luminal A	SBT & Luminex
MaCa 101	Tumor	1.10	5.40	Yes	Luminal A	SBT & Luminex
MaCa 102	Tumor	1.40	1.30	Yes	Luminal A	SBT & Luminex
MaCa 104	Tumor	0.88	0.03	Yes	Luminal B, Her2-	SBT & Luminex
	Adj. benign	1.30	0.49	Yes		
MaCa 105	Tumor	2.60	2.75	Yes	Luminal A	SBT & Luminex
	Adj. benign	1.90	1.20	Yes		
MaCa 107	Tumor	1.50	0.30	Yes	Luminal B, Her2-	SBT & Luminex
	Adj. benign	1.50	0.87	Yes		
MaCa 108	Tumor	0.59	0.04	Yes	Luminal B, Her2-	Ghosh
	Adj. benign	0.83	0.12	Yes		
MaCa 109	Tumor	2.70	2.90	Yes	Luminal B, Her2-	SBT & Luminex
MaCa 110	Tumor	1.30	1.07	Yes	Luminal B, Her2-	SBT & Luminex
	Adj. benign	1.20	0.18	Yes		
MaCa 111	Tumor	0.20	0.05	Yes	Luminal A	SBT & Luminex
	Adj. benign	1.50	3.50	Yes		
MaCa 113	Tumor	1.90	0.94	Yes	Luminal A	Ghosh
	Adj. benign	1.50	3.50	Yes		
MaCa 114	Tumor	0.40	0.10	Yes	Luminal B, Her2-	SBT & Luminex
	Adj. benign	0.93	0.08	Yes		
MaN 115	Benign	2.19	2.11	Yes		SBT & Luminex
MaCa 116	Tumor	0.40	0.10	Yes	Luminal B, Her2-	Ghosh
	Adj. benign	0.93	0.80	Yes		
MaCa 121	Tumor	0.51	-	Yes	Her2 +	Ghosh
	Adj. benign	0.79	-	Yes		
MaN 122	Benign	7.50	26.51	Yes		SBT & Luminex
MaCa 123	Tumor	1.20	-	Yes	Luminal A	SBT & Luminex
	Adj. benign	3.50	-	Yes		
MaCa 125	Tumor	0.15	-	Yes	Luminal A	Ghosh
	Adj. benign	0.72	-	Yes		
MaCa 127	Tumor	1.66	3.14	Yes	TNBC	SBT & Luminex
MaN 128	Benign	3.36	-	Yes		SBT & Luminex
MaCa 129	Tumor	3.01	-	Yes	TNBC	SBT & Luminex
MaCa 135	Tumor	0.46	-	Yes	Luminal B, Her2-	Ghosh
	Adj. benign	0.16	-	Yes		
MaCa 138	Tumor	1.13	-	Yes	Luminal B, Her2-	Ghosh
	Adj. benign	1.18	-	Yes		
MaCa 144	Tumor	0.65	-	Yes	Luminal A	Ghosh
	Adj. benign	1.62	-	Yes		
MaCa 145	Tumor	2.45		Yes	Luminal B, Her2-	Ghosh
	Adj. benign	2.11	1.81	Yes		
MaCa 149	Tumor	1.17	-	Yes	Luminal A	Ghosh
	Adj. benign	1.38	-	Yes		
MaCa 151	Tumor	1.26	-	Yes	Luminal A	Ghosh
	Adj. benign	1.33	-	Yes		
MaCa 152	Tumor	1.05	-	Yes	Luminal B, Her2-	Ghosh
	Adj. benign	1.58	-	Yes		
MaCa 154	Tumor	0.55	-	Yes	TNBC	Ghosh
	Adj. benign	0.47	-	Yes		
MaCa 156 (TN)	Tumor	1.80	-	Yes	TNBC	Ghosh
	Adj. benign	0.87	-	Yes		
MaCa158	Tumor	0.39	-	Yes	Luminal A	Ghosh
	Adj. benign	0.76	-	Yes		
MaCa159	Tumor	0.39	-	Yes	Luminal A	Ghosh
	Adj. benign	0.71	-	Yes		
MaN160	Benign	4.10	21.00	Yes		Ghosh
MaN161	Benign	2.78	2.00	Yes		Ghosh
MaCa162	Tumor	0.95	-	Yes	Luminal A	Ghosh

Subject	Dignity	Mass IP [g]	Remainig mass [g]	MS	Tumor Subclass	Typing
	Adj. benign	2.16	-	Yes		
MaCa163	Tumor	0.36	-	Yes	Luminal A	Ghosh
	Adj. benign	1.13	-	Yes		

