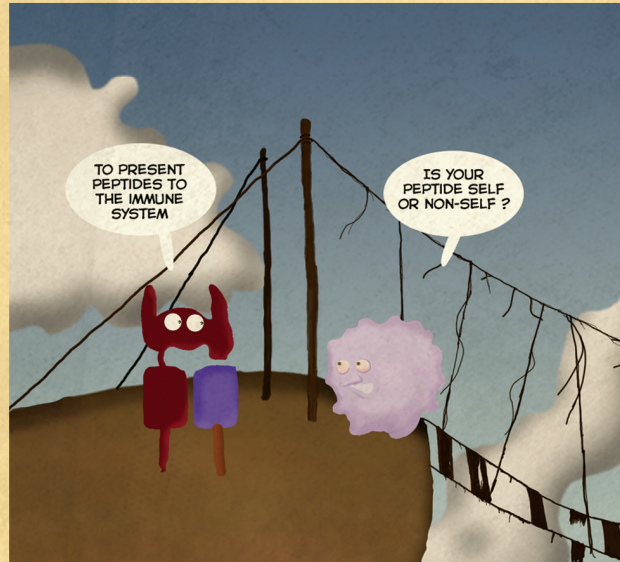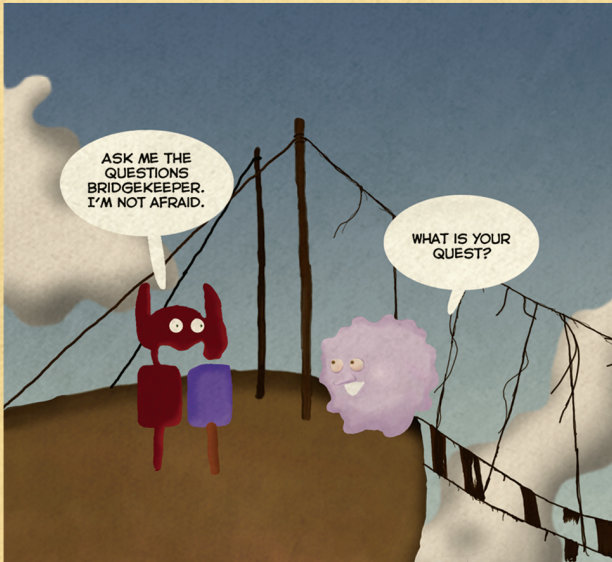# Improved immunopeptidomics analysis methods for cancer antigen discovery

"What is your quest?" - "To seek the Holy Grail"

Laura C. Demmers

Improved immunopeptidomics analysis methods for cancer antigen discovery

-"What is your quest?"

"To seek the Holy Grail."-

Laura Cornelia Demmers

Author: Laura. C. Demmers

Improved immunopeptidomics analysis methods for cancer antigen discovery

-"What is your quest?"

"To seek the Holy Grail."-

Verbeterde immunopeptidomics analyse methoden voor het ontdekken van kanker antigenen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 31 januari 2022 des middags te 2.15 uur

door

Laura Cornelia Demmers

geboren op 6 maart 1991
te Waalwijk

Promotor:

Prof. dr. A.J.R. Heck

Copromotor:

Dr. W. Wu

"You want weapons? We're in a library! Books! The best weapons in the world!"

"It's a new dawn

It's a new day

It's a new life…for me

And I'm feeling good"

-Muse-

Table of contents

Chapter 1

Introduction

Our immune system eliminates pathogenic microbes, toxins and allergenic proteins on a daily basis, thereby allowing the day-to-day maintenance of homeostasis. It is thus not surprising that the immune system is complex and well-coordinated with multiple protective mechanisms. In general, all systems rely on detection of pathogenic features and foreign proteins, which are distinct from endogenous proteins of the host. In this way, pathogens are cleared effectively and specifically without damaging the host tissue.

The Human Leukocyte Antigen (HLA) system

The human immune system consists of two different pathways. Firstly, the innate immune system, the oldest evolutionary defense strategy known. It functions as a first line of defense against bacteria, viruses and parasites by recruiting diverse immune cells to the infection site through cytokines (1). In addition, the innate immune system can activate the adaptive immune system through antigen presentation by human leukocyte antigen (HLA) class I and class II molecules. Antigens presented by these systems are small peptides, of either endogenous or exogenous origin, that can be recognized by T-cells through the T-cell receptor (TCR). Subsequently, a distinct molecular cascade is triggered, depending on HLA class I or class II, in which immunological memory against the antigen is formed (2). Thus, HLA functions at the intersection between innate and adaptive immune responses.

HLA molecules can be divided into two classes: HLA class I and HLA class II. For each of these classes, there are six different loci in the genome. HLA-A/B/C (classical) and HLA-E/F/G (non-classical) for HLA class I and HLA-DPA/B, HLA-DQA/B and HLA-DRA/B for HLA class II. On the cell surface, where the final destination of the presented antigens is, the HLA class I α-chain is stabilized by one molecule of β-2-microglobulin, whereas for HLA class II, one α-chain and one β-chain form a stable molecule (Figure 1).

Although structurally and functionally related, the two classes of HLA molecules have their own different biological function. HLA class I is important in maintaining 'healthy' cells and destroying potential cancer cells. HLA class II on the other hand, is more important in clearance of infected cells (3). The effector T-cells are also different for both classes of HLA molecules. HLA class I presents antigens to CD8[+] T-cells (cytotoxic T-cells) and HLA class II presents antigens to CD4[+] T-cells (T-helper cells). When a respective T-cell does not recognize an antigen as being 'self' (from a host protein), the target cell will be eliminated.

Molecules of both HLA classes are extremely polymorphic (especially the classical ones) at the gene and protein level, with several thousands of alleles found all over the population (table 1 and table 2).
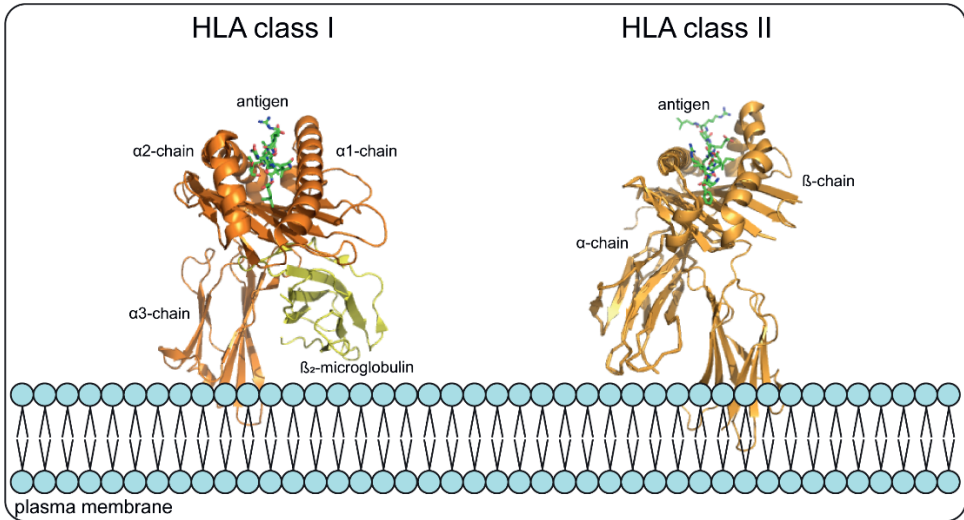
Figure 1: Schematic representation of HLA molecules on the cell surface. HLA-A*02:01 (left) presenting an antigen from the SARS-CoV-2 nucleoprotein (222 – 230, PDB: 7KGQ) and HLA-DRB1 (right) presenting an antigen from histone H2B (PDB: 6BIY).

Table 1: Number of alleles and proteins per HLA class I type assigned as of March 2021

	HLA-A	HLA-B	HLA-C	HLA-E	HLA-F	HLA-G
Alleles	6766	7967	6620	271	45	82
Proteins	4064	4962	3831	110	6	22

Table 2: Number of alleles and proteins per HLA class II type assigned as of March 2021

	HLA-DRA	HLA-DRB	HLA-DQA	HLA-DQB	HLA-DPA	HLA-DPB
Alleles	29	3701	306	1991	258	1749
Proteins	2	155	7	86	7	88

The HLA polymorphisms are most frequently found in the peptide binding region, which results in changes of the peptide binding groove, and thereby allowing a wide range of distinct peptides to bind. In the peptide binding groove, antigens anchor with specific amino acids (depending on the HLA allele) into the HLA molecule (Figure 2). Changing these anchor residues results in changes in the HLA-bound repertoire, which may then impact the defense against malignant cells and invading pathogens.
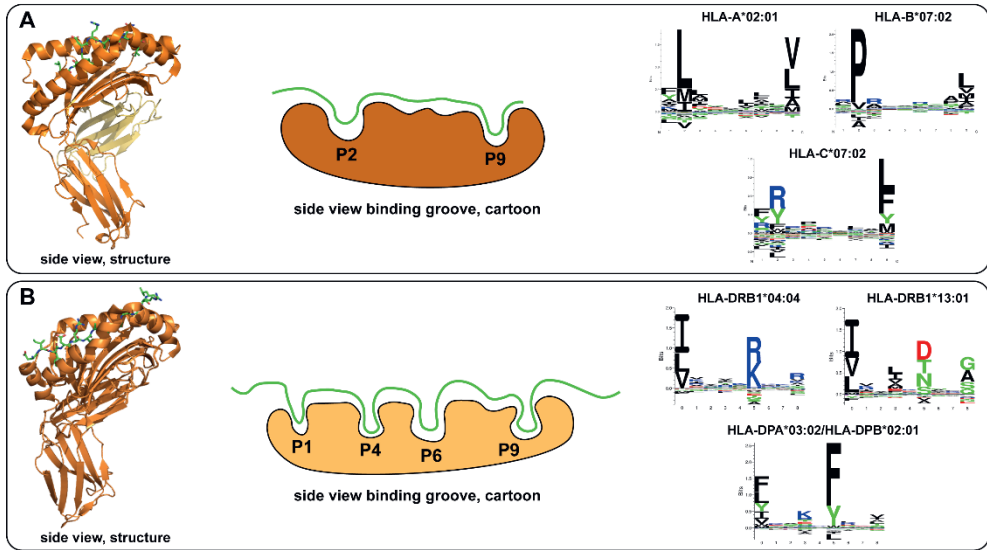
Figure 2: Antigenic peptide binding into the peptide binding groove of HLA class I and HLA class II molecules. A) HLA class I structure and schematic of the binding groove followed by three examples of possible peptides with respective binding motif. **B)** HLA class II structure and schematic of the binding groove followed by three examples of possible peptides with respective binding motif.

Molecular mechanisms of HLA processing and presentation

Compared to other intricate aspects of HLA biology that are far from complete, antigen processing and presentation have been most extensively studied. A wealth of knowledge has been accumulated pertaining to the similarities and differences in these two pathways.

Molecular mechanisms of HLA class I antigen presentation

HLA class I is present on the plasma membrane of every nucleated cell. The antigens are predominantly sourced endogenously and presented to CD8+ T-cells (cytotoxic T-cells). The peptides originate from intracellular protein degradation by the proteasome or from defective ribosomal products, which are derived from translation errors and folding mistakes (4, 5) (Figure 3).

HLA class I antigens arise as by-products from intracellular protein homeostasis and are created by proteasomes. Proteasomes can diffuse freely through the cytoplasm and nucleus to find their ubiquitinated substrates by mere chance (6). Evolution of the adaptive immune system led to evolution of alternate forms of proteasomes each with their own specialized functions. For example the immunoproteasome has different active subunits and is constitutively expressed only in dendritic cells and

lymphocytes. The expression of immunoproteasome in other cell types is largely induced by interferon-gamma signaling only upon infection (7-9).
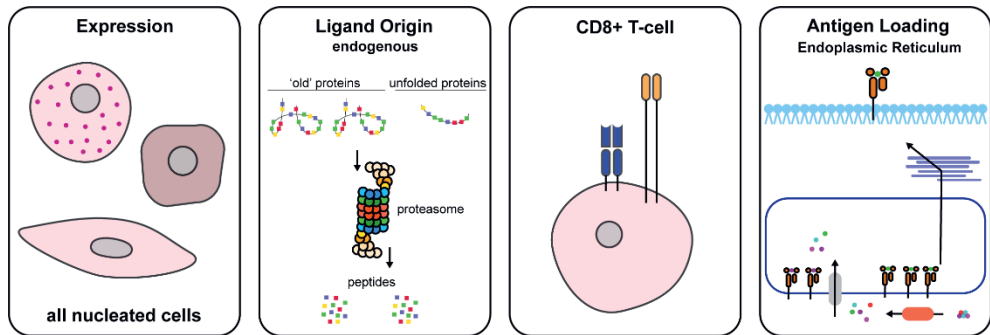
Figure 3: HLA class I characteristics. HLA class I is expressed by every nucleated cell. The antigens are derived from endogenous proteins, using the ubiquitin-dependent proteasomal degradation pathway, loaded in the endoplasmic reticulum and eventually recognized by CD8[+] T-cells.

The immunoproteasome's differential active subunits can lead to differential cleavage sites from the regular proteasome, thereby generating peptides with different amino acid ends and composition compared to peptides generated by the generic proteasome (10). After proteolysis by any of these proteasomes, the peptides generated are released into the cytosol. Approximately 60% of all proteasome-generated peptides are too small to be loaded on HLA class I molecules (11); only 15% of these fragments are with the optimal length (8 to 12 amino acids). The remaining pool of peptides are too long for loading and need to be further trimmed by endogenous proteases in the cytosol before they fit into an HLA class I molecule. Many different proteases like ERAP1 (12), are involved in this peptide trimming, and each protease may have a unique specificity to create and/or destroy antigenic peptides. The trimmed peptides can subsequently be then translocated into the endoplasmic reticulum (ER) via the antigen transporter 1 and/or 2 in the ER membrane (4, 13).

Once the peptides have entered the ER, loading on empty HLA class I molecules proceeds via the peptide loading complex (PLC). This complex consists of the proteins tapasin, calreticulin and ERp57 (14). Once peptides are loaded, the peptides with high affinity towards the peptide binding groove constantly compete off the peptides with low affinity towards the peptide binding groove. Upon loading of a high affinity peptide, a glycan on the HLA class I molecules triggers release form the ER whereby the formed complex moves towards the Golgi apparatus and subsequently to the plasma membrane where the peptide is presented (15) (Figure 4). Only an estimated 0.02% of all generated peptides by the proteasome will eventually survive long enough to be presented by HLA class I molecules on the plasma membrane (16).
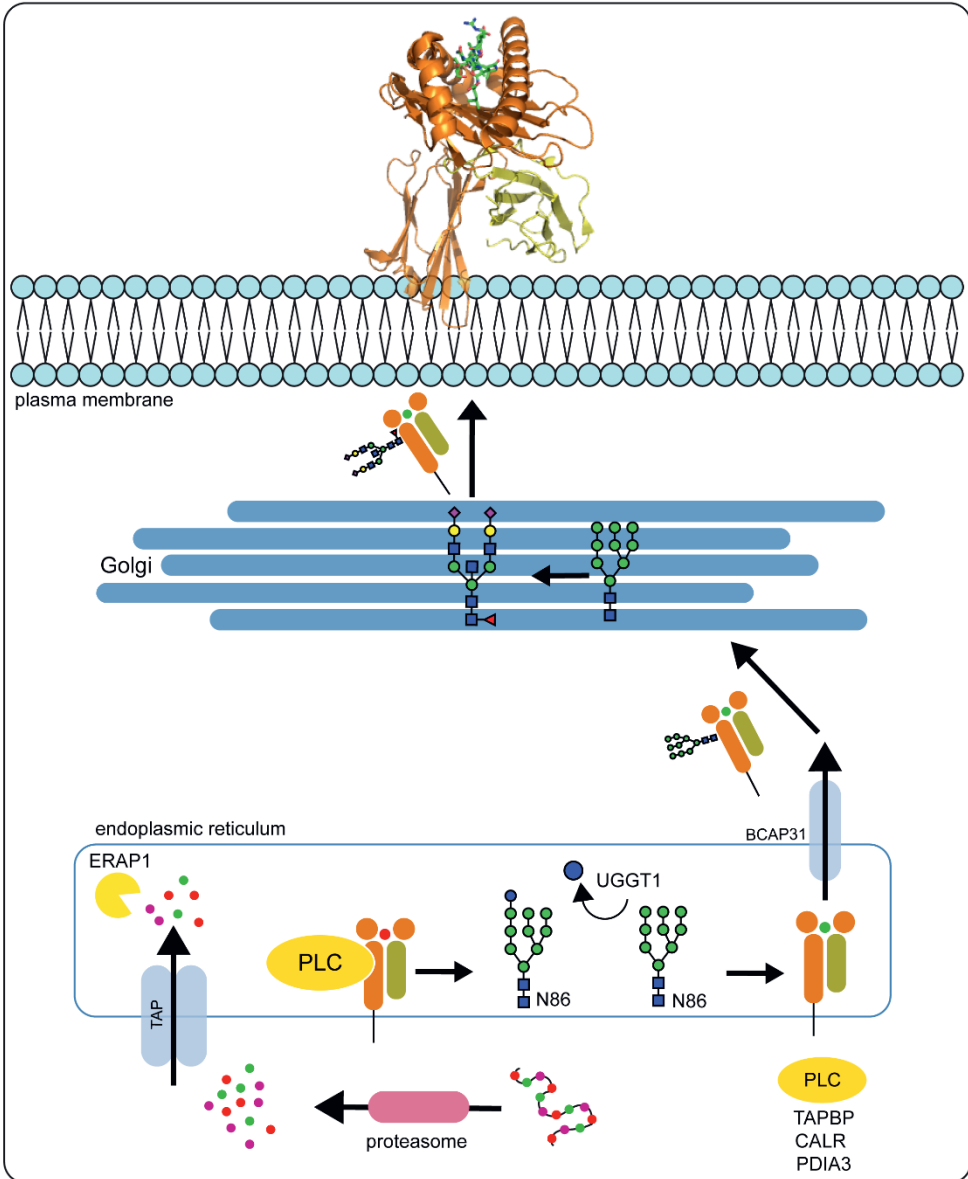
Figure 4: Schematic representation of HLA class I antigen presentation. Proteins are degraded by the proteasome. The resulting peptides are subsequently transported into the endoplasmic reticulum where they are loaded on empty HLA class I molecules. After binding of a high affinity peptide, a glucose is removed from the glycan at position N86 which triggers release from the ER towards the Golgi, where the glycan is matured. Subsequently, the HLA class I complex is stably inserted on the plasma membrane.

Molecular mechanisms of HLA class II antigen presentation

The HLA class II antigen processing and presentation pathway is inherently different from the HLA class I pathway. HLA class II molecules are only expressed on immune

cells such as dendritic cells and B-cells. Expression on epithelial cells can be induced upon infection via interferon-gamma signaling. Since the HLA class II peptide loading groove is open on both sides, HLA class II peptides can also extend out of the peptide binding groove. The peptides are sampled from exogenous proteins through the endosomal pathway and from endogenous proteins via autophagy. On the plasma membrane, the antigens are presented to CD4$^+$ T-cells (T-helper cells) (17) (Figure 5).
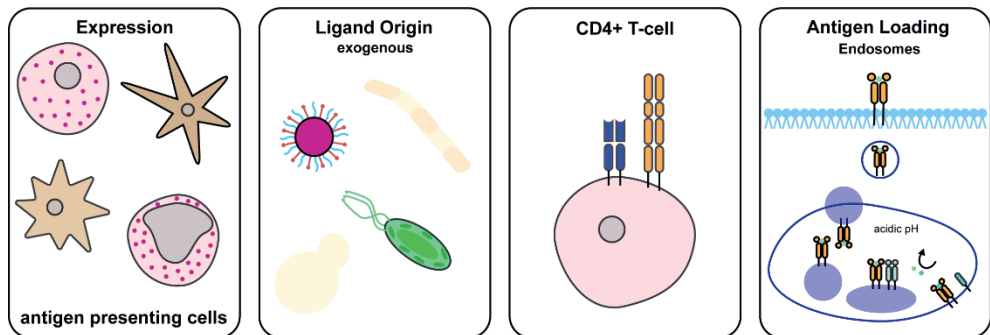
Figure 5: HLA class II characteristics. HLA class I is expressed on every nucleated cell. The antigens are derived from endogenous proteins, loaded in the endoplasmic reticulum and eventually recognized by CD4$^+$ T-cells.

Similar to HLA class I, HLA class II complexes are assembled in the ER, where, instead of binding a high affinity antigenic peptide, a fragment of the CD74 protein (invariant chain) binds into the peptide binding groove and targets the HLA class II molecule to the endosomal pathway (3, 18). Subsequently, the HLA class II molecules encounter antigenic peptides in the MIIC compartment (19), where the invariant chain is sequentially cleaved by the proteases legumain, cathepsin S, cathepsin L and cathepsin F, leaving behind a small peptide fragment called CLIP.

This CLIP peptide that remains can be exchanged out of the HLA class II molecule for a high-affinity antigenic peptide with the help of the chaperones HLA-DM and HLA-DO (20, 21). After loading, the HLA class II complexes with antigenic peptides are transported to the plasma membrane through vesicular transport, where they are stably inserted (20, 22, 23) (Figure 6).
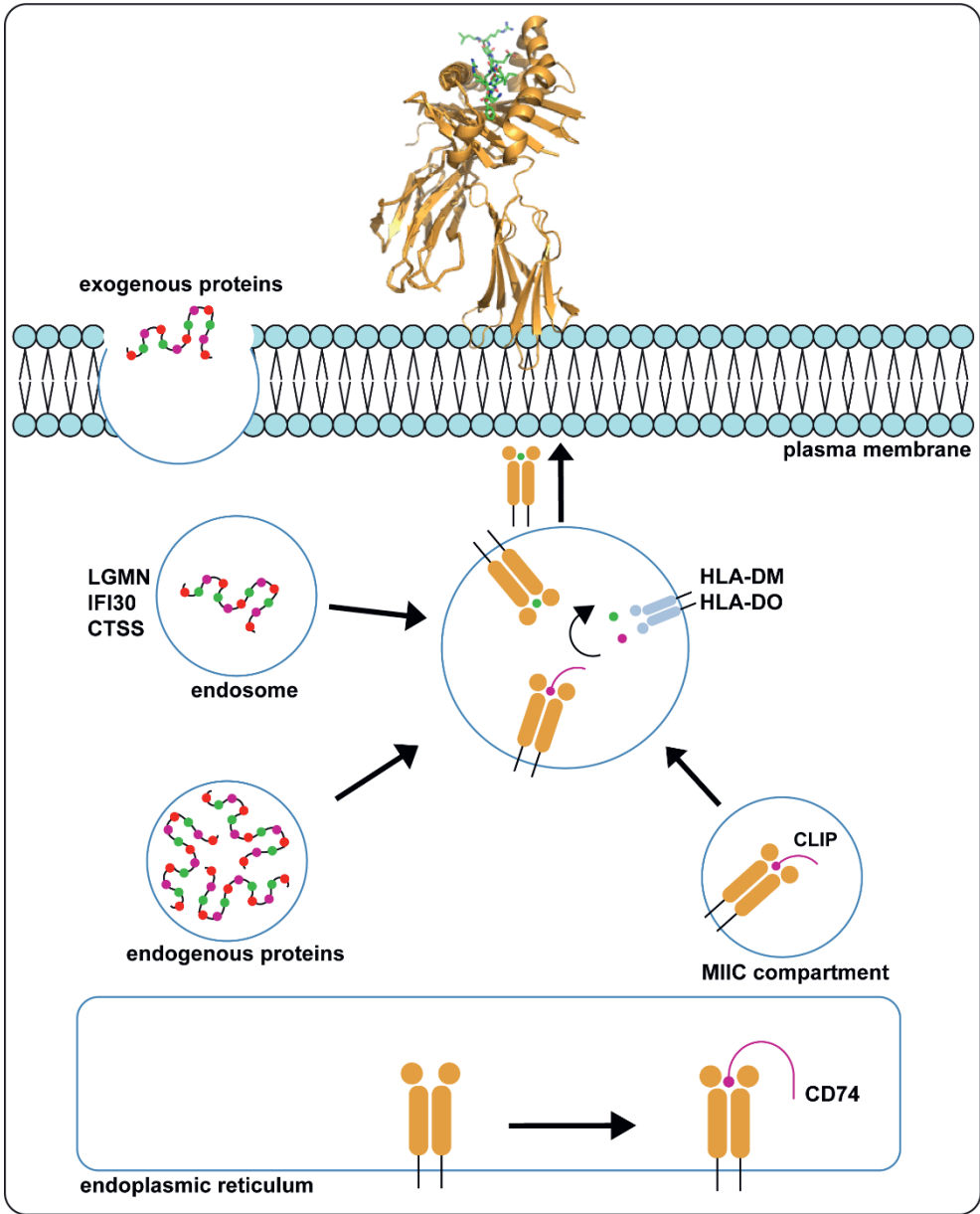
Figure 6: Schematic representation of HLA class II antigen presentation. Proteins from exogenous or endogenous origin are cleaved in the endosome by a variety of proteases. The peptides are transported to the MIIC complex where, with the chaperones HLA-DO and HLA-DM, the CLIP peptide fragment can be exchange for a higher affinity antigen. Subsequently, the HLA class II complex is stably inserted on the plasma membrane.

HLA cross-presentation

The HLA class I and HLA class II pathways overlap via a process termed HLA cross-presentation. Cross-presentation results in the presentation of exogenous peptides by HLA class I molecules, instead of by HLA class II molecules. Cross-presentation can be achieved via a vacuolar or cytosolic route. In the cytosolic route, phagocytosed proteins are exported to the cytosol where they are degraded by the proteasome. Subsequently, they can follow the 'standard' HLA class I presentation pathway or they can be transported into phagosomes and/or endosomes through the TAP transporter and be loaded there (24, 25). HLA class I molecules reside in endosomes after internalization of the molecule through invagination of the plasma membrane (26), a process mediated by a conserved tyrosine residue in the cytoplasmic tail of HLA class I molecules. In the vacuolar pathway, exogenous proteins are directly processed by proteases (cathepsin S) in the phagosome and subsequently loaded into HLA class I molecules (27, 28).

Antigenic peptide splicing

In addition to presentation of HLA class II antigens by HLA class I molecules, spliced peptides can be presented by HLA class I molecules (29-31). These peptides arise after cleavage and subsequent ligation within the proteasome. When the spliced peptides originate from a single protein it's called a cis-proteasomal spliced peptide. Ligation of the peptide fragments can be either from N-terminus to C-terminus (normal) or from C-terminus to N-terminus (reversed). When the spliced antigenic peptide is composed of peptide fragments from two different proteins, it is called a trans-proteasomal spliced peptide. The frequency and abundance of such splicing events are still heavily debated, whereby some groups have claimed that approximately 10% to 30% of all antigens are spliced (32, 33), whereby it has been shown that these spliced antigens could also be immunogenic (34, 35).

With extensive efforts in studying the molecular pathways underlying HLA class I and HLA class II antigen processing and presentation, many aspects of these pathways are becoming more clearly understood. Nevertheless, the exact mechanisms and function of HLA cross-presentation and the presentation of spliced antigens, and the biology associated with these atypical processes, still remain to be further defined.

HLA in human disease

HLA is one of the strongest predisposing factors in human diseases and is associated with almost every autoimmune disease known to date. Such

observations exemplify the importance of understanding HLA function, in the hope of designing better therapeutic regimes to rectify or prevent these conditions.

HLA and autoimmune disease

The most well-known cases of HLA predisposition in autoimmune diseases are diabetes type-1 (HLA-DQ6/8), arthritis (HLA-B27, HLA-DR4) and celiac disease (HLA-DQ2). However, these diseases manifest themselves in very different ways. The most common determinant is the presentation of aberrant antigens to the immune system (36). For all three examples, the aberrant antigens arise differently. In diabetes type-1 for example, "mimic" peptides are presented by pancreatic islet cells which are recognized as 'non-self' after which the tissue is destroyed (37). On the other hand, in arthritis, arginine residues in antigens are deamidated to citrullinated residues. These residues are subsequently recognized as non-self, resulting in destruction of the cells which present the antigen (38). In celiac disease, a novel binding motif is created. Transglutaminase converts glutamines of the gluten protein to glutamic acid allowing a whole new variety of antigens to be presented on the cell surface. There is no T-cell tolerance towards these antigens and therefore the presenting cells are destroyed (39).

HLA and infection

Apart from the crucial role in autoimmune diseases, the HLA mediated immune response is also necessary for clearance of bacterial and viral infections. Different pathways can lead to presentation of exogenous antigens. Upon infection, viral and/or bacterial proteins that enter the host cell are processed similarly to host proteins and therefore also get degraded by the host proteasome. After degradation, these peptides will follow the host HLA class I antigen processing and presentation pathway to be presented on HLA molecules on the cell surface. Exogenous antigens can also be loaded directly onto HLA class II molecules after elimination of pathogens via autophagy in the endosomal pathway (40).

Presentation of pathogenic antigens from invading pathogens can be exploited for the development of anti-viral/bacterial vaccines. In practice, a protein or peptide mixture may be used to vaccinate healthy individuals. After vaccination, the protein/peptides are internalized by antigen presenting cells (APCs) and subsequently proteolyzed further and presented by HLA class II molecules. Then, CD4[+] T-cells recognize the pathogenic epitope and orchestrate expansion of the reactive T-cells clone and will in parallel stimulate B-cells to produce antibodies, so that upon subsequent infection with the same pathogen, it can be cleared quickly and efficiently without systemic impacts (41).

**Methods to study HLA class I and HLA class II antigens**

Dissecting HLA biology in health and disease is crucial to understand HLA associated diseases and treatments. Given the strong implications in human health, two main approaches have emerged to investigate which peptides will be or are presented by HLA molecules.

The first approach is based on *in silico* antigen prediction. Based on the HLA allele (thus antigen binding motif), peptides are predicted with computational tools that can simulate proteasomal cleavage sites (42, 43) and with predictive tools that can calculate binding affinity of a peptide towards each specific HLA allele (44-46). Computational tools that calculate binding affinity are artificial neural networks trained with empirical measurements from the immune epitope database (IEDB), greatly increasing their potential (47). These predicted antigens can be directly used in biological assays to determine their immunogenicity *in vitro* and *in vivo*. The major advantage of *in silico* prediction of antigens is the enormous amount of peptides that can be predicted at once. The major drawback of this high throughput method is that a predicted peptides is not necessarily presented at the cell surface by HLA molecules and therefore less accessible for T-cell targeting.

An alternative approach is to directly measure and sequence the peptide antigens presented by affinity purification and mass spectrometry (47, 48). Before peptides can be measured, the HLA molecules and bound peptides need to be purified from a sample of interest. The gold standard to purify antigens is immuno-affinity purification.

Immuno-affinity purification (IAP) utilizes antibodies against HLA to purify the complexes, which enables specific enrichment of HLA class I and HLA class II allotypes. Furthermore, for IAP, cells are lysed prior to purification, which allows both purification of frozen cells and purification of intracellular HLA complexes. After purification, HLA complexes are separated by ultrafiltration (49, 50) (Figure 7).

There are also possibilities to directly elute bound antigens from the cell surface by mild acid elution. This method is used by a small minority of labs in the community. During mild acid elution, fresh cells are resuspended in an acidic buffer, which releases HLA class I complexes and peptide ligands from the cells. By subsequent ultracentrifugation and ultrafiltration, the peptide ligands and HLA complexes can be separated from the cell suspension (51, 52) (Figure 7). All advantages and disadvantages of mild acid elution and immuno-affinity purification are summarized in table 3.
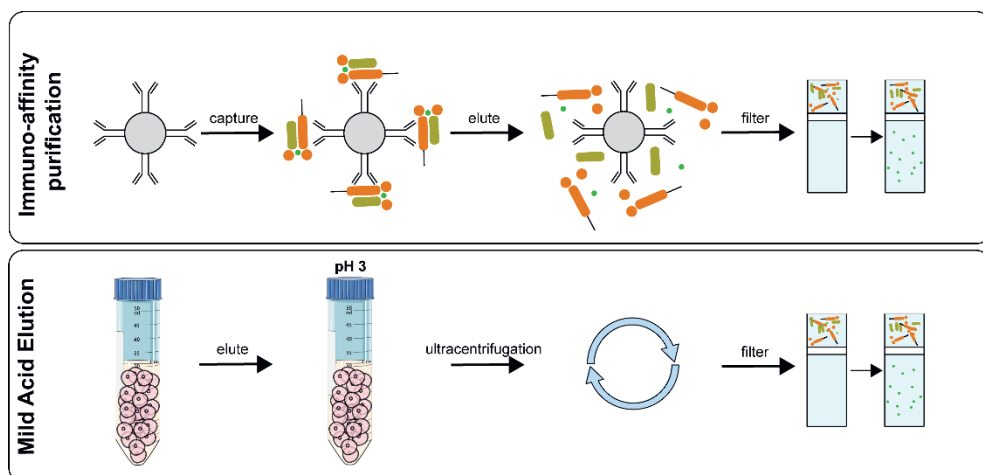
Figure 7: Schematic overviews of the immune affinity purification mild acid elution methods for isolation of HLA class I ligands. During mild acid elution, HLA complexes are released in an acidic buffer and purified with subsequent ultracentrifugation and ultrafiltration. During immune affinity purification, HLA complexes are purified with specific antibodies and eluted from those with an acidic buffer. By subsequent ultrafiltration the complexes are separated from the antigens.

Table 3: Advantages and disadvantages of mild acid elution and immuno-affinity purification.

	Immuno-affinity purification	Mild acid elution
Costs	++	-
Equipment	-	+
Expertise	+	-
Sample size	++	-
Contaminants	-	+
Specificity	+	-
Frozen sample	++	-

After purification of the HLA bound peptide ligands, they are separated by liquid chromatography (LC) and subsequently measured by mass spectrometry (MS).

LC-MS strategies for HLA peptide ligands

The inherent low abundance and enormous diversity of HLA peptide ligands makes it very challenging to analyze them by LC-MS/MS. Fortunately, LC-MS/MS methods have been substantially optimized for analyzing these samples.

Due to low inherent abundance of HLA peptide antigens, sensitivity of the mass spectrometer is a strong determinant of antigen sequencing depth. In this respect, coverage of the highly complex immunopeptidome can also benefit from advances made in shot-gun proteomics. With the advent of strong cation exchange (SCX) (53, 54) and high-pH reversed phase fractionation (55-59) for proteomics, these have also been applied with great success to HLA peptide ligandomics, significantly increasing the detection depth of the ligandome.

In addition to optimizing LC methods, it is crucial to select the most suitable mass spectrometer to detect the peptide ligands. To overcome detection problems caused by the inherent low abundance of the peptide ligands, it is preferred to measure with an Orbitrap mass analyzer instead of a time of flight mass analyzer. This is because the Orbitrap mass analyzer enables trapping and accumulation of ions before analyzing them. In this way, more ions are detected allowing analysis of low abundant peptides (59). Aside from the mass analyzer, the choice between an untargeted and a targeted approach for precursor selection is important. In general, the untargeted approach works best when analyzing a large repertoire of ligands and when there is no specific interest for a particular peptide ligand. Despite the clear advantage in antigen sequencing breadth, untargeted MS measurements tend to rely on intensity-based precursor prioritization, which favors the sequencing of high-abundance peptides. This biases against the detection of low-abundant neo-antigens, which may be better picked up with scheduled parallel reaction monitoring (PRM).