Table 10-2: MaCa associated HLA-I ligands

Nr.	Peptide	Gene Name	# MaCa samples	# other Tumors	HLA	Positive Samples	Molecular Subtype
1	YVYQNNIYL	FAP	12	24	A*02:01; C*02:02; C*03:04; C*07:02; C*14:02; C*16:01	MaCa149; MaCa138; MaCa159; MaCa102; MaCa113; MaCa145; MaCa109; MaCa114; MaCa107; MaCa152; MaCa154; MaCa156	5xLuminal A; 5xLuminal B Her2-; 2xTNBC
2	FLIDSSEGV	COL6A3	9	18	A*02:01	MaCa100; MaCa102; MaCa107; MaCa114; MaCa145; MaCa149; MaCa152; MaCa156; MaCa159	5xLuminal A; 3xLuminal B Her2-; 1xTNBC
3	SLLQATDFMSL	THY1	8	12	A*02:01	MaCa102; MaCa109; MaCa138; MaCa145; MaCa149; MaCa152; MaCa154; MaCa156	2xLuminal A; 4xLuminal B Her2-; 2xTNBC
4	ATEGVEVFY	GGT5	7	17	A*01:01	MaCa104; Maca110; MaCa114; MaCa123; MaCa145; MaCa151; MaCa162	2xLuminal A; 5xLuminal B Her2-
5	ILIRYEPQL	LAMB1	7	24	A*02:01	MaCa100; MaCa102; MaCa105; MaCa107; MaCa109; MaCa152; MaCa156	4xLuminal A; 2xLuminal B Her2-; 1xTNBC
6	LLDFADVTY	DSE	7	8	A*01:01	MaCa101; MaCa104; Maca110; MaCa123; MaCa144; MaCa151; MaCa162	4xLuminal A; 3xLuminal B Her2-
7	VTEFEDIKSGY	SET	7	13	A*01:01	MaCa101; MaCa104; MaCa129; MaCa135; MaCa144; MaCa151; MaCa154	3xLuminal A; 2xLuminal B Her2-; 2xTNBC
8	IALLPLLQA	NDUFA13	6	11	A*02:01; B*51:01; C*03:04; C*12:03; C*16:01	MaCa149MaCa105MaCa109MaCa100MaCa107MaCa145	4x Luminal A; 3xLuminal B Her2-;
9	YTAEQLVYL	RNF213	6	16	A*02:01; C*02:02; C*03:04; C*05:01; C*16:01	MaCa149MaCa152MaCa114MaCa107MaCa138MaCa154	2x Luminal A; 3x 4xLuminal B Her2-; 1xTNBC
10	KAYEQALQY	FOXA1	6	15	A*03:01; B*15:01; B*15:17; C*12:03	MaCa105MaCa144MaCa151MaCa101MaCa129MaCa163	5x Luminal A; 1xTNBC
11	SVGDTFLYL	GLB1L	6	8	C*03:04; C*05:01; C*07:02; C*16:01	MaCa149MaCa104MaCa145MaCa102MaCa152MaCa109	2x Luminal A; 4xLuminal B Her2-
12	TATTSQPVL	ABR	6	27	C*03:03; C*03:04; C*16:01	MaCa149MaCa129MaCa100MaCa101MaCa114MaCa151	4x Luminal A; 1xLuminal B Her2-; 1xTNBC
13	IAQGSYIAL	NOMO1; NOMO2; NOMO3	6	21	C*03:03; C*03:04	MaCa100MaCa101MaCa109MaCa114MaCa129MaCa151	3x Luminal A; 2x Luminal B Her2-; 1x TNBC
14	AILETAPKEV	RRBP1	6	26	A*02:01	MaCa102; MaCa109; MaCa138; MaCa145; MaCa152; MaCa154	1x Luminal A; 4xLuminal B Her2-; 1xTNBC
15	DSELQREGVSHY	PIGT	6	15	A*01:01	MaCa101; Maca110; MaCa114; MaCa135; MaCa144; MaCa154	2x Luminal A; 3xLuminal B Her2-; 1x TNBC
16	FLLSLRGAGA	HLA-DPA1	6	15	A*02:01	MaCa102; MaCa105; MaCa107; MaCa109; MaCa138; MaCa156	3x Luminal A; 2xLuminal B Her2-; 1x TNBC
17	GLGELAGLTV	STRN	6	26	A*02:01	MaCa107; MaCa109; MaCa138; MaCa145; MaCa149; MaCa152	2x Luminal A; 4xLuminal B Her2-
18	GLWEDGRSTLL	CREB3L1	6	22	A*02:01	MaCa100; MaCa102; MaCa104; MaCa107; MaCa109; MaCa149	4x Luminal A; 2x Luminal B Her2-
19	TTDIIEKY	DDX60L	6	25	A*01:01	MaCa101; Maca110; MaCa144; MaCa151; MaCa154; MaCa162	3x Luminal A; 2xLuminal B Her2-; 1x TNBC
20	VTDESIPSY	AKAP9	6	16	A*01:01	MaCa101; Maca110; MaCa123; MaCa144; MaCa145; MaCa151	4x Luminal A; 2xLuminal B Her2-
21	IFYLKLEDL	NR4A1; NR4A2; NR4A3	5	23	B*08:01; C*07:01; C*07:02; C*12:03; C*14:02	MaCa154; MaCa156; MaCa159; MaCa105; MaCa107	3x Luminal A; 2x TNBC
22	AETEARFGAQL	KRT19	5	21	B*40:01; B*40:02; B*44:02	MaCa100; MaCa101; MaCa114; MaCa107; MaCa149	4x Luminal A; 1xLuminal B Her2-
23	LSFVDTRTL	COL1A2	5	29	B*57:01; C*03:04; C*15:02	MaCa104; MaCa123; MaCa145; MaCa101; MaCa111	3x Luminal A; 2x Luminal B Her2-;
24	VVDKTLLLV	C11orf24	5	16	A*01:01; C*04:01; C*05:01	MaCa145; MaCa113; MaCa149; MaCa102; MaCa152	3x Luminal A; 2x Luminal B Her2-;

Nr.	Peptide	Gene Name	# MaCa samples	# other Tumors	HLA	Positive Samples	Molecular Subtype
25	ALPPPPPEL	RAPH1	5	5	A*02:01; C*04:01	MaCa100; MaCa109; MaCa138; MaCa145; MaCa113	2x Luminal A; 3xLuminal B Her2-
26	FLLDGSINF	COL6A3	5	9	A*02:01; C*02:02	MaCa104; MaCa105; MaCa109; MaCa145; MaCa107	2x Luminal A; 3x Luminal B Her2-;
27	ASEASRLAHY	H2BC1; H2BC13; H2BU1	5	35	A*01:01	MaCa101; MaCa129; MaCa135; MaCa145; MaCa162	1x Luminal A; 3x Luminal B Her2-; 1x TNBC;
28	ATEEEILVY	PSAP	5	4	A*01:01	MaCa101; MaCa123; MaCa144; MaCa14; 5MaCa154	3x Luminal A; 1x 1xLuminal B Her2-; 1x TNBC;
29	FLDALDHAGY	DDX58	5	7	A*01:01	MaCa101; Maca110; MaCa123; MaCa144; MaCa162	3x Luminal A; 2x Luminal B Her2-
30	FLLDGSEGV	COL6A3	5	7	A*02:01	MaCa100; MaCa145; MaCa149; MaCa152; MaCa156	2x Luminal A; 2x Luminal B Her2-; 1xTNBC;
31	FQDQLHQLY	STAT2	5	8	A*01:01	MaCa101; MaCa123; MaCa129; MaCa151; MaCa162	3x Luminal A; 1x Luminal B Her2-; 1x TNBC
32	GLLDQDTGLVL	MACF1	5	16	A*02:01	MaCa107; MaCa109; MaCa113; MaCa138; MaCa145	2x Luminal A; 3x Luminal B Her2-
33	GLLGAGGTVSV	TIMM50	5	20	A*02:01	MaCa102; MaCa107; MaCa109; MaCa145; MaCa152	2x Luminal A; 3x Luminal B Her2-;
34	GLTDNIHLV	MXRA5	5	9	A*02:01	MaCa100; MaCa102; MaCa152; MaCa156; MaCa159	3x Luminal A; 1x Luminal B Her2-; 1x TNBC
35	HSDQITASSQY	NRP1	5	7	A*01:01	MaCa101; MaCa123; MaCa129; MaCa145; MaCa162	2x Luminal A; 2x Luminal B Her2-; 1x TNBC
36	LLAGQTYHV	COL6A3	5	7	A*02:01	MaCa102; MaCa107; MaCa109; MaCa145; MaCa152	2x Luminal A; 3x Luminal B Her2-;
37	LLFNDVQTL	PLEC	5	14	A*02:01	MaCa102; MaCa107; MaCa109; MaCa138; MaCa152	2x Luminal A; 3x Luminal B Her2-
38	LLQEEVTKV	POSTN	5	5	A*02:01	MaCa102; MaCa109; MaCa138; MaCa149; MaCa152	2x Luminal A; 3x Luminal B Her2-
39	NSDGYGGNY	DDX3X; DDX3Y	5	29	A*01:01	MaCa101; MaCa129; MaCa144; MaCa145; MaCa162	2x Luminal A; 2xLuminal B Her2-; 1x TNBC
40	SLIEDLILL	SMYD3	5	13	A*02:01	MaCa105; MaCa107; MaCa109; MaCa152; MaCa154	2x Luminal A; 2x Luminal B Her2-; 1x TNBC;
41	SLLEHFNTV	TRPS1	5	5	A*02:01	MaCa102; MaCa107; MaCa109; MaCa138; MaCa152	2x Luminal A; 3xLuminal B Her2-
42	SLLTEPALV	TBC1D9	5	12	A*02:01	MaCa102; MaCa138; MaCa145; MaCa149; MaCa152	2x Luminal A; 3x Luminal B Her2-; 1x TNBC
43	YSDVAKGILQY	SPATA20	5	15	A*01:01	MaCa101; MaCa114; MaCa135; MaCa151; MaCa154	2x Luminal A; 2x Luminal B Her2-; 1x TNBC;
44	YSEETLRARFY	IMMT	5	10	A*01:01	MaCa104; MaCa110; MaCa123; MaCa135; MaCa162	1x Luminal A; 4x Luminal B Her2-;

Table 10-3: HLA-I MaCa associated HLA-I source proteins

Nr.	Gene Name	Peptide	# MaCa samples	# other Tumors	HLA	Positive Samples	Molecular Subtypes
1	SMAD9; SMAD9L	ISDTLGQVY	8	24	A*01:01	MaCa101; MaCa110; MaCa114; Maca123; MaCa135; MaCa145; MaCa151; MaCa162	3x Luminal A; 5x Luminal B Her2-
2	RASGEF1B; RASGEF1C	LLIKDIYFL	7	30	A*02:01	MaCa109; MaCa138; MaCa145; MaCa149; MaCa152; MaCa154; MaCa156	1x Luminal A; 4x Luminal B Her2-; 2x TNBC
3	GAL3ST1; GAL3ST3; GAL3ST4	KAYEQALQY	6	17	A*03:01; B*15:01; B*15:17; C*12:03	MaCa101; MaCa105; MaCa129; MaCa144; MaCa151; MaCa163	4x Luminal A; 1x TNBC
4	PLSCR1; PLSCR2	ILIHQQIEL	6	28	A0201	MaCa102; MaCa107; MaCa109; MaCa114; MaCa138; MaCa145	2x Luminal A; 4x Luminal B Her2-
5	KCNJ3; KCNJ5	YLSDLFTTL	6	1	C*02:02; A*02:01	MaCa107; MaCa109; MaCa138; MaCa145; MaCa149; MaCa156	2x Luminal A; 3x Luminal B Her2-; 1x TNBC
6	KAT5; KAT7	LSDLGLLSY	6	20	A*01:01	Maca123; MaCa129; MaCa135; MaCa144; MaCa145; MaCa154	2x Luminal A; 2x Luminal B Her2-; 2x TNBC
7	FOXA1	YFDESLVLL	6	20	C*04:01; C*05:01	MaCa113; MaCa116; MaCa125; MaCa127; MaCa144; MaCa149	4x Luminal A; 1x Luminal B Her2-; 1x TNBC
8	CTSD; CTSE; PGA3; PGA4; PGA5; PGC	VVDKTLLLV	5	22	A*01:01; C*04:01; C*05:01	MaCa102; MaCa113; MaCa145; MaCa149; MaCa152	2x Luminal A; 2x Luminal B Her2-
9	NR4A1; NR4A2; NR4A3	EEVDVPIKLY; EEVDVPIKL	5	23	B*44:02; B*44:03	MaCa101; MaCa110; MaCa149; MaCa152; MaCa156	2x Luminal A; 2x Luminal B Her2-; 1 TNBC
10	EPHB1; EPHB2	YRPLTVLTF	5	41	C*07:01; C*07:02	MaCa102; MaCa109; MaCa145; MaCa156; MaCa163	2x Luminal A; 1x Luminal B Her2-; 2x TNBC
11	TMTC3; TMTC4	IFYLKLEDL	5	21	B*08:01; C*07:01; C*07:02; C*12:03; C*14:02	MaCa105; MaCa107; MaCa154; MaCa156; MaCa159	3x Luminal A; 2x TNBC
12	C11orf24	AIVDTGTSL	5	11	C*03:03; C*03:04	MaCa100; MaCa101; MaCa109; MaCa129; MaCa151	3x Luminal A; 1x Luminal B Her2-; 1x TNBC
13	AQP5	AIAGAGILYGV; FYLLFPNSL; LSLSERVAI; SALPTILQI	4	9	C*16:01	MaCa110; MaCa127; MaCa138; MaCa154	2x Luminal B Her2neg; 2x TNBC
14	CD276	GEGEGSKTAL	4	26	B*40:01; B*40:02	MaCa100; MaCa101; MaCa107; MaCa114	3x Luminal A; 1x Luminal B Her2neg
15	PODNL1	FYLLFPNSL; LSLSERVAI; SALPTILQI	4	8	B*35:02; B*35:02	MaCa127; MaCa129; MaCa145; MaCa149	1x Luminal A; 1x Luminal B Her2-; 2 TNBC
16	OR13C3	LDKLISLFY	4	61	A*01:01	MaCa101; MaCa110; MaCa114; Maca123	2x Luminal A; 2x Luminal B Her2-
17	DCLK3	EPDAALMIMDL; SEILIIQSL; SPRNPTQEL	4	16	C*05:01	MaCa109; MaCa144; MaCa145; MaCa154	1x Luminal A; 2x Luminal B Her2-; 1 TNBC
18	KIF26A; KIF26B	YEIDDVERL	4	12	B*40:01; B*40:02	MaCa100; MaCa101; MaCa107; MaCa114	1x Luminal A; 3x Luminal B Her2-
19	COL1A1; COL1A2; COL2A1	AEGNSRFTY	4	8	B*44:02; B*44:03	MaCa102; MaCa110; MaCa149; MaCa152	2x Luminal A; 2x Luminal B Her2-
20	TUBB2A; TUBB2B; TUBB4A; TUBB4B; TUBB6; TUBB8	LQLERINVY; QLERINVYY	4	37	B*15:01	MaCa151; MaCa129; MaCa114; MaCa109	1x Luminal A; 2x Luminal B Her2neg; 1xTNBC
21	CBWD1; CBWD2; CBWD3; CBWD5; CBWD6	IYLDGIITI; NEATRQVAL	4	19	B*14:01; B*14:02	MaCa110; Maca123; MaCa127	1x Luminal A; 1x Luminal B Her2-; 1x TNBC
22	MATN3	AFQDKVSSY; EPLEEHVFY	3	22	C*14:02	MaCa111; MaCa116; MaCa138	1x Luminal A; 2x Luminal B He2neg
23	COMP	TESQVRLLW; RELQETNAAL	3	59	B*44:02	MaCa100; MaCa101; MaCa152	2x Luminal A; 1x Luminal B Her2-