Finally, last but not least, a rational choice on the method of peptide fragmentation can still be made to boost sequence confidence in the antigen peptides detected. Fragmentation can be induced by higher collisional dissociation (HCD), which is a very fast fragmentation method ideal for sequencing a large repertoire of ligands quickly and in a high throughput manner (60). HCD fragmentation results in formation of *b*-ions and *y*-ions (61). Since fragmentation efficiency is peptide-dependent, *b*-ions and *y*-ions alone are sometimes not sufficient to unambiguously annotate the full peptide sequence. This may be complemented by electron transfer dissociation (ETD), which also generates *c*-ions and *z*-ions. When used in combination, for instance in a dual fragmentation method, EThcD, the spectra obtained are richer in information (62). Although EThcD is slower with a longer cycle time (63, 64), the gain in diverse kinds of ions could still boost the confident identification of HLA peptides.

To achieve more information rich spectra, EThcD fragmentation can be used, though it comes at the expense of a longer cycling time. EThcD is a dual fragmentation method consisting of electron transfer dissociation (ETD) with supplemental HCD. EThcD fragmentation results in generation of not only *b*-ions and *y*-ions, but also *c*-ions and *z*-ions, which together make up an information rich mass spectrum. Subsequently, these spectra are used for peptide sequence identification using automated database searches.

Post-acquisition data handling for HLA peptide ligands

The generated mass spectra are subsequently used for peptide sequence identification. In short, a list of theoretical masses (*m/z*) based on the length of HLA class I and HLA class II peptide ligands, is calculated with *in silico* non-enzymatic digestion. Then, the *m/z* values and charge states of the precursor masses are compared to every *in silico* predicted sequences. Peptides that match with a certain mass tolerance are kept for MS/MS comparisons. Theoretical fragment ion masses are calculated for each peptide and compared the fragmentation spectra. Each peptide is then assigned a probability score, and those with highest score are reported as peptide spectrum match, which eventually will identify the peptide.

Mass spectrometry based peptide and protein identification relies on matching measured spectra to theoretical spectra generated from a user specified database. For exploratory studies with cell lines, a generic Swissprot protein database is a good standard, which is sufficient to identify most peptides from a sample. Nonetheless, Swissprot alone may be less than sufficient for studies from patient-derived material, as the lack of representation for patient-specific SNPs in the generic Swissprot database may cause such peptides harboring coding changes to be missed in the database search altogether. As such, a patient-specific database is critically needed, and especially so in neo-antigen discovery, where ideal HLA peptides contain a tumor-specific change in sequence. This type of patient-specific databases can be derived from whole genome sequencing of the same patient-derived material and will therefore contain all single nucleotide polymorphisms specific to the patient under investigation. Given the highly patient-specific nature of such an approach, it is possible to identify antigens with personal mutational signatures as well.

Even with these two types of databases, it is not possible to for example identify spliced peptides. Since these peptides are derived from different parts of proteins, it is necessary to either create a specific database which covers all splice possibilities, or a *de novo* peptide sequencing approach can be used.

Novel applications of HLA research

With the development of improved ultrasensitive LC-MS methods and search strategies for HLA peptide ligands, it is now possible to translate HLA research into a more clinical setting.

Recently, targeting cell-surface antigens became a potential way to treat certain types of cancer. Tumors can present tumor associated antigens (TAA) at the cell surface which may be derived from overexpressed proteins or other aberrantly expressed, processed, or degraded cellular proteins. Antigen-based vaccines targeting these TAA's have already showed their potential in a clinical setting. For

example, three antigen-based vaccines are currently explored to treat melanoma. One vaccine is targeted against peptides from the MAGE-1 protein (65). The two other vaccines are targeted against tyrosinase (66) peptides or MART-1 peptides (67).

Capitalizing on these advances, and further incorporating rational thoughts on the presence of tumor-specific coding mutations, it is also now possible to screen for mutation-containing neo-antigens, which may further eliminate possible peripheral tolerance during therapy. The medical potential of neo-antigens is tremendous, if only the bottleneck on screening such low-abundant entities can be alleviated. In this respect, targeted mass spectrometry on predicted neo-antigens based on tumor whole genome sequencing may be a promising strategy in the near future.

Taken together, mass spectrometry based neo-antigen detection is increasingly becoming the method of choice for future immune-therapy based personalized cancer treatments.

1

Scope of this thesis

In this thesis, I describe several advances I have made in isolation and measurement techniques applied to HLA ligandomics, allowing us to obtain new mechanistic insights into HLA loading, as well as the critical application of these techniques for antigen peptide discovery.

In **chapter 2**, I describe the first technical advancement. I show that high-pH reversed phase fractionation can be used to increase the detection depth of the HLA ligandome. Using complementary SCX pre-fractionation, the detection depth is even further increased. Furthermore, I show that high-pH fractionation is crucial for detection of citrullinated residues and that the pre-fractionation method of choice should always carefully be chosen based on the chemical properties of the HLA peptides of interest.

In **chapter 3**, I describe the mechanism behind HLA class II re-shaping of the ligandome upon a 'mimicked fever'. By applying a heat stimulus to B-cells, the proteome is re-shaped in such a way that B-cells are prepared for infection. Furthermore, we show that mainly the HLA class II chaperone CD74 is greatly affected upon this heat stimulus.

In **chapter 4**, we apply glycoproteomics to analyze the composition of HLA class I glycans in three different B-cell lines. We show that the composition is depending on the HLA type. Furthermore, we show that the glycan composition can be used as localization marker of the HLA complex inside or at the surface of the cell.

In **chapter 5**, we use single-cell derived colorectal cancer organoids to study tumor clonal heterogeneity at the proteome and ligandome level. We were able to recapitulate the original single cell features, but also found clone specific changes at the proteome and ligandome level. Furthermore, we believe the most promising therapeutic targets arise from proteins that are actively degraded by tumor cells such as DNA damage repair proteins.

In **chapter 6**, we discuss remaining challenges and future prospective in the field of immunopeptidomics.

References

1. Chaplin, D. D., Overview of the immune response. J Allergy Clin Immunol 2010, 125, (2 Suppl 2), S3-23.

2. Farber, D. L.; Netea, M. G.; Radbruch, A.; Rajewsky, K.; Zinkernagel, R. M., Immunological memory: lessons from the past and a look to the future. Nat Rev Immunol 2016, 16, (2), 124-8.

3. Neefjes, J.; Jongsma, M. L.; Paul, P.; Bakke, O., Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol 2011, 11, (12), 823-36.

4. Michalek, M. T.; Grant, E. P.; Gramm, C.; Goldberg, A. L.; Rock, K. L., A role for the ubiquitin-dependent proteolytic pathway in MHC class I-restricted antigen presentation. Nature 1993, 363, (6429), 552-4.

5. Bourdetsky, D.; Schmelzer, C. E.; Admon, A., The nature and extent of contributions by defective ribosome products to the HLA peptidome. Proc Natl Acad Sci U S A 2014, 111, (16), E1591-9.

6. Reits, E. A.; Benham, A. M.; Plougastel, B.; Neefjes, J.; Trowsdale, J., Dynamics of proteasome distribution in living cells. EMBO J 1997, 16, (20), 6087-94.

7. Monaco, J. J.; Nandi, D., The genetics of proteasomes and antigen processing. Annu Rev Genet 1995, 29, 729-54.

8. Gaczynska, M.; Rock, K. L.; Goldberg, A. L., Gamma-interferon and expression of MHC genes regulate peptide hydrolysis by proteasomes. Nature 1993, 365, (6443), 264-7.

9. Groettrup, M.; Kraft, R.; Kostka, S.; Standera, S.; Stohwasser, R.; Kloetzel, P. M., A third interferon-gamma-induced subunit exchange in the 20S proteasome. Eur J Immunol 1996, 26, (4), 863-9.

10. Toes, R. E.; Nussbaum, A. K.; Degermann, S.; Schirle, M.; Emmerich, N. P.; Kraft, M.; Laplace, C.; Zwinderman, A.; Dick, T. P.; Muller, J.; Schonfisch, B.; Schmid, C.; Fehling, H. J.; Stevanovic, S.; Rammensee, H. G.; Schild, H., Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. J Exp Med 2001, 194, (1), 1-12.

11. Cascio, P.; Hilton, C.; Kisselev, A. F.; Rock, K. L.; Goldberg, A. L., 26S proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide. EMBO J 2001, 20, (10), 2357-66.

12. Serwold, T.; Gonzalez, F.; Kim, J.; Jacob, R.; Shastri, N., ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. Nature 2002, 419, (6906), 480-3.

13. Reits, E.; Griekspoor, A.; Neijssen, J.; Groothuis, T.; Jalink, K.; van Veelen, P.; Janssen, H.; Calafat, J.; Drijfhout, J. W.; Neefjes, J., Peptide diffusion,

protection, and degradation in nuclear and cytoplasmic compartments before antigen presentation by MHC class I. Immunity 2003, 18, (1), 97-108.

14. Cresswell, P.; Bangia, N.; Dick, T.; Diedrich, G., The nature of the MHC class I peptide loading complex. Immunol Rev 1999, 172, 21-8.

15. Moremen, K. W.; Tiemeyer, M.; Nairn, A. V., Vertebrate protein glycosylation: diversity, synthesis and function. Nat Rev Mol Cell Biol 2012, 13, (7), 448-62.

16. Yewdell, J. W.; Reits, E.; Neefjes, J., Making sense of mass destruction: quantitating MHC class I antigen presentation. Nat Rev Immunol 2003, 3, (12), 952-61.

17. Suri, A.; Lovitch, S. B.; Unanue, E. R., The wide diversity and complexity of peptides bound to class II MHC molecules. Curr Opin Immunol 2006, 18, (1), 70-7.

18. Cresswell, P., Invariant chain structure and MHC class II function. Cell 1996, 84, (4), 505-7.

19. Neefjes, J., CIIV, MIIC and other compartments for MHC class II loading. Eur J Immunol 1999, 29, (5), 1421-5.

20. Unanue, E. R.; Turk, V.; Neefjes, J., Variations in MHC Class II Antigen Processing and Presentation in Health and Disease. Annu Rev Immunol 2016, 34, 265-97.

21. Ghosh, P.; Amaya, M.; Mellins, E.; Wiley, D. C., The structure of an intermediate in class II MHC maturation: CLIP bound to HLA-DR3. Nature 1995, 378, (6556), 457-62.

22. Boes, M.; Cerny, J.; Massol, R.; Op den Brouw, M.; Kirchhausen, T.; Chen, J.; Ploegh, H. L., T-cell engagement of dendritic cells rapidly rearranges MHC class II transport. Nature 2002, 418, (6901), 983-8.

23. Kleijmeer, M.; Ramm, G.; Schuurhuis, D.; Griffith, J.; Rescigno, M.; Ricciardi-Castagnoli, P.; Rudensky, A. Y.; Ossendorp, F.; Melief, C. J.; Stoorvogel, W.; Geuze, H. J., Reorganization of multivesicular bodies regulates MHC class II antigen presentation by dendritic cells. J Cell Biol 2001, 155, (1), 53-63.

24. Houde, M.; Bertholet, S.; Gagnon, E.; Brunet, S.; Goyette, G.; Laplante, A.; Princiotta, M. F.; Thibault, P.; Sacks, D.; Desjardins, M., Phagosomes are competent organelles for antigen cross-presentation. Nature 2003, 425, (6956), 402-6.

25. Guermonprez, P.; Saveanu, L.; Kleijmeer, M.; Davoust, J.; Van Endert, P.; Amigorena, S., ER-phagosome fusion defines an MHC class I cross-presentation compartment in dendritic cells. Nature 2003, 425, (6956), 397-402.

26. Lizee, G.; Basha, G.; Tiong, J.; Julien, J. P.; Tian, M.; Biron, K. E.; Jefferies, W. A., Control of dendritic cell cross-presentation by the major histocompatibility complex class I cytoplasmic domain. Nat Immunol 2003, 4, (11), 1065-73.

27. Shen, L.; Sigal, L. J.; Boes, M.; Rock, K. L., Important role of cathepsin S in generating peptides for TAP-independent MHC class I crosspresentation in vivo. Immunity 2004, 21, (2), 155-65.

28. Bertholet, S.; Goldszmid, R.; Morrot, A.; Debrabant, A.; Afrin, F.; Collazo-Custodio, C.; Houde, M.; Desjardins, M.; Sher, A.; Sacks, D., Leishmania antigens are presented to CD8+ T cells by a transporter associated with antigen processing-independent pathway in vitro and in vivo. J Immunol 2006, 177, (6), 3525-33.

29. Berkers, C. R.; de Jong, A.; Schuurman, K. G.; Linnemann, C.; Meiring, H. D.; Janssen, L.; Neefjes, J. J.; Schumacher, T. N.; Rodenko, B.; Ovaa, H., Definition of Proteasomal Peptide Splicing Rules for High-Efficiency Spliced Peptide Presentation by MHC Class I Molecules. J Immunol 2015, 195, (9), 4085-95.

30. Textoris-Taube, K.; Keller, C.; Liepe, J.; Henklein, P.; Sidney, J.; Sette, A.; Kloetzel, P. M.; Mishto, M., The T210M Substitution in the HLA-a*02:01 gp100 Epitope Strongly Affects Overall Proteasomal Cleavage Site Usage and Antigen Processing. J Biol Chem 2015, 290, (51), 30417-28.

31. Dalet, A.; Vigneron, N.; Stroobant, V.; Hanada, K.; Van den Eynde, B. J., Splicing of distant peptide fragments occurs in the proteasome by transpeptidation and produces the spliced antigenic peptide derived from fibroblast growth factor-5. J Immunol 2010, 184, (6), 3016-24.

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

33. Liepe, J.; Marino, F.; Sidney, J.; Jeko, A.; Bunting, D. E.; Sette, A.; Kloetzel, P. M.; Stumpf, M. P.; Heck, A. J.; Mishto, M., A large fraction of HLA class I ligands are proteasome-generated spliced peptides. Science 2016, 354, (6310), 354-358.

34. Hanada, K.; Yewdell, J. W.; Yang, J. C., Immune recognition of a human renal cancer antigen through post-translational protein splicing. Nature 2004, 427, (6971), 252-6.

35. Dalet, A.; Robbins, P. F.; Stroobant, V.; Vigneron, N.; Li, Y. F.; El-Gamil, M.; Hanada, K.; Yang, J. C.; Rosenberg, S. A.; Van den Eynde, B. J., An antigenic peptide produced by reverse splicing and double asparagine deamidation. Proc Natl Acad Sci U S A 2011, 108, (29), E323-31.

36.   Thorsby, E.; Lie, B. A., HLA associated genetic predisposition to autoimmune diseases: Genes involved and possible mechanisms. Transpl Immunol 2005, 14, (3-4), 175-82.

37.   Durinovic-Bello, I., Autoimmune diabetes: the role of T cells, MHC molecules and autoantigens. Autoimmunity 1998, 27, (3), 159-77.

38.   Hill, J. A.; Southwood, S.; Sette, A.; Jevnikar, A. M.; Bell, D. A.; Cairns, E., Cutting edge: the conversion of arginine to citrulline allows for a high-affinity peptide interaction with the rheumatoid arthritis-associated HLA-DRB1*0401 MHC class II molecule. J Immunol 2003, 171, (2), 538-41.

39.   Koning, F., The molecular basis of celiac disease. J Mol Recognit 2003, 16, (5), 333-6.

40.   van Montfoort, N.; van der Aa, E.; Woltman, A. M., Understanding MHC class I presentation of viral antigens by human dendritic cells as a basis for rational design of therapeutic vaccines. Front Immunol 2014, 5, 182.

41.   Malonis, R. J.; Lai, J. R.; Vergnolle, O., Peptide-Based Vaccines: Current Progress and Future Challenges. Chem Rev 2020, 120, (6), 3210-3229.

42.   Nussbaum, A. K.; Kuttler, C.; Hadeler, K. P.; Rammensee, H. G.; Schild, H., PAProC: a prediction algorithm for proteasomal cleavages available on the WWW. Immunogenetics 2001, 53, (2), 87-94.

43.   Stevanovic, S., Identification of tumour-associated T-cell epitopes for vaccine development. Nat Rev Cancer 2002, 2, (7), 514-20.

44.   Schultze, J. L.; Vonderheide, R. H., From cancer genomics to cancer immunotherapy: toward second-generation tumor antigens. Trends Immunol 2001, 22, (9), 516-23.

45.   Shastri, N.; Schwab, S.; Serwold, T., Producing nature's gene-chips: the generation of peptides for display by MHC class I molecules. Annu Rev Immunol 2002, 20, 463-93.

46.   Admon, A.; Barnea, E.; Ziv, T., Tumor antigens and proteomics from the point of view of the major histocompatibility complex peptides. Mol Cell Proteomics 2003, 2, (6), 388-98.

47.   Andreatta, M.; Nielsen, M., Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics 2016, 32, (4), 511-7.

48.   van Els, C. A.; Herberts, C. A.; van der Heeft, E.; Poelen, M. C.; van Gaans-van den Brink, J. A.; van der Kooi, A.; Hoogerhout, P.; Jan ten Hove, G.; Meiring, H. D.; de Jong, A. P., A single naturally processed measles virus peptide fully dominates the HLA-A*0201-associated peptide display and is mutated at its anchor position in persistent viral strains. Eur J Immunol 2000, 30, (4), 1172-81.

49. Falk, K.; Rotzschke, O.; Stevanovic, S.; Jung, G.; Rammensee, H. G., Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature 1991, 351, (6324), 290-6.

50. Deres, K.; Beck, W.; Faath, S.; Jung, G.; Rammensee, H. G., MHC/peptide binding studies indicate hierarchy of anchor residues. Cell Immunol 1993, 151, (1), 158-67.

51. Caron, E.; Kowalewski, D. J.; Chiek Koh, C.; Sturm, T.; Schuster, H.; Aebersold, R., Analysis of Major Histocompatibility Complex (MHC) Immunopeptidomes Using Mass Spectrometry. Mol Cell Proteomics 2015, 14, (12), 3105-17.

52. Sugawara, S.; Abo, T.; Kumagai, K., A simple method to eliminate the antigenicity of surface class I MHC molecules from the membrane of viable cells by acid treatment at pH 3. J Immunol Methods 1987, 100, (1-2), 83-90.

53. Mommen, G. P.; Frese, C. K.; Meiring, H. D.; van Gaans-van den Brink, J.; de Jong, A. P.; van Els, C. A.; Heck, A. J., Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc Natl Acad Sci U S A 2014, 111, (12), 4507-12.

54. Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C., Orthogonality of separation in two-dimensional liquid chromatography. Anal Chem 2005, 77, (19), 6426-34.

55. Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C., Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. J Sep Sci 2005, 28, (14), 1694-703.

56. Essader, A. S.; Cargile, B. J.; Bundy, J. L.; Stephenson, J. L., Jr., A comparison of immobilized pH gradient isoelectric focusing and strong-cation-exchange chromatography as a first dimension in shotgun proteomics. Proteomics 2005, 5, (1), 24-34.

57. Delmotte, N.; Lasaosa, M.; Tholey, A.; Heinzle, E.; Huber, C. G., Two-dimensional reversed-phase x ion-pair reversed-phase HPLC: an alternative approach to high-resolution peptide separation for shotgun proteome analysis. J Proteome Res 2007, 6, (11), 4363-73.

58. Yang, F.; Shen, Y.; Camp, D. G., 2nd; Smith, R. D., High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. Expert Rev Proteomics 2012, 9, (2), 129-34.

59. Wang, H.; Sun, S.; Zhang, Y.; Chen, S.; Liu, P.; Liu, B., An off-line high pH reversed-phase fractionation and nano-liquid chromatography-mass spectrometry method for global proteomic profiling of cell lines. J Chromatogr B Analyt Technol Biomed Life Sci 2015, 974, 90-5.

60. Yates, J. R.; Ruse, C. I.; Nakorchevsky, A., Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng 2009, 11, 49-79.

61.   Paizs, B.; Suhai, S., Fragmentation pathways of protonated peptides. Mass Spectrom Rev 2005, 24, (4), 508-48.

62.   McAlister, G. C.; Phanstiel, D.; Good, D. M.; Berggren, W. T.; Coon, J. J., Implementation of electron-transfer dissociation on a hybrid linear ion trap-orbitrap mass spectrometer. Anal Chem 2007, 79, (10), 3525-34.

63.   Frese, C. K.; Altelaar, A. F.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J.; Mohammed, S., Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. Anal Chem 2012, 84, (22), 9668-73.

64.   Frese, C. K.; Zhou, H.; Taus, T.; Altelaar, A. F.; Mechtler, K.; Heck, A. J.; Mohammed, S., Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). J Proteome Res 2013, 12, (3), 1520-5.

65.   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 1991, 254, (5038), 1643-7.

66.   Brichard, V.; Van Pel, A.; Wolfel, T.; Wolfel, C.; De Plaen, E.; Lethe, B.; Coulie, P.; Boon, T., The tyrosinase gene codes for an antigen recognized by autologous cytolytic T lymphocytes on HLA-A2 melanomas. J Exp Med 1993, 178, (2), 489-95.

67.   Kawakami, Y.; Eliyahu, S.; Sakaguchi, K.; Robbins, P. F.; Rivoltini, L.; Yannelli, J. R.; Appella, E.; Rosenberg, S. A., Identification of the immunodominant peptides of the MART-1 human melanoma antigen recognized by the majority of HLA-A2-restricted tumor infiltrating lymphocytes. J Exp Med 1994, 180, (1), 347-52.

my HLA petides under all the polymer and ammonium acetate signal

HELP! HELP! I'M BEING REPRESSED!

Chapter 2

Pre-fractionation extends, but also creates a bias in the detectable HLA class I ligandome

Laura C. Demmers[1, 2], Albert J. R. Heck[1, 2], and Wei Wu[1, 2]

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
[2] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

Abstract

HLA class I molecules can present a range of cellular alterations (mutations, changes in protein copy number, aberrant post-translational modifications or pathogen proteins) to $CD8^+$ T lymphocytes, in the form of HLA peptide ligands. At any given moment, tens of thousands of different self and foreign HLA class I peptides may be presented on the cell surface by HLA class I complexes. Due to the enormous biochemical diversity and low abundance of each of these peptides, HLA ligandome analysis presents unique challenges. Even with advances in enrichment strategies and MS instrumentation and fragmentation, sufficient ligandome depth for identification of viral pathogens and immuno-therapeutically important tumor neo-antigens is still not routinely achievable. In this study, we evaluated two pre-fractionation techniques, high pH reversed phase and strong cation exchange, for complementary analyses of HLA class I peptide ligands. We observe that pre-fractionation substantially extends the detectable HLA class I ligandome, but also creates an identification bias. We thus advocate a rational choice between high pH reversed phase or strong cation exchange pre-fractionation for deeper HLA class I ligandome analysis, depending on the HLA locus, allele or peptide ligand modification in question.

Introduction

One critical function of the immune system is to counteract cancer development by identifying abnormal cells for destruction, through recognition of neo-antigens presented on these cells by human leukocyte antigen (HLA) molecules. Promising strategies in cancer immunotherapy target these neo-antigens, which arise from proteolytic processing of tumor-specific and/or mutated proteins (1). While some of these cancer signatures can be picked up by next generation sequencing at the DNA or RNA level, mass spectrometry (MS) remains the method of choice that enables direct measurement and identification of neo-antigens that are presented on the tumor cell surface. Therefore, ultra-sensitive MS profiling methodologies are needed to delve deeper into the HLA class I ligandome, to support the rational design of immunotherapy at a personalized level (2, 3).

HLA class I peptide ligands are challenging to analyze due to enormous diversity in the repertoire and the inherently low abundance of most of these species. Through the advances in mass spectrometry over the recent decades, the HLA class I peptide ligandome can now be analyzed in depth, reaching to the detection of hundreds up to thousands of unique peptide ligands per cell line or tissue sample. We previously demonstrated that the use of complementary peptide fragmentation techniques, especially EThcD as an alternative 'spectra-rich' fragmentation method, can further boost the identification of HLA class I peptide ligands, including those post-translationally modified by arginine methylation, glycosylation and phosphorylation (4-6). Sample pre-fractionation, notably by strong cation exchange (SCX), has also been shown to further increase the detection depth of the HLA class I ligandome (7). The outcomes of these deep HLA class I peptide ligand profiling experiments have prominently altered the developmental path of immuno-therapeutics, but also indicate that we still have plenty to gain by delving deeper into the HLA class I peptide ligandome.

The analysis of HLA class I ligandomes benefit in general from advances in shot-gun proteomics, as not all, but many experimental aspects of these analyses are shared. Since the advent of high pH reversed phase (RP) fractionation in 2012 (8-11) in shot-gun proteomics, it has become widely accepted as a robust and high-performance alternative to SCX (12-14). Given the ideal compatibility of high pH RP pre-fractionation to the second dimension, low pH RP-LC-MS, without the need for additional desalting procedures, many laboratories have adopted high pH RP for routine proteomics analyses. We reasoned that reducing purification steps between HLA class I peptide ligand isolation and MS analysis could further minimize the loss of low-abundant HLA class I peptide species and extend the identification depth. To this end, we benchmarked the performance of high pH RP pre-fractionation and SCX pre-fractionation against 'no-fractionation', by using an Epstein-Barr virus transformed immortalized B lymphoblastoid cell line JY. This cell line is a routinely

used model cell line in the field of immuno-peptidomics, that is homozygous for HLA-A*02:01, HLA-B*07:02 and HLA-C*07:02.

We describe here high pH RP pre-fractionation on HLA class I peptide ligands as a reliable alternative approach for detailed HLA class I ligandome profiling, that is highly complementary to the conventional SCX. We observed that high pH RP performed comparably to SCX in terms of total HLA class I peptide ligand identification, but presents a distinct allele-specific identification bias, that could be exploited for focused studies of hydrophobic peptide ligands and post-translationally modified peptide ligands, especially for charge-reducing modifications such as phosphorylation and citrullination. We conclude that the choice between these two pre-fractionation methods will depend on rational consideration of the research question surrounding a particular HLA locus, allele or peptide ligand post-translational modification.

Materials and Methods

Cell culture

The HLA class I homozygous cell line JY (HLA-A*02:01, HLA-B*07:02, HLA-C*07:02) was cultured in RPMI 1640 supplemented with 10% fetal bovine serum, 10mM L-glutamine, 50U/ml penicillin and 50μg/ml streptomycin in a humidified atmosphere at 37ºC with 5% CO_2.

Immuno-affinity purification

Per immuno-affinity purification (IP), $5*10^8$ cells were harvested by centrifugation and washed three times with cold PBS. Cells were lysed for 1,5h at 4ºC in 10ml lysis buffer per gram cell pellet. The lysis buffer consisted of Pierce IP lysis buffer (Thermo Fischer Scientific) supplemented with 1x complete protease inhibitor cocktail (Roche Diagnostics), 50μg/ml DNase I (Sigma-Aldrich) and 50μg/ml RNase A (Sigma-Aldrich). Subsequently, the lysate was cleared by centrifugation for 1h at 18,000g at 4ºC. Protein concentration was determined with the Bradford assay (Bio-Rad). HLA class I complexes and peptide ligands were immunoprecipitated using 0.5mg W6/32 antibody (15) coupled to 125μl Protein A/G beads (Santa Cruz) from 25mg whole cell lysate. Antibodies were crosslinked to protein A/G beads to prevent co-elution. Incubation took place at 4ºC for approximately 16h. After immunoprecipitation, the beads were washed with 40ml cold PBS. HLA class I complexes and peptide ligands were subsequently eluted with 10% acetic acid. Peptide ligands were separated from

HLA class I complexes using 10kDa molecular weight cutoff filters (Millipore). The flowthrough containing the HLA class I peptide ligands was dried by vacuum centrifugation.

Peptide fractionation

HLA class I peptide ligands from six IPs were mixed and divided into two equal samples for either C18 reversed-phase fractionation at high pH or strong cation exchange. In high pH reversed phase fractionation, peptides were loaded on C18 STAGE-tips in 200mM ammonium formate pH 10, and eluted into 12 fractions with 11-100% acetonitrile. For strong cation exchange, peptides were loaded on SCX SPE cartridges (1mg, Supelco) in 20% acetonitrile with 0.1% formic acid, and eluted into 12 fractions with 50-500mM ammonium acetate. All samples were dried by vacuum centrifugation, and reconstituted in 10% formic acid prior to LC-MS/MS analyses.

LC-MS/MS

The data was acquired with an UHPLC 1290 system (Agilent) coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fischer Scientific). Peptides were trapped (Dr Maisch Reprosil C18, 3µM, 2cm x 100µM) for 5 min in solvent A (0.1% formic acid in water) before being separated on an analytical column (Agilent Poroshell, EC-C18, 2.7µm, 50cm x 75µm). Solvent B consisted of 0.1% formic acid in 80% acetonitrile. For high pH reversed phase samples (fraction 1 and 2), the gradient was as follows: first 5 minutes trapping, followed by 85 minutes gradient from 12-30% solvent B and subsequently 10 minutes wash with 100% solvent B and 10 minutes re-equilibration with 100% solvent A. For fraction 3 and 4 the gradient was from 15-32% solvent B. For fraction 5 and 6 the gradient was from 18-36% solvent B. For fraction 7 to 10 the gradient was from 20-38% solvent B and for fraction 11 and 12 from 22-44% solvent B. For the SCX fractions, the gradient was as follows: first 5 minutes trapping, followed by 85 minutes gradient from 7-35% solvent B and subsequently 10 minutes wash with 100% solvent B and 10 minutes re-equilibration with 100% solvent A. The mass spectrometer operated in data-dependent mode. Full scan MS spectra from *m/z* 400-650 were acquired at a resolution of 60,000 after accumulation to a target value or $4*10^5$ or a maximum injection time of 50ms. Up to 3 most intense precursors with a charge state of 2+ or 3+ starting at *m/z* 100 were chosen for fragmentation. EThcD fragmentation was performed at 35% normalized collision energy on selected precursors with 18s dynamic exclusion after accumulation of $5*10^4$ ions or a maximum injection time of 250ms. MS/MS spectra were acquired at a resolution of 15,000.

Data analysis

Raw files were searched using Sequest HT in Proteome Discoverer 2.2, against the Swissprot human database (20258 entries, downloaded on 02-02-2018) appended with the 20 most abundant FBS contaminants (16). The search was set to unspecific with a minimum precursor mass of 797 Da to a maximum precursor mass of 1950 Da corresponding to peptides between 8 and 12 amino acids long. Identified peptides were filtered to a 1% false discovery rate (FDR) using the percolator algorithm, 5% peptide FDR and Xcorr > 1. Cysteine cysteinylation and methionine oxidation were set as variable modifications. Serine phosphorylation and arginine citrullination were set as variable modifications in separate workflows. From the identified peptides, FBS contaminants were removed. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (17) partner repository with the dataset identifier PXD011257. Binding affinity of HLA class I peptide ligands was predicted using the NetMHC 4.0 algorithm (18) with a stringent binder cutoff of IC50 <500nM. The data was visualized with Graphpad PRISM 7.

Results

To compare the effectiveness of offline high pH reversed phase and strong cation exchange pre-fractionation in extending coverage of the HLA class I ligandome, we started with a 'master pool' of immuno-affinity purified peptide ligands, and separated each half into 12 fractions for further LC-MS/MS analyses (Figure 1).

Fractionation reaches deeper into the HLA class I ligandome

Without any pre-fractionation, we identified approximately 6500±100 peptides per single LC-MS injection, of which 5700±100 were HLA class I peptide ligands. Repeated injections of the same unfractionated sample did not substantially increase the cumulative identification further; with an excellent identification overlap of >95% between injection replicates, the same peptide ligand species were fragmented over 12 repeated injections (Figure 2A-B). This implied that the limit in identification was not due to stochastic sampling (Figure 2A), making it also very clear that for deeper analysis of the HLA class I ligandome, sample pre-fractionation is essential.
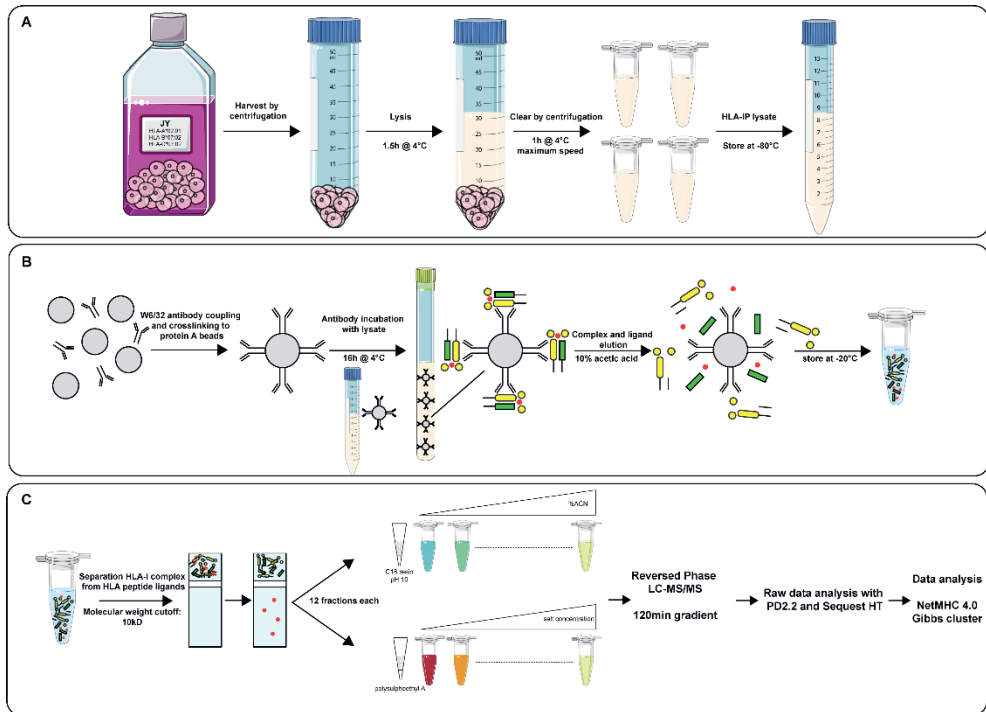
Figure 1: Experimental workflow applied for HLA class I peptide ligand analysis. A) Cell harvest and lysis steps for lysate preparation for HLA class I immuno-affinity purification. **B)** HLA class I complex and peptide ligand immuno-affinity purification using W6/32 antibodies coupled to protein A/G beads. **C)** HLA peptide ligand purification, fractionation and identification.