Nr.	Gene Name	Peptide	# MaCa samples	# other Tumors	HLA	Positive Samples	Molecular Subtypes
24	LY6E	AADGGLRASVTL; LLLSLLPAL	3	35	A*02:01	MaCa107; MaCa109; MaCa152	1x Luminal A; 2x Luminal B Her2-
25	ARL6IP4	VEALPGPSL	3	19	B*40:01	MaCa100; MaCa101; MaCa114	2x Luminal A; 1x Luminal B Her2-
26	SLC39A4	SLYEDDITF	3	14	B*15:01	MaCa109; MaCa129; MaCa151	1x Luminal A; 1x Luminal B Her2-; 1x TNBC
27	TMEM219	HEGLVLTKL; SHEGLVLTKL	3	12	B*40:01	MaCa101; MaCa114; MaCa144	2x Luminal A; 1x Luminal B her2neg
28	SMPDL3B	YYVYNSVSY; IYNQIAELW	3	14	A*33:01	MaCa110; MaCa127; MaCa162	2x Luminal B Her2-; 1x TNBC
29	CTHRC1	EVVDLYNGM; SPQRLRGLL; VLFSGSLRL	3	11	A*26:01	MaCa105; MaCa109; MaCa145	1x Luminal A; 2x Luminal B Her2-
30	MS4A6E	LSLMLVSTV	3	16	A*02:01; B*51:01; C*05:01	MaCa138; MaCa152; MaCa154	2x Luminal B Her2-; 1x TNBC
31	CMTM3	AESGLSFITF; ISITAIAKY	3	6	B*44:02	MaCa101; MaCa109; MaCa152	1x Luminal A; 2x Luminal B Her2-
32	C1QTNF6	NQKEAVILY	3	8	B*15:01	MaCa109; MaCa129; MaCa151	1x Luminal A; 1x Luminal B Her2-; 1 TNBC
33	CHIA	LLTGLVLIL	3	6	A*02:01	MaCa102; MaCa109; MaCa159	2x Luminal A; 1x Luminal B Her2-
34	MANBAL	RYGLFLGAIF; SPPEVPEPTF	3	13	B*07:02; B3502	MaCa113; MaCa127; MaCa129	1x Luminal A; 2x TNBC
35	ZNF358	KAFGQASSL; KAFGQSSAL	3	16	C*03:04, C0701	MaCa100; MaCa101; MaCa135	2x Luminal A; 1x Luminal B Her2-
36	TBC1D7	QYLDVLHAL; SVYYEKVGF	3	17	B1501	MaCa127; MaCa129; MaCa162	1x Luminal B Her2-; 2x TNBC
37	LZTS1	SPESASHQL	3	22	B*07:02; B*35:02; B*35:01	MaCa113; MaCa127; MaCa129	1x Luminal A; 2x TNBC
38	LOX L2	FGFPGERTY	3	26	C*12:03	MaCa105; MaCa144; MaCa151	3x Luminal A
40	TBL1X; TBL1Y	TSDEVNFLVY	3	172	A*01:01	MaCa129; MaCa144; MaCa154	1x Luminal A; 2x TNBC
41	LYPLA1; LYPLA2	FSQGGALSL	3	7	C*03:04	MaCa100; MaCa101; MaCa109	1x Luminal A; 2x Luminal B Her2-
42	ABCA1; ABCA2	GRSVVLTSH; TVEEHIWFY	3	3	A*01:01	MaCa162; MaCa144; MaCa101	2x Luminal A, 1x Luminal B Her2-
43	AK1; AK5	FLIDGYPREV	3	16	A*02:01	MaCa102; MaCa104; MaCa152	1x Luminal A; 2x Luminal B Her2-
44	C1QB; C1QC	FTCKVPGLYY	3	10	A*01:01	MaCa101; MaCa110; MaCa162	1x Luminal A; 2x Luminal B Her2-
45	HSPA1A; HSPA1B	TRIPKVQKL	3	117	C*06:02	Maca123; MaCa138; MaCa145	1x Luminal A; 2x Luminal B Her2-
46	FPR2; FPR3	LPTSLERAL	3	5	B*07:02; B*35:01; B*35:01	MaCa113; MaCa127; MaCa129	1x Luminal A; 2x TNBC
47	KDM5A; KDM5C	MPVLEQSVL	3	29	B*35:02; B*35:01	MaCa127; MaCa129; MaCa149	1x Luminal A; 2x TNBC
48	PRCP; DPP7	ALADFAELI	3	19	A*02:01	MaCa100; MaCa149; MaCa159	3x Luminal A
49	KRT17; KRT19	AADDFRTKF; RLAADDFRTKF	3	86	C*05:01	MaCa152; MaCa144; MaCa102	1x Luminal A; 1x LuminalB Her2-; 1x TNBC
50	HERC3; HERC6	SLEGIPLAQV	3	16	A*02:01	MaCa114; MaCa138; MaCa152	3x Luminal B Her2-
51	IFI16; NOC3L	MSKLISEM	3	37	C*12:03	MaCa105; MaCa144; MaCa151	3x Luminal A
52	ITCH; WWP2	VTEENKEEY	3	13	A*01:01	MaCa101; MaCa145; MaCa162	1x Luminal A; 2x Luminal B Her2-
53	ABCA2; ABCA1	TVEEHLWFY	3	11	A*01:01	MaCa110; MaCa135; MaCa145	3x Luminal A
54	PARVB; PARVA	VLLDWINDV	3	10	A*02:01	MaCa100; MaCa107; MaCa109	2x Luminal A; 1x Luminal B Her2-
55	SLC23A1; SLC23A2	SRNLFVLGF; TLFGMITAV	3	21	A*02:01	MaCa162; MaCa154; MaCa145	2x Luminal B Her2-; 1x TNBC
56	CADPS; CADPS2	TLDHRLNDSY	3	19	A*01:01	MaCa101; MaCa135; MaCa145	1x Luminal A; 2x Luminal B Her2-
57	VAMP1; VAMP2; VAMP3	SELDDRADAL	3	14	B*40:01	MaCa100; MaCa101; MaCa114	2x Luminal A; 1x Luminal B Her2-

Nr.	Gene Name	Peptide	# MaCa samples	# other Tumors	HLA	Positive Samples	Molecular Subtypes
58	SRSF4; SRSF5; SRSF6	DADDAVYEL	3	8	B*35:02	MaCa113; MaCa127; MaCa129	1x Luminal A; 2xTNBC
59	GBP4; GBP7; GBP6	FFPDFIWTV	3	8	C*06:02	MaCa105; MaCa107; MaCa129	2x Luminal A; 1 TNBC
60	IGKV1-17; IGKV1D-16; IGKV1-17; IGKV1-39	LQSGVPSRF	3	10	B*38:01	MaCa109; MaCa144; MaCa151	2x Luminal A; 1x Luminal B Her2-
61	MLH1; MLH3; PMS1; SLC25A42	YGFRGEAL	3	8	B*08:01	MaCa101; MaCa109; MaCa114	1x Luminal A; 2x Luminal B Her2-
62	TBXT; EOMES; TBR1; TBX19; TBX21	TQFIAVTAY	3	11	B*15:01	MaCa109; MaCa129; MaCa151	1x Luminal A; 1x Luminal B Her2-; 1x TNBC

Table 10-4: MaCa-associated HLA-II Ligands

Nr.	Peptide	Gene Name	# MaCa samples	# other Tumors	Positive Samples	Molecular Subtypes
1	GNWKIIRSENFEEL	CRABP2	12	28	MaCa100; MaCa101; MaCa102; MaCa104; MaCa107; MaCa109; MaCa110; MaCa114; MaCa123; MaCa127; MaCa159; MaCa163;	7x Luminal A; 4x Luminal B, Her2-; 1x TNBC
2	DTHAYNVADFESL	COL12A1	9	24	MaCa100; MaCa101; MaCa102; MaCa104; MaCa107; MaCa109; MaCa110; MaCa114; MaCa121; MaCa123; MaCa127; MaCa156; MaCa159;	6x Luminal A; 4x Luminal B, Her2-; 1x Her2+; 2x TNBC
3	KPSRLPFLDIAPLDIGG	COL1A2	9	23	MaCa101; MaCa108; MaCa109; MaCa121; MaCa123; MaCa129; MaCa138; MaCa151; MaCa162;	3x Luminal A; 4x Luminal B, 1x Her2-; Her2+
4	RHVFIVDDFESFEK	COL12A1	9	28	MaCa100; MaCa101; MaCa102; MaCa104; MaCa107; MaCa109; MaCa123; MaCa145; MaCa156;	5x Luminal A; 3x Luminal B, Her2-; 1x TNBC
5	RLPIIDLAPVDVGGTD	COL5A2	9	28	MaCa101; MaCa109; MaCa111; MaCa121; MaCa129; MaCa138; MaCa151; MaCa158; MaCa162;	4x Luminal A; 3x Luminal B, Her2-; 1x Her2+
6	APAVVHIELFRKLPFSKRE	HTRA1	8	15	MaCa105; MaCa110; MaCa111; MaCa121; MaCa125; MaCa138; MaCa144; MaCa151;	5x Luminal A; 2x Luminal B, Her2-; 1x Her2+
7	DQPRSAPSLRPKDYEVDATLKSLNN	COL1A2	8	26	MaCa100; MaCa107; MaCa109; MaCa113; MaCa129; MaCa145; MaCa151; MaCa162;	4x Luminal A; 3x Luminal B, Her2-; 1x TNBC
8	DTHAYNVADFESLSR	COL12A1	8	23	MaCa100; MaCa101; MaCa102; MaCa104; MaCa107; MaCa113; MaCa135; MaCa156;	5x Luminal A; 2x Luminal B, Her2-; 1x TNBC
9	GNWKIIRSENFEELLK	CRABP2	8	19	MaCa100; MaCa101; MaCa104; MaCa109; MaCa110; MaCa123; MaCa159; MaCa163;	5x Luminal A; 3x Luminal B, Her2-
10	KPSRLPFLDIAPLDIGGADQ	COL1A2	8	33	MaCa101; MaCa109; MaCa121; MaCa129; MaCa138; MaCa151; MaCa158; MaCa162;	3x Luminal A; 3x Luminal B, Her2-; 1x Her2+; 1x TNBC
11	LPIIDLAPVDVGGT	COL5A2	8	25	MaCa101; MaCa109; MaCa121; MaCa129; MaCa138; MaCa151; MaCa158; MaCa162;	3x Luminal A; 3x Luminal B, Her2-; 1x Her2+; 1x TNBC
12	LPIIDLAPVDVGGTD	COL5A2	8	36	MaCa101; MaCa111; MaCa121; MaCa129; MaCa138; MaCa151; MaCa158; MaCa162;	4x Luminal A; 2x Luminal B, Her2-; 1x Her2+; 1x TNBC
13	PGVPAYRFVRFDYIPPV	THBS2	8	27	MaCa100; MaCa101; MaCa104; MaCa107; MaCa109; MaCa110; MaCa121; MaCa156;	3x Luminal A; 3x Luminal B, Her2-; 1x Her2+; 1x TNBC
14	SRLPIIDVAPLDVGAPDQE	COL1A1	8	51	MaCa101; MaCa109; MaCa111; MaCa129; MaCa138; MaCa151; MaCa158; MaCa162;	4x Luminal A; 3x Luminal B, Her2-; 1x TNBC