By pre-fractionation of the immuno-purified HLA class I ligandome into 12 fractions, with either SCX or high pH RP, we were able to extend the number of identifications to above 10,000 peptides, of which about 8600 were HLA class I peptide ligands (Figure 2C-D). Considering the total number of identifications, either pre-fractionation method could increase identifications by ~50% compared to without fractionation. However, the overlap in identified peptide ligands between high pH RP fractionation and SCX fractionation turned out to be only ~40%, alluding to the idea that these fractionation methods are very complementary in ligandome analyses (Figure 2E). Indeed, when we combined the data from both fractionation methods, we identified 13,500 peptides, of which 11,200 HLA class I peptide ligands (Figure 2F), representing an increase of ~100% in identified HLA class I peptide ligands compared to the analysis of the same samples but unfractionated. Therefore, we demonstrate here for the first time that high pH reversed phase fractionation is a good alternative to SCX for deeper analysis of HLA class I peptide ligands. High pH reversed phase and SCX fractionation also appear to be complementary in the

analysis of HLA class I peptide ligands, both reaching much deeper into the HLA class I peptide ligandome, when compared to unfractionated analysis.
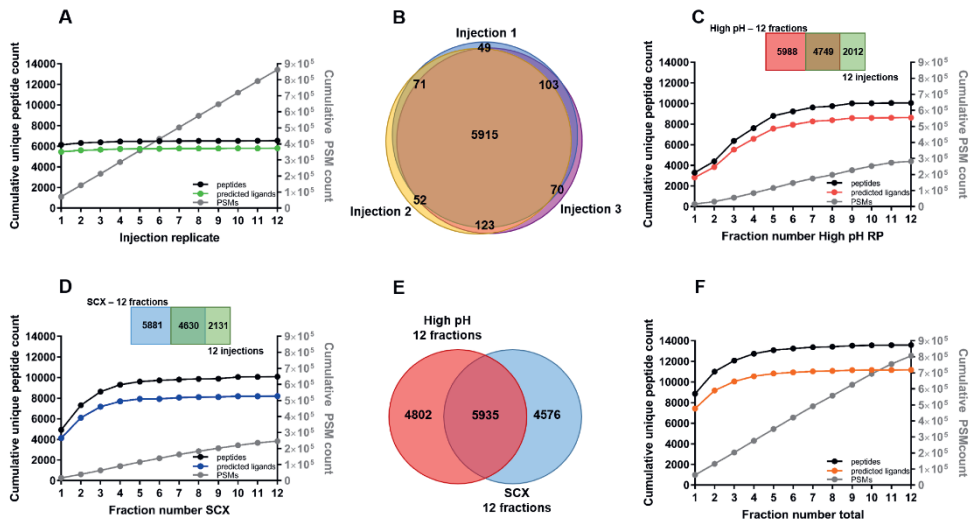
Figure 2: HLA class I peptide ligand identification. A) Cumulative unique peptides (black), predicted peptide ligands (green) and PSM counts (grey) over repeated injections of the same unfractionated sample. **B)** Peptide identification overlap between repeated injections of the same unfractionated sample. **C)** Cumulative unique peptides (black), predicted peptide ligands (red) and PSM (grey) counts over all analyzed high pH RP fractions. **D)** Cumulative unique peptides (black), predicted peptide ligands (blue) and PSM (grey) counts over all analyzed SCX fractions. **E)** Overlap between identified peptides in high pH RP (n=10737, red) and SCX (n=10511, blue). **F)** Cumulative unique peptides (black), predicted peptide ligands (orange) and PSM (grey) count over combined high pH RP and SCX fractions.

Boosting HLA class I peptide ligand extraction

While pre-fractionation boosts identification of HLA class I peptide ligands considerably (Figure 2), more starting material is also needed than in single-shot LC-MS analyses. This is often a limiting bottleneck, especially in the context of ligandome profiling from patient material. Since obtaining larger biopsies is not feasible, maximizing the extraction of HLA class I complexes and peptide ligands from limited material is then crucial to maximize ligandome coverage, for more in-depth patient-specific analyses.

It has been demonstrated that immuno-affinity purifications cannot fully deplete target protein complexes in one cycle, largely due to affinity characteristics of antibody-based capture and equilibrium dissociation (19, 20). Therefore, we further tested if recycling the flowthrough from immuno-affinity purification for sequential enrichments could still increase the yield of HLA class I peptide ligands. We performed sequential immuno-affinity purifications using the same starting cell

lysate, injected these samples in separate LC-MS runs with the same parameters and then analyzed the purity of HLA class I peptide ligands measured. As shown in Figure 3A and 3B, approximately 5000 peptides were identified from the 1st and 2nd IP, of which about 4500 were HLA class I peptide ligands. In the 3rd IP, the purity was still good, albeit the total identification number was about 10% lower than in the first two IPs. By the 4th reuse of the same lysate, peptide ligand identification declined to approximately 2600, with a marginal increase of unassigned peptides (15%), suggesting the lysate was starting to become depleted of HLA class I complexes. Collectively, these data demonstrate that the same starting material for HLA class I peptide ligand immuno-affinity purification may be re-used for several times for repeated extraction of the same HLA class I complexes. We further verified that the HLA class I peptide ligands identified in sequential IPs share a large overlap of ~80%, validating that repeated use of the same starting material does not introduce profiling bias or sampling artifacts associated with extended incubation times (Figure 3C).

Peptide characteristics of HLA class I peptide ligands detected with high pH reversed phased or SCX pre-fractionation

By pooling HLA class I peptide ligands obtained from two sets of immuno-affinity purifications, each performed for three sequential times, we created a common pool of HLA class I peptide ligands which was then used to compare and contrast the properties of peptide ligands identified by either high pH RP or SCX fractionation. The goal was to rationalize the substantial non-overlap in the HLA peptide ligand identifications between these two complementary pre-fractionation approaches (Figure 2E).
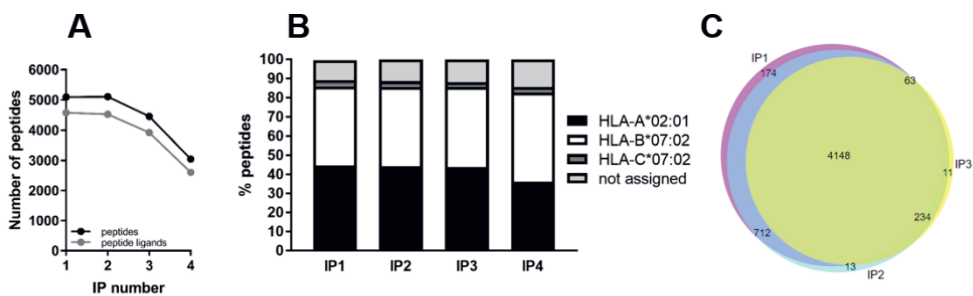

Figure 3: HLA class I peptide ligand identification in sequential immuno-affinity purifications using single LC-MS/MS runs. A) Unique peptide (black) and predicted peptide ligand (grey) identifications. **B)** Proportion of HLA-A*02:01, HLA-B*07:02 and HLA-C*07:02 peptide ligands. **C)** Identification overlap of the first three sequential IPs.

Specifically, we examined peptide characteristics within predicted ligands for each HLA allele (HLA-A*02:01, HLA-B*07:02, HLA-C*07:02), to check for potential bias in peptide lengths (Figure 4A) and sequence motifs (Figure 4B). HLA class I peptide ligands are typically between 8 and 12 amino acids long, with 9-residue ligands being the most frequent. By comparing the peptide ligands identified with and without pre-fractionation, we found no significant difference in the peptide length distribution (Figure 4A). Amongst HLA-A*02:01 peptide ligands, ~70% were 9-mers, ~20% 10-mers and <10% were 11-mers regardless of pre-fractionation approach. Comparable trends were also observed in HLA-B*07:02 and HLA-C*07:02. In addition, Gibbs clustering sequence logos (21) of the peptide ligands identified with or without pre-fractionation were almost indistinguishable, with strong conservation of anchor residues at position 2 and the C-terminus (Figure 4B). Regardless of the fractionation approach used, HLA-A*02:01 ligands anchor with leucine at position 2 and a C-terminal leucine/valine; HLA-B*07:02 ligands show enrichment for proline at position 2 and a C-terminal leucine/valine; whereas HLA-C*07:02 ligands over-represent arginine at position 2 and a C-terminal leucine/phenylalanine.

Taken together, our data suggest that within ligands of the same HLA allele, increased identification through pre-fractionation is not related to unique peptide properties that induce preferential binding and/or separation in either high pH RP or SCX.

High pH reversed phase or SCX fractionation each introduces an allele-specific identification bias

We next sought to explain the substantial non-overlap in HLA class I peptide ligand identification in high pH RP or SCX, by calculating the contribution of peptide ligands, from each allele, to total ligand identification in the JY cell line (Figure 5). Specifically, we examined separately the proportion of HLA-A*02:01, HLA-B*07:02 and HLA-C*07:02 peptide ligands in high pH RP or SCX fractionation to check for any allele-specific bias. Using binding affinity of peptides, predicted by NetMHC 4.0 with a stringent cutoff of IC50 <500nM, we binned identified peptide ligands by allele specificity. As shown in Figure 5, except for a small percentage of HLA-C*07:02 ligands and 11% non-assigned peptides, HLA-A*02:01 and HLA-B*07:02 ligands were represented almost evenly without pre-fractionation (41.0% and 44.1% respectively), using the data from 12 repeated injections of the same sample (Figure 5A). However, with pre-fractionation, HLA-A*02:01 ligands appear to be over-represented in high pH RP while HLA-B*07:02 ligands appear to be identified more frequently with SCX. Since HLA-C*07:02 ligands represent only a small and consistent proportion in either fractionation approach, we visualized only the binary distribution of HLA-A*02:01 and HLA-B*07:02 ligands in figure 5B, where a clear bias in HLA allele specificity was observed compared to without pre-fractionation.
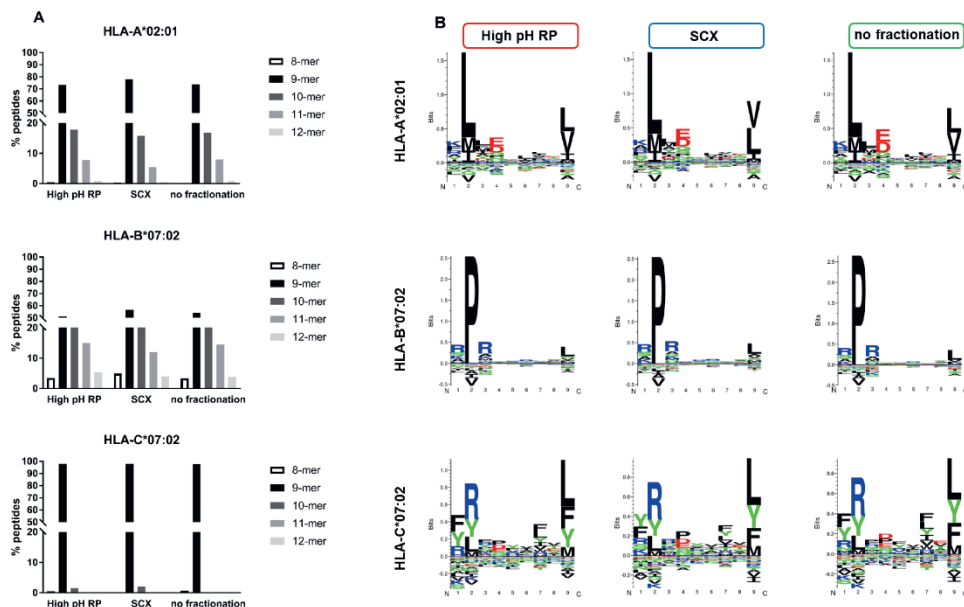
Figure 4: Comparable HLA class I peptide ligand characteristics in high pH RP, SCX and no-fractionation. A) Length distribution of HLA-A*02:01 peptide ligands (high-pH RP n=4705, SCX n=3419, no fractionation n=2921), HLA-B*07:02 peptide ligands (high pH RP n=4075, SCX n=4825, no fractionation n=2804) and HLA-C*07:02 peptide ligands (high-pH RP n=389, SCX n=290, no fractionation n=252). **B)** Gibbs clustering sequence motifs for HLA-A*02:01, HLA-B*07:02 and HLA-C*07:02 peptides identified with high-pH RP, SCX or without pre-fractionation.

By following cumulative peptide ligand identification over either 12 high pH RP or 12 SCX fractions, we further rationalized the lower identification of HLA-A*02:01 ligands in SCX. As shown in Figure 5C, the cumulative increase in unique HLA-A*02:01 ligands was restricted to the first 3 SCX fraction in a linear salt gradient, while HLA class I peptide ligands continue to be identified over the 12 fractions (Figure 2D, cumulative PSM count). This indicates that HLA-A*02:01 binders are not optimally separated by using the charge selective SCX gradient, likely due to higher peptide hydrophobicity attributed to the leucine anchor and prevalent leucine/valine C-terminus. We assessed the isoelectric point distribution of HLA-A*02:01 peptides, which revealed a striking majority of peptides with an isoelectric point between 3 and 7. These peptides would be negatively charged at pH 3 and therefore not optimally separated on a positive-charge selective SCX gradient (Figure 5D). On the other hand, HLA-B*07:02 ligands feature prominently arginine residues flanking the proline anchor at position 2. This makes HLA-B*07:02 ligands more charged in low pH SCX, and therefore better separated and consequently better identified in SCX across all 12 fractions when compared to high pH RP (Figure 5E). Assessment of the

isoelectric point distribution of these peptides again confirms that more HLA-B*07:02 peptide ligands have isoelectric points between 8 and 12 (Figure 5F). Based on these observations we therefore conclude that in high pH RP there is preferential identification of HLA-A*02:01 peptide ligands, whereas SCX performs better in the analysis of HLA-B*07:02 peptide ligands.
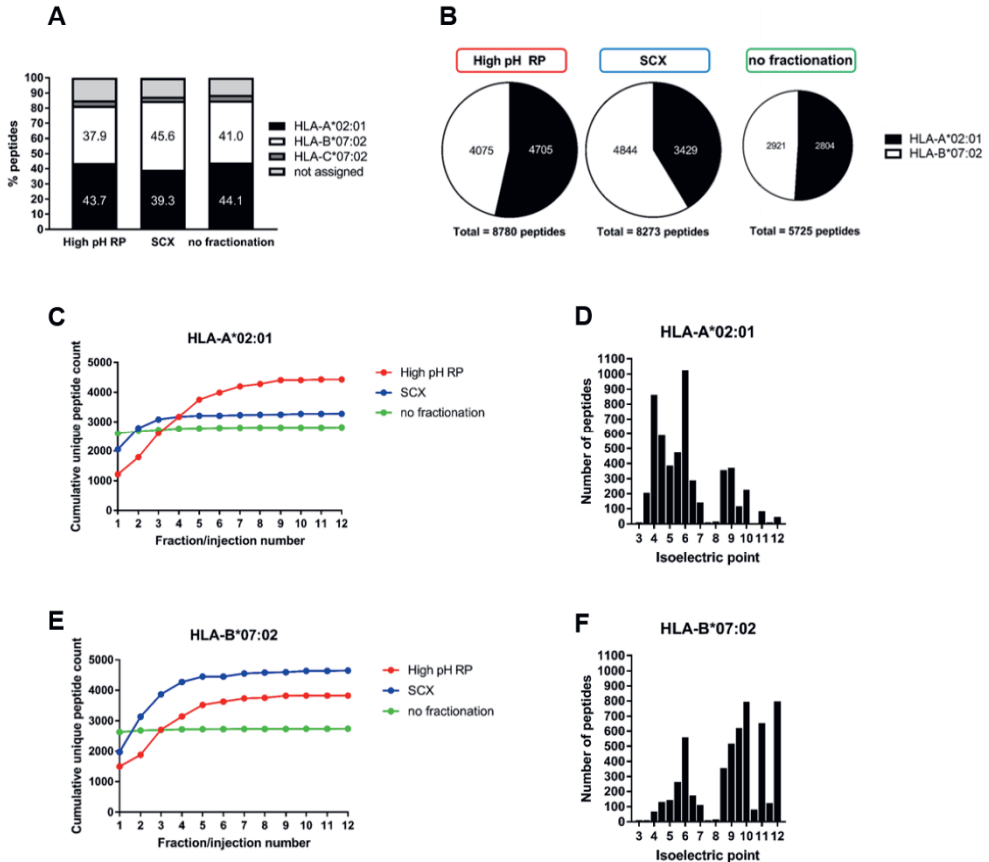
Figure 5: Bias in allele-specific HLA class I peptide ligand identification in high pH RP, SCX compared to no fractionation. A) Identification of HLA-A*02:01, HLA-B*07:02 and HLA-C*07:02 peptide ligands per fractionation method. **B)** Identification bias in HLA-A*02:01, HLA-B*07:02 peptide ligands when using high pH RP fractionation, SCX fractionation or no fractionation. **C)** Cumulative unique HLA-A*02:01 peptide ligand count in high pH RP (red), SCX (blue) or without fractionation (green). **D)** Isoelectric point distribution of all identified HLA-A*02:01 peptide ligands. **E)** Cumulative unique HLA-B*07:02 peptide ligand count in high pH RP (red), SCX (blue) or without fractionation (green). **F)** Isoelectric point distribution of all identified HLA-B*07:02 peptide ligands.

High pH reversed phase fractionation enables in-depth profiling of HLA class I peptide ligands containing serine phosphorylations
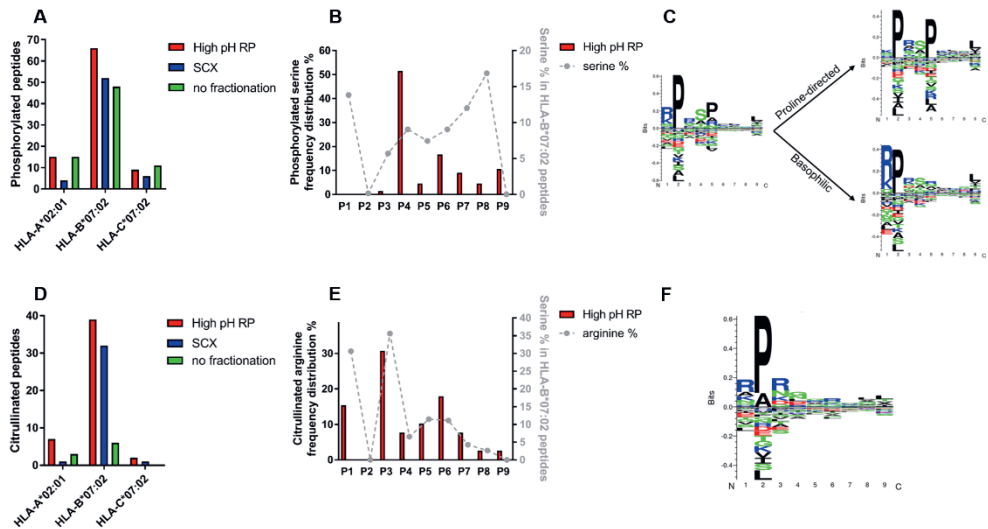
Serine phosphorylation on HLA class I peptide ligands have been reported previously and can be viewed as neo-antigens arising from aberrant phosphorylation in tumors (6, 22, 23). Since serine phosphorylation reduces the charge of peptides at low pH, we reasoned serine phosphorylated HLA class I peptide ligands would be better separated and identified using high pH RP pre-fractionation compared to SCX (24). We therefore attempted to detect and localize serine phosphorylation on HLA class I peptide ligands with either high pH RP or SCX fractionation (Figure 6A-C). In total, we detected 45% more phosphopeptides using the high pH RP workflow than SCX (90 phospho-HLA peptides vs 62 respectively; Supporting Table S1), in strong agreement with our theoretical understanding. Remarkably, the SCX workflow even provided a lower number of phosphorylated peptide ligands than the unfractionated workflow (62 phospho-HLA peptides versus 74 respectively, Supporting Table S1). Also shown in Figure 6A, serine phosphorylation occurs predominantly on HLA-B*07:02 peptide ligands and is indeed most frequently detected in high pH RP fractionation.

We next focused on the characteristics of the phosphorylated HLA-B*07:02 peptide ligands detected with high pH RP. Interestingly, although positions 1 and 8 of HLA-B*07:02 peptide ligands are most often occupied by serine residues, almost no phosphorylations are observed at these sites (Figure 6B). On the contrary, phosphorylated serines are most frequently localized to position 4; in ~50% of all the serine phosphorylated HLA-B*07:02 peptides. Deeper analysis of serine phosphorylated peptides revealed the presence of two distinct kinase motifs (Figure 6C). Hence, it seems that the majority of phosphorylated peptide ligands for loading on HLA-B*07:02 may have been provided by proline-directed kinases with SP consensus and basophilic kinases with RxxS consensus motif (25).

Pre-fractionation is essential for detection of citrullinated HLA class I peptide ligands

Extending the depth in analysis of the HLA class I ligandome by pre-fractionation leads not only to increased identification of phosphorylated HLA class I peptide ligands, but has also been shown to improve detection of HLA class I peptide ligands containing other post-translational modifications, e.g. OGlcNAcylation (4) or arginine (di)methylation (5). In our data, we next looked for HLA class I peptide ligands harboring arginine citrullination, another important modification linked to rheumatoid arthritis (26-30). In contrast to serine phosphorylation, arginine citrullination is detected rather poorly without pre-fractionation. In our analysis, we were able to detect 48 and 34 citrullinated HLA-B*07:02 peptide ligands with high pH RP and SCX respectively compared to 16 in unfractionated samples (Figure 6D, Supporting Table

S2). This implies that pre-fractionation is almost a necessity for detailed study of this modification. In our data, arginine citrullination is mostly detected on HLA-B*07:02 peptide ligands. This is not surprising as binders of HLA-B*07:02 contain more arginine residues on average.

Figure 6: Serine phosphorylation and arginine citrullination on JY HLA class I peptide ligands are preferred on HLA-B*07:02. A) HLA assigned peptides containing phosphoserine distributed per allele per fractionation method. **B)** Serine phosphorylation per position in HLA-B*07:02 peptide ligands identified with high pH RP compared to overall serine percentage per position in HLA-B*07:02 peptide ligands. **C)** Gibbs cluster sequence logo from HLA-B*07:02 phosphorylated peptides (n = 66) and motifs specific for proline directed kinases (n = 32) and basophilic kinases (n= 34). **D)** HLA assigned peptides containing citrullinated arginine distributed per allele per fractionation method. E) Arginine citrullination per position in HLA-B*07:02 peptide ligands identified with high pH RP compared to overall arginine percentage per position in HLA-B*07:02 peptide ligands. F) Gibbs cluster sequence logo from HLA-B*07:02 citrullinated peptides (n = 39).

In contrast to the preference for serine phosphorylation at position 4, the distribution of arginine citrullination per position in the peptide ligands followed largely the overall arginine frequency per position in HLA-B*07:02 peptide ligands (Figure 6E-F). Cumulatively, we detected here the highest number of citrullinated HLA peptide ligands reported to date (Supporting table S2). Notably, some of these citrullinated peptides originate from the filaggrin source protein, which has been reported to harbor several citrullinated rheumatoid arthritis-specific epitopes (31). Some of citrullinated peptides detected in this work overlap with the rheumatoid arthritis epitopes reported earlier, while we also detect novel arginine citrullination sites on filaggrin.

Discussion

Here, we profiled HLA class I peptide ligands from the HLA homozygous cell line JY to benchmark the utility of two complementary fractionation approaches, namely high pH RP and SCX. Compared to without fractionation, each of these strategies expanded the HLA class I ligandome coverage by about 50%, but significant non-overlap between these two strategies also provides a cumulative gain in peptide identification of >100% compared to without fractionation. This significant boost in ligandome coverage is however highly dependent on intrinsic properties of peptide ligands. Hence, we clearly demonstrate that the choice of pre-fractionation approach can introduce an allele specific analytical bias. We showed experimentally that HLA class I peptide ligands with largely hydrophobic residues (e.g. HLA-A*02:01, motif xLxxxLLx(V/L)) are better pre-fractionated on a less charge selective high pH gradient, whereas charged HLA class I peptide ligands containing arginines (HLA-B*07:02, motif RPRxxRxx(L/V)) are better pre-fractionated and thus detected with SCX. Since this bias is strongly influenced by conserved properties of the HLA class I peptide ligands, we think there is a critical need to rationalize which fractionation approach to use, depending on both the HLA locus to be investigated and the allele specificity within the locus.

On the other hand, this apparent bias presents an opportunity to further improve allele-specific ligandome coverage. For instance, a large majority of HLA-A*02 alleles have hydrophobic ligands that feature predominantly in leucine and valine residues (e.g. HLA-A*02:03, motif xLAxx(L/V)xx(L/V)l HLA-A*02:06, motif (xLLxxLxx(V/L)). This implies that high pH RP pre-fractionation would also be the better choice for most HLA-A*02 peptide ligands to maximize specific coverage in the respective ligandomes (Supplementary Figure S1). Conversely, just like for HLA-B*07:02, ligands of other HLA-B alleles (e.g. HLA-B*27:02, motif RRLxxxxxL; HLA-B*27:20, motif RRxxxxxRL) or HLA-A alleles with affinity for charged peptides (e.g. HLA-A*30:01, motif RPRxxRxx(L/V); HLA-A*31:01, motif RTRxxxxxR) would also be better analyzed by using SCX. In view of these considerations, we put forth high pH RP as a valuable alternative analytical strategy to choose from, to further expand the coverage of hydrophobic HLA class I peptide ligands. It is key to note that employing high pH RP alone instead of SCX already profiles the ligandome to similar depth with comparable total identifications, but in addition with a bigger proportion of hydrophobic peptide ligands. This further implies that by choosing the appropriate strategy between high pH RP and SCX, the allele-specific ligandome space could be expanded.

In addition, post-translation modifications on HLA class I peptide ligands can also alter the biophysical and electrostatic properties of these peptides, to further impact the ideal choice of analytical strategy. For instance, phospho-modifications will reduce the net charge of peptide ligands, and theoretically makes these peptides better retained and separated on a non-charge selective high pH RP gradient. We

validated this experimentally with indeed more phosphorylated peptides detected using the high pH RP workflow. In fact, fractionating phosphorylated peptide ligands on a sub-optimal charge-selective gradient results in even fewer identifications compared to without any pre-fractionation. This further affirms the importance of a rational choice of pre-fractionation method, and shows that an inappropriate strategy will defeat the purpose of fractionation altogether. With the same rational thought, modifications that involve removal of a charge, such as citrullination, would also be better analyzed on high pH RP, as we documented here experimentally.

While pre-fractionation offers distinct advantages in allele-specific and PTM-specific HLA peptide ligand identification, more starting material is inherently needed to harness these benefits when compared to the unfractionated workflow. To tackle this, we show here that it is possible to obtain more HLA peptide ligands by re-using the flowthrough from immuno-affinity purifications, without increasing the initial input material. This can also further relieve the specimen bottleneck on patient-specific HLA peptide ligand analyses. We verify here that repeated use of the lysate does not compromise the quality and purity of the immuno-affinity purification, even to the point of HLA complex depletion, and that a large overlap of >80% in MS identification is still possible between sequential re-use.

Utilizing all strategies and considerations described above, we deeply profiled the phosphorylated HLA-B*07:02 ligandome by high pH RP, to examine the preferred site of serine phosphorylation against serine occupancy in the peptide ligand sequence. We found a strong preference for phospho-serine at position 4, but not at positions 1 and 8, despite higher occurrence of serine in the latter. This, we further rationalized against the loading model proposed previously; HLA-B antigens are collectively stabilized at position 1 by π-π stacking at position 1 with the R62 guanidinium group, hydrophobic interaction with the W167 indole group and the salt bridge with the N163 carboxyl group on the HLA-B backbone (6). Phosphorylation at serine in position 1 is likely to critically destabilize these docking interactions, such that peptides serine-phosphorylated at position 1 can no longer be loaded, whereas serine phosphorylations at position 4 can be stabilized through contact with R62 and a water-mediated reaction with the carboxyl group of E163 in the HLA-B backbone (6), and thus are more prominently observed. Thus our data strongly supports the loading preference of HLA-B peptide ligands reported previously, where a phospho-serine neo-antigen at P4 can extend out of the binding groove, to bind putatively to T-cells in a PTM-dependent manner (22, 23, 32-34).

Taken together, we show in this work that detection of HLA class I peptide ligands can be improved tremendously by a carefully deliberated choice, or complementary use of high pH RP and SCX fractionation. We demonstrate here that the physical and chemical properties of HLA class I peptide ligands can strongly influence the choice of analytical strategy, and that with a research question in mind surrounding a particular HLA locus, allele or peptide ligand PTM, a rational consideration of which

pre-fractionation to adopt will meaningfully expand coverage in the ligandome space of interest, and potentially boost identification of the much sought-after tumor neo-antigens.

Acknowledgements

Supporting information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.8b00821.

References

1.   Vinay, D. S.; Ryan, E. P.; Pawelec, G.; Talib, W. H.; Stagg, J.; Elkord, E.; Lichtor, T.; Decker, W. K.; Whelan, R. L.; Kumara, H.; Signori, E.; Honoki, K.; Georgakilas, A. G.; Amin, A.; Helferich, W. G.; Boosani, C. S.; Guha, G.; Ciriolo, M. R.; Chen, S.; Mohammed, S. I.; Azmi, A. S.; Keith, W. N.; Bilsland, A.; Bhakta, D.; Halicka, D.; Fujii, H.; Aquilano, K.; Ashraf, S. S.; Nowsheen, S.; Yang, X.; Choi, B. K.; Kwon, B. S., Immune evasion in cancer: Mechanistic basis and therapeutic strategies. Semin Cancer Biol 2015, 35 Suppl, S185-S198.

2.   Schumacher, T. N.; Schreiber, R. D., Neoantigens in cancer immunotherapy. Science 2015, 348, (6230), 69-74.

3.   Bassani-Sternberg, M.; Coukos, G., Mass spectrometry-based antigen discovery for cancer immunotherapy. Curr Opin Immunol 2016, 41, 9-17.

4.   Marino, F.; Bern, M.; Mommen, G. P. M.; Leney, A. C.; van Gaans-van den Brink, J. A. M.; Bonvin, A.; Becker, C.; van Els, C.; Heck, A. J. R., Extended

O-GlcNAc on HLA Class-I-Bound Peptides. J Am Chem Soc 2015, 137, (34), 10922-10925.

5. Marino, F.; Mommen, G. P.; Jeko, A.; Meiring, H. D.; van Gaans-van den Brink, J. A.; Scheltema, R. A.; van Els, C. A.; Heck, A. J., Arginine (Di)methylated Human Leukocyte Antigen Class I Peptides Are Favorably Presented by HLA-B*07. J Proteome Res 2017, 16, (1), 34-44.

6. Alpizar, A.; Marino, F.; Ramos-Fernandez, A.; Lombardia, M.; Jeko, A.; Pazos, F.; Paradela, A.; Santiago, C.; Heck, A. J.; Marcilla, M., A Molecular Basis for the Presentation of Phosphorylated Peptides by HLA-B Antigens. Mol Cell Proteomics 2017, 16, (2), 181-193.

7. Mommen, G. P.; Frese, C. K.; Meiring, H. D.; van Gaans-van den Brink, J.; de Jong, A. P.; van Els, C. A.; Heck, A. J., Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc Natl Acad Sci U S A 2014, 111, (12), 4507-12.

8. Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C., Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. J Sep Sci 2005, 28, (14), 1694-703.

9. Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C., Orthogonality of separation in two-dimensional liquid chromatography. Anal Chem 2005, 77, (19), 6426-34.

10. Essader, A. S.; Cargile, B. J.; Bundy, J. L.; Stephenson, J. L., Jr., A comparison of immobilized pH gradient isoelectric focusing and strong-cation-exchange chromatography as a first dimension in shotgun proteomics. Proteomics 2005, 5, (1), 24-34.

11. Delmotte, N.; Lasaosa, M.; Tholey, A.; Heinzle, E.; Huber, C. G., Two-dimensional reversed-phase x ion-pair reversed-phase HPLC: an alternative approach to high-resolution peptide separation for shotgun proteome analysis. J Proteome Res 2007, 6, (11), 4363-73.

12. Yang, F.; Shen, Y.; Camp, D. G., 2nd; Smith, R. D., High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. Expert Rev Proteomics 2012, 9, (2), 129-34.

13. Wang, H.; Sun, S.; Zhang, Y.; Chen, S.; Liu, P.; Liu, B., An off-line high pH reversed-phase fractionation and nano-liquid chromatography-mass spectrometry method for global proteomic profiling of cell lines. J Chromatogr B Analyt Technol Biomed Life Sci 2015, 974, 90-5.

14. Batth, T. S.; Francavilla, C.; Olsen, J. V., Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. J Proteome Res 2014, 13, (12), 6176-86.

15. Barnstable, C. J.; Bodmer, W. F.; Brown, G.; Galfre, G.; Milstein, C.; Williams, A. F.; 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.

16. Zheng, X.; Baker, H.; Hancock, W. S.; Fawaz, F.; McCaman, M.; Pungor, E., Jr., Proteomic analysis for the assessment of different lots of fetal bovine serum as a raw material for cell culture. Part IV. Application of proteomics to the manufacture of biological drugs. Biotechnol Prog 2006, 22, (5), 1294-300.

17. Vizcaino, J. A.; Csordas, A.; del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; Xu, Q. W.; Wang, R.; Hermjakob, H., 2016 update of the PRIDE database and its related tools. Nucleic Acids Res 2016, 44, (D1), D447-56.

18. Andreatta, M.; Nielsen, M., Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics 2016, 32, (4), 511-7.

19. Talmage, D. W., The primary equilibrium between antigen and antibody. Ann N Y Acad Sci 1957, 70, (1), 82-93.

20. Reverberi, R.; Reverberi, L., Factors affecting the antigen-antibody reaction. Blood Transfus 2007, 5, (4), 227-40.

21. Andreatta, M.; Alvarez, B.; Nielsen, M., GibbsCluster: unsupervised clustering and alignment of peptide sequences. Nucleic Acids Res 2017, 45, (W1), W458-W463.

22. Zarling, A. L.; Ficarro, S. B.; White, F. M.; Shabanowitz, J.; Hunt, D. F.; Engelhard, V. H., Phosphorylated peptides are naturally processed and presented by major histocompatibility complex class I molecules in vivo. J Exp Med 2000, 192, (12), 1755-62.

23. Zarling, A. L.; Polefrone, J. M.; Evans, A. M.; Mikesh, L. M.; Shabanowitz, J.; Lewis, S. T.; Engelhard, V. H.; Hunt, D. F., Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. Proc Natl Acad Sci U S A 2006, 103, (40), 14889-94.

24. Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P., Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci U S A 2004, 101, (33), 12130-5.

25. Amanchy, R.; Periaswamy, B.; Mathivanan, S.; Reddy, R.; Tattikota, S. G.; Pandey, A., A curated compendium of phosphorylation motifs. Nat Biotechnol 2007, 25, (3), 285-6.

26. James, E. A.; Moustakas, A. K.; Bui, J.; Papadopoulos, G. K.; Bondinas, G.; Buckner, J. H.; Kwok, W. W., HLA-DR1001 presents "altered-self" peptides derived from joint-associated proteins by accepting citrulline in three of its binding pockets. Arthritis Rheum 2010, 62, (10), 2909-18.

2

27.   Nguyen, H.; James, E. A., Immune recognition of citrullinated epitopes. Immunology 2016, 149, (2), 131-8.

28.   Scally, S. W.; Petersen, J.; Law, S. C.; Dudek, N. L.; Nel, H. J.; Loh, K. L.; Wijeyewickrema, L. C.; Eckle, S. B.; van Heemst, J.; Pike, R. N.; McCluskey, J.; Toes, R. E.; La Gruta, N. L.; Purcell, A. W.; Reid, H. H.; Thomas, R.; Rossjohn, J., A molecular basis for the association of the HLA-DRB1 locus, citrullination, and rheumatoid arthritis. J Exp Med 2013, 210, (12), 2569-82.

29.   Seward, R. J.; Drouin, E. E.; Steere, A. C.; Costello, C. E., Peptides presented by HLA-DR molecules in synovia of patients with rheumatoid arthritis or antibiotic-refractory Lyme arthritis. Mol Cell Proteomics 2011, 10, (3), M110 002477.