Nr.	Peptide	Gene Name	# MaCa samples	# other Tumors	Positive Samples	Molecular Subtypes
15	ARLPIIDLAPVDVGGTD	COL5A2	6	23	MaCa101; MaCa109; MaCa111; MaCa121; MaCa138; MaCa158;	3x Luminal A; 2x Luminal B, Her2-; 1x Her2+
16	DGNPRVRSLVQDTRIHLVPSLNPDG	AEBP1	6	11	MaCa100; MaCa104; MaCa114; MaCa154; MaCa156; MaCa162;	1x Luminal A; 3x Luminal B, Her2-; 2x TNBC
17	DPDDTHAYNVADFESLSRIV	COL12A1	6	13	MaCa100; MaCa101; MaCa104; MaCa107; MaCa109; MaCa110;	3x Luminal A; 3x Luminal B, Her2-
18	DQPRSAPSLRPKDYEVDATLKS	COL1A2	6	22	MaCa107; MaCa109; MaCa116; MaCa129; MaCa151; MaCa162;	2x Luminal A; 3x Luminal B, Her2-; 1x TNBC
19	DRDLEVDTTLKSLS	COL1A1	6	19	MaCa101; MaCa109; MaCa111; MaCa129; MaCa145; MaCa162;	2x Luminal A; 3x Luminal B, Her2-; 1x TNBC
20	EITYINRDTKIILETK	SFRP2	6	16	MaCa100; MaCa104; MaCa114; MaCa154; MaCa156; MaCa162;	1x Luminal A; 3x Luminal B, Her2-; 2x TNBC
21	ENDVIISINGQSVVS	HTRA1	6	18	MaCa105; MaCa108; MaCa116; MaCa123; MaCa127; MaCa135;	2x Luminal A; 3x Luminal B, 1x TNBC
22	EQHLYYQDQLLPVSR	TPSAB1, TPSD1	6	14	MaCa100; MaCa113; MaCa114; MaCa116; MaCa135; MaCa159;	3x Luminal A; 3x Luminal B, Her2-
23	FPSYSAYRIQKNAFVNQPT	COL12A1	6	6	MaCa105; MaCa114; MaCa116; MaCa135; MaCa138; MaCa156;	1x Luminal A; 4x Luminal B, 1x TNBC
24	GAIHIFREIIKPAEKSLH	POSTN	6	8	MaCa102; MaCa107; MaCa116; MaCa125; MaCa149; MaCa151;	5x Luminal A; 1x Luminal B, Her2-
25	GDGRTIVDLEGTPVVSPD	FNDC1	6	12	MaCa104; MaCa105; MaCa116; MaCa123; MaCa127; MaCa135;	2x Luminal A; 3x Luminal B, Her2-; 1x TNBC
26	IAVGYVDDTQFVR	HLA-A, HLA-B, HLA-C	6	19	MaCa100; MaCa107; MaCa110; MaCa111; MaCa151; MaCa162;	4x Luminal A; 2x Luminal B, Her2-
27	IDENTVHMSWAKPVDPI	COL12A1	6	26	MaCa102; MaCa123; MaCa145; MaCa149; MaCa151; MaCa158;	5x Luminal A; 1x Luminal B, Her2-
28	IEYKTTKTSRLPII	COL1A1	6	28	MaCa121; MaCa123; MaCa138; MaCa145; MaCa151; MaCa158;	3x Luminal A; 2x Luminal B, Her2-; 1x Her2+
29	ITYINRDTKIIL	SFRP2	6	15	MaCa100; MaCa104; MaCa114; MaCa154; MaCa156; MaCa162;	1x Luminal A; 2x Luminal B, Her2-; 2x TNBC
30	KADIALIKIDHQGKLP	HTRA1	6	15	MaCa101; MaCa111; MaCa121; MaCa138; MaCa151; MaCa158;	4x Luminal A; 1x Luminal B, Her2-; 1x Her2+
31	KPLVIIAEDVDGEAL	HSPD1	6	19	MaCa101; MaCa109; MaCa121; MaCa129; MaCa151; MaCa162;	2x Luminal A; 2x Luminal B, Her2-; 1x Her2+; 1x TNBC
32	KTTKTSRLPIIDVAPLDVGAPD	COL1A1	6	29	MaCa101; MaCa111; MaCa121; MaCa129; MaCa138; MaCa158;	3x Luminal A; 1x Luminal B, Her2-; 1x Her2+; 1x TNBC
33	LPDETEVVEETVAEVTE	SPARC	6	19	MaCa100; MaCa123; MaCa127; MaCa138; MaCa151; MaCa162;	3x Luminal A; 2x Luminal B, Her2-; 1x TNBC
34	LPFLDIAPLDIGG	COL1A2	6	7	MaCa101; MaCa109; MaCa121; MaCa138; MaCa151; MaCa162;	2x Luminal A; 3x Luminal B, Her2-; 1x Her2+
35	LPIIDVAPLD	COL1A1	6	29	MaCa101; MaCa121; MaCa138; MaCa151; MaCa158; MaCa162;	3x Luminal A; 2x Luminal B, Her2-; 1x Her2+
36	PIIDVAPLDVGAPDQE	COL1A1	6	28	MaCa101; MaCa121; MaCa138; MaCa151; MaCa158; MaCa162;	3x Luminal A; 2x Luminal B, Her2-; 1x Her2+
37	SSPVVIDASTAIDAPS	FN1	6	36	MaCa100; MaCa101; MaCa105; MaCa107; MaCa114; MaCa162;	4x Luminal A; 2x Luminal B, Her2-
38	TGSMSIIFFLPLK	SERPINF1	6	53	MaCa101; MaCa104; MaCa110; MaCa135; MaCa144; MaCa156;	2x Luminal A; 3x Luminal B, Her2-; 1x TNBC
39	TKTSRLPIIDVAPLDVGAPD	COL1A1	6	21	MaCa101; MaCa111; MaCa121; MaCa129; MaCa151; MaCa158;	4x Luminal A; 3x Luminal B, Her2-; 1x Her2+; 1x TNBC
40	TNKPSRLPFLDIAPLDIGGADQ	COL1A2	6	30	MaCa101; MaCa109; MaCa121; MaCa129; MaCa151; MaCa162;	2x Luminal A; 2x Luminal B, Her2-; 1x Her2+; 1x TNBC
41	TTKTSRLPIIDVAPLDVGAPD	COL1A1	6	32	MaCa101; MaCa109; MaCa138; MaCa151; MaCa162;	2x Luminal A; 3x Luminal B, Her2-

Nr.	Peptide	Gene Name	# MaCa samples	# other Tumors	Positive Samples	Molecular Subtypes
42	VKEYILSYAPALKPF	FNDC1	6	10	MaCa100; MaCa107; MaCa114; MaCa135; MaCa156; MaCa159;	3x Luminal A; 2x Luminal B, Her2-; 1x TNBC
43	VPISDTIIPAVPPP	FN1	6	23	MaCa100; MaCa101; MaCa107; MaCa113; MaCa114; MaCa127;	4x Luminal A; 1x Luminal B, Her2-; 1x TNBC
44	VPLDPKGTTQIDPNWVIR	THBS2	6	15	MaCa100; MaCa104; MaCa105; MaCa107; MaCa110; MaCa114;	3x Luminal A; 3x Luminal B, Her2-
45	VRDRDLEVDTTLKSLSQQ	COL1A1	6	28	MaCa100; MaCa101; MaCa104; MaCa129; MaCa145; MaCa162;	2x Luminal A; 3x Luminal B, Her2-; 1x TNBC
46	VVNVYSVVEDEYSEPL	COL12A1	6	24	MaCa108; MaCa109; MaCa129; MaCa138; MaCa145; MaCa162;	5x Luminal B, Her2-; 1x TNBC

Table 10-5: Frequent MaCa-associated HLA-II Source Proteins

Nr.	Gene Name	Peptide	# MaCa samples	# other Tumors	Positive Samples	Molecular Subtype
1	COL11A1	DIQQFLITGDPKAAY; DIQQFLITGDPKAAYD; EKTVIEINTPKIDQ; QQFLITGDPKAAYDY	8	59	MaCa100; MaCa101; MaCa104; MaCa109; MaCa111; MaCa129; MaCa145; MaCa162	3x Luminal A; 4x Luminal B; Her2-; 1x Triple-e
2	FAP	AEYFRNVDYLLIH; AEYFRNVDYLLIHG; AEYFRNVDYLLIHGT; SKLAYVYQNNIYLKQRPG; VRIFIIDTTYPAYVGP;	8	27	MaCa100; MaCa101; MaCa104; MaCa109; MaCa110; MaCa114; MaCa135; MaCa156	2x Luminal A; 5x Luminal B; Her2-; 1x Triple-e;
3	ADAMTS7	DYFIEPLDSAPARPG; LAPGFVSETRRRGGLG; SKLAYVYQNNIYLKQRPG	5	21	MaCa100; MaCa101; MaCa107; MaCa110; MaCa138	3x Luminal A; 2xLuminal B; Her2-;
4	POU3F3	AAAAAAAAHLPSMAGGQQPPPQSLL	4	11	MaCa113; MaCa116; MaCa145; MaCa149	2x Luminal A; 2x Luminal B; Her2-
5	MYO3B	GDPPLFDMHPVKTLFKIPRNPPP; QKKLSDFRLPEEKPPR	4	8	MaCa121; MaCa123; MaCa138; MaCa151	2x Luminal A; 1x Luminal B; Her2-; 1x Her2+;
6	SCUBE2	DKKLIKALFDVLAHP; DKKLIKALFDVLAHPQ; KKLIKALFDVLAHP; NHQEILKDKKLIKALF	4	18	MaCa100; MaCa101; MaCa108; MaCa135	2x Luminal A; 2x Luminal B; Her2-
7	GPR34	DRYIKINRSIQQRK; DRYIKINRSIQQR	4	21	MaCa100; MaCa107; MaCa110; MaCa113	3x Luminal A; 1x Luminal B; Her2-
8	GRP	LREYIRWEEAARNL	3	7	MaCa135; MaCa145; MaCa149	1x Luminal A; Luminal B; Her2-
9	TGFB3	IKKKRVEAIRGQILSK; IRGQILSKLRLTSPP	3	3	MaCa123; MaCa127; MaCa151	2x Luminal A; 1x TNBC
10	HTR3A	SGERVSFKITLLL	3	9	MaCa104; MaCa156; MaCa162	2x Luminal B; Her2-; 1x TNBC
11	UBE2D1	RIQKELSDLQRDPP; RIQKELSDLQRDPPA	3	16	MaCa114; MaCa123; MaCa151	1x Luminal A; 2x Luminal B; Her2-
12	TNFAIP2	LAEIIRLQDPSAIK; LPPRQIRLLEATFLSSE; TVEELKAALERGQLE;	3	12	MaCa101; MaCa127; MaCa162	1x Luminal A; 1x Luminal B; Her2-; 1x TNBC
13	SKIV2L	AGKTVVAEYAIALAQK; SPVGDFYRLIPQPAFQ	3	17	MaCa123; MaCa127; MaCa135	1x Luminal A; 1x Luminal B; Her2-; 1x TNBC;
14	PKN1	AKNVLRLLSAAKAPDR; LERRLGELPADHPK	3	18	MaCa123; MaCa127; MaCa135	1x Luminal A; 1x Luminal B; Her2-; 1x TNBC
15	RIMS1	APPSAPPDRSKGAEPSQQALGPEQ; KILEPKWNQTFVYSHVHRRDFRERM; KVETMLRNDSLSSDQSESVRPSP;	3	9	MaCa101; MaCa108; MaCa127	1x Luminal A; 1x Luminal B; Her2-; 1x TNBC
16	SPATA2L	TGGRAWEPPAEELPQ	3	15	MaCa101; MaCa129; MaCa135	1x Luminal A; 1x Luminal B; Her2-; 1x TNBC
17	PCP2	PVGSKDGAQKRAGTLSPQ	3	27	MaCa107; MaCa145; MaCa162	1x Luminal A; 2x Luminal B; Her2-
18	LILRA2	FPIPSITWEHAGRYH; ITGQFYDRPSLSVQPVP; SNPYLLSLPSDPLE;	3	13	MaCa113; MaCa135; MaCa138	1x Luminal A; 2x Luminal B; Her2-
19	SHROOM3	AHAAREDSLPEESSAPD; QPDASMMQISQGMIGPP	3	9	MaCa107; MaCa135; MaCa162	1x Luminal A; 2x Luminal B; Her2-
20	LMLN	GLRASATSTPVSLGSSP	3	21	MaCa108; MaCa114; MaCa129	2x Luminal B; Her2-; 1x TNBC
21	HAUS8	HGTAPPDLDLSAINDKSI	3	9	MaCa109; MaCa145; MaCa162	3x Luminal B; Her2-