30.   Ting, Y. T.; Petersen, J.; Ramarathinam, S. H.; Scally, S. W.; Loh, K. L.; Thomas, R.; Suri, A.; Baker, D. G.; Purcell, A. W.; Reid, H. H.; Rossjohn, J., The interplay between citrullination and HLA-DRB1 polymorphism in shaping peptide binding hierarchies in rheumatoid arthritis. J Biol Chem 2018, 293, (9), 3236-3251.

31.   Union, A.; Meheus, L.; Humbel, R. L.; Conrad, K.; Steiner, G.; Moereels, H.; Pottel, H.; Serre, G.; De Keyser, F., Identification of citrullinated rheumatoid arthritis-specific epitopes in natural filaggrin relevant for antifilaggrin autoantibody detection by line immunoassay. Arthritis Rheum 2002, 46, (5), 1185-95.

32.   Andersen, M. H.; Bonfill, J. E.; Neisig, A.; Arsequell, G.; Sondergaard, I.; Valencia, G.; Neefjes, J.; Zeuthen, J.; Elliott, T.; Haurum, J. S., Phosphorylated peptides can be transported by TAP molecules, presented by class I MHC molecules, and recognized by phosphopeptide-specific CTL. J Immunol 1999, 163, (7), 3812-8.

33.   Mohammed, F.; Cobbold, M.; Zarling, A. L.; Salim, M.; Barrett-Wilt, G. A.; Shabanowitz, J.; Hunt, D. F.; Engelhard, V. H.; Willcox, B. E., Phosphorylation-dependent interaction between antigenic peptides and MHC class I: a molecular basis for the presentation of transformed self. Nat Immunol 2008, 9, (11), 1236-43.

34.   Petersen, J.; Wurzbacher, S. J.; Williamson, N. A.; Ramarathinam, S. H.; Reid, H. H.; Nair, A. K.; Zhao, A. Y.; Nastovska, R.; Rudge, G.; Rossjohn, J.; Purcell, A. W., Phosphorylated self-peptides alter human leukocyte antigen class I-restricted antigen presentation and generate tumor-specific epitopes. Proc Natl Acad Sci U S A 2009, 106, (8), 2776-81.

Supplementary Figure S1
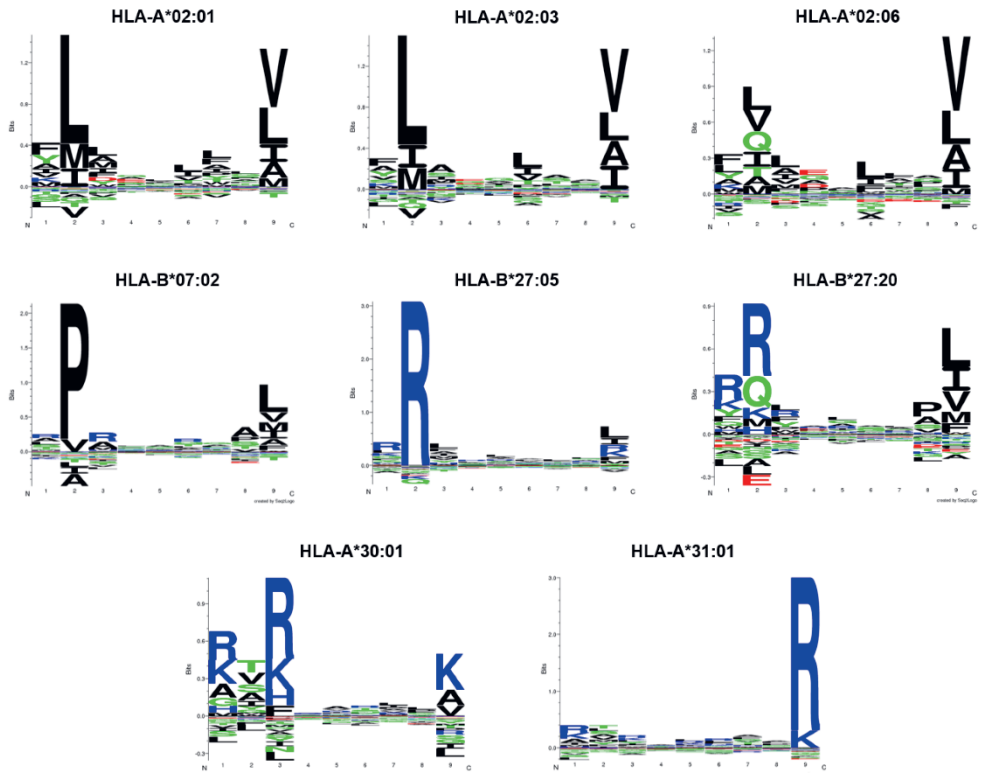
Supplementary Figure S1: Gibbs clustering motifs for nine different HLA types.

Chapter 3

HLA class II presentation is specifically altered at elevated temperatures in the B-lymphoblastic cell line JY

Laura C. Demmers[1, 2], Wei Wu[1, 2] and Albert J. R. Heck[1, 2]

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
[2] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

Abstract

HLA class I and HLA class II molecules play critical roles in our adaptive immune system by signaling a cell's health status to the immune system, through presentation of small peptides. Understanding HLA biology is important due to its prominent implication in auto-immune diseases and role in cancer immuno-therapy. Although both the HLA class I and class II antigen processing and presentation pathways have been studied extensively, the fundamental rules in HLA class II antigen presentation still remain less understood. To clarify the mechanistic and adaptive differences between the HLA class I and class II systems, we challenged a B lymphoblast cell line (JY), widely used as model system in studying antigen presentation, with a high temperature treatment, to mimic a 'fever-like state', representing one of the most common physiological responses to infection. In the absence of real invading pathogenic peptides to present, we could focus on delineating the intrinsic HLA pathway adaptations in response to high temperature in this particular cell line. Following a three-pronged approach we performed quantitative analyses of the proteome, the HLA class I ligandome, as well as the HLA class II ligandomes. The data reveal that elevated temperature may already prepare these cells for an immune-like response, through increased HLA class II presentation capacity and specific release of, from the invariant chain originating, CLIP peptides. Interestingly, at high temperature, prominent changes in the composition of the CLIP repertoire were observed, with enrichment of peptides containing C-terminal extensions beyond the CLIP-core region. Collectively, these illustrate intriguing temperature sensitive adaptations in this B cell line.

Introduction

Human leukocyte antigen (HLA) class I and class II molecules are transmembrane proteins that present small peptide ligands (8- to 12-mers, and 13- to 18-mers respectively) from either endogenous or exogenous proteins to the immune system. Thereby, they play a crucial role in our adaptive immune defense (1). These peptides also reflect the health status of the presenting cells. Usually, the presentation of a 'self' peptide can be tolerated by the immune system, whereas cells presenting a 'non-self' peptide, such as viral, bacterial or mutated peptide, will be destroyed when the immune system is triggered. This prevents the further propagation of viral and/or bacterial infections and puts a stop to cancer initiation. Since failure to eliminate such diseased cells can lead to severe consequences, it is critical to fully understand HLA class I and HLA class II peptide ligand processing and presentation.

HLA peptide processing and presentation is increasingly studied in view of its potential therapeutic role in personalized anti-cancer therapy (2-5). While similarities exist between the HLA class I and HLA class II presentation pathways, there are also key differences that distinguish these pathways. HLA class I peptide ligands are presented to CD8[+] T-cells and predominantly sourced endogenously, from peptides that arise from intracellular protein degradation or from defective ribosomal products (6, 7). Some of these peptides may then be translocated to the endoplasmic reticulum (ER) via antigen peptide transporter 1 and/or 2 in the ER membrane (6, 8). In the ER, the peptide loading on empty HLA class I molecules proceeds via a complex containing tapasin, calreticulin and ERp57 (9), but also higher affinity peptides constantly compete off lower affinity ones. Upon loading of a high affinity peptide, the glycan on the HLA class I molecule triggers release from the ER towards the Golgi apparatus and subsequently to the plasma membrane to present its peptide (10). Recently, a tapasin analogue, TAPBPR has been reported. Although these proteins are analogues, TAPBPR does not associate with the peptide loading complex, does not have to reside in the ER and is mutually exclusive with tapasin and therefore plays an alternative role in peptide loading (11). The fact that this pathway is only recently discovered highlights that, despite intense past investigations, there is still much more to discover about HLA peptide processing and presentation.

The HLA class II presentation pathway is inherently different. HLA class II peptide ligands are presented to CD4[+] T-cells and predominantly sampled from exogenous proteins and endogenous proteins produced via autophagy in the endosomal pathway (12). HLA class II complexes are assembled in the ER, where, instead of binding of a high affinity antigenic peptide, a part of CD74 (invariant chain) binds into the peptide binding groove and targets the molecule into the endosomal pathway (13, 14). HLA class II molecules come in contact with antigenic peptides in the MIIC compartment (15), where the invariant chain is cleaved sequentially by legumain (LGMN), cathepsin S, cathepsin L and cathepsin F, leaving behind a smaller peptide

fragment called CLIP. This HLA class II bound CLIP peptide can be exchanged out for a higher affinity antigenic peptide with help of the chaperones HLA-DM and HLA-DO (16, 17). After loading, the HLA class II complexes with peptide ligands are transported to the plasma membrane through vesicular transport where they are stably inserted (18-20).

While these HLA class I and class II mechanisms have been studied for a long time in normal homeostatic conditions, still very little is known about how specific cellular stress conditions could alter these pathways, similarly or distinctively. Therefore, we induced heat stress on JY B-cells by a three-day exposure to high temperature, to monitor how HLA processing and presentation are affected. Physiologically, this would be akin to a fever state (21), without actual invading pathogens. From this, we discovered that heat-induced cellular stress alone seemed to re-shape the B-cell proteome and prepares B-cells for an immune response, thereby specifically modulating the ultimate step in HLA class II presentation, and stimulating release of specific CD74 CLIP peptides.

Materials and Methods

Cell culture

The B-lymphoblastoid cell line JY (HLA-A*02:01, HLA-B*07:02, HLA-C*07:02, HLA-DPA1*01:03, HLA-DPB1*02:01/04:02, HLA-DQA1*01:03/03:01, HLA-DQB1*03:02/06:03, HLA-DRA*01, HLA-DRB1*04:04/13:01, HLA-DRB4*04, HLA-DRB5*02) was cultured in RPMI 1640 medium (+glutamine, Gibco, United States) supplemented with 10% fetal bovine serum, 50 U/ml penicillin and 50µg/ml streptomycin in a humidified incubator at 37ºC with 5% CO_2. Three days before harvest, the cells were split and grown at either 37ºC or 40ºC in a humidified incubator with 5% CO_2.

Proteomics

JY cells were lysed in 8M Urea in 50mM ammonium bicarbonate supplemented with 1x complete EDTA-free protease inhibitor cocktail (Roche Diagnostics, Switzerland), 50µg/ml DNAse I (Sigma-Aldrich, United States) and 50µg/ml RNAse A (Sigma-Aldrich). The lysate was cleared by centrifugation for 1h at 18000g at 15ºC. The protein concentration was determined with the Bradford assay (Bio-Rad, United States). For each sample, 50µg of whole cell lysate was reduced, alkylated and

digested sequentially with Lys-C (1:100) and trypsin (1:75). The digested peptides were acidified to 0.1% formic acid and purified by SepPak C18 columns (Supelco, United States). Peptide elution was performed with 80% acetonitrile in 0.1% formic acid. The samples were dried by vacuum centrifugation and reconstituted in 2% formic acid prior to LC-MS/MS analysis. Per sample, three technical replicates were measured by LC-MS/MS.

HLA class I and HLA class II ligandomics

Per condition, 5×10^8 cells were harvested by centrifugation and washed three times with phosphate-buffered saline (PBS) which had been incubated at the respective growth temperatures. The pelleted cells were disrupted in 10ml lysis buffer per gram cell pellet for 1.5h at 4ºC, on gentle end-to-end rotation. The lysis buffer consisted of Pierce IP lysis buffer (Thermo Fischer Scientific, United States) supplemented with 1x complete protease inhibitor cocktail (Roche Diagnostics), 50µg/ml DNAse I (Sigma-Aldrich) and 50µg/ml RNAse A (Sigma-Aldrich). The lysate was then cleared by centrifugation for 1h at 18000g at 4ºC. The protein concentration of the supernatant was determined with the BCA assay (Pierce, United States).

HLA class I immuno-affinity purification was performed as previously described (22). Briefly, HLA class I complexes were purified from 25 mg of lysate, using 0.5mg W6/32 antibody (23) coupled to 125µl protein A/G beads (Santa Cruz, United States). To prevent co-elution, the antibodies were cross-linked to protein A/G beads. For retrieval of HLA class II complexes, we used an HLA-DR specific antibody (B8-11-2, Bioceros/Polpharma Biologics, The Netherlands), and performed the pulldown using the HLA class I depleted lysate as input. For both immuno-affinity purifications, incubation took place at 4ºC for approximately 16h. After immuno-affinity purification, the beads were washed with 40ml of cold PBS. HLA class I and HLA class II complexes and peptide ligands were eluted with 10% acetic acid. The peptide ligands were separated from the HLA molecules using 10kDa molecular weight cutoff filters (Millipore, United States) for HLA class I and 30kDa molecular weight cutoff filters (Millipore) for HLA class II. The flowthrough containing the HLA class I or HLA class II peptide ligands was freeze-dried, reconstituted in 0.1% formic acid for further cleanup by C18 STAGE tips (Thermo Fischer Scientific), and eluted from C18 STAGE tips with 80% acetonitrile, 0.1% formic acid. The samples were dried by vacuum centrifugation and reconstituted in 2% formic acid prior to LC/MS-MS analysis. Per sample, three technical replicates were measured by LC-MS/MS. The 10kDa and 30kDa retentate containing HLA class I and HLA class II proteins were resuspended in 8M Urea, and retained for gel analyses.

Proteome LC-MS/MS analysis

The data was acquired with an UHPLC 1290 system (Agilent) coupled to a Q-Exactive HFX mass spectrometer (Thermo Fischer Scientific). The peptides were trapped (Dr Maisch Reprosil C18, 3μM, 2cm x 100μM) for 5min in solvent A (0.1% formic acid in water) before being separated on an analytical column (Agilent Poroshell, EC-C18, 2.7μM, 50cm x 75μM). Solvent B consisted of 80% acetonitrile in 0.1% formic acid. The gradient was as follows: 5min trapping, followed by 155min gradient from 10% to 36% solvent B. Subsequently, 10min of washing with 100% solvent B and 10min re-equilibration with 100% solvent A. The mass spectrometer operated in data-dependent mode. Full scan MS spectra from *m/z* 375-1600 were acquired at a resolution of 60,000 to a target value of $3x10^6$ or a maximum injection time of 20ms. MS/MS spectra were acquired at a resolution of 15,000. The top 15 most intense precursors with a charge state of 2+ to 5+ were chosen for fragmentation. HCD fragmentation was performed at 27% normalized collision energy on selected precursors with 16s dynamic exclusion at a 1.4*m/z* isolation window after accumulation to $1x10^6$ ions or a maximum injection time of 50ms.

Ligandome LC-MS/MS analysis

The data was acquired with an UHPLC 1290 system (Agilent) coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fischer Scientific). Peptides were trapped (Dr Maisch Reprosil C18, 3 μM, 2 cm × 100 μM) for 5 min in solvent A (0.1% formic acid in water) before being separated on an analytical column (Agilent Poroshell, EC-C18, 2.7 μm, 50 cm × 75 μm). Solvent B consisted of 80% acetonitrile in 0.1% formic acid. The gradient was as follows: first 5min of trapping, followed by 90min gradient from 7% to 35% solvent B. Subsequently 10min of washing with 100% solvent B and 10min re-equilibration with 100% solvent A. The mass spectrometer operated in data-dependent mode. Full scan MS spectra from *m/z* 400-650 (HLA class I) or *m/z* 300-1500 (HLA class II) were acquired at a resolution of 60,000 after accumulation to a target value of $4x10^5$ or a maximum injection time of 50ms (HLA class I) or 250ms (HLA class II). Tandem mass spectrometry (MS/MS) spectra were acquired at a resolution of 15,000. Up to three most intense precursors with a charge state of 2+ or 3+ starting at *m/z* 100 (HLA class I) or charge state 2+ to 5+ (HLA class II) were chosen for fragmentation. For peptide identification EThcD fragmentation (24) was performed at 35% normalized collision energy on selected precursors with 18s dynamic exclusion (HLA class I) or 60s dynamic exclusion (HLA class II) after accumulation of $5x10^4$ ions or a maximum injection time of 250ms (HLA class I) or 1500ms (HLA class II).

Proteome data analysis

Raw files were searched using MaxQuant version 1.6.10.0 and the Andromeda search engine against the human Uniprot database (20431 entries, downloaded in December 2019) edited with the JY-specific HLA proteins. Enzyme specificity was set to trypsin and up to 2 missed cleavages were allowed. Cysteine carbamidomethylation was set as fixed modification. Methionine oxidation and N-terminal acetylation were set as variable modifications. Precursor mass tolerance was set to 20ppm. Fragment ion tolerance was set to 4.5ppm. False discovery rate (FDR) was restricted to 1% in both protein and peptide quantification. For quantitative comparisons, label-free quantification (based on unique + razor peptides) was enabled with "match between runs". For HLA protein quantification, the label-free quantification was based on unique peptides only. Data normalization and statistics were performed with Perseus version 1.6.7.0. Gene ontology (GO) analysis was performed with Database for Annotation, Visualization and Integrated Discovery version 6.8 (DAVID (25)). The data was visualized with Graphpad Prism 8.0.

Ligandome data analysis

Raw files were searched using Sequest HT in Proteome Discoverer 2.2 against the Swiss-Prot human database (20258 entries, downloaded in February 2018) edited with JY specific HLA proteins and 20 most abundant FBS contaminants (26). The search was set to unspecific with a minimum precursor mass of 797Da to a maximum precursor mass of 1950Da (HLA class I) or a minimum precursor mass of 350Da to a maximum precursor mass of 5000Da (HLA class II) with a mass tolerance of 10ppm. Fragment ion tolerance was set to 0.02Da. The identified peptides were filtered against 1% FDR using the Percolator algorithm, 5% peptide FDR and Xcorr >1. Cysteine cysteinylation and methionine oxidation were set as variable modifications. From the identified peptides, FBS contaminants were removed. Binding affinity of HLA class I peptide ligands was predicted with NetMHCpan-4.0 (27) with a binder cutoff at rank 2. Binding affinity of HLA class II ligands was predicted with NetMHCIIpan-4.0 (28) with a binder cutoff of <1000nM. Alignments were made using the msa R package. The data was visualized with Graphpad Prism 8.0.

Experimental design and statistical rationale

For each proteome and HLA peptidome biological sample, three technical replicates were measured. These samples were injected from separate injection wells to prevent evaporation and thereby concentration of the samples. Proteome LFQ

intensities were extracted by MaxQuant and were Log2 transformed in Perseus. The proteome identifications were filtered for at least 2 valid values in at least one of the conditions. Missing values were imputed based on a normal distribution. Pairwise comparisons were performed using a student's t-test (two-sided, adjusted p-value <0.05).

Results

Composition of JY HLA class I and class II ligandomes

To investigate how the B-cell proteome and ligandome changes upon cellular stress, JY cells were grown at 37ºC and subsequently split into two and further grown for three days at either 37ºC or at 40ºC before analysis. From these paired materials, we purified HLA class I and HLA class II complexes and peptide ligands sequentially from the JY cell line, grown at 37 ºC. The JY cell line is homozygous for HLA-A*02:01, HLA-B*07:02, HLA-C*07:02, which simplifies the interpretation of HLA class I peptide presentation, and makes it a widely used benchmarking cell line in immunopeptidomics (22, 29-35). In terms of HLA class II, JY cells have the alleles HLA-DPA1*01:03, HLA-DPB1*02:01/04:02, HLA-DQA1*01:03/03:01, HLA-DQB1*03:02/06:03, HLA-DRA*01, HLA-DRB1*04:04/13:01, HLA-DRB4*04 and HLA-DRB5*02, as verified by HLA typing via next generation sequencing. By means of Coomassie staining, we verified the specific immunoaffinity capture of HLA class I and HLA class II proteins on SDS-PAGE at the respective molecular weights (Figure 1A). This gave us confidence that sequential purification of HLA class I and class II complexes from the same lysate was feasible, with a detectable yield.

From HLA class I purification, we identified >10,000 unique peptide ligands (10380 ± 24, Figure 1B), over three technical replicates. Within this, 93% of all peptides were predicted to bind to HLA-A*02:01, HLA-B*07:02, or HLA-C*07:02. The smaller number of HLA-C binders is likely due to lower protein copy number of HLA-C relative to the other class I HLAs (36, 37). In contrast to the large identification of HLA class I peptide ligands, only about 2100 HLA class II peptides (2173 ± 10, Figure 1C) were identified. This lower identification number, though still with high specificity of 81%, is expected and consistent with the lower expression level of HLA class II on JY cells. Here we only predicted the binding affinity against HLA-DR alleles, since we performed the HLA class II immunoaffinity purification with a DR-specific antibody, unlike in the HLA class I purification where a pan-HLA class I antibody (W6/32) was used. The combined peptide numbers for HLA-A/B/C and HLA-DRB1 exceed 100% as the binding motifs for the alleles can be similar. In that way,

peptides can be assigned to multiple alleles. As expected, HLA class II peptide ligands were also substantially longer with a mode of 15 residues, compared to 9 residues for HLA class I (Figure 1D).

Interestingly, HLA class II peptides sampled from the top ten source proteins made up almost 70% in intensity in each measurement, suggesting that fragments of these ten source proteins heavily dominate the HLA class II peptide ligandome, in the absence of pathogenic peptides to present. Amongst these was CD74, the invariant chain precursor protein of the CLIP peptide(s), which is known to bind and aid the folding of HLA class II molecules. Peptides originating from class I HLA proteins were also amongst the top contributors to HLA class II ligandome presentation. This is likely due to the obligatory retrograde recycling of plasma membrane HLA class I molecules during HLA class II presentation (38, 39). In a striking cross comparison, the top ten source proteins only contributed about 10% of the HLA class I peptide ligands (Figure 1E), reflecting fundamental differences in the sampling space of HLA class I and class II peptide presentation.
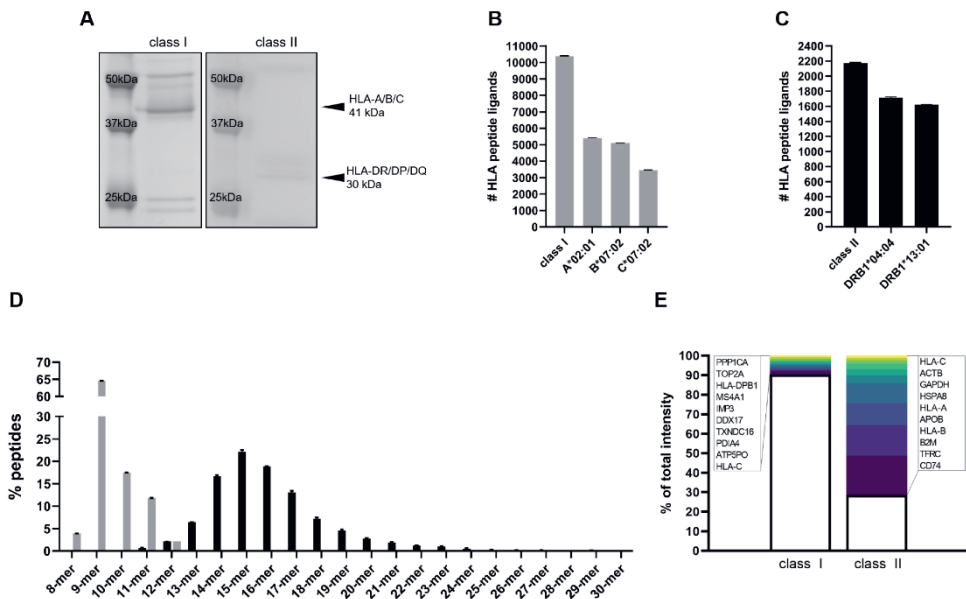
Figure 1: HLA class I and class II peptide ligand characteristics. A) Coomassie stained gel of immunoprecipitated HLA class I and HLA class II proteins. The HLA class I and HLA class II bands are visible at 41kDa and 30kDa respectively. **B)** Total number of HLA class I peptide ligands identified. First bar shows the total number, the other bars the distribution per HLA type. **C)** Total number of HLA class II peptide ligands identified and their distribution over the different HLA types. **D)** HLA class I (grey) and HLA class II (black) peptide ligand length distributions. **E)** Abundance based top 10 source proteins contributing to the HLA class I and HLA class II peptide ligandomes.

The CD74 peptide cluster is the most sampled self-protein in the HLA class II repertoire

Given the high occupancy of CD74 peptides on HLA class II molecules (~20%, Figure 1E), we examined the sequence features and intensities of all these CD74 peptides more closely. CD74 is a 296 amino acid protein that plays a key role in HLA class II peptide loading. During the early steps of HLA class II peptide loading, the CLIP region of CD74 (Figure 2A) is known to be inserted into the HLA class II binding groove to stabilize the complex and prevent unspecific binding of other peptides, until the molecule has reached the endosomal compartment where the intended peptide cargo is picked up(14, 40-42). With an open HLA class II peptide groove that can accommodate both N-terminal and C-terminal peptide protrusions(43), the HLA class II binding groove was reported to contain four anchoring positions (P1, P4, P6 and P9) where M107, A110, P112 and M115 from the CLIP core region are supposed to bind (17, 44).

In our JY HLA class II peptide ligandome, a total of 32 different peptides from the CD74 CLIP region were detected (Figure 2B). By aligning these peptides against the CD74 CLIP sequence, we noticed that only M115 (out of the four documented anchor positions) is perfectly conserved in all the CLIP peptides detected, whereas almost half of the peptides (15 out of 32) did not include M107. This affirms that what is classically described as the 'CLIP peptide' is not a single sequence, but a set of different sequences. Moreover, almost all the peptides from CD74 featured extensions outside the core region; 17 peptides were detected with N-terminal extensions, whereas 15 were C-terminally extended. By means of NetMHCIIpan predictions, we further verified that all these 32 peptides detected could bind to DRB1*04:04/13:01, and that each of these peptides can bind to at least one of the two DRB1 alleles of the JY cell line, with <1000nM affinity. Despite apparent sequence laddering, the intensity distribution of these 32 CLIP peptides was far from uniform. For instance, a LPKPPKPVSKMRMATPLLMQALPM peptide contributed about 50% in intensity, and 8 other peptides together constitute another 38% in intensity (Figure 2C, Supplementary Table I), suggesting that in addition to sequence variation, intensity, which may be used as a proxy for HLA class II groove occupancy, could also differ. Collectively this data consolidates the CLIP peptide repertoire and proportion in normal growth conditions at 37 ºC.
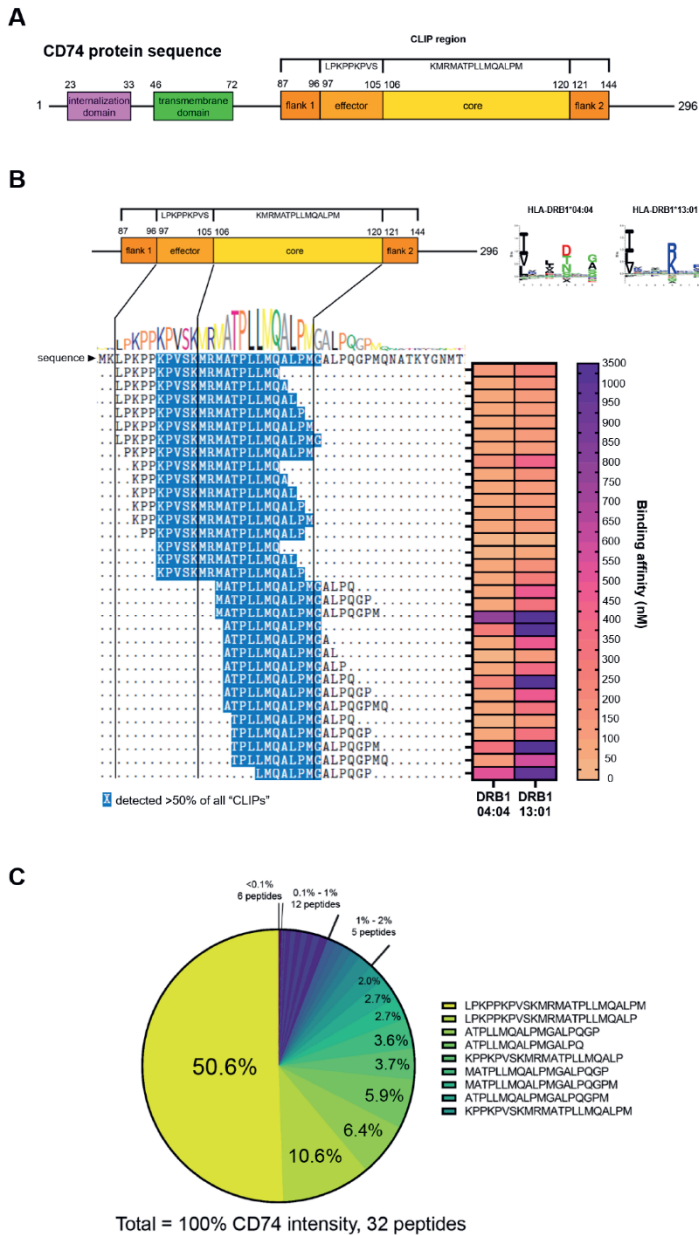
Figure 2: CLIP is a repertoire of peptides. A) Domain schematic of the CD74 protein, annotating the internalization, transmembrane and CLIP regions. **B)** Alignment of the identified CLIP peptide cluster with their corresponding binding affinity (nM) for HLA-DRB1*04:04 and HLA-DRB1*13:01.**C)** Pie chart of all identified CLIP peptides with contributing intensity as a fraction of total intensity of peptides originating from the CD74 CLIP region for cells grown at 37ºC.

High temperature prepares B cells for an HLA class II immune response

With extensive HLA class I and class II peptide ligandome characterization of the JY cell line, and resting characteristics of the CLIP peptide repertoire fully documented, we next challenged the model system with a three-day heat treatment at 40ºC, to simulate a prolonged fever response though in the absence of real invading pathogens. We then analyzed the total proteome together with the HLA class I and class II peptides, for temperature-induced changes along the antigen presentation pathway and consequences on the ligandome. Without the competition for loading and interference of foreign antigens, we aimed to better focus on the intrinsic response and adaptations on antigen presentation.

As shown in the volcano plot (Figure 3A, Supplementary Table II) summarizing changes in 3478 proteins that were quantifiable in at least 2 out of 3 technical replicates, relatively few proteins were regulated by drastic fold-changes, suggesting proteome changes were not widespread but likely quite specific in JY cells subjected to high temperature. Indeed, the proteins that did change in abundance by more than 2-fold (177 upregulated; 103 down-regulated) were significantly enriched for functional processes involving regulation of T cell proliferation and activation, and interferon-gamma mediated signaling, both of which are consistent with the biological function of B cells during an infection (45-47) (Figure 3B-C). Amongst the upregulated proteins, numerous cell surface localized cluster of differentiation (CD) proteins indicating B cell functional activation were also detected (Figure 4A-B). The presence of CD20 and CD22 has been linked to B-cell activation and differentiation (48-50), while CD86, CD48, CD70, CD47 and CD166 are markers of functional B cells that can engage in T cell co-simulation (51-55). CD59 is the major protective protein against the membrane attack complex (complement system), ensuring that only invading pathogens are lysed (56) and CD40 is required for immunoglobulin production (57). In the gene ontology enrichment performed on up-regulated proteins (Figure 3B), the most significant enrichment was found in the biological process of HLA class II antigen presentation (Figure 3B-C). This affirmed the relevance of heat treatment and our model system to study temperature-induced changes in antigen presentation.

A detailed examination of specific changes in HLA class II antigen processing and presentation pathway revealed that only the abundance of HLA class II molecules seemed to change significantly (>2-fold) on heat treatment, while no significant changes were observed elsewhere along the HLA class II antigen presentation route (Figure 5A). A mild elevation in CD74 protein level (of 1.6 fold) might also have correlated additionally with the need to bind more copies of HLA class II proteins *en route* to the endosome. Collectively, this demonstrates that the adaptations in JY cells, induced by high temperature, likely occur in the final step of HLA class II presentation, but not earlier in the capacity to invaginate and breakdown exogenous pathogens (for instance LGMN, IFI30, CTSS, Figure 5A).
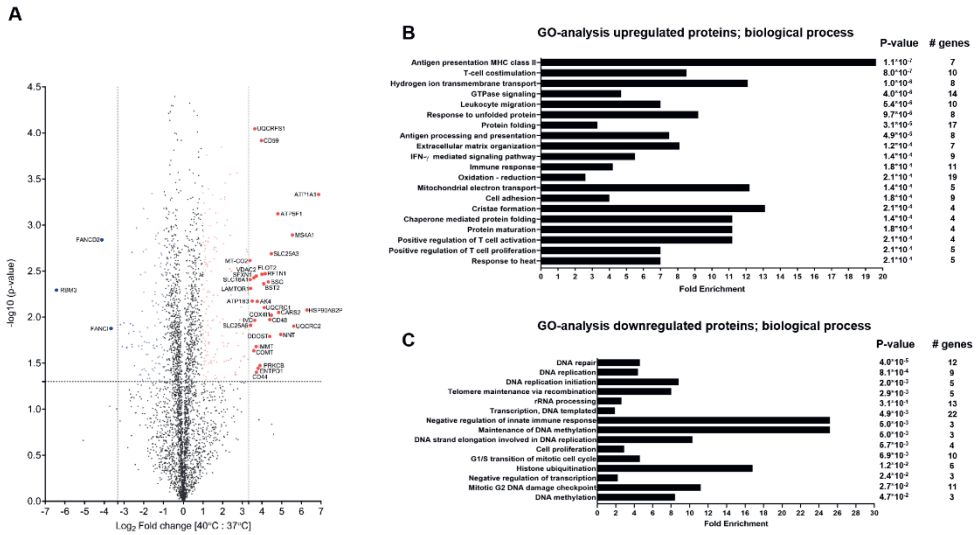
Figure 3: Proteome alterations at 40ºC. A) Volcano plot of all changing proteins (n=3478) at 40ºC *versus* 37ºC. All proteins with a >10-fold increase are indicated in red and with a >10-fold decrease in blue. **B)** Gene Ontology biological process analysis of all upregulated proteins at 40ºC using the full proteome as background. **C)** Gene Ontology biological process analysis of all downregulated proteins at 40ºC using the full proteome as background.
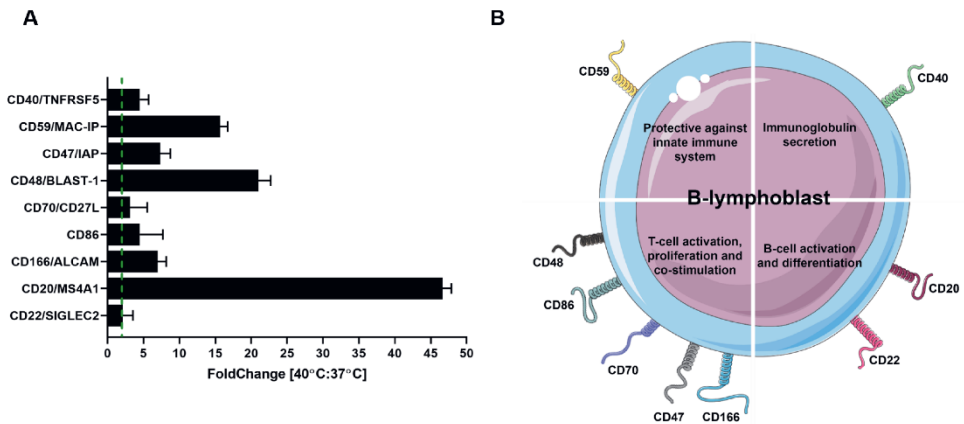
Figure 4: B-cell preparation at 40ºC. A) Protein fold change (cutoff at fold change 2) at 40ºC *versus* 37ºC of cell surface cluster of differentiation (CD) proteins. **B)** Functional categories of up-regulated CD proteins.

In contrast, the HLA class I pathway appeared to be minimally altered at the elevated temperature (Figure 5B), both at the processing and presentation levels. Although a marginal increase in ER transporter BCAP31 was detected, the ER transporter function of BCAP31 is not specific only to the export of loaded HLA class I complexes, but also involving general export of many other membrane proteins. This proteome comparison of the presentation machinery thus provides clear evidence that the regulation of HLA class I and class II pathways are controlled distinctly and independently, such that either system may be triggered separately. In addition, the specific up-regulation of HLA class II proteins and CD74 appear to be priming steps following high temperature that might have consequences such as boosting HLA class II presentation when pathogens are indeed present in the system.

To evaluate this adaptation in high temperature further, at the level of HLA peptides presented, we overlapped either class I or class II peptide ligand species identified at 37°C and 40°C (Figure 6A-B). Since the temperature adaptation did not involve increased expression of HLA class I proteins, it was logical that the total number of HLA class I peptide ligands presented did not change significantly (9589 ± 71 peptide ligands), and a large peptide species overlap of 91% was observed. HLA class II peptide ligand species on the other hand were increased by 17% in total (2547 ± 12 peptide ligands). Considering that the abundance of every HLA class II protein increased by more than 100% (Figure 5A), the increase in HLA peptide species seemed low. This implies that at 40°C, more HLA class II proteins are loaded with seemingly the same repertoire of peptide ligands. This is also evident from Figure 6C, where the top 10 source proteins contributing peptides to HLA class I and class II presentation remain largely the same as in the case of normal growth at 37°C (Figure 1F). Specifically, EIF4G and TMED9 (for class I) and FCER2 (for class II) in Figure 6C were not amongst the top 10 source proteins at 37°C (Figure 1E), but all still within the top 40 major source proteins providing peptides for HLA class I and class II loading. Collectively, these data lead us to conclude that at 40°C, more HLA class II proteins are made, but the HLA class II peptide ligandome repertoire does not change drastically with expanded sampling from more proteins.

C-terminal extended CLIP-core peptides are overrepresented at 40°C

As shown earlier, 34 different laddered CLIP peptides were detectable in the HLA class II peptidome of cells cultured at 37°C (Figure 2B), making CD74 one of the most heavily sampled self-proteins in JY cells. Intriguingly, high temperature induced shifts in sampling from CD74, both in terms of individual CLIP peptide intensity, as well as the composition of the CLIP repertoire. As shown in Figure 7A, the total occupancy of CLIP on the HLA class II molecules decreased from 20% to 10% on exposure to high temperature. The most abundant species out of the CLIP repertoire at 37°C, LPKPPKPVSKMRMATPLLMQALPM (51%), remained the most abundant
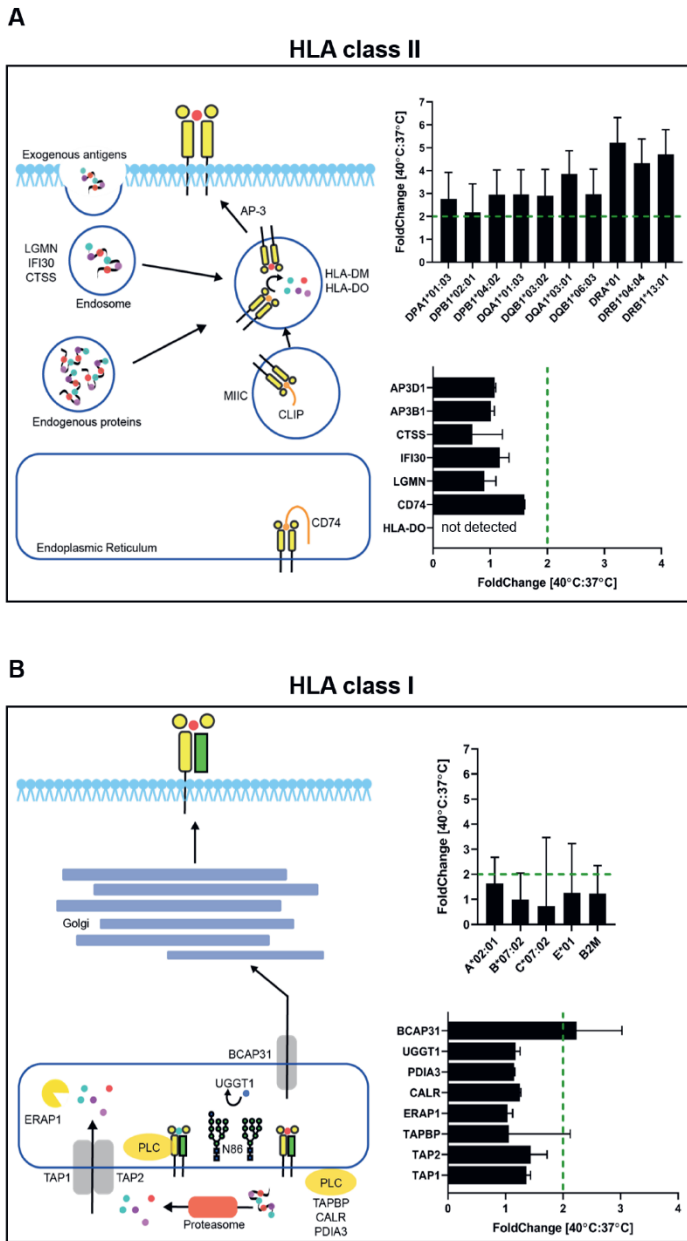
Figure 5: HLA class II and HLA class I antigen processing and presentation at 40ºC. A) Schematic of the HLA class II processing and presentation pathways. Annotated are important proteins involved in this process with the respective, substantial, fold change they undergo at elevated temperature (40ºC *versus* 37ºC). **B)** Schematic of the HLA class I processing and presentation pathways. Annotated are important proteins involved in this process with the respective, negligible, fold change they undergo at elevated temperature (40ºC *versus* 37ºC).

at 40ºC, but its total contribution to the CLIP repertoire was halved to 28% (Figure 7B). In addition, the next six most abundant CLIP variants contributed instead 60% of the intensity to the CLIP cluster. As such, the CLIP repertoire at 40 ºC was more balanced, and consisted of multiple equally abundant variants (Figure 7B) instead of the at 37 ºC dominant LPKPPKPVSKMRMATPLLMQALPM peptide (Figure 2C). Quantitatively, N-terminal and C-terminal extended CLIP peptides were also differentially presented by HLA class II proteins when cells were exposed to high temperature (Figure 7C-E). Collectively, these data outline both the qualitative and quantitative changes in the HLA class II associated CLIP peptide repertoire when JY cells are exposed to high temperature.
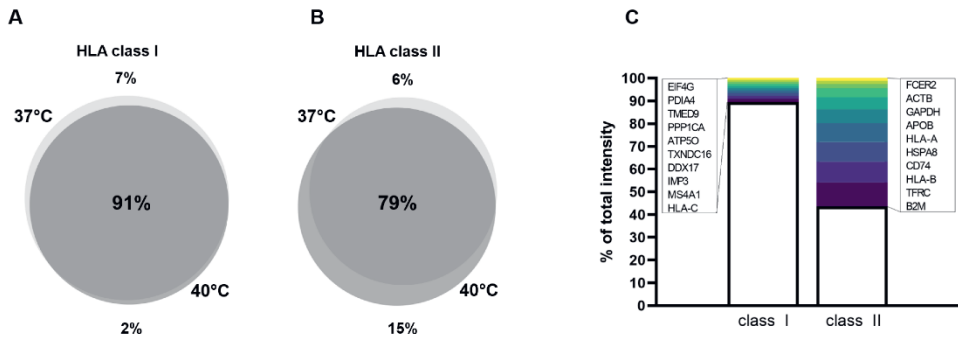
Figure 6: HLA class I and class II peptide ligand characteristics at 40ºC compared to 37 ºC. A) Overlap of identified HLA class I peptide ligands between 37ºC and 40ºC. **B)** Overlap of identified HLA class II peptide ligands between 37ºC and 40ºC. **C)** Top 10 source proteins of HLA class I and HLA class II peptide ligandomes at 40ºC.

Taken together, the up-regulation in HLA class II proteins (Figure 5A) together with decreased total CLIP occupancy seems to suggest additional priming mechanisms in preparation to present peptides from exogenous pathogens. It is intriguing that elevated temperature alone is sufficient to trigger these changes. The peculiar skew towards C-terminal extension on the other hand seemed to imply that CLIP peptide trimming may be influenced by temperature.
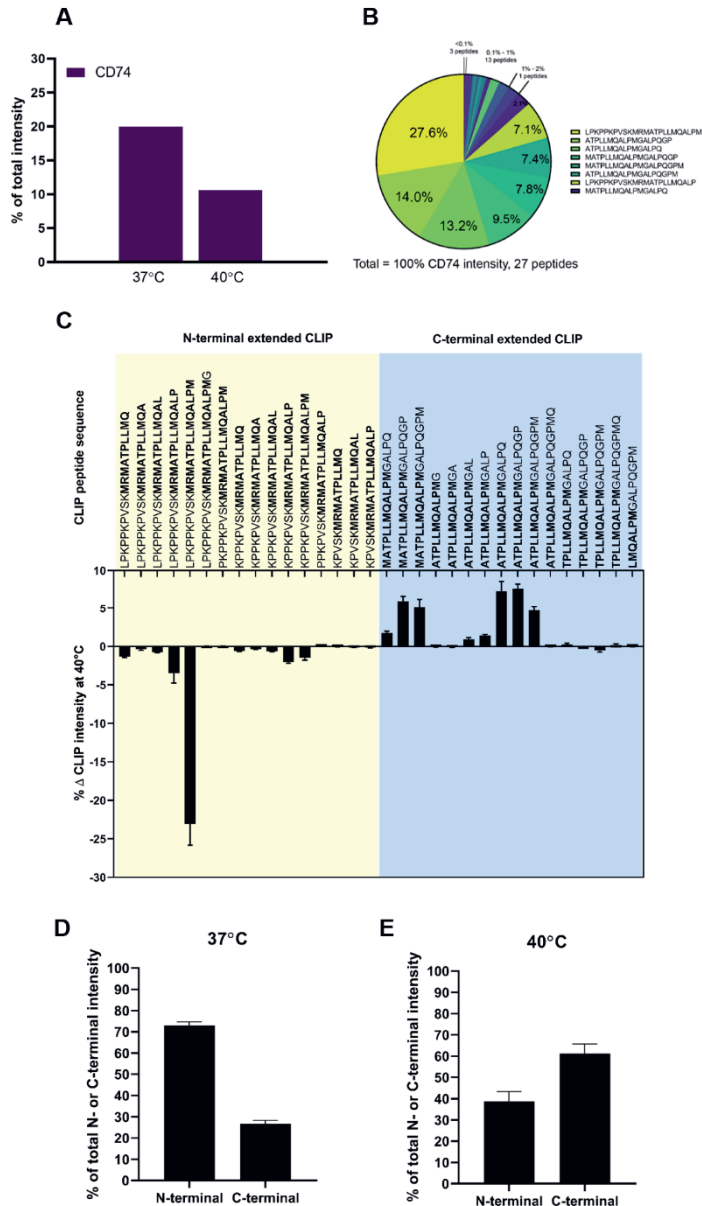
Figure 7: Temperature induced changes in the HLA bound CLIP peptide repertoire. A) Contribution of CD74 CLIP peptides to the total HLA class II ligandome at 37ºC and 40ºC. **B)** Pie chart of all identified CLIP peptides at 40ºC, with the contributing intensity as a fraction to the total intensity of all CD74 peptides. **C)** Percentage change in CLIP intensity at 40ºC versus 37ºC. **D)** Percentage of N-terminal and C-terminal CLIP intensity at 37ºC. **E)** Percentage of N-terminal and C-terminal CLIP intensity at 37ºC.

Discussion

In this study, we returned to some of the basics of in HLA peptide presentation to expand our understanding of the fundamental rules underlying this process. By analyzing the proteome and the HLA class I and class II peptide ligandomes from the same batch of JY cells, we could directly compare and contrast the antigen presentation via class I and class II mechanisms in the same biological system. This enabled over 10,000 HLA class I and more than 2100 HLA class II peptide ligands to be purified and identified from the same batch of JY material (Figure 1). While the resting HLA class I ligandome consisted of diverse peptide fragments from many more intracellular proteins, the HLA class II ligandome is dominated by peptide ligands originating from only a dozen different source proteins, contributing ~70% to the total MS intensity. In concordance with a previous report (58), we also observed that the dominant peptides loaded on HLA class II molecules were derived from mostly membrane proteins and HLA class I proteins, as well as CD74, thereby reflecting the unique biochemical processing and loading paths specific to HLA class II presentation.

In the absence of invading pathogens, peptide fragments of CD74 dominate the HLA class II ligandome of resting JY cells. While the contribution of the CLIP peptide was estimated to be ~60% in previous reports (44, 59), we found this proportion to be much more modest, at about 20% in our experiment. We believe this discrepancy arises largely from the lack of instrument sensitivity, such that other low-intensity class II ligands were not detectable, leading to a skewed quantitative estimate of CLIP peptide contribution. Indeed, with significantly increased MS detection sensitivity, we now also discover a larger repertoire of CLIP peptides, characterized by sequence laddering, particularly with both N-terminal and C-terminal extensions in addition to the CLIP-core sequence (Figure 2).This supports the notion that CLIP is a set of different sequences centered around a di-leucine motif (13, 60). Even though CLIP was classically thought to anchor in the open peptide groove of HLA class II via four residues, we found M107 amongst these four to be dispensable in half of the CLIP peptides detected from our model system. On the contrary, the di-leucine sequence is conserved in 31 out of 32 CLIP peptides, suggesting that this region may be way more critical for binding to HLA class II molecules, and that these four additional anchor positions may only play a subsidiary role to assist in the loading of the CLIP peptide repertoire. The importance of these additional anchor sites may then also vary depending on the HLA class II allele and specific peptide loading groove. Indeed, in support of our model, CLIP peptide sliding and flipped docking have been sporadically reported (61).

By employing a pathogen-free and artificially-induced high temperature treatment in culture, we mimicked a three-day fever state (21) to probe changes to the HLA class

I and class II antigen presentation systems in B cells. This experimental setup may seem simplistic, but was put together with much thought to focus on identifying host adaptations at the proteome and antigen levels. Even in the absence of real invading exogenous pathogens, we observed that our model system became adapted in three days for B-cell activation, and various B-cell specific functions (Figure 3, Figure 4), suggesting that temperature alone could trigger the cells and partly prepare them to tackle infections by means of upregulation of HLA class II proteins. In addition, the absence of real invading exogenous pathogens allowed us to observe the spontaneous loss of CLIP peptide binding to HLA class II proteins, as opposed to these peptides being competed off the HLA class II molecules by peptides originating from pathogenic proteins. Such a spontaneous loss of CLIP peptide binding would not have been discernable in a pathogen challenged system. We thus believe, conceptually, this is a very interesting observation and physiologically intriguing adaptation in the immune system that we show here to be highly specific to HLA class II presentation.

More intriguingly, we also observed that the CLIP peptide repertoire loaded on HLA class II proteins can change in composition at elevated temperatures (Figure 7B). In our opinion, this is unlikely due to global changes in the class II peptide processing pathways, since a very large proportion of the other HLA class II peptide ligands remain unchanged (Figure 6B), the HLA class II peptide processing proteins remain largely un-regulated (Figure 5A) and the top 10 source proteins also did not change substantially (Figure 6C). In fact, amongst these top 10 proteins, CD74, the precursor to the CLIP peptide repertoire is the only protein to become significantly less sampled at the higher temperature (Figure 7A). By aligning these pieces of evidence, high temperature appears to specifically increase the expression of HLA class II proteins (Figure 5A), and change the occupancy of specific CLIP peptides on HLA class II molecules. This is also consistent with the critical role of CLIP in protecting the peptide loading groove from pre-mature antigen loading before HLA class II proteins reach the endosome (13, 60). As such, the release of CLIP peptides at high temperature may indeed signal a preparatory step to present peptides of exogenous pathogens. Moreover, not only is the collective occupancy of CLIP peptides on HLA class II proteins halved at 40 ºC (from ~20% to ~10%), the distribution of CLIP variants within the CLIP repertoire was also altered, in favor of peptides with C-terminal extensions beyond the CLIP-core (Figure 7D). Specifically, the proportion of C-terminally extended CLIP peptides increased from 39% to 61% during high temperature treatment. This may further hint at temperature-induced changes in HLA class II peptide trimming, which may be interesting to explore in the future.

To further examine the coherence of our CLIP repertoire against other HLA class II ligandomes, we also analyzed the ligands from three publicly available HLA class II datasets (PXD004894(4); PXD020011(62); PXD012308(63)). PXD004894 contains HLA class II ligandomes of 21 different melanoma cell lines. PXD020011 contains

immature and mature dendritic cell ligandomes from three different donors as well as the paired CD19+ B cells and CD4+ and CD8+ T cells. PXD012308 contains HLA class II ligand data from 7 different B cells (also JY), one T-cell line and 15 different meningioma cell lines. In all cases where CLIP peptides were documented, they were derived from N-terminal extended CD74 peptides, in agreement with our observation that these CLIP species dominate in normal temperature conditions.

Although still marginally when compared to work on HLA class I ligandomes, in recent years also quite a few mass spectrometry-based HLA class II peptide profiling studies have been reported (64-69). This growing interest may be linked in view of the strong disease correlation. Nonetheless, the fundamental rules in HLA class II antigen presentation still remain much less clear than for HLA class I. The fever state is one of the most common physiological responses that accompany infections, which may be accompanied by changes in preference for the CLIP repertoire, as we show here. However, the current study is limited to a single commonly-used B cell line grown in vitro. In addition, HLA-DO and HLA-DM proteins were also below detection in the JY cell line, which precluded further investigations into the mechanism and efficiency of CLIP peptide exchange in this cell line system. Future work should establish if these temperature adaptations are conserved in more cell lines, and look into if these HLA-II proteins loaded with C-terminal CLIP are still inserted in the cell surface, although the approach we take here requires a significant input of cell material, which may not be compatible with tissue-level profiling. More fundamentally, it would be interesting to identify the mechanism of shift in CLIP presentation, as such information may enable a specific boost to the presentation of foreign antigens to intensify the trigger of the host immune system. This may lead to quicker natural clearance of such disease-causing pathogens, without too much interference on the normal HLA class I and class II peptide ligandome.

Acknowledgements

Data availability

The mass spectrometry proteomics and peptidomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (70) partner repository with the data set identifier PXD022930 or DOI: 10.6019/PXD022930. The raw file to sample mapping is described in Supplementary Table III.

Supporting information

The Supporting Information is available free of charge on the Elsevier Inc (on behalf of American Society for Biochemistry and Molecular Biology) website at DOI: https://doi.org/10.1016/j.mcpro.2021.100089

References

1. Chaplin, D. D., Overview of the immune response. J Allergy Clin Immunol 2010, 125, (2 Suppl 2), S3-23.

2. Rammensee, H. G.; Singh-Jasuja, H., HLA ligandome tumor antigen discovery for personalized vaccine approach. Expert Rev Vaccines 2013, 12, (10), 1211-7.

3. Newey, A.; Griffiths, B.; Michaux, J.; Pak, H. S.; Stevenson, B. J.; Woolston, A.; Semiannikova, M.; Spain, G.; Barber, L. J.; Matthews, N.; Rao, S.; Watkins, D.; Chau, I.; Coukos, G.; Racle, J.; Gfeller, D.; Starling, N.; Cunningham, D.; Bassani-Sternberg, M.; Gerlinger, M., Immunopeptidomics of colorectal cancer organoids reveals a sparse HLA class I neoantigen landscape and no increase in neoantigens with interferon or MEK-inhibitor treatment. J Immunother Cancer 2019, 7, (1), 309.

4. Bassani-Sternberg, M.; Braunlein, E.; Klar, R.; Engleitner, T.; Sinitcyn, P.; Audehm, S.; Straub, M.; Weber, J.; Slotta-Huspenina, J.; Specht, K.; Martignoni, M. E.; Werner, A.; Hein, R.; D, H. B.; Peschel, C.; Rad, R.; Cox, J.; Mann, M.; Krackhardt, A. M., Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. Nat Commun 2016, 7, 13404.

5. Demmers, L. C.; Kretzschmar, K.; Van Hoeck, A.; Bar-Epraim, Y. E.; van den Toorn, H. W. P.; Koomen, M.; van Son, G.; van Gorp, J.; Pronk, A.; Smakman, N.; Cuppen, E.; Clevers, H.; Heck, A. J. R.; Wu, W., Single-cell

derived tumor organoids display diversity in HLA class I peptide presentation. Nat Commun 2020, 11, (1), 5338.

6. Michalek, M. T.; Grant, E. P.; Gramm, C.; Goldberg, A. L.; Rock, K. L., A role for the ubiquitin-dependent proteolytic pathway in MHC class I-restricted antigen presentation. Nature 1993, 363, (6429), 552-4.

7. Bourdetsky, D.; Schmelzer, C. E.; Admon, A., The nature and extent of contributions by defective ribosome products to the HLA peptidome. Proc Natl Acad Sci U S A 2014, 111, (16), E1591-9.

8. Reits, E.; Griekspoor, A.; Neijssen, J.; Groothuis, T.; Jalink, K.; van Veelen, P.; Janssen, H.; Calafat, J.; Drijfhout, J. W.; Neefjes, J., Peptide diffusion, protection, and degradation in nuclear and cytoplasmic compartments before antigen presentation by MHC class I. Immunity 2003, 18, (1), 97-108.

9. Cresswell, P.; Bangia, N.; Dick, T.; Diedrich, G., The nature of the MHC class I peptide loading complex. Immunol Rev 1999, 172, 21-8.

10. Moremen, K. W.; Tiemeyer, M.; Nairn, A. V., Vertebrate protein glycosylation: diversity, synthesis and function. Nat Rev Mol Cell Biol 2012, 13, (7), 448-62.

11. Hermann, C.; Trowsdale, J.; Boyle, L. H., TAPBPR: a new player in the MHC class I presentation pathway. Tissue Antigens 2015, 85, (3), 155-66.

12. Suri, A.; Lovitch, S. B.; Unanue, E. R., The wide diversity and complexity of peptides bound to class II MHC molecules. Curr Opin Immunol 2006, 18, (1), 70-7.

13. Neefjes, J.; Jongsma, M. L.; Paul, P.; Bakke, O., Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol 2011, 11, (12), 823-36.

14. Cresswell, P., Invariant chain structure and MHC class II function. Cell 1996, 84, (4), 505-7.

15. Neefjes, J., CIIV, MIIC and other compartments for MHC class II loading. Eur J Immunol 1999, 29, (5), 1421-5.

16. Unanue, E. R.; Turk, V.; Neefjes, J., Variations in MHC Class II Antigen Processing and Presentation in Health and Disease. Annu Rev Immunol 2016, 34, 265-97.

17. Ghosh, P.; Amaya, M.; Mellins, E.; Wiley, D. C., The structure of an intermediate in class II MHC maturation: CLIP bound to HLA-DR3. Nature 1995, 378, (6556), 457-62.

18. Wubbolts, R.; Fernandez-Borja, M.; Oomen, L.; Verwoerd, D.; Janssen, H.; Calafat, J.; Tulp, A.; Dusseljee, S.; Neefjes, J., Direct vesicular transport of MHC class II molecules from lysosomal structures to the cell surface. J Cell Biol 1996, 135, (3), 611-22.

19. Boes, M.; Cerny, J.; Massol, R.; Op den Brouw, M.; Kirchhausen, T.; Chen, J.; Ploegh, H. L., T-cell engagement of dendritic cells rapidly rearranges MHC class II transport. Nature 2002, 418, (6901), 983-8.

20. Kleijmeer, M.; Ramm, G.; Schuurhuis, D.; Griffith, J.; Rescigno, M.; Ricciardi-Castagnoli, P.; Rudensky, A. Y.; Ossendorp, F.; Melief, C. J.; Stoorvogel, W.; Geuze, H. J., Reorganization of multivesicular bodies regulates MHC class II antigen presentation by dendritic cells. J Cell Biol 2001, 155, (1), 53-63.

21. Pritchard, M. T.; Ostberg, J. R.; Evans, S. S.; Burd, R.; Kraybill, W.; Bull, J. M.; Repasky, E. A., Protocols for simulating the thermal component of fever: preclinical and clinical experience. Methods 2004, 32, (1), 54-62.

22. Demmers, L. C.; Heck, A. J. R.; Wu, W., Pre-fractionation Extends but also Creates a Bias in the Detectable HLA Class Iota Ligandome. J Proteome Res 2019, 18, (4), 1634-1643.

23. Barnstable, C. J.; Bodmer, W. F.; Brown, G.; Galfre, G.; Milstein, C.; Williams, A. F.; 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.

24. Mommen, G. P.; Frese, C. K.; Meiring, H. D.; van Gaans-van den Brink, J.; de Jong, A. P.; van Els, C. A.; Heck, A. J., Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc Natl Acad Sci U S A 2014, 111, (12), 4507-12.

25. Huang da, W.; Sherman, B. T.; Lempicki, R. A., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009, 4, (1), 44-57.

26. Zheng, X.; Baker, H.; Hancock, W. S.; Fawaz, F.; McCaman, M.; Pungor, E., Jr., Proteomic analysis for the assessment of different lots of fetal bovine serum as a raw material for cell culture. Part IV. Application of proteomics to the manufacture of biological drugs. Biotechnol Prog 2006, 22, (5), 1294-300.

27. 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.

28. Reynisson, B.; Barra, C.; Kaabinejadian, S.; Hildebrand, W. H.; 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 2020, 19, (6), 2304-2315.

29. Zarling, A. L.; Ficarro, S. B.; White, F. M.; Shabanowitz, J.; Hunt, D. F.; Engelhard, V. H., Phosphorylated peptides are naturally processed and presented by major histocompatibility complex class I molecules in vivo. J Exp Med 2000, 192, (12), 1755-62.

30. Cobbold, M.; De La Pena, H.; Norris, A.; Polefrone, J. M.; Qian, J.; English, A. M.; Cummings, K. L.; Penny, S.; Turner, J. E.; Cottine, J.; Abelin, J. G.; Malaker, S. A.; Zarling, A. L.; Huang, H. W.; Goodyear, O.; Freeman, S. D.; Shabanowitz, J.; Pratt, G.; Craddock, C.; Williams, M. E.; Hunt, D. F.; Engelhard, V. H., MHC class I-associated phosphopeptides are the targets of memory-like immunity in leukemia. Sci Transl Med 2013, 5, (203), 203ra125.

31. Hassan, C.; Kester, M. G.; de Ru, A. H.; Hombrink, P.; Drijfhout, J. W.; Nijveen, H.; Leunissen, J. A.; Heemskerk, M. H.; Falkenburg, J. H.; van Veelen, P. A., The human leukocyte antigen-presented ligandome of B lymphocytes. Mol Cell Proteomics 2013, 12, (7), 1829-43.

32. Tafuro, S.; Meier, U. C.; Dunbar, P. R.; Jones, E. Y.; Layton, G. T.; Hunter, M. G.; Bell, J. I.; McMichael, A. J., Reconstitution of antigen presentation in HLA class I-negative cancer cells with peptide-beta2m fusion molecules. Eur J Immunol 2001, 31, (2), 440-9.

33. Bassani-Sternberg, M.; Pletscher-Frankild, S.; Jensen, L. J.; Mann, M., Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. Mol Cell Proteomics 2015, 14, (3), 658-73.

34. Chong, C.; Marino, F.; Pak, H.; Racle, J.; Daniel, R. T.; Muller, M.; Gfeller, D.; Coukos, G.; Bassani-Sternberg, M., High-throughput and Sensitive Immunopeptidomics Platform Reveals Profound Interferongamma-Mediated Remodeling of the Human Leukocyte Antigen (HLA) Ligandome. Mol Cell Proteomics 2018, 17, (3), 533-548.

35. Petersen, J.; Wurzbacher, S. J.; Williamson, N. A.; Ramarathinam, S. H.; Reid, H. H.; Nair, A. K.; Zhao, A. Y.; Nastovska, R.; Rudge, G.; Rossjohn, J.; Purcell, A. W., Phosphorylated self-peptides alter human leukocyte

antigen class I-restricted antigen presentation and generate tumor-specific epitopes. Proc Natl Acad Sci U S A 2009, 106, (8), 2776-81.

36.   Apps, R.; Meng, Z.; Del Prete, G. Q.; Lifson, J. D.; Zhou, M.; Carrington, M., Relative expression levels of the HLA class-I proteins in normal and HIV-infected cells. J Immunol 2015, 194, (8), 3594-600.

37.   Kaur, G.; Gras, S.; Mobbs, J. I.; Vivian, J. P.; Cortes, A.; Barber, T.; Kuttikkatte, S. B.; Jensen, L. T.; Attfield, K. E.; Dendrou, C. A.; Carrington, M.; McVean, G.; Purcell, A. W.; Rossjohn, J.; Fugger, L., Structural and regulatory diversity shape HLA-C protein expression levels. Nat Commun 2017, 8, 15924.

38.   Gromme, M.; Uytdehaag, F. G.; Janssen, H.; Calafat, J.; van Binnendijk, R. S.; Kenter, M. J.; Tulp, A.; Verwoerd, D.; Neefjes, J., Recycling MHC class I molecules and endosomal peptide loading. Proc Natl Acad Sci U S A 1999, 96, (18), 10326-31.

39.   Montealegre, S.; van Endert, P. M., Endocytic Recycling of MHC Class I Molecules in Non-professional Antigen Presenting and Dendritic Cells. Front Immunol 2018, 9, 3098.

40.   Bertolino, P.; Rabourdin-Combe, C., The MHC class II-associated invariant chain: a molecule with multiple roles in MHC class II biosynthesis and antigen presentation to CD4+ T cells. Crit Rev Immunol 1996, 16, (4), 359-79.

41.   Blum, J. S.; Wearsch, P. A.; Cresswell, P., Pathways of antigen processing. Annu Rev Immunol 2013, 31, 443-73.

42.   Rock, K. L.; Reits, E.; Neefjes, J., Present Yourself! By MHC Class I and MHC Class II Molecules. Trends Immunol 2016, 37, (11), 724-737.

43.   Stern, L. J.; Brown, J. H.; Jardetzky, T. S.; Gorga, J. C.; Urban, R. G.; Strominger, J. L.; Wiley, D. C., Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature 1994, 368, (6468), 215-21.

44.   Sette, A.; Southwood, S.; Miller, J.; Appella, E., Binding of major histocompatibility complex class II to the invariant chain-derived peptide, CLIP, is regulated by allelic polymorphism in class II. J Exp Med 1995, 181, (2), 677-83.

45.   Leon, B.; Ballesteros-Tato, A.; Misra, R. S.; Wojciechowski, W.; Lund, F. E., Unraveling effector functions of B cells during infection: the hidden world

beyond antibody production. Infect Disord Drug Targets 2012, 12, (3), 213-21.

46. Mizoguchi, A.; Bhan, A. K., A case for regulatory B cells. J Immunol 2006, 176, (2), 705-10.

47. Mauri, C.; Bosma, A., Immune regulatory function of B cells. Annu Rev Immunol 2012, 30, 221-41.

48. Tedder, T. F.; Boyd, A. W.; Freedman, A. S.; Nadler, L. M.; Schlossman, S. F., The B cell surface molecule B1 is functionally linked with B cell activation and differentiation. J Immunol 1985, 135, (2), 973-9.

49. Li, H.; Ayer, L. M.; Lytton, J.; Deans, J. P., Store-operated cation entry mediated by CD20 in membrane rafts. J Biol Chem 2003, 278, (43), 42427-34.

50. Clark, E. A.; Giltiay, N. V., CD22: A Regulator of Innate and Adaptive B Cell Responses and Autoimmunity. Front Immunol 2018, 9, 2235.

51. Lanier, L. L.; O'Fallon, S.; Somoza, C.; Phillips, J. H.; Linsley, P. S.; Okumura, K.; Ito, D.; Azuma, M., CD80 (B7) and CD86 (B70) provide similar costimulatory signals for T cell proliferation, cytokine production, and generation of CTL. J Immunol 1995, 154, (1), 97-105.

52. Elishmereni, M.; Levi-Schaffer, F., CD48: A co-stimulatory receptor of immunity. Int J Biochem Cell Biol 2011, 43, (1), 25-8.

53. Arens, R.; Nolte, M. A.; Tesselaar, K.; Heemskerk, B.; Reedquist, K. A.; van Lier, R. A.; van Oers, M. H., Signaling through CD70 regulates B cell activation and IgG production. J Immunol 2004, 173, (6), 3901-8.

54. Oldenborg, P. A., CD47: A Cell Surface Glycoprotein Which Regulates Multiple Functions of Hematopoietic Cells in Health and Disease. ISRN Hematol 2013, 2013, 614619.

55. Hassan, N. J.; Barclay, A. N.; Brown, M. H., Frontline: Optimal T cell activation requires the engagement of CD6 and CD166. Eur J Immunol 2004, 34, (4), 930-40.

56. Farkas, I.; Baranyi, L.; Ishikawa, Y.; Okada, N.; Bohata, C.; Budai, D.; Fukuda, A.; Imai, M.; Okada, H., CD59 blocks not only the insertion of C9 into MAC but inhibits ion channel formation by homologous C5b-8 as well as C5b-9. J Physiol 2002, 539, (Pt 2), 537-45.

57.   Elgueta, R.; Benson, M. J.; de Vries, V. C.; Wasiuk, A.; Guo, Y.; Noelle, R. J., Molecular mechanism and function of CD40/CD40L engagement in the immune system. Immunol Rev 2009, 229, (1), 152-72.

58.   Chicz, R. M.; Urban, R. G.; Gorga, J. C.; Vignali, D. A.; Lane, W. S.; Strominger, J. L., Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. J Exp Med 1993, 178, (1), 27-47.

59.   Riberdy, J. M.; Newcomb, J. R.; Surman, M. J.; Barbosa, J. A.; Cresswell, P., HLA-DR molecules from an antigen-processing mutant cell line are associated with invariant chain peptides. Nature 1992, 360, (6403), 474-7.

60.   Cresswell, P.; Roche, P. A., Invariant chain-MHC class II complexes: always odd and never invariant. Immunol Cell Biol 2014, 92, (6), 471-2.

61.   Gunther, S.; Schlundt, A.; Sticht, J.; Roske, Y.; Heinemann, U.; Wiesmuller, K. H.; Jung, G.; Falk, K.; Rotzschke, O.; Freund, C., Bidirectional binding of invariant chain peptides to an MHC class II molecule. Proc Natl Acad Sci U S A 2010, 107, (51), 22219-24.