Nr.	Gene Name	Peptide	# MaCa samples	# other Tumors	Positive Samples	Molecular Subtype
22	RPAP3	KKVIIEETGNLIQT; KPLKKVIIEETGNLIQT; TEGERKQIEAQQNKQ;	3	9	MaCa101; MaCa123; MaCa127	2x Luminal A; 1x TNBC
23	WDR26	EDHKVYIWHKRSELPI; SDEDVIRLIGQHLNG	3	19	MaCa105; MaCa123; MaCa127	2x Luminal A
24	OSGIN2	YNSNIPVIHVFRRRVTDP	3	16	MaCa100; MaCa110; MaCa113	1x Luminal A; 1x Luminal B; Her2-
25	SBNO1	AADLIDAEQRMKKSMWGQFWSAH; EIVMKSIVNLDSPMVSPP	2	11	MaCa114; MaCa127	1x Luminal B; Her2-; 1x TNBC
26	SFI1	YYEQRLLRKVFEEWKEEWWVFQHEW	2	28	MaCa100; MaCa107	2x Luminal A
27	MAPKBP1	QLPVSSLFQGPENLQPPPPEKTPNP	2	12	MaCa104; MaCa158	1x Luminal A; 1x Luminal B; Her2-
28	SRPX2	GEHVIRYTAYDR; IDRDRYMEPVTPEE	2	8	MaCa101; MaCa114	1x Luminal A; 1x Luminal B; Her2-
29	NOS1AP	RRIRYEFKAKNIKKKK; VAQVHLLKDQLAAE;	2	5	MaCa101; MaCa138	1x Luminal A; 1x Luminal B; Her2-
30	STC2	CPAIREMVSQLQRE; KDLLLHEPYVDL	2	3	MaCa101; MaCa102	2x Luminal A
31	GOSR1	RHRDILQDYTHEFHK; VNNRRTELFLKEHDH	2	19	MaCa100; MaCa151	2x Luminal A
32	SCGB1D2	DQMSLQKRSLIAEVLVK; KRSLIAEVLVKILKK	2	0	MaCa135; MaCa138	2x Luminal B; Her2-
33	IGKV1D-12	APKLLIYSASSL; SGTEFTLTISSLQPED	2	38	MaCa114; MaCa162	2x Luminal B; Her2-
34	KRT6B	NMQDLVEDLKNKYEDEI	2	52	MaCa156; MaCa162	1x Luminal B; Her2-; 1x TNBC
35	INHBB	GDRWNMVEKRVDLKR; VTALRKLHAGKVREDG	2	7	MaCa102; MaCa127	1x Luminal A; 1x TNBC
36	PTHLH	EGRYLTQETNKVET; KPGKRKEQEKKKRRTRSAWLDSGVT	2	8	MaCa114; MaCa162	2x Luminal B; Her2-
37	HOXB8	EQSPSPTQLFPWMRPQ; EQSPSPTQLFPWMRPQA	2	10	MaCa151; MaCa162	1x Luminal A; 1x Luminal B; Her2-
38	FPR3	AQNHRTMSLAKRVMT; ERLIRSLPTSLERA	2	7	MaCa101; MaCa138	1x Luminal A; 1x Luminal B; Her2-
40	GRPR	DRYKAIVRPMD; SADRYKAIVRPMD	2	3	MaCa102; MaCa151	2x Luminal A
41	FCGR2C	EPPGRQMIAIRKRQPE	2	147	MaCa105; MaCa110	1x Luminal A; 1x Luminal B; Her2-
42	CBS	LDQYRNASNPLAH; VDVLRALGAEIVRTP	2	16	MaCa127; MaCa129	2x TNBC
43	HIRA	NPEMLKYQRRQQQQQ	2	20	MaCa101; MaCa107	2x Luminal A
44	GFRA1	MTPNYIDSSSLSVAPW; TPNYIDSSSLSVAPW	2	3	MaCa138; MaCa151	1x Luminal A; Luminal B; Her2-
45	C1QTNF8	EILKGEKGEAGVRGRAG	2	13	MaCa101; MaCa110	1x Luminal A; 1x Luminal B; Her2-
46	PRKCZ	GLIIHVFPSTPEQPG; RSPFDIITDNPD	2	2	MaCa114; MaCa158	1x Luminal A; 1x Luminal B; Her2-
47	PPFIA1	DDKTTIKCETSPPSSPR; LPEVEAELAQRVAAL	2	6	MaCa101; MaCa123	2x Luminal A
48	OASL	GGSYAYAINPNSFI; KEEVLDAVRTVEEFL	2	7	MaCa101; MaCa162	1x Luminal A; 1x Luminal B; Her2-
49	CSPP1	KDLELRVAASGAQDP	2	6	MaCa116; MaCa125	1x Luminal A; 1x Luminal B; Her2-
50	AP5B1	EAVHVPCLCPGRPARPLLLPLQ	2	4	MaCa107; MaCa110	1x Luminal A;1xLuminal B; Her2-
51	FNBP1L	AFFNILNELNDYAGQR; SPEGSYTDDANQEVR	2	4	MaCa100; MaCa105	2x Luminal A
52	C6orf132	APGSADYGFAPAAGRSPYT; LPLPPSFHIRPASQVYPD	2	4	MaCa113; MaCa138	1x Luminal A; 1x Luminal B; Her2-
53	RARS2	KYNFDTMIYVTDKGQKKHFQQVFQM; LPPENLITSISAVPISQK	2	3	MaCa101; MaCa127	1x Luminal A; 1x TNBC
54	ATF7IP2	KVANSEAMILDKNL	2	4	MaCa107; MaCa151	2x Luminal A
55	STRIP1	EPAVGGPGPLIVNNKQPQPPPP; LPPLPEDSIKVIRNMRAASPPA	2	11	MaCa101; MaCa123	2x Luminal A
56	DARS2	NTEIGFLQDALSKP; VSGTVISRPAGQENPK	2	7	MaCa101; MaCa102	2x Luminal A
57	MUC6	VTNEFVSEEGKFLEPH	2	10	MaCa135; MaCa144	1x Luminal A; 1x Luminal B; Her2-
58	PIP4P1	RSPLLSEPIDGGAG	2	12	MaCa101; MaCa158	2x Luminal A
59	ENOX1	KEALLIGII; QARDDFYEWECKQRMRAREERHRRK	2	6	MaCa114; MaCa162	2x Luminal B; Her2-
60	SSH3	RQSFAVLRGAVLGL	2	4	MaCa105; MaCa135	1x Luminal A; 1x Luminal B; Her2-
61	UGT2B28	EAIYHGIPMVGIPLFWD; KNMIYVLYFDFWFQMCDMKKWDQF	2	13	MaCa149; MaCa156	1x Luminal A; 1x TNBC

Nr.	Gene Name	Peptide	# MaCa samples	# other Tumors	Positive Samples	Molecular Subtype
62	CHIA	LLTGLVLIL	2	5	MaCa105; MaCa109	1x Luminal A; 1x Luminal B; Her2-
63	CEP44	VNPEITALQTMLA	2	2	MaCa123; MaCa127	1x Luminal A; 1x TNBC
64	CHRD	AHMLLQNELFLNVGTK; GVVKDLEPELLRH	2	2	MaCa110; MaCa123	1x Luminal A; 1x Luminal B; Her2-;
65	JPT2	EPKSDLKAARSIPAGAE; EPKSDLKAARSIPAGAEPG	2	4	MaCa121; MaCa151	1x Luminal A; 1x Her2+;
66	EXOSC4	DHLERVLEAAAQAAR; GPHEIRGSRARALPD	2	7	MaCa123; MaCa162	1x Luminal A; 1x Luminal B; Her2-
67	METTL5	KIKIDIIAELRYDLP; RYDLPASYKFHKKKSVDIEVDL	2	57	MaCa129; MaCa138	1x Luminal B; Her2-; 1x TNBC;
68	TMEM167B	TNVYSLDGILVFG	2	19	MaCa127; MaCa162	1x Luminal B; Her2-; 1x TNBC
69	ARHGAP35	EEQRFKALQKLQAER; QEIDGRFTSIPCSQPQHKLEIFH	2	23	MaCa100; MaCa123	2x Luminal A
70	CZIB	WQYIRLMDSVALKGGR	2	2	MaCa108; MaCa123	1x Luminal A; 1x Luminal B; Her2-
71	RAB20	APKQVQLEDAVALYKK; SPRAPKQVQLEDAVALYKK	2	6	MaCa123; MaCa127	1x Luminal A; 1x TNBC
72	TAS2R4	SASLLIHSLRRHIQKMQKNATGFW	2	1	MaCa109; MaCa156	1x Luminal B; Her2-; 1x TNBC
73	RELCH	AKETIQALRANLTKAA	2	32	MaCa123; MaCa127	1x Luminal A; 1x TNBC
74	ABCF2	HGQELLSDTKLELNSGR; STLLKLLTGELLPT	2	13	MaCa100; MaCa135	1x Luminal A;1xLuminal B; Her2-
75	METRN	RRAYEAARAAHLHP	2	5	MaCa138; MaCa144	1x Luminal A; 1x Luminal B; Her2-;