62.   Marino, F.; Semilietof, A.; Michaux, J.; Pak, H. S.; Coukos, G.; Muller, M.; Bassani-Sternberg, M., Biogenesis of HLA Ligand Presentation in Immune Cells Upon Activation Reveals Changes in Peptide Length Preference. Front Immunol 2020, 11, 1981.

63.   Racle, J.; Michaux, J.; Rockinger, G. A.; 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.

64.   Wendorff, M.; Garcia Alvarez, H. M.; Osterbye, T.; ElAbd, H.; Rosati, E.; Degenhardt, F.; Buus, S.; Franke, A.; Nielsen, M., Unbiased Characterization of Peptide-HLA Class II Interactions Based on Large-Scale Peptide Microarrays; Assessment of the Impact on HLA Class II Ligand and Epitope Prediction. Front Immunol 2020, 11, 1705.

65.   Chen, B.; Khodadoust, M. S.; Olsson, N.; Wagar, L. E.; Fast, E.; Liu, C. L.; Muftuoglu, Y.; Sworder, B. J.; Diehn, M.; Levy, R.; Davis, M. M.; Elias, J. E.; Altman, R. B.; Alizadeh, A. A., Predicting HLA class II antigen presentation through integrated deep learning. Nat Biotechnol 2019, 37, (11), 1332-1343.

66.   Barra, C.; Alvarez, B.; Paul, S.; Sette, A.; Peters, B.; Andreatta, M.; Buus, S.; Nielsen, M., Footprints of antigen processing boost MHC class II natural ligand predictions. Genome Med 2018, 10, (1), 84.

67.   Abelin, J. G.; Harjanto, D.; Malloy, M.; Suri, P.; Colson, T.; Goulding, S. P.; Creech, A. L.; Serrano, L. R.; Nasir, G.; Nasrullah, Y.; McGann, C. D.; Velez, D.; Ting, Y. S.; Poran, A.; Rothenberg, D. A.; Chhangawala, S.; Rubinsteyn, A.; Hammerbacher, J.; Gaynor, R. B.; Fritsch, E. F.; Greshock, J.; Oslund, R. C.; Barthelme, D.; Addona, T. A.; Arieta, C. M.; Rooney, M. S., Defining HLA-II Ligand Processing and Binding Rules with Mass Spectrometry Enhances Cancer Epitope Prediction. Immunity 2019, 51, (4), 766-779 e17.

68.   Kampstra, A. S. B.; van Heemst, J.; Janssen, G. M.; de Ru, A. H.; van Lummel, M.; van Veelen, P. A.; Toes, R. E. M., Ligandomes obtained from different HLA-class II-molecules are homologous for N- and C-terminal residues outside the peptide-binding cleft. Immunogenetics 2019, 71, (8-9), 519-530.

69.   Mommen, G. P.; Marino, F.; Meiring, H. D.; Poelen, M. C.; van Gaans-van den Brink, J. A.; Mohammed, S.; Heck, A. J.; van Els, C. A., Sampling From the Proteome to the Human Leukocyte Antigen-DR (HLA-DR) Ligandome Proceeds Via High Specificity. Mol Cell Proteomics 2016, 15, (4), 1412-23.

70.   Vizcaino, J. A.; Csordas, A.; del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; Xu, Q. W.; Wang, R.; Hermjakob, H., 2016 update of the PRIDE database and its related tools. Nucleic Acids Res 2016, 44, (D1), D447-56.

Supplementary Figure 1
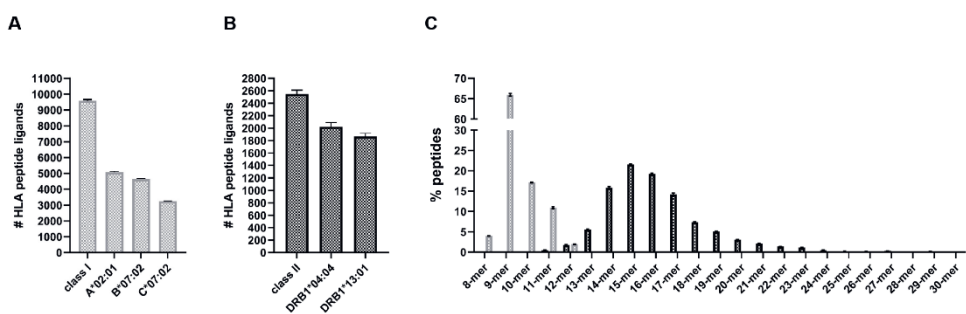
Figure S1: HLA class I and class II peptide ligand characteristics at 40ºC. A) Total number of HLA class I peptide ligands identified and their distribution over the different HLA class I types. **B)** Total number of HLA class II peptide ligands identified and their distribution over the different HLA class II types. **C)** HLA class I (grey) and HLA class II (black) peptide ligand length distributions.

IS IT BEHIND THE GLYCAN?

NO, IT IS THE GLYCAN!
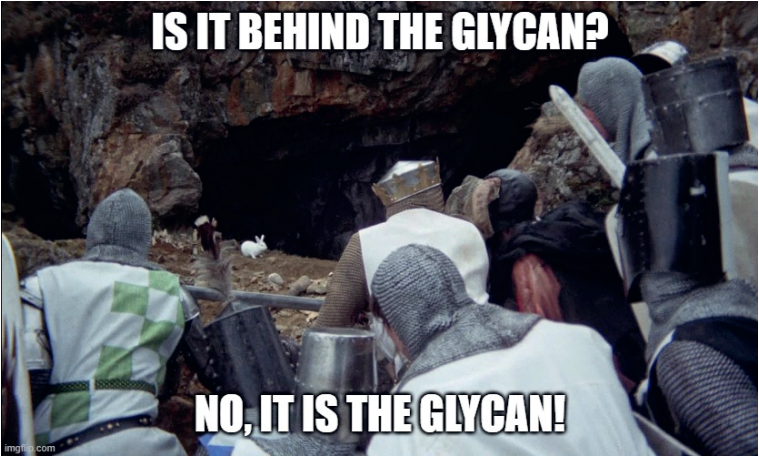
Chapter 4

Allotype-specific glycosylation and cellular localization of human leukocyte antigen class I proteins

Max Hoek*[1, 2], Laura C. Demmers*[1, 2], Wei Wu[1, 2] and Albert J. R. Heck[1, 2]

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
[2] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

*these authors contributed equally

Abstract

Presentation of antigens by human leukocyte antigen (HLA) complexes at the cell surface is a key process in the immune response. The α-chain, containing the peptide binding groove, is one of the most polymorphic proteins in the proteome. All HLA class I α-chains carry a conserved N-glycosylation site, but little is known about its nature and function. Here, we report an in-depth characterization of N-glycosylation features of HLA class I molecules. We observe that different HLA-A α-chains carry similar glycosylation, distinctly different from the HLA-B, HLA-C and HLA-F α-chains. Whereas HLA-A displays the broadest variety of glycan characteristics, HLA-B α-chains carry mostly mature glycans, HLA-C and HLA-F α-chains carry predominantly high-mannose glycans. We expected these glycosylation features to be directly linked to cellular localization of the HLA complexes. Indeed, analyzing HLA class I complexes from crude plasma and inner membrane enriched fractions confirmed that most HLA-B complexes can be found at the plasma membrane, while most HLA-C and HLA-F molecules reside in the ER and Golgi membrane and HLA-A molecules are more equally distributed over these cellular compartments. This allotype-specific cellular distribution of HLA molecules should be taken into account when analyzing peptide antigen presentation by immunopeptidomics.

Introduction

Major histocompatibility complex (MHC) class I molecules, in humans termed human leukocyte antigen (HLA) complexes, play a key role in our immune system, as they provide a means to present foreign peptide antigens at the cellular surface, providing a signal to our T-cells to eliminate them. Following their initial discovery as main factor in defining successful organ transplants, HLA complexes and peptide presentation processes have been studied extensively over the last decades (1-3). As a result of this large body of research, we have quite a coherent picture of how HLA class I molecules function, and largely understand their critical role in transplantation (4, 5), autoimmunity (6-9), bacterial and viral infections (10-13), and more recently tumor immunotherapy (14, 15).

The cellular processes underlying HLA class I antigen presentation are well understood (16). Briefly, the endogenously synthesized class I peptide ligands of antigens are generated primarily by cytosolic and nuclear proteasomes. These peptides are translocated to the endoplasmic reticulum (ER) where they are loaded into the peptide binding groove of HLA class I molecules by means of the peptide-loading complex (PLC). After an antigen has been loaded, the stable and functional HLA class I complex comprises a heavy α-chain, a β2-microglobulin (β2m) chain and the loaded peptide. These complexes traverse out of the ER, into the Golgi, *en route* to the cell-surface. Upon incorporation into the plasma membrane, the peptide can be presented to CD8[+] T cells, thereby allowing the T cells to identify and eliminate pathogen-infected cells or cancer cells (17, 18).

Beyond this generic description, HLA class I complexes are actually extremely diverse. This diversity starts at the DNA level, as in the human genome 6 genes encode for 6 different HLA class I heavy chains; the more abundantly expressed classical genes A, B, and C and the non-classical genes E, F, and G. The classical HLA-A, B and C genes are highly polymorphic, where several thousands of alleles have been identified across the population (19). The most frequent polymorphisms are found in the sequence regions that encode the peptide-binding groove, highlighted by green boxes (20, 21) (see Figure 1A). These polymorphisms allow for a wide range of specificities for peptide binding, which consequently allows a diverse set of peptides to be presented at the cell-surface. Figure 1A summarizes sequence polymorphisms that are observed in some of the most common HLA A, B and C allotypes (22). The sequence homology of these proteins is also visualized through the phylogenetic tree (depicted in Figure 1B), from which it is clear that allotypes belonging to separate classical genes do cluster more together. A much more extensive phylogenetic tree analysis of HLA allotypes is provided in Figure S1.

Notwithstanding these widespread and frequently occurring polymorphisms, a particular region in the sequence, conserved over all reported alleles (19) is highlighted in purple in Figure 1A. This highlighted NxS/T motif is a well-known motif

for N-linked glycosylation. It has been established that the asparagine 86 site is indeed glycosylated (23), however the nature of the glycans attached and their functional role have not been extensively studied. In Figure 1C a structural model of the HLA class I complex is depicted, with attached to it, modeled to scale, a commonly observed mature glycan structure. It is apparent from this structural model that the N-glycan is localized in the proximity of the peptide binding groove, and likely when the molecule is at the plasma membrane the glycan can be exposed to the outside of the cell and interact with external factors, including of T cell origin.

Cellular protein N-glycosylation is a complex multi-step process in which many enzymes in the ER and Golgi are involved. Therefore, HLA complexes will undergo their glycosylation while traversing through the ER and Golgi compartments. The rate of this traversal has been studied by pulse-chase assays which use endoglycosidase H (Endo H), an enzyme that digests specifically high-mannose and hybrid/asymmetric glycans (on HLA) to assess the rate through the ER and early Golgi (24-27). To describe the N-glycosylation pathway in short, during this process a Glc3Man9GlcNAc2 glycan is first transferred to the target protein, during or shortly after protein synthesis and translocation into the ER (28). Subsequently, this structure is trimmed, losing two Glc residues to form Glc1Man9GlcNAc2. Next, if the protein is folded correctly the last Glc residue is removed, and the protein can continue to the Golgi. In the ER or early cis-Golgi, additional Man residues can be removed. This leaves only the Man3GlcNAc2 core that is found in nearly all N-linked glycans. Subsequently in the medial- and trans-Golgi, processing occurs to form complex type glycans with antennae that can include several GlcNAc, Fuc, Gal and SA residues (29-31). It is expected that a protein is transported to the cell-surface only after full glycan maturation has occurred.

Here, we monitored in detail the HLA gene specific glycosylation patterns of HLA class I molecules of three different cell lines, all homozygous for the classical HLA genes A, B and C. The cell lines and the allotypes they harbor are shown in Figure 1C. Using a glycopeptide-centric proteomics approach for each distinct HLA heavy-chain, unique glycopeptides could be identified harboring distinct glycoforms. Quantifying these different glycoforms revealed that each of the HLA allotypes was modified with a distinct glycoform profile, which we expected to be directly linked to their differential cellular distribution. When further dissected with subcellular fractionation, we could indeed observe distinctive glycosylations that correlate with differential plasma membrane presentation. In the studied JY cells for instance, our data indicate that HLA-B complexes reside nearly exclusively at the plasma membrane, whereas a large part of the HLA-A complexes are also observed in inner membrane compartments. This correlates with our findings that HLA-B has exclusively mature bisected and extended glycans, while HLA-A has a considerable fraction of immature high-mannose glycans. As far as we know, this is the first time that HLA class I glycoform profiling is taken as an indicator to assess HLA cellular

localization. Although the correlation between glycosylation and localization was expected, our approach allows us to directly sample subcellular localization of each HLA allotype considering glycosylation as localization marker, and shows that each HLA allotype can have a unique subcellular distribution despite their very high molecular and structural similarities.
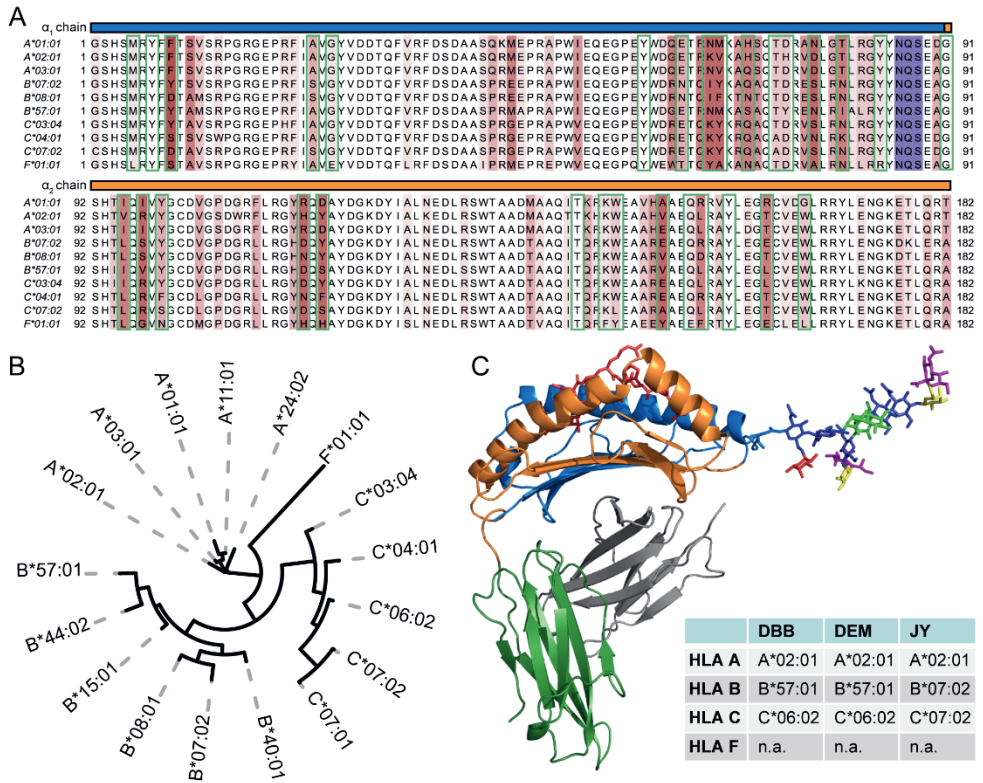
Figure 1: Sequence and Structural Characteristics of HLA class I molecules. A) Multiple sequence alignment of the mature form of several common classical HLA A, B and C genes and HLA F*01:01. The α1 and α2 domains of the sequences are shown starting at position 1 of the mature sequence and ending at position 182. Above the amino acid sequence, the α1 and α2 domains are highlighted in blue and orange. Positions that display amino acid variation between allotypes are highlighted in red with darker tints indicating a higher degree of sequence variation. The conserved N-linked glycosylation site motif is highlighted in purple. Amino acids involved in forming the peptide binding groove are boxed in green **B)** Phylogenetic tree visualization of the multiple sequence alignment depicted in 1A. Although all HLA class I genes display high homology, the allotypes still cluster per distinct classical gene type. See Figure S1 for a much more extensive phylogenetic analysis of HLA class I heavy chain allotypes. **C)** Structural model of the HLA-A*02:01 molecule (PDB: 1I4F). Chains within the HLA molecule are colored: α1-Blue, α2-Orange, α3-Green, β2m-Grey and Binding Peptide-Red. Attached to Asn86 is modeled on scale a reported detected bisected glycan structure.

Materials and Methods

Cell culture

The B-lymphoblastoid cell line JY (containing HLA-A*02:01, HLA-B*07:02, HLA-C*07:02) was cultured in RPMI 1640 medium (+glutamine, Gibco) supplemented with 10% fetal bovine serum, 50 U/ml penicillin (Gibco) and 50 µg/ml streptomycin (Gibco) in a humidified atmosphere at 37 °C with 5% CO_2. The B-lymphoblastoid cell lines DEM (containing HLA-A*02:01, HLA-B*57:01, HLA-C*06:02) and DBB (containing HLA-A*02:01, HLA-B*57:01, HLA-C*06:02) were cultured in RPMI 1640 medium (+glutamine) supplemented with 15% fetal bovine serum, 1 mM sodium pyruvate (Gibco), 50 U/ml penicillin and 50 µg/ml streptomycin in a humidified atmosphere at 37 °C with 5% CO_2.

HLA class I immuno-affinity purification and protein digestion

Per cell line, $5x10^8$ cells were harvested by centrifugation and washed three times with phosphate-buffered saline (PBS). The cells were lysed in 10 ml Pierce IP lysis buffer (Thermo Scientific) supplemented with 1x complete protease inhibitor cocktail (Roche Diagnostics), 50 µg/ml DNAse I (Sigma-Aldrich) and 50 µg/ml RNAse A (Sigma-Aldrich) per gram cell pellet for 1.5 h at 4 °C. The lysate was cleared by centrifugation for 1 h at 18000 g at 4 °C. The protein concentration was determined using the BCA assay (Pierce). HLA class I immuno-affinity purification was performed as described by Demmers et al. (32). In short, HLA class I complexes were immunopurified using 0.5 mg W6/32 antibody (33) coupled to 125 µl protein A/G beads (Santa Cruz) from 25 mg whole cell lysate. For the plasma membrane and inner membrane fractions 0.16 mg W6/32 antibody coupled to 40µl protein A/G beads was used for ~160 µg and ~60 µg input respectively. To prevent co-elution, the antibodies were cross-linked to protein A/G beads. Incubation took place at 4 °C for approximately 16 h. After immuno-affinity purification, the beads were washed with 40 ml of cold PBS. HLA class I complexes and peptide ligands were eluted with 10% acetic acid. The peptide ligands were separated from the HLA molecules using 10 kDa molecular weight cutoff filters (Millipore). The HLA molecules were denatured in 8 M Urea in 500 mM ammonium bicarbonate with 1x complete EDTA-free protease inhibitor cocktail (Roche Diagnostics). The HLA proteins were reduced, alkylated and digested with 50 ng trypsin. The digested peptides were loaded onto C18 SEPPAK columns (Supelco) in 0.1% formic acid and eluted after clean-up with 80% acetonitrile in 0.1% formic acid. The samples were dried by vacuum centrifugation and reconstituted in 2% formic acid prior to LC-MS/MS analysis.

Cellular Fractionation

Plasma membrane fractionation was performed by gentle homogenization in an isotonic environment, in the absence of detergents. Homogenized JY cells were applied to a dextran gradient for phase-separation (by differential partitioning) into the plasma membrane fraction (low density top fraction) or inner membrane fraction (high density bottom layer. Intermediate fractions where liquid-liquid mixing could occur were carefully and generously discarded. A similar approach has been published previously (34).

Mass spectrometry analysis

Glycoproteomics data were acquired using an Agilent 1290 UHPLC (Agilent) coupled to an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer (Thermo Scientific). Proteomics data was acquired using an equivalent UHPLC setup on a Q-exactive™ HF (Thermo Scientific). Peptides were first trapped on a 2cm x 100 µm Reprosil C18 trap column (3 µm particle size), followed by separation on a 50 cm x 75 µm Poroshell EC-C18 analytical column (2.7 µm). Trapping was performed for 5 minutes 0.1% formic acid (solvent A) and eluted with a gradient using 80% ACN with 0.1% formic acid (Solvent B). Gradient: 13% B to 40% B over 48 min. 40% B to 100% over 1 min holding at 100% B for 4 min.

For proteomics the mass spectrometer was operated in data dependent mode using a top 15 method. Full MS scans were captured using 60000 resolution at 200 *m/z*, mass range 310 - 1600 *m/z*, with an AGC target of 3e6, and max injection time of 20 ms. MS/MS scans were triggered on charge states 2 – 5 and excluded for 12 sec after selection. A selection window of 1.4 *m/z* was used followed by HCD fragmentation at 27 normalized collisional energy. Fragmentation scans were captured at 30000 resolution, with a fixed first mass of 120 *m/z*, with an AGC target of 1e5 and 50 ms max injection time.

For glycoproteomics the mass spectrometer was operated in a product ion triggered data dependent mode using a 3 sec cycle type. Full MS scans were captured on the Orbitrap at 60000 resolution at 200 *m/z*, mass range 350 – 2000 *m/z*, with an AGC target of 4e5, and max injection time of 50 ms. MS/MS scans were triggered on charge states 2 – 8 and excluded for 30 sec after selection. A selection window of 1.6 *m/z* was used followed by fragmentation used HCD at 30% collisional energy. Fragmentation scans were captured on the Orbitrap at 30000 resolution, with mass range 120 - 4000 *m/z*. For stepping HCD and EThcD triggering methods a targeted mass trigger was configured with a mass list of common glycan oxonium ion fragments as reported by Reiding, K.R., et al. (35). When at least 3 ions were detected another MS/MS scan was triggered on the same precursor. These scans used the same isolation and resolution parameters. For EThcD calibrated charge

dependent ETD parameters were used and 25% supplemental activation, stepping HCD used collisional energies of 10, 25 and 40%. EThcD and Stepping HCD scans used AGC target of 400% and 250 ms max injection time.

Data analysis

Data analysis of proteomics data was performed using MaxQuant (v.1.6.17.0) and the Andromeda search engine. Data was searched against a Swissprot human database (20431 entries, downloaded on Sept 18, 2019) appended with the specific HLA allotypes found in each cell line. Enzyme specificity was set to trypsin and up to 2 missed cleavages were allowed. Fixed modification of carbamidomethyl at C and variable modifications oxidation at M and N-terminal acetylation were used. An FDR rate of 0.01 was set for both protein and peptide identification. Label free quantification was performed using IBAQ (36, 37).

Data analysis of glycoproteomics data was performed using PMi-Byonic (Protein Metrics) (v3.6). Byonic settings were as follows, fully specific c-terminal of RK with up to 2 missed cleavages. 10 ppm precursor mass tolerance both HCD and EThcD were applicable with fragment mass tolerance of 20 ppm. Modifications, oxidation variable at M, carbamidomethyl fixed at C, and using the Byonic N-glycan 132 human database for identification of glycosylation. Precursor isotope off by x was set to 'Too high or low (narrow)' with a maximum precursor mass of 10,000 Da. Precursor assignment was computed from MS with a maximum of 2 precursors per MS/MS. A 1 % FDR (or 20 reverse count) cutoff was used with decoys added in the database. For the protein database, a fasta file was created containing all HLA class I and class II allotypes for the specific cell lines. Subsequent analysis was performed with Python 3.8 using Pandas 1.1.3 (38), Numpy 1.19.2 (39), Matplotlib 3.3.2 (40) and Seaborn 0.11. Glycopeptides were selected with a minimum Byonic score of 150 and absolute log probability higher than 1. In addition, a glycan needed to be detected at least 6 times per HLA protein per cell line. Glycans were assigned categories based on compositional requirements. Paucimannose: HexNAc ≤ 2, Hex ≤ 3. High mannose: HexNAc = 2, Hex > 3, Hybrid/Asymmetric: HexNAc = 3, Diantennary: HexNAc = 4, Bisected: HexNAc ≥ 5, Hex ≤ 5, Extended: HexNAc ≥ 5, Hex > 5. Distributions of glycan categories were calculated based on glycan peptide spectral match (PSM) counts.

Results

Abundances of HLA class I proteins and their glycosylation profiles in whole cell lysates

To investigate the glycosylation of HLA class I heavy chain molecules in detail, we selected a panel of three human cell lines namely DBB, DEM and JY. These cell lines are all Epstein-Barr virus transformed immortalized B lymphoblastoid cells of the ECACC HLA-Typed collection, and were selected as they all express HLA class I at a reasonably high level. All three cell lines are homozygous for HLA Class I, meaning that each only expresses a single A, B and C allotype. DBB and DEM both express HLA-A*02:01, B*57:01 and C*06:02. JY expresses HLA-A*02:01, B*07:02 and C*07:02 (Figure 1C). These allotypes are amongst some of the most common allotypes found in the Caucasian population and these cell lines are routinely used as model systems for studying HLA molecules and their peptide ligandomes (41-43).

The analysis was started by enriching all HLA class I molecules from whole cell lysates by immuno-affinity purification using the pan-HLA class I W6/32 antibody (33). The purified HLA complexes were digested by trypsin and all resulting peptides were subjected to LC-MS/MS. In a first set of shotgun proteomics analyses, the relative abundance of the different HLA molecules was assessed in the 3 studied cell lines, using only the unique non-glycosylated HLA peptides. To achieve relative quantification, the IBAQ intensities of the HLA proteins in a sample were summed and the intensity of each HLA protein in that sample was normalized to this summed intensity. Clearly, this data revealed that in all cell lines the order of abundance in the whole cell lysates was HLA-A > HLA-B > HLA-C > HLA-F (see Figure 2A and the Supplementary Tables S1-S3, unique peptides used for quantification are shown in Supplementary Tables S6-S8). In contrast to the typical data-dependent shotgun proteomics approaches, a glycopeptide targeted product-ion triggered fragmentation strategy was next employed. Precursor ions were selected and fragmented using higher-energy collisional dissociation (HCD) as is also the case in standard shotgun methods. However, if the resulting MS/MS scan contained glycan specific oxonium ions, an additional EThcD or stepping-energy HCD scan was triggered on the same, likely glycopeptide, precursor ion (35). In this targeted approach and enabled by the high sequence coverage, we could confidently assign a variety of glycopeptides for each HLA allotype, based on the small but unique protein sequence features. Since the HLA class I molecules were amongst the most abundant proteins in the sample, as a result of the immuno-affinity enrichment, no further enrichment of glycopeptides was needed to analyze the glycopeptides.

Based on the mass shift induced by the glycans to the peptide backbone, combined with the known biosynthetic knowledge about human protein glycosylation, we next annotated glycan compositions to all identified HLA glycopeptides, which were assigned to a smaller number of categories as depicted and color-coded,

schematically in Figure 2B. Among these 6 categories high-mannose glycans (green) are characterized by having only Man residues extending the Man3GlcNAc2 core shared by all N-glycans. Complex type glycans instead have branches extending the core that are initiated by GlcNAc, which can still be further elongated. Additionally, complex type glycans often contains one or more Fuc and SA residues. Hybrid/asymmetric glycans (purple) display one Man extended branch and one complex type branch. Diantennary glycans (light blue) instead have two complex type branches. Extended glycans (red) have more than two complex type branches. Bisected glycans (orange) are complex type glycans carrying a bisecting GlcNAc residue attached to the β-Man of the core. Finally, paucimannose and truncated glycans (dark blue) are small mannose glycans or truncated glycans that do not have the complete N glycan core.
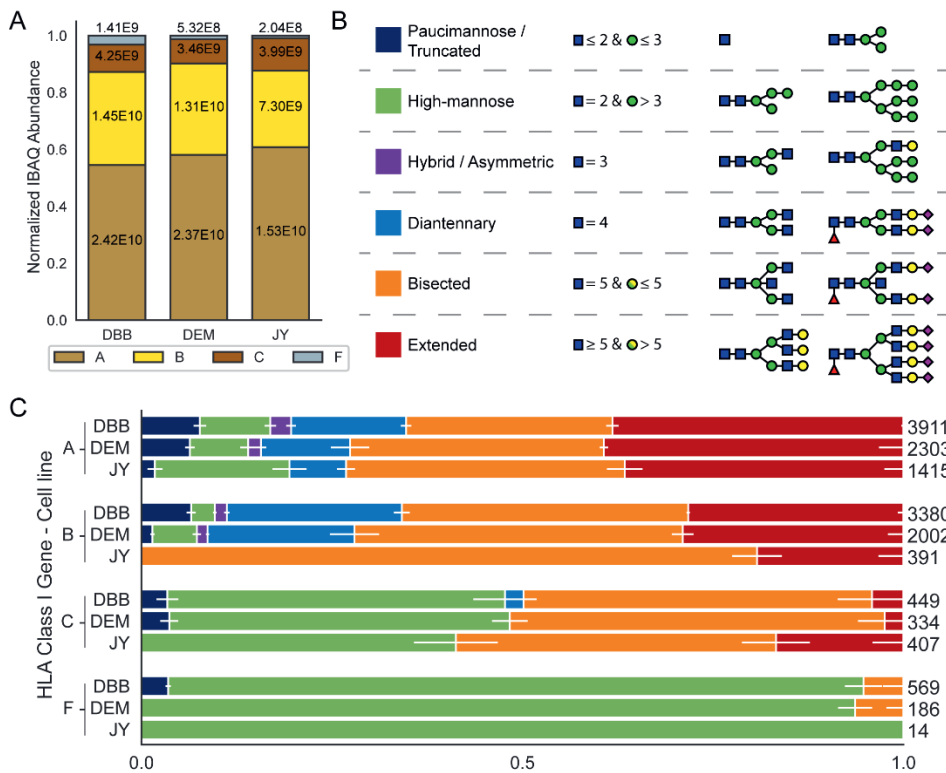
Figure 2: HLA abundances and glycosylation in the cell lines DBB, DEM and JY. A) Normalized protein abundances of different HLA class I molecules immunopurified from whole cell lysates determined by proteomics, using solely unique non-glyco peptides for quantification. Within the bars, the non-normalized values are depicted. **B)** Categorization and color-coding of glycan classes. The 6 glycan categories are paucimannose (dark blue), high-mannose (green), hybrid/asymmetric (purple), diantennary (light blue), bisected (orange) and extended (red). **C)** Stacked bar plots depicting the distribution of glycans for each HLA class I gene in each cell line. Within the bars, the horizontal white depicts the standard deviation averaged over 9 injection replicates. On the right of each bar, values indicate the cumulative number of glycopeptides detected across all injection replicates.

We next assessed the relative abundance of these different categories per individual HLA glycopeptide based on the number of spectral counts (peptide-spectra matches; PSMs) we observed in our LC MS/MS analysis (27), the detected glycopeptides and glycans are listed in Supplementary Tables S11-S13 and S16-S18 respectively. These results are depicted in Figure 2C using the color codes for the different glycan categories, with errors bars obtained from replicate measurements. A more detailed overview of the observed glycan compositions is provided in Figure S2. At first glance, although there is some similarity, this analysis revealed that glycoprofiles of allotypes can be quite distinct, even when they are expressed in the same cell. It should be noted that the site occupancy of the glycan is 100%, a fact we could also reproduce by measuring the intact mass of HLA-A by LC-MS. Consequently, non-glycosylated versions of the glycopeptides were not detected in any experiments.

HLA-A*02:01 is abundantly expressed by all three cell lines we investigated. Our data reveal that the glycosylation profile on HLA-A*02:01 is nearly identical in all three cell lines and dominated by bisected (orange) and extended (red) glycans. However, a considerable fraction (25-30% based on spectral counts) of the HLA-A molecules carry smaller and/or simpler glycans, mostly high mannose (green) but also some paucimannose (dark-blue) glycans. The DBB and DEM cell lines express B*57:01, the glycosylation of which in these two cell lines is nearly identical and resembles closely that of HLA-A*02:01. In sharp contrast, HLA-B*07:02 in JY cells did harbor a distinct glycosylation pattern, that features almost exclusively bisected (orange) and extended (red) glycans. Distinctively, many more HLA-C molecules, i.e. HLA-C*07:02 in JY cells, and HLA-C*06:02 in both DBB and DEM cells (estimated to be around 50% based on spectral counts), are detected with high-mannose glycans (green). The HLA-C glycosylation patterns are quite similar in all three studied cell lines, but different from the other HLA genes. Finally, we also detected, while substantially less abundant, unique glycopeptides originating from HLA-F in all three cell lines. However, we were not able to annotate the exact HLA-F allele. HLA-F, being one of the non-classical HLA genes, is much less polymorphic (only 6 different forms have been reported), and all these variants share the same sequence for the part covering the tryptic glycopeptides studied here. In addition, according to our proteomics data, HLA-F expression is quite low, with the DBB cell line forming a noticeable exception (Figure 1A). HLA-F independent from the source cell of origin was found to be nearly exclusively harboring high-mannose type glycans (green). We did not consider the two other non-classical HLA genes; HLA-E and HLA-G. Sporadically a few unique peptides of HLA-E were detected in our analyses, although their abundance were substantially lower compared to the other HLAs. No unique glycopeptides were detected for HLA-E. No unique peptides for HLA-G were detected in any of our analyses.

Annotation of glycan structures on HLA glycopeptides

LC-MS/MS on glycopeptides is well suited to determine the most likely glycan composition. How these carbohydrate moieties are linked together in the glycan structure, however, is much harder to reveal by mass spectrometric means. Nonetheless, in exceptional cases LC-MS/MS can also shed light on how a glycan is structured, as nicely described for bisecting glycans by Dang et al. (44), using diagnostic fragment ions. Following a similar strategy, we depict in Figure 3A, an EThcD fragmentation spectrum for a glycopeptide derived from the heavy chain of HLA B*57:01, observing diagnostic fragment ions indicating glycan bisection. In detail, in Figure 3A at *m/z* 1168.7 (z=5+) the intact precursor is shown, annotated with "M". From this data, we determined the intact glycan to have the SA2Gal2Man3GlcNAc5Fuc1 composition. In the EThcD spectrum a fragment at *m/z* 1051.5 (z=4+) is detected, assigned to the peptide plus the glycan lacking 4 Hexoses, 2 HexNAcs and 2 NeuAcs. The mass loss observed for this fragment ion indicates the loss of both antennae on the glycan. After cleaving off both antennae only a peptide+GlcNAc3Man1Fuc1 fragment remains. This fragment is unique to bisecting glycans, suggesting the bisecting structure of the glycan we observed on HLAs.
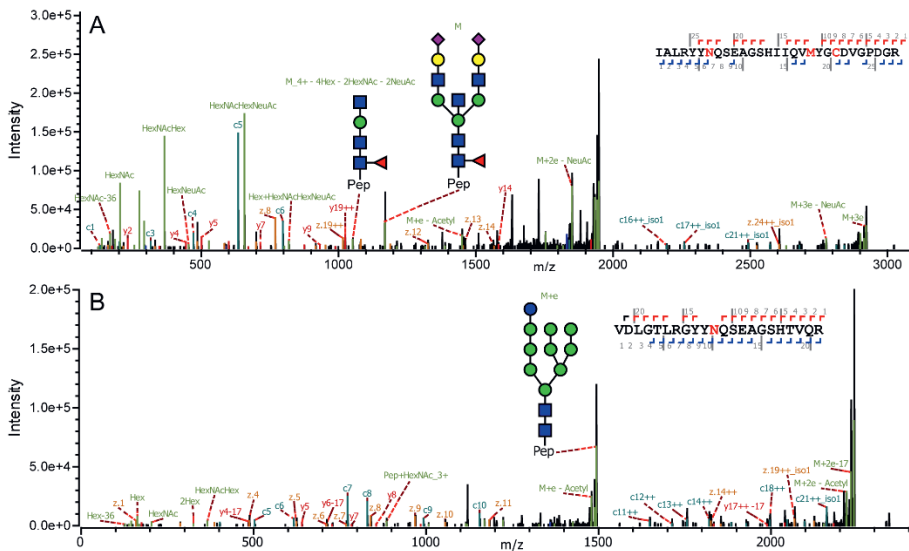
Figure 3: Examples of annotated HLA heavy chain glycopeptides. A) EThcD MS/MS spectrum of a HLA-B*57:01 derived bisected glycopeptide (DBB cell line). The precursor glycopeptide (M), *m/z* 1168.7 (z=5+), together with it annotated glycan structure are indicated. The fragment ion at *m/z* 1051.5 (z=4+), with its annotated glycan structure, is verifying the proposed bisected structure. Several additional peptide backbone and glycan fragment ions used for the identification of the glycopeptide are annotated. **B)** EThcD MS2 spectrum of a HLA-A*02:01 derived high-mannose glycopeptide (DBB cell line). The annotated glycan structure is indicated. Several additional peptide backbone and glycan fragment ions used for the identification of the glycopeptide are annotated. See Figure S3 for additional MS/MS spectra of different HLA glycopeptides.

In Figure 3B, the EThcD spectrum of a HLA-A*02:01 derived glycopeptide harboring the high-mannose glycan Glc1Man9GlcNAc2 is shown. High-mannose glycans can have up to 9 Man residues, thus the 10th hexose is accredited to a Glc. The high-mannose glycan Glc1Man9GlcNAc2 is thought to be important for binding of the HLA heavy chain to the folding chaperone Calreticulin in the ER. Calreticulin interacts optimally with monoglucosylated HLA class I heavy chains, whatever their state of assembly with light chains and peptide and inhibits their aggregation (45, 46). It is expected that proteins harboring these glycans are localized in the ER. As our experiments were carried out on whole cell lysates, this data hints at that we analyze both internal as well as cell-surface expressed HLA molecules. In Figure S3, several more MS/MS spectra are provided highlighting different glycopeptides detected in our measurements.

Relative abundances of HLA class I proteins and their glycosylation profiles in the plasma membrane versus the inner membrane

Following their initial synthesis by the ribosome, HLA molecules reside in and traverse through the various inner-membrane compartments (ER and Golgi) of the cell, where they are properly folded and loaded with peptide on their way to the plasma membrane. During this process, maturation of the HLA heavy chain glycan occurs by the various glyco-enzymes present in these compartments (29-31). The data presented above were all resulting from HLA affinity enrichments performed on whole cell lysates. However, our data did indicate that HLA molecules within a single cell may be differentially glycosylated, which we expect to be linked to the compartmental localization of the HLA molecules in and on the cell. In essence, we utilize the glycosylation characteristic as a proxy for the subcellular distribution of the HLA molecules. Therefore, we extended our analysis performing the HLA affinity enrichment separately on (crude) plasma membrane and inner membrane (mostly ER and Golgi) fractions of the cells. This pre-fractionation comes at the expense of sensitivity, as the pull-downs generally require quite some starting material. Consequently, in the fractionated samples we were able to detect only the highest abundant glycopeptides, yielding a less accurate representation of the glycan repertoire. Although some sensitivity is lost, it should still show whether the glycosylation accurately represents the distributions of HLA molecules. We performed this analysis on the JY cells, as in the whole cell lysates of these cells we observed the most striking differences in glycosylation patterns between HLA-A, B, C and F (Figure 2). A proteomics-based evaluation of the subcellular fractionation, as described previously (47), assessing the enrichment of protein UniProt keywords in the inner and plasma membrane fractions (Figure S4) revealed successful fractionation.
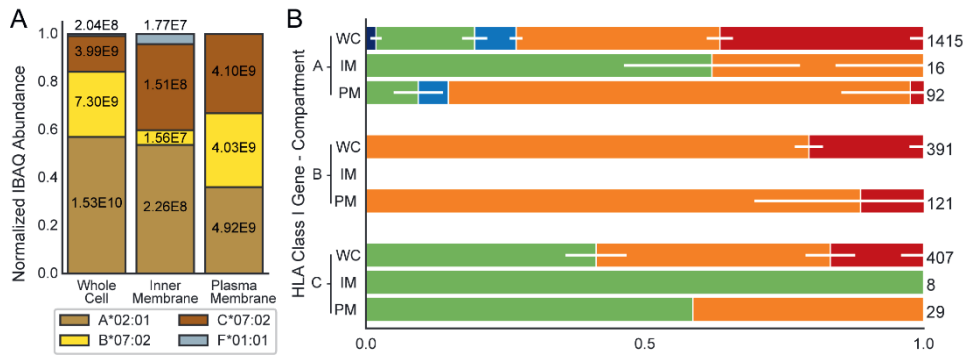
Figure 4: HLA abundances and glycosylation in different compartments of the JY cells. A) Normalized protein abundances of different HLA class I genes in whole cell lysates, and the inner membrane (ER and Golgi) and plasma membrane fractions determined by proteomics, using solely unique non-glyco peptides for quantification. Within the bars, the non-normalized values are depicted. **B)** Stacked bar plots depicting the distribution of glycans for each HLA class I gene in each of the three analyzed fractions. Abbreviations for whole cell (WC), inner membrane (IM) and plasma membrane (PM) are used to specify the compartments. Within the bars, the horizontal white lines depict the standard deviation averaged over 9 injection replicates for the whole cell and 3 injection replicates for inner- and plasma membrane preparations. On the right of each bar, values indicate the cumulative number of glycopeptides detected across all injection replicates. Due to the low abundance of HLA-B in the inner membrane fraction, and for HLA-F, no glycopeptides could be detected following the cellular fractionation. The whole cell data presented in this figure is identical to the JY cell data presented in Figure 2.

Again, as described above, we assessed by standard shotgun proteomics, quantifying by solely using unique peptides, the relative abundance of the different HLA molecules in the plasma and inner-membrane fraction and compared that to that observed in the whole cell lysates (Figure 4A, and Supplementary Tables S3-S5 unique peptides used for quantification are shown in Supplementary Tables S8-S10). A list of other abundantly detected proteins is provided as supplementary table. This data clearly revealed that while HLA-A is quite abundant in both the plasma and inner membrane fractions, HLA-B is nearly uniquely enriched in the plasma membrane. In contrast, HLA-C was found to be enriched in the inner membrane fraction. Moreover, while HLA-F could not be detected in the whole cell and plasma membrane, it was reasonably abundant in the inner membrane fraction.

Next, we focused on potential compartment specific HLA heavy chain glycosylation (Figure 4B), and for the glycopeptide analysis we again performed immuno-affinity purification on the HLA molecules. We were only able to make that comparison for the classical HLA class I genes A, B and C, as the abundance of HLA-F was too low for detection of sufficient glycopeptides. Additionally, although we detected a few HLA-B unique peptides in the shotgun experiments for the inner-membrane fraction we did not detect glycopeptides of these molecules in the inner membrane fraction, likely due to the low abundance of these molecules in that fraction (Figure 4A). In Figure 4B the same glycan categorization and color-coding is used as in Figure 2B

and additionally the data for the whole cell lysate (WC) shown in Figure 2B is replicated in Figure 4B for comparison with the data obtained for the inner membrane (IM) and plasma membrane (PM) fractions. The data presented in Figure 4B clearly show that complex glycans represented by the bisected (orange) and extended (red) categories are enriched in the PM fraction, especially for HLA-A and HLA-B. Conversely, high-mannose glycans are found to be enriched in the IM fraction, especially for HLA-A and HLA-C. The fact that we do not detect paucimannose and diantennary glycosylated glycopeptides in the IM and PM fractions may likely be attributed to our lower sensitivity in these experiments due to the lower amount of starting material prior to the affinity purification. The detected glycopeptides and glycans are given in Supplementary Tables S13-S15 and S18-S20. Notwithstanding these sensitivity issues, the data presented in Figure 4 clearly reveal that the pool of HLA molecules is indeed distributed over the different compartments, and these HLA molecules carry compartment "specific" glycans.

Discussion

Here we report an in-depth characterization of the N-glycosylation features of HLA class I molecules, affinity purified from whole cell lysates and enriched plasma membrane and inner membrane fractions. We make the striking observation that different allotypes also display different glycosylation patterns, which is noteworthy considering that HLA molecules of different allotypes are nearly identical in sequence and structure, except for a few key residues in the peptide-binding groove. Cumulatively, the data presented here hint at a substantial diversity in the distribution of HLA class I complexes over the different membrane compartments of the cell. By using glycosylation as a proxy for subcellular localization we are able to show that different allotypes are distinctly distributed over the compartments and that a considerable amount of HLA molecules may reside intracellularly.

Based on the observed glycosylation patterns it appears that HLA-B complexes, especially in JY cells, are mostly rather mature in their glycosylation and reside largely on the plasma membrane of the cell. In sharp contrast, the population of HLA-C molecules are largely enriched in less mature high-mannose glycans, and mostly residing in the inner-membrane ER/cis-Golgi compartments. Even more so HLA-F is exclusively modified with high-mannose glycans and seems largely retained in the inner membrane compartments of the cell, as proposed previously (48). The population of HLA-A complexes in the cell display a glycan distribution ranging from paucimannose, high-mannose, diantennary, to fully matured bisected and extended glycans, and seems also to be more widely distributed over the plasma membrane and the inner-membrane compartments. The distinctive glycosylation and cellular

localization of HLA class I proteins as observed in this study is schematically summarized in Figure 5. These observations are made in our model JY cells, but are also conserved to some extent in the DBB and DEM cells that we analyzed in parallel.
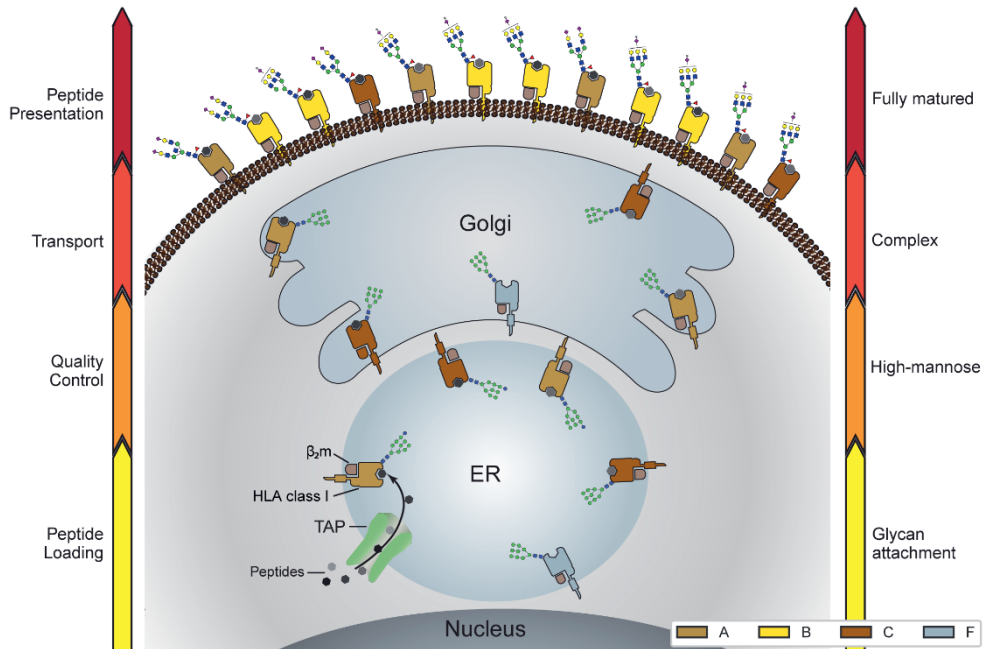
Figure 5. Distinctive glycosylation and cellular localization of HLA class I proteins. Starting as nascent chains synthesized by cellular ribosomes HLA heavy chains traverse through specific inner membrane compartments to get properly folded, trimmed, associated with the β2m chain, loaded with the peptide antigens and glycosylated. These processes occur largely sequential and require chaperones, the peptide-loading complex and a variety of glycoenzymes, residing in the different sub compartments of the ER and Golgi. Fully assembled, peptide-loaded and maturely glycosylated HLA complexes make it to the cell surface becoming embedded in the plasma membrane, where they present the peptides to the T-cell receptors of CD8+ T-cells. The arrows on either side of the schematic indicate the stages of HLA antigen expression (left) and glycosylation maturation (right). The HLA complexes of different class I genes are color coded corresponding to the colors used in figures 2A and 4A. The number of copies presented illustrate roughly their relative distribution over the inner membrane compartments and plasma membrane in JY cells. Moreover, prototypical HLA glycans observed uniquely in the inner membrane fraction and plasma membrane fractions are depicted as well.

Previous reports have also looked at the functional role for the N-linked glycosylation on HLA heavy chains. It has been reported that the glycan Glc1Man9GlcNAc2 is the specific structure that is recognized by the folding chaperones calnexin and calreticulin (49), which form a part of the PLC (50). Consequently, absence of the HLA glycan, accomplished by mutating the asparagine 86, has been shown to completely eliminate PLC activity in vitro (51). A specific functional role of the glycan

when the HLA molecule is at the cell surface has not been reported. Except for the discovery that sialylation modulates cell surface stability (52). This observation is not surprising considering the generalized role of sialic acid for extracellular proteins (53, 54).

While it is well known that HLA complexes can reside either at the plasma membrane and at the inner membranes of the ER and Golgi compartments, the exact distribution of different class I allotypes over these compartments has been not been investigated in detail. Evidently, this distribution is determined by a variety of factors. Formation of sub-optimally loaded HLA molecules, e.g. loaded with low affinity peptides, can be retained intracellularly or exhibit shorter a half-life (55). Furthermore, certain sequences exhibit a preference for TAP which consequently promotes PLC formation, ultimately leading to increased cell surface expression (56). There is also the TAP independent route for peptide loading utilizing the protein TAPBPR that has different affinities for certain HLA allotypes (57). An extreme case has been identified where a single natural polymorphism between two allotypes B*44:02 and B*44:05, has strong implications for peptide loading, where the presence of tapasin even lowers the affinity of peptides that can be loaded on B*44:05 (25, 26). However, whether such factors could contribute to the substantial observed differences in subcellular localization between allotypes as shown here is unclear.

In support of our observations, Ryan and Cobb summarized that there are noticeable differences in glycosylation between HLA-A/HLA-B and HLA-C (58). They argue that the distinct N-glycans found on HLA-C may correlate with a functional role separate from those of HLA-A and HLA-B. They argued that, in line with our data, HLA-C is less abundant at the cell surface, but also binds a more restricted repertoire of peptide antigens, suggesting a reduced role in antigen presentation to CD8+ T cells. Indeed, HLA-C may play a more prominent role in interaction with NK-cell receptors (59). However, they did not report a difference between HLA-A and HLA-B glycosylation. It has been mostly assumed that HLA-A and HLA-B are functionally very alike. Therefore, we assume that the observed differences are caused by a variance in efficiency at which these allotypes traverse through the secretory pathway.

A well-defined functional role for HLA-F has yet to be established. The few reports available hint at its predominant cellular presence in the inner membrane compartments and being dominantly decorate with high-mannose/hybrid N-glycans (48). However, it has also been reported that HLA-F is a high-affinity ligand for NK-cell receptor KIR3DS1, which suggests that HLA-F could play a role similar to HLA-C (60), although we did not detect it at the cell-surface.

Although we describe primarily the differences in glycosylation of HLA molecules originating from different class I genes, there may also be differences within allotypes

4

of a specific class I gene. For instance, our data show a substantial difference in glycosylation between the allotypes of HLA-B (Figure 2). Whereas B*07:02 in the JY cell harbors exclusively mature bisected (orange) and extended (red) glycans, the HLA-B*57:01 heavy chains molecules in both the DBB and DEM cells, show a broader distribution. Next to harboring these bisected and extended mature glycans, a substantial proportion of (up to 25% by spectral count) diantennary (light blue) and high mannose (green) glycans. Evidently, this could additionally be a result of the different cell lines used in this study, with their different compartmental organization and glycoenzymes they may harbor. In that sense, it may be interesting to extend our glycoprofiling analysis to a well-defined cell line expressing just a single HLA allele as reported by Abelin et al (61). Nevertheless, it is also tempting to speculate about the observed differences with respect to HLA-B*57:01 and B*07:02, which differ in the Bw motifs they contain. The Bw motif is a public epitope (an epitope shared by multiple allotypes), for HLA molecules that plays a major role in transplant rejection (62). Two motifs are recognized, Bw4 and Bw6 (63), which are found on all HLA-B and some HLA-A heavy chains, and is determined by the amino acid sequence surrounding the N-linked glycosylation site. HLA-B*57:01 of the DBB and DEM cells displays a Bw4 motif while B*07:02 of JY cells displays a Bw6 motif. HLA molecules harboring the Bw4 motif have been shown to elicit a stronger NK-cell response (64) and the glycan attached to the HLA heavy chain is required for the NK-cell interaction (65). There may thus be a link between this Bw motif, the HLA glycosylation pattern and the interaction with NK-cell KIR cells.

Our findings should impact the important field of HLA peptide ligandome analyses. In these analysis, the focus lies on pathogen-derived or cancer related neo-antigens, presented by HLA molecules. These antigens are of great importance in the development of pathogen- and anti-tumor targeting vaccines (14). The two main experimental approaches to retrieve peptide-antigens from HLA molecules are immunoprecipitation from whole cell extracts or mild acid elution on intact cells (66). In the latter approach only the ligandome present at the cell surface should be sampled, whereas with the former approach the intercellular HLA complexes are probed as well. Based on our observations, the two approaches could chart distinctive ligandomes particularly for HLA allotypes, which can have a considerable part of their molecules residing intra-cellularly. Each approach is clearly associated with different strengths and potential bias (67); also a definitive consensus is still elusive, and may be context-dependent.

In conclusion, through an in-depth analysis of the glycosylation profiles on HLA class I molecules, originating from 3 different cell lines, covering 4 different class I genes and 6 different allotypes, we observed that they all exhibit different glycosylation profiles. It would be interesting to expand such analysis to more cell lines covering a broader range of allotypes, and/or even primary cells, to probe whether such observations are general or very much cell type specific. We show that the HLA

glycan-signature reflects the cellular organization, and is distinct for the same HLA molecules residing at the plasma membrane or the inner membranes (ER and Golgi). Although, this is fully in line with what would be expected based on our knowledge of protein cellular glycosylation, analyzing glycoform profiles may thus provide a direct quantitative assessment of the cellular distribution of specific HLA class I molecules. As the peptide repertoire presented at the cell surface is sampled by T-cells, our observations are significant for the analysis and interpretation of HLA peptide ligandomes.

Acknowledgments

Data Availability

4

The mass spectrometry proteomics and glycoproteomics data and accompanying processed search files have been deposited to the ProteomeXchange Consortium via the PRIDE (68) partner repository with the data set identifier PXD023684.

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: https://doi.org/10.1021/acs.jproteome.1c00466

References

1.   Vyas, J. M.; Van der Veen, A. G.; Ploegh, H. L., The known unknowns of antigen processing and presentation. Nat Rev Immunol 2008, 8, (8), 607-18.

2.   Neefjes, J.; Jongsma, M. L.; Paul, P.; Bakke, O., Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol 2011, 11, (12), 823-36.

3.   Klein, J.; Sato, A., The HLA system. New England Journal of Medicine 2000, 343, (10), 702-709.

4.   Bradley, B. A., The role of HLA matching in transplantation. Immunol Lett 1991, 29, (1-2), 55-9.

5.   Montgomery, R. A.; Tatapudi, V. S.; Leffell, M. S.; Zachary, A. A., HLA in transplantation. Nat Rev Nephrol 2018, 14, (9), 558-570.

6.   Gough, S. C.; Simmonds, M. J., The HLA Region and Autoimmune Disease: Associations and Mechanisms of Action. Curr Genomics 2007, 8, (7), 453-65.

7.   Zanelli, E.; Breedveld, F. C.; de Vries, R. R., HLA association with autoimmune disease: a failure to protect? Rheumatology (Oxford) 2000, 39, (10), 1060-6.

8.   Bodis, G.; Toth, V.; Schwarting, A., Role of Human Leukocyte Antigens (HLA) in Autoimmune Diseases. Rheumatol Ther 2018, 5, (1), 5-20.

9.   Muniz-Castrillo, S.; Vogrig, A.; Honnorat, J., Associations between HLA and autoimmune neurological diseases with autoantibodies. Auto Immun Highlights 2020, 11, (1), 2.

10.   Mehra, N. K.; Kaur, G., MHC-based vaccination approaches: progress and perspectives. Expert Rev Mol Med 2003, 5, (7), 1-17.

11.   Blackwell, J. M.; Jamieson, S. E.; Burgner, D., HLA and infectious diseases. Clin Microbiol Rev 2009, 22, (2), 370-85, Table of Contents.

12.   Ochoa, E. E.; Huda, R.; Scheibel, S. F.; Nichols, J. E.; Mock, D. J.; El-Daher, N.; Domurat, F. M.; Roberts, N. J., Jr., HLA-associated protection of lymphocytes during influenza virus infection. Virol J 2020, 17, (1), 128.

13.   Goulder, P. J.; Walker, B. D., HIV and HLA class I: an evolving relationship. Immunity 2012, 37, (3), 426-40.

14. Schumacher, T. N.; Schreiber, R. D., Neoantigens in cancer immunotherapy. Science 2015, 348, (6230), 69-74.

15. Bassani-Sternberg, M.; Coukos, G., Mass spectrometry-based antigen discovery for cancer immunotherapy. Curr Opin Immunol 2016, 41, 9-17.

16. Neefjes, J. J.; Schumacher, T. N.; Ploegh, H. L., Assembly and intracellular transport of major histocompatibility complex molecules. Curr Opin Cell Biol 1991, 3, (4), 601-9.

17. Demmers, L. C.; Kretzschmar, K.; Van Hoeck, A.; Bar-Epraim, Y. E.; van den Toorn, H. W. P.; Koomen, M.; van Son, G.; van Gorp, J.; Pronk, A.; Smakman, N.; Cuppen, E.; Clevers, H.; Heck, A. J. R.; Wu, W., Single-cell derived tumor organoids display diversity in HLA class I peptide presentation. Nat Commun 2020, 11, (1), 5338.

18. Leone, P.; Shin, E. C.; Perosa, F.; Vacca, A.; Dammacco, F.; Racanelli, V., MHC class I antigen processing and presenting machinery: organization, function, and defects in tumor cells. J Natl Cancer Inst 2013, 105, (16), 1172-87.

19. Robinson, J.; Halliwell, J. A.; Hayhurst, J. D.; Flicek, P.; Parham, P.; Marsh, S. G., The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res 2015, 43, (Database issue), D423-31.

20. Parham, P.; Lomen, C. E.; Lawlor, D. A.; Ways, J. P.; Holmes, N.; Coppin, H. L.; Salter, R. D.; Wan, A. M.; Ennis, P. D., Nature of polymorphism in HLA-A, -B, and -C molecules. Proc Natl Acad Sci U S A 1988, 85, (11), 4005-9.

21. van Deutekom, H. W.; Kesmir, C., Zooming into the binding groove of HLA molecules: which positions and which substitutions change peptide binding most? Immunogenetics 2015, 67, (8), 425-36.

22. Gonzalez-Galarza, F. F.; McCabe, A.; Santos, E.; Jones, J.; Takeshita, L.; Ortega-Rivera, N. D.; Cid-Pavon, G. M. D.; Ramsbottom, K.; Ghattaoraya, G.; Alfirevic, A.; Middleton, D.; Jones, A. R., Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. Nucleic Acids Res 2020, 48, (D1), D783-D788.

23. Barber, L. D.; Patel, T. P.; Percival, L.; Gumperz, J. E.; Lanier, L. L.; Phillips, J. H.; Bigge, J. C.; Wormwald, M. R.; Parekh, R. B.; Parham, P., Unusual uniformity of the N-linked oligosaccharides of HLA-A, -B, and -C glycoproteins. J Immunol 1996, 156, (9), 3275-84.

24.   Scull, K. E.; Dudek, N. L.; Corbett, A. J.; Ramarathinam, S. H.; Gorasia, D. G.; Williamson, N. A.; Purcell, A. W., Secreted HLA recapitulates the immunopeptidome and allows in-depth coverage of HLA A*02:01 ligands. Mol Immunol 2012, 51, (2), 136-42.

25.   Williams, A. P.; Peh, C. A.; Purcell, A. W.; McCluskey, J.; Elliott, T., Optimization of the MHC class I peptide cargo is dependent on tapasin. Immunity 2002, 16, (4), 509-520.

26.   Bailey, A.; Dalchau, N.; Carter, R.; Emmott, S.; Phillips, A.; Werner, J. M.; Elliott, T., Selector function of MHC I molecules is determined by protein plasticity. Sci Rep 2015, 5, 14928.

27.   Lundgren, D. H.; Hwang, S. I.; Wu, L.; Han, D. K., Role of spectral counting in quantitative proteomics. Expert Rev Proteomics 2010, 7, (1), 39-53.

28.   Aebi, M.; Bernasconi, R.; Clerc, S.; Molinari, M., N-glycan structures: recognition and processing in the ER. Trends Biochem Sci 2010, 35, (2), 74-82.

29.   Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Darvill, A. G.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H., Essentials of Glycobiology. Essentials of Glycobiology 2015-2017, p Bookshelf ID: NBK310274.

30.   Moremen, K. W.; Tiemeyer, M.; Nairn, A. V., Vertebrate protein glycosylation: diversity, synthesis and function. Nat Rev Mol Cell Biol 2012, 13, (7), 448-62.

31.   Stanley, P., Golgi glycosylation. Cold Spring Harb Perspect Biol 2011, 3, (4), a005199.

32.   Demmers, L. C.; Heck, A. J. R.; Wu, W., Pre-fractionation Extends but also Creates a Bias in the Detectable HLA Class Iota Ligandome. J Proteome Res 2019, 18, (4), 1634-1643.

33.   Barnstable, C. J.; Bodmer, W. F.; Brown, G.; Galfre, G.; Milstein, C.; Williams, A. F.; 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.

34.   Mezzadra, R.; Sun, C.; Jae, L. T.; Gomez-Eerland, R.; de Vries, E.; Wu, W.; Logtenberg, M. E. W.; Slagter, M.; Rozeman, E. A.; Hofland, I.; Broeks, A.; Horlings, H. M.; Wessels, L. F. A.; Blank, C. U.; Xiao, Y.; Heck, A. J. R.; Borst, J.; Brummelkamp, T. R.; Schumacher, T. N. M., Identification of

CMTM6 and CMTM4 as PD-L1 protein regulators. Nature 2017, 549, (7670), 106-110.

35. Reiding, K. R.; Bondt, A.; Franc, V.; Heck, A. J. R., The benefits of hybrid fragmentation methods for glycoproteomics. Trac-Trends in Analytical Chemistry 2018, 108, 260-268.

36. Schwanhausser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M., Global quantification of mammalian gene expression control. Nature 2011, 473, (7347), 337-42.

37. Cox, J.; Hein, M. Y.; Luber, C. A.; Paron, I.; Nagaraj, N.; Mann, M., Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics 2014, 13, (9), 2513-26.

38. McKinney, W. In Data structures for statistical computing in python, Proceedings of the 9th Python in Science Conference, 2010; Austin, TX: 2010; pp 51-56.

39. Walt, S. v. d.; Colbert, S. C.; Varoquaux, G., The NumPy array: a structure for efficient numerical computation. Computing in Science & Engineering 2011, 13, (2), 22-30.

40. Hunter, J. D., Matplotlib: A 2D graphics environment. Computing in Science & Engineering 2007, 9, (3), 90-95.

41. Chong, C.; Marino, F.; Pak, H.; Racle, J.; Daniel, R. T.; Muller, M.; Gfeller, D.; Coukos, G.; Bassani-Sternberg, M., High-throughput and Sensitive Immunopeptidomics Platform Reveals Profound Interferongamma-Mediated Remodeling of the Human Leukocyte Antigen (HLA) Ligandome. Mol Cell Proteomics 2018, 17, (3), 533-548.

42. Bassani-Sternberg, M.; Pletscher-Frankild, S.; Jensen, L. J.; Mann, M., Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. Mol Cell Proteomics 2015, 14, (3), 658-73.

43. Petersen, J.; Wurzbacher, S. J.; Williamson, N. A.; Ramarathinam, S. H.; Reid, H. H.; Nair, A. K.; Zhao, A. Y.; Nastovska, R.; Rudge, G.; Rossjohn, J.; Purcell, A. W., Phosphorylated self-peptides alter human leukocyte antigen class I-restricted antigen presentation and generate tumor-specific epitopes. Proc Natl Acad Sci U S A 2009, 106, (8), 2776-81.

44. Dang, L.; Shen, J.; Zhao, T.; Zhao, F.; Jia, L.; Zhu, B.; Ma, C.; Chen, D.; Zhao, Y.; Sun, S., Recognition of Bisecting N-Glycans on Intact

Glycopeptides by Two Characteristic Ions in Tandem Mass Spectra. Anal Chem 2019, 91, (9), 5478-5482.

45. Wearsch, P. A.; Jakob, C. A.; Vallin, A.; Dwek, R. A.; Rudd, P. M.; Cresswell, P., Major histocompatibility complex class I molecules expressed with monoglucosylated N-linked glycans bind calreticulin independently of their assembly status. J Biol Chem 2004, 279, (24), 25112-21.

46. Culina, S.; Lauvau, G.; Gubler, B.; van Endert, P. M., Calreticulin promotes folding of functional human leukocyte antigen class I molecules in vitro. J Biol Chem 2004, 279, (52), 54210-5.

47. Armony, G.; Heck, A. J.; Wu, W., Extracellular crosslinking mass spectrometry reveals HLA class I–HLA class II interactions on the cell surface. Molecular Immunology 2021, 136, 16-25.

48. Lee, N.; Ishitani, A.; Geraghty, D. E., HLA-F is a surface marker on activated lymphocytes. Eur J Immunol 2010, 40, (8), 2308-18.

49. Kozlov, G.; Pocanschi, C. L.; Rosenauer, A.; Bastos-Aristizabal, S.; Gorelik, A.; Williams, D. B.; Gehring, K., Structural basis of carbohydrate recognition by calreticulin. J Biol Chem 2010, 285, (49), 38612-20.

50. Thomas, C.; Tampe, R., Proofreading of Peptide-MHC Complexes through Dynamic Multivalent Interactions. Front Immunol 2017, 8, 65.

51. Wearsch, P. A.; Peaper, D. R.; Cresswell, P., Essential glycan-dependent interactions optimize MHC class I peptide loading. Proc Natl Acad Sci U S A 2011, 108, (12), 4950-5.

52. Silva, Z.; Ferro, T.; Almeida, D.; Soares, H.; Ferreira, J. A.; Deschepper, F. M.; Hensbergen, P. J.; Pirro, M.; van Vliet, S. J.; Springer, S.; Videira, P. A., MHC Class I Stability is Modulated by Cell Surface Sialylation in Human Dendritic Cells. Pharmaceutics 2020, 12, (3), 249.

53. Morell, A. G.; Gregoriadis, G.; Scheinberg, I. H.; Hickman, J.; Ashwell, G., The role of sialic acid in determining the survival of glycoproteins in the circulation. J Biol Chem 1971, 246, (5), 1461-7.

54. Bork, K.; Horstkorte, R.; Weidemann, W., Increasing the sialylation of therapeutic glycoproteins: the potential of the sialic acid biosynthetic pathway. J Pharm Sci 2009, 98, (10), 3499-508.

55. Springer, S., Transport and quality control of MHC class I molecules in the early secretory pathway. Curr Opin Immunol 2015, 34, 83-90.

56.   Yarzabek, B.; Zaitouna, A. J.; Olson, E.; Silva, G. N.; Geng, J.; Geretz, A.; Thomas, R.; Krishnakumar, S.; Ramon, D. S.; Raghavan, M., Variations in HLA-B cell surface expression, half-life and extracellular antigen receptivity. Elife 2018, 7.

57.   Morozov, G. I.; Zhao, H.; Mage, M. G.; Boyd, L. F.; Jiang, J.; Dolan, M. A.; Venna, R.; Norcross, M. A.; McMurtrey, C. P.; Hildebrand, W.; Schuck, P.; Natarajan, K.; Margulies, D. H., Interaction of TAPBPR, a tapasin homolog, with MHC-I molecules promotes peptide editing. Proc Natl Acad Sci U S A 2016, 113, (8), E1006-15.

58.   Ryan, S. O.; Cobb, B. A., Roles for major histocompatibility complex glycosylation in immune function. Semin Immunopathol 2012, 34, (3), 425-41.

59.   Blais, M. E.; Dong, T.; Rowland-Jones, S., HLA-C as a mediator of natural killer and T-cell activation: spectator or key player? Immunology 2011, 133, (1), 1-7.

60.   Garcia-Beltran, W. F.; Holzemer, A.; Martrus, G.; Chung, A. W.; Pacheco, Y.; Simoneau, C. R.; Rucevic, M.; Lamothe-Molina, P. A.; Pertel, T.; Kim, T. E.; Dugan, H.; Alter, G.; Dechanet-Merville, J.; Jost, S.; Carrington, M.; Altfeld, M., Open conformers of HLA-F are high-affinity ligands of the activating NK-cell receptor KIR3DS1. Nat Immunol 2016, 17, (9), 1067-74.

61.   Abelin, J. G.; Keskin, D. B.; Sarkizova, S.; Hartigan, C. R.; Zhang, W.; Sidney, J.; Stevens, J.; Lane, W.; Zhang, G. L.; Eisenhaure, T. M.; Clauser, K. R.; Hacohen, N.; Rooney, M. S.; Carr, S. A.; Wu, C. J., Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction. Immunity 2017, 46, (2), 315-326.

62.   Lutz, C. T., Human leukocyte antigen Bw4 and Bw6 epitopes recognized by antibodies and natural killer cells. Curr Opin Organ Transplant 2014, 19, (4), 436-41.

63.   Van Rood, J. J.; Van Leeuwen, A., Leukocyte Grouping. A Method and Its Application. J Clin Invest 1963, 42, (9), 1382-90.

64.   Carr, W. H.; Pando, M. J.; Parham, P., KIR3DL1 polymorphisms that affect NK cell inhibition by HLA-Bw4 ligand. J Immunol 2005, 175, (8), 5222-9.

65.   Salzberger, W.; Garcia-Beltran, W. F.; Dugan, H.; Gubbala, S.; Simoneau, C.; Gressens, S. B.; Jost, S.; Altfeld, M., Influence of Glycosylation Inhibition on the Binding of KIR3DL1 to HLA-B*57:01. PLoS One 2015, 10, (12), e0145324.

66.   Freudenmann, L. K.; Marcu, A.; Stevanovic, S., Mapping the tumour human leukocyte antigen (HLA) ligandome by mass spectrometry. Immunology 2018, 154, (3), 331-345.

67.   Sturm, T.; Sautter, B.; Worner, T. P.; Stevanovic, S.; Rammensee, H. G.; Planz, O.; Heck, A. J. R.; Aebersold, R., Mild Acid Elution and MHC Immunoaffinity Chromatography Reveal Similar Albeit Not Identical Profiles of the HLA Class I Immunopeptidome. J Proteome Res 2021, 20, (1), 289-304.

68.   Vizcaíno, J. A.; Csordas, A.; Del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T., 2016 update of the PRIDE database and its related tools. Nucleic acids research 2016, 44, (D1), D447-D456.

Supplementary Figure 1
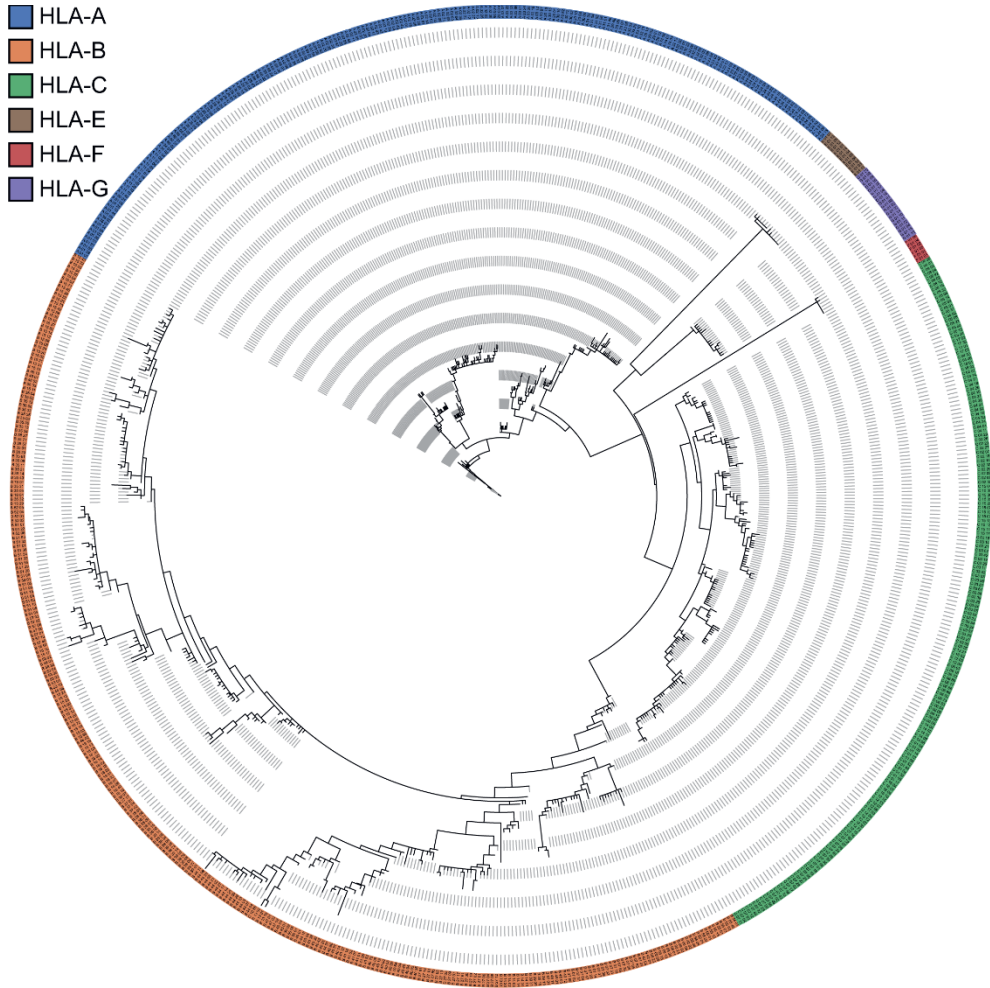
Figure S1: Phylogenetic tree visualization of the most observed HLA class I heavy chain allotypes in the European population. The allotypes were selected in order of abundance from http://allelefrequencies.net/. For multiple sequence alignment of the HLA Class I heavy chain molecules, only the α1 and α2 domains were used, because many allotypes are missing sequence information on the α3 domain. Additionally, the α3 domain does not contribute to the peptide-binding groove, thus sequence variation in this region is less relevant. As also shown in Figure 1B for a smaller number of sequences, the HLA subtypes cluster together even when including a large number of sequences.

| Gene	Line	N1	N2H2F1	N2H7	N2H8	N2H9	N2H10	N4H4F1	N4H5F1S1	N4H5F1S2	N4H5F2S2	N5H3F1	N5H5S2	N5H5F1	N5H5F1S1	N5H5F1S2	N5H5F2S1	N5H6S2	N5H6F1	N5H6F1S1	N5H6F1S2	N6H5F2S1	N6H6F1	N6H6F1S1	N6H6F1S2	N6H6F1S3	N6H7F1S1
A	DBB	3.6	2.3	1.4	2.4	2.7	1.7	2.1	4.2	3.9		0.8	2.4	4.3	6.3	3.4	2	1.3	2.9	4.6	5.3	1.5	4.6	3	3.6	0.7	1.6
A	DEM	4.3	1.4	0.6	2.3	3.3	0.9	0.7	2.7	5.4		1.2	1.4	3.1	8.4	9.8	3	2.7	0.4	4.7	6.7	3.6	0.3	5	5.6	1.3	1.4
A	JY	1.8		2.2	4	7.2	1.8			5.5		2.3	1.7		9.9	14.2	3.7	2.8		1.1	8.3	4.8		5.6	7.3	5.1	
B	DBB	1.4	5.1		1.2	2.6		1.1	8.3	10.1		0.7	8.2	14	8.6	1.1	1.2	1.7	5.4	4.2	0.6	4.6	4	1.4			2.6
B	DEM		1.4		1.9	3.9			2	12.7	2.3	1	3	8.3	23		3	3.7		3.8	9.9	3.9		5	1.6		1.1
B	JY														80.8			6			9						
C	DBB	3.4			16.5	32.7		2.4					13.3	21.9	5.7									4.1			
C	DEM	3.7		2.4	20.2	31.4								4.9	30.2	4.7					2.4						
C	JY			9.6	12.4	16.9	2.3								29.9	7					11.4	5.1		5.4			
F	DBB	3.5			38.9	57.6																					
F	DEM				32	67.9																					
F	JY					100																					

Figure S2: Abundance distribution of glycans for each HLA class I gene in each investigated cell line. Each cell shows a percentage value of the relative abundance for a glycan in a particular class I gene and cell line (across each row). The horizontal axis shows several observed glycan compositions. Only compositions with at least >2% relative abundance are shown, consequently each horizontal row does not add up to 100% exactly. The cells are colored according to the color bar on the right to visually distinguish higher abundant glycans. This figure shows in greater detail that the bisected glycan N5H5F1S(0-2) is the most abundant glycan on HLA-A and B for all cell lines, while on HLA-C and F the high-mannose glycans are more abundant. It also reveals in greater detail than Figure 2 that HLA-A has a more diverse glycosylation pattern than the other genes.

Supplementary Figure 3

A)
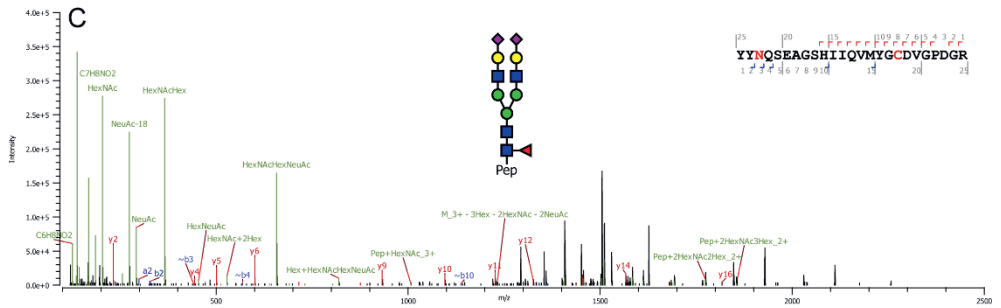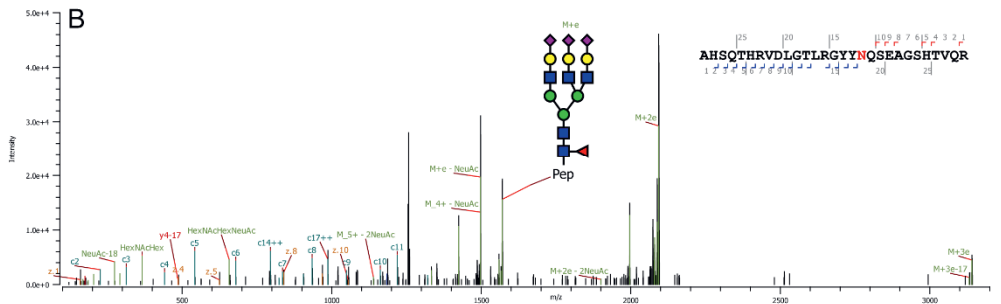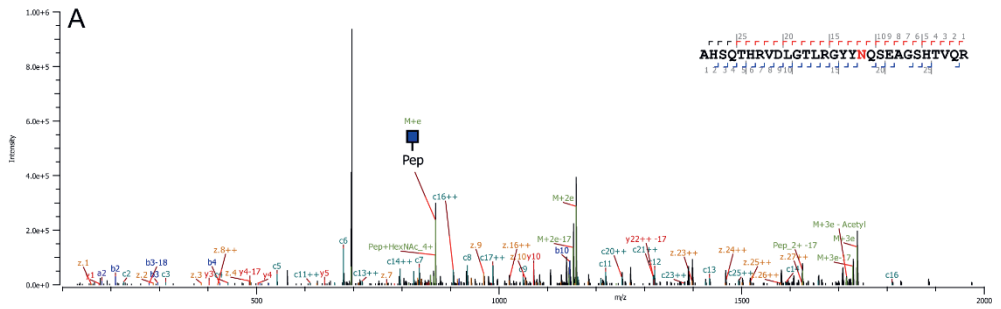
B)
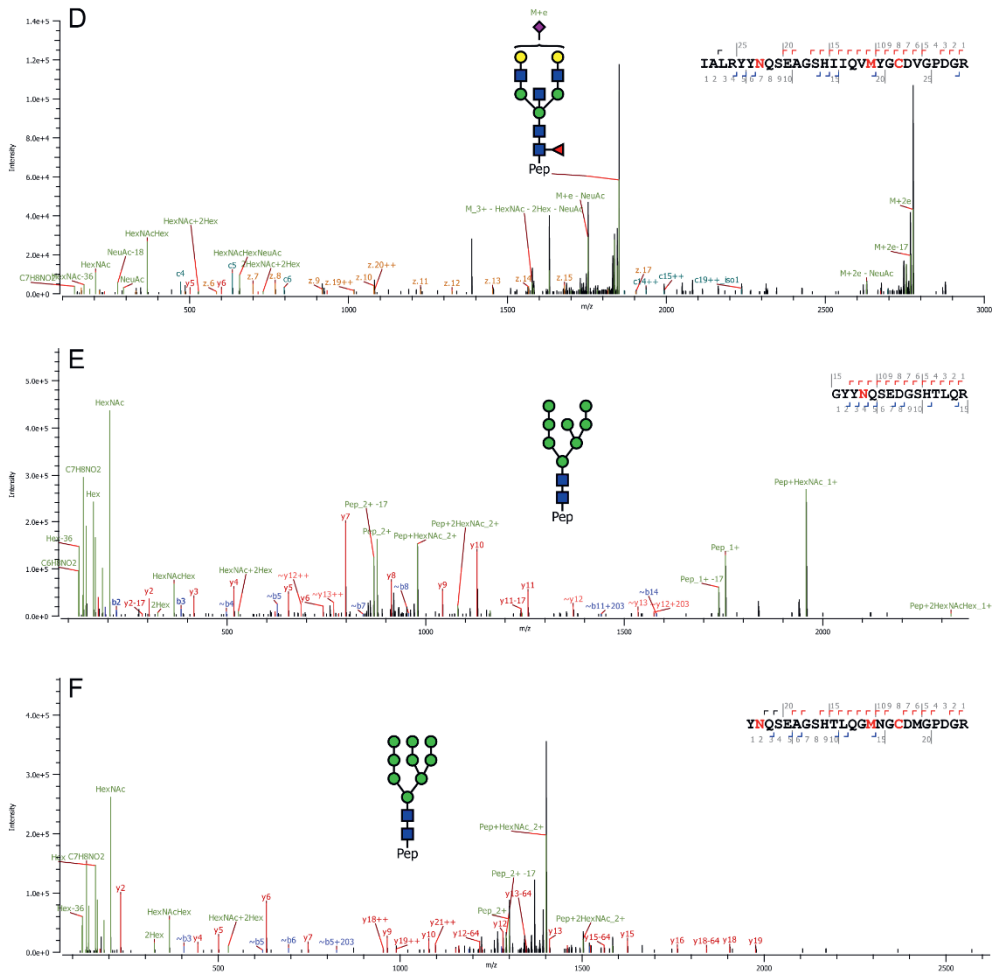
Figure S3: Additional examples of annotated MS/MS spectra of HLA heavy chain glycopeptides.
All spectra display peptide backbone and glycan fragments used for identification and annotation of the
glycopeptides. **A)** EThcD MS/MS spectrum of a HLA-A*02:01 glycopeptide, expressing a paucimannose
type glycan (JY cell line). **B)** EThcD MS/MS spectrum of a HLA-A*02:01 glycopeptide, expressing a
complex type extended glycan (JY cell line). **C)** Stepping energy HCD MS/MS spectrum of a HLA-B*57:01
derived diantennary glycopeptide (DEM cell line). Unlike the EThcD MS/MS spectra, in HCD and stepping
HCD MS/MS spectra the intact precursor ion (annotated as "M") is usually not observed. However, the
accompanying MS1 scan (not shown here) still provides the intact glycopeptide mass to support the
identification. **D)** EThcD MS/MS spectrum of a HLA-B*57:01 glycopeptide expressing a bisected glycan
(DBB cell line). The diagnostic fragment ion for bisection is not observed, but due to the confirmation of
other bisecting glycans as shown in figure 3A, a bisecting structure is assumed. **E)** HCD MS/MS spectrum
of a HLA-C*07:02 derived high mannose glycopeptide (JY cell line). **F)** Stepping HCD MS/MS spectrum
of a HLA-F glycopeptide, expressing a high mannose glycan (DEM cell line).

Supplementary Figure S4
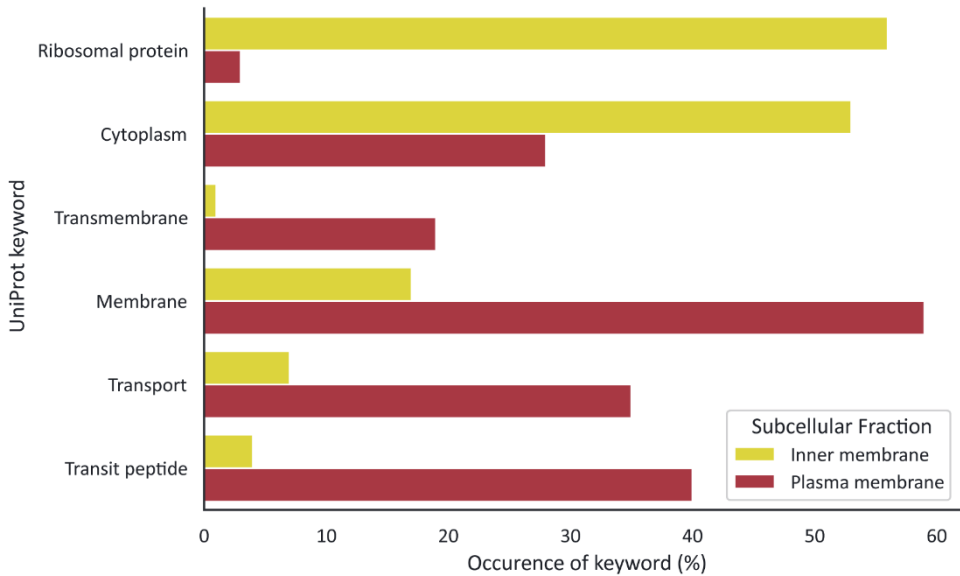
Figure S4: Proteomics analysis reveals the distinct protein keywords enriched in the Inner membrane and Plasma membrane fractions. For both the inner and plasma membrane fractions UniProt keywords for the top 100 most abundant proteins as determined by IBAQ were assessed. The occurrence of keywords related to the inner and plasma membrane compartments are shown as a percentage where a value of 100% would indicate that all top 100 proteins are assigned with the keyword.

Chapter 5

Single-cell derived tumor organoids display diversity in HLA class I peptide presentation

Laura C. Demmers[1,2], Kai Kretzschmar[3,4], Arne Van Hoeck[5], Yotam E. Bar-Epraïm[3,4], Henk W. P. van den Toorn[1,2], Mandy Koomen[3,4], Gijs van Son[3,4], Joost van Gorp[6], Apollo Pronk[7], Niels Smakman[7], Edwin Cuppen[5,8], Hans Clevers[3,4,9], Albert J. R. Heck [1,2] & Wei Wu [1,2]

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.
[2] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands.
[3] Oncode Institute, Hubrecht Institute, 3584 CT Utrecht, The Netherlands.
4 Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences and University Medical Centre Utrecht, 3584 CT Utrecht, The Netherlands.
[5] Center for Molecular Medicine and Oncode Institute, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands.
[6] Department of Pathology, St. Antonius Hospital, 3543 AZ Utrecht, The Netherlands.
[7] Department of Surgery, Diakonessenhuis Hospital, 3582 KE Utrecht, The Netherlands.
[8] Hartwig Medical Foundation, 1098 XH Amsterdam, The Netherlands.
[9] Princess Máxima Center for Pediatric Oncology, 3584 CS Utrecht, The Netherlands

Abstract

Tumor heterogeneity is a major cause of therapeutic resistance. Immunotherapy may exploit alternative vulnerabilities of drug-resistant cells, where tumor-specific human leukocyte antigen (HLA) peptide ligands are promising leads to invoke targeted anti-tumor responses. Here, we investigate the variability in HLA class I peptide presentation between different clonal cells of the same colorectal cancer patient, using an organoid system. While clone-specific differences in HLA peptide presentation were observed, broad inter-clone variability was even more prevalent (15-25%). By coupling organoid proteomics and HLA peptide ligandomics, we also found that tumor-specific ligands from DNA damage control and tumor suppressor source proteins were prominently presented by tumor cells, coinciding likely with the silencing of such cytoprotective functions. Collectively, these data illustrate the heterogeneous HLA peptide presentation landscape even within one individual, and hint that a multi-peptide vaccination approach against highly conserved tumor suppressors may be a viable option in patients with low tumor-mutational burden.

Introduction

Clonal diversity and tumor evolution are key drivers of cancer progression. As individual cancer cells within the same tumor tissue can divide and independently acquire beneficial genetic and epigenetic traits, the tumor becomes increasingly heterogeneous and diversely adapted to survive chemotoxic stress. As such, differential resistance to treatment is most often attributed to molecular signatures that are tumor clone specific (1, 2). Immunotherapy may then be administered in combination to eliminate residual disease (3).

Central to the concept of such a sequential or combinatorial therapeutic regime, is shifting the treatment focus away from the 'selection' of a chemo-resistant phenotype (4-6), towards exploiting unique immune vulnerabilities in drug-resistant tumor clones (7, 8).

Immunotherapy relies on the patient's own immune system to induce an anti-tumor response. Next to checkpoint blockade inhibitors, monoclonal antibodies and anti-tumor vaccines may be raised against tumor cells (9) or tumor-specific cell surface antigens (HLA class I peptide ligands), which arise as proteasome-generated byproducts from protein homeostasis (10). While HLA class I peptide ligand presentation mirrors the 'health-status' of each cell, the immense repertoire of tumor-surface peptide antigens has become a rich source of inspiration for rational design of highly personalized peptide vaccines (11-15).

One of the key questions that still needs to be addressed, is the extent of heterogeneity in HLA class I peptide ligand presentation, between tumor cells in the same environment or individual. This is important, because if substantial heterogeneity also exists at the level of tumor cell-surface ligand presentation, tumor cells that present less to the immune system could also evade from the efficacy of monoclonal antibodies and anti-tumor vaccines, just like drug-resistant cancer cells can evade from pharmacological inhibition.

In the growing field of immunopeptidomics, we and others have contributed to drastically improve HLA peptide ligand detection, through enhanced ligand isolation, preparation(16, 17), chromatographic separation(16, 18), the use of improved mass spectrometry based peptide fragmentation/sequencing strategies(18) and bioinformatics(19). Despite these notable advances in analytics, it remains difficult to probe for variation in HLA class I peptide ligand presentation in tumor clonal sub-populations, or even ideally in single tumor cells. The major bottleneck remains the combination of the extreme low-abundance in neo-antigens (20) and the lack of 'PCR-equivalents' for proteins and HLA peptides. Here, we amplified single-cell patient material into clonal organoids for deeper clonal proteome and HLA class I peptide ligand analyses, starting from a low mutational burden microsatellite-stable (MSS) colorectal cancer (CRC) patient harboring the HLA-A*02:01, HLA-B*15:01/57:01 and HLA-C*03:04/06:02 alleles. Organoids grown in controlled

culture are known to be genetically stable, and should retain the molecular signatures and surface marks of the originating cells(21-23), making the organoid technology in our investigation an ideal system to also amplify the protein and HLA class I peptide ligands.

Using this approach, we detected clear inter clonal proteome and HLA ligandome heterogeneity, detecting a large amount of HLA class I peptide ligands (about 7000) across four different colorectal cancer clones, that had been isolated concurrently from the same individual patient in vivo, and also amplified using the organoid technology under the same in vitro conditions. Comparing against peptide ligands presented by normal colon organoids, also from the same donor, about 300 HLA class I peptide ligands were presented exclusively by the tumor clones. We further identified a considerable number of clone-specific antigens and tumor-shared antigens originating from several notorious oncogenic proteins. Interestingly, these unique peptide ligand signatures were largely uncoupled from the trends observed at the proteome level within the same clones, exemplifying again current limits in predicting antigen presentation based on DNA, RNA or protein-level regulation.

To our knowledge, this is the first investigation on intra-patient clonal diversity in HLA class I peptide presentation. The findings we distill here should ultimately contribute towards improving therapeutic considerations and efficacy in personalized immunotherapy.

Materials and Methods

Human material and informed consent

The organoid lines used in this study were derived from biopsies (provided by Departments of Surgery and Pathology of the Diakonessenhuis hospital, Utrecht, The Netherlands) of a CRC and adjacent healthy colon mucosal epithelium taken from colon tissue resected during a left hemicolectomy to remove a CRC from a female patient (71 years of age). Tissue collection was approved by the medical ethical committee (METC) of the Diakonessenhuis hospital, in agreement with the declaration of Helsinki and according to Dutch and European Union legislation. The samples were collected under METC protocol 12/093 HUB-Cancer following written informed consent at the Diakonessenhuis hospital Utrecht.

Organoid generation and cultures

Colonic epithelial organoid lines were derived as previously described (24, 25). Briefly, crypts of the healthy portion of the colonic epithelium (at least 10cm away from the tumor side) were isolated by digestion of the intestinal mucosa in chelation solution (5.6 mM Na_2HPO_4, 8.0 mM KH_2PO_4, 96.2 mM NaCl, 1.6 mM KCl, 43.4 mM Sucrose, and 54.9 mM D-Sorbitol, Sigma) supplemented with dithiotreitol (0.5mM, Sigma) and EDTA (2mM, in-house) for 30 minutes at 4°C. Colonic crypts were subsequently plated in basement membrane extract (BME; Cultrex PC BME RGF type 2, Amsbio). Organoids were grown in human intestinal stem cell medium (HISCM), which is composed of Advanced Dulbecco's modified Eagle medium/F12 supplemented with penicillin/streptomycin, 10mM HEPES and Glutamax (all Gibco, Thermo Fischer Scientific) with 50% Wnt3a conditioned medium (in-house), 20% R-Spondin1 conditioned medium (in-house), 10% Noggin conditioned medium (in-house), 1x B27, 1.25mM N-acetyl cysteine, 10mM nicotinamide, 50ng/ml human EGF, 10nM Gastrin, 500nM A83-01, 3µM SB202190, 10nM prostaglandine E2 and 100µg/ml Primocin (Invivogen). Tumor biopsies were digested into single cells using collagenase II (1mg/ml, Gibco, Thermo Fischer Scientific), supplemented with hyaluronidase (10µg/ml) and LY27632 (10µM) for 30 minutes at 37°C while shaking. Dissociated tumor cells were plated in BME and organoids were cultured in HICS minus Wnt conditioned medium and supplemented with 10µM LY27632 at 37°C.

Clonal organoid derivation and amplification

To generate clonal tumor organoid lines, early passage tumor organoids were dissociated into single cells by TryPLE express (Thermo Fischer Scientific), washed and suspended in FACS buffer (PBS with 2mM ETDA and 5% FCS). Prior to FACS purification, DAPI was added to the FACS buffer. HLA-A2+ single cells were sorted into separate wells of 96-wells plates containing 100µl HISCM with 10µM LY27632 and coated with BME. Sorted cells were then covered with 10µl BME and placed in the incubator at 37°C. LY27632 was added to the HISCM for the first week after sorting. Clonal tumor organoids were then expanded in HICS minus WNT conditioned medium.

Whole genome sequencing and somatic analysis of clonal organoid lines

Organoids were dissociated and DNA was isolated using the QiaSymphony DSP DNA Mini Kit (Qiagen; 937236). Libraries were prepared using the Illumina TruSeq DNA Nano Library Prep Kit (20015964). Paired-end sequencing of the organoid lines was performed (2 × 150 bp) on the generated libraries with 30× coverage using the Illumina HiSeq X Ten sequencing system at the Hartwig Medical Foundation. Somatic mutations were analysed by the HMF somatic mutation workflow from https://github.com/hartwigmedical/pipeline which was installed the pipeline locally

using GNU Guix with the recipe from https://github.com/UMCUGenetics/guix-additions. Full pipeline description is explained elsewhere (26). Details and settings of all the tools can be found at their Github page. Briefly, sequence reads were mapped against human reference genome GRCh37 using Burrows-Wheeler Alignment (BWA-MEM) v0.7.5a. Subsequently, somatic single base substitutions (SBSs) and small insertions and deletions (INDELS) were determined by providing the genotype and tumor (or organoid for in-vitro analysis) sequencing data to Strelka v1.0.14 with adjustments as described in (26).

Organoid expansion and quantitative proteomics

Normal colon organoids and clonal tumor organoid lines were expanded in their respective media to about 5×10^8 cells per organoid line. In the last medium change 3 days prior to harvest for peptidome analysis, the clonal tumor organoids lines received standard HISCM. On the day of harvest, organoids were removed from the 6-well culture plates using medium and P1000 pipettes and spun down for 8 min at 500*g with the brakes off. Cell pellets were then incubated with Cell Recovery Solution (Roche Diagnostics) for 30min on ice to remove excess BME. Cells liberated from BME were then washed three times in excess PBS to remove any residual BME. After the last wash, all PBS was removed and the tube opening was quickly dried using a paper towel. Cell pellets were snap frozen in dry ice and stored at -80°C.

Organoids were lysed by gentle vortexing in 8M Urea in 50mM ammonium bicarbonate supplemented with 50µg/ml DNAse I (Sigma-Aldrich), 50µg/ml RNAse A (Sigma-Aldrich) and 1x complete EDTA-free protease inhibitor cocktail (Roche Diagnostics). Subsequently, the lysate was cleared by centrifugation for 1h at 18000*g at 15°C. Protein concentration was determined with the Bradford assay (Bio-Rad). For each sample, 20µg of total protein was reduced, alkylated and digested sequentially with Lys-C (1:100) and trypsin (1:75). For high-pH reversed phase fractionation, peptide digests were loaded on C18 STAGE-tips in 200mM ammonium formate at pH 10 and eluted into 5 fractions with 11-80% acetonitrile. All samples were dried by vacuum centrifugation and reconstituted in 10% formic acid prior to LC-MS/MS analyses. Per fraction, three technical replicates were measured by LC-MS/MS.

Immuno-affinity purification

Organoids were lysed as described in Demmers et al, 2019 (16). In short, pHLA complexes were immunoprecipitated (nine IP equivalents per organoid clone) using 0.7mg W6/32 antibody (27) coupled to 175µl protein A/G beads (Santa Cruz) from

35mg organoid lysate. Antibodies were cross-linked to protein A/G beads to prevent co-elution. Incubation took place for approximately 16h at 4°C. After immunoprecipitation, the beads were washed with 40ml cold PBS. pHLA complexes were subsequently eluted with 10% acetic acid. Peptide ligands were separated from HLA class I complexes using 10kD molecular weight cutoff filters (Millipore). The fraction containing HLA class I peptide ligands was dried by vacuum centrifugation and reconstituted in 10% formic acid prior to LC-MS/MS analyses. Per organoid clone, six technical replicates were measured of which three with EThcD fragmentation (Exp 1) and three with HCD fragmentation (Exp 2).

Proteome LC-MS/MS analyses

The data was acquired with an UHPLC 1290 system (Agilent) coupled to a Q-Exactive HF mass spectrometer (Thermo Fischer Scientific). Peptides were trapped (Dr Maisch Reprosil C18, 3µM, 2cm x 100µM) for 5min in solvent A (0.1% formic acid in water) before being separated on an analytical column (Agilent Poroshell, EC-C18, 2.7µM, 50cm x 75µM). Solvent B consisted of 0.1% formic acid in 80% acetonitrile. For high-pH reversed phase samples (fraction 1) the gradient was as follows: first 5min of trapping, followed by 85min of gradient from 12 to 30% solvent B and, subsequently, 10min of washing with 100% solvent B and 10min of re-equilibration with 100% solvent A. For fraction 2 the gradient was from 15 to 32% solvent B. For fraction 3 the gradient was from 18 to 36% solvent B. For fraction 4 the gradient was from 20 to 38% solvent B and for fraction 5 from 22 to 44% solvent B. The mass spectrometer operated in data-dependent mode. Full scan MS spectra from *m/z* 375 – 1600 were acquired at a resolution of 60.000 to a target value of $3x10^6$ or a maximum injection time of 20ms. The top 15 most intense precursors with a charge state of 2+ to 5+ were chosen for fragmentation. HCD fragmentation was performed at 27% normalized collision energy on selected precursors with 16s dynamic exclusion at a 1.4m/z isolation window after accumulation to $1x10^5$ ions or a maximum injection time of 50ms. Tandem mass spectrometry (MS/MS) spectra were acquired at a resolution of 15.000.

Ligandome LC-MS/MS analyses

The data was acquired with an UHPLC 1290 system (Agilent) coupled to an Orbitrap Fusion Lumos Tribrid (for EThcD fragmentation) mass spectrometer or a Q-Exactive HF-X (for HCD fragmentation) mass spectrometer (Thermo Fischer Scientific). Trapping and running conditions were similar as described above, with the exception of a 7 to 40% solvent B gradient. The mass spectrometer operated in data-dependent mode. Full scan MS spectra from *m/z* 400-650 were acquired at a resolution of 60.000 after accumulation to a target value of $4x10^5$ or a maximum

injection time of 50ms. Up to 3 (EThcD) or 15 (HCD) most intense precursors with a charge state of 2+ or 3+ starting at *m/z* 100 were chosen for fragmentation. EThcD fragmentation and HCD fragmentation were both performed at 35% normalized collision energy on selected precursors with 18s (EThcD) or 16s (HCD) dynamic exclusion after accumulation of $5x10^4$ ions or a maximum injection time of 250ms. Tandem mass spectrometry (MS/MS) spectra were acquired at a resolution of 15.000.

Proteome data analysis

Raw files were searched using MaxQuant version 1.5.3.30 and the Andromeda search engine against the human uniprot database (147854 entries, downloaded in January 2016). Enzyme specificity was set to trypsin and up to 2 missed cleavages were allowed. Cysteine carbamidomethylation was set as fixed modification. Methionine oxidation and N-terminal acetylation were set as variable modifications. The false discovery rate (FDR) was restricted to 1% in both protein and peptide identification. For quantitative comparisons, label-free quantification (LFQ) was performed with "match between runs" enabled. Data normalization, imputation and statistics were performed with Perseus version 1.6.2.2. The data was visualized with Graphpad PRISM 8.

Ligandome data analyses

Raw files were searched using Sequest HT in Proteome Discoverer 2.2 against the Swissprot human database (20258 entries, downloaded in Feb 2018) appended with the 20 most abundant FBS contaminants (28). The search was set to unspecific with a minimum precursor mass of 797 Da to a maximum precursor mass of 1950 Da corresponding to peptides between 8 and 12 amino acids long. Identified peptides were filtered to a 1% FDR using the percolator algorithm, 5% peptide FDR and Xcorr >1. Cysteine cysteinylation and methionine oxidation were set as variable modifications. From the identified peptides, FBS contaminants were removed. Binding affinity of HLA class I peptide ligand was predicted using the NetMHC 4.0 pan algorithm (29). The data was visualized with Graphpad PRISM 8.

Results

*Single-cell models of colorectal cancer by using organoid amplification*

Our first objective was to generate patient-derived organoid clones, from single tumor cells that retain heterogeneity and recapitulate the hallmarks of CRC (schematic, Figure 1A). From the same patient, four tumor clones were isolated and maintained in organoid culture, alongside a normal colon organoid line generated from tumor-free colon mucosal tissue biopsied from the same patient. As shown in Figure 1B, the tumor clones T1, T3, T4 and T5 are morphologically distinct from normal organoids of the same patient. Somatic mutation analysis of each CRC organoid clone, against the normal colon organoid line as germline reference, revealed that many exonic mutations are shared between the CRC clones (Figure 1C and Supplementary Figure 1). However, every CRC organoid clone also harbored unique mutations that recapitulate intratumor heterogeneity and reveal that the tumor clones are not genetically identical. DNA copy number analysis also revealed a conserved duplication of chromosome 7 and 8 and chromosomal arm 13q (Supplementary Figure 1), which are established characteristics for CRC (30).

Single-cell amplified CRC organoid proteomes reveal clonal heterogeneity

Next to somatic mutation analysis of each CRC organoid clone, we further compared the tumor clones by deep label-free quantitative proteomics, with abundance comparisons made across ~6,000 proteins. As shown in Figure 2A, proteome signatures of tumor clones were noticeably different from the paired normal organoid line. Cluster analysis in Figure 2A reveals proteome characteristics shared by all four tumor clones, which includes numerous proteins involved in chromosomal segregation (Figure 2B-C) and mTOR signaling (Figure 3A). Compared to the normal clone, a statistically significant decrease in numerous proteins that mediate mitotic spindle assembly, regulate spindle assembly checkpoint(31, 32) or form the chromosomal passenger complex (33) (schematic in Figure 2B) was observed. The protein level down-regulation of these key regulators of chromosomal integrity was found to be highly consistent across all four tumor organoid lines, with respect to the normal control (Figure 2C), for instance, the downregulation of Aurora B (up to 25-fold), Survivin (up to 45-fold) and INCENP (up to 78-fold). Phenotypic defects in chromosomal segregation in these CRC organoids were further corroborated by subsequent karyotyping analysis. Karyotype analysis of the end-stage organoid clones revealed aneuploidy in all four CRC clones, with >50% of cells analyzed having a chromosome number outside the normal range of 44-46 (Supplementary Figure 2). This is consistent with aneuploidy documented in the majority of solid tumors (34). The differential degree of aneuploidy between the tumor clones on the other hand also demonstrates diversity in chromosomal aberrations between the tumor clones, likely pre-existing at the point of isolation.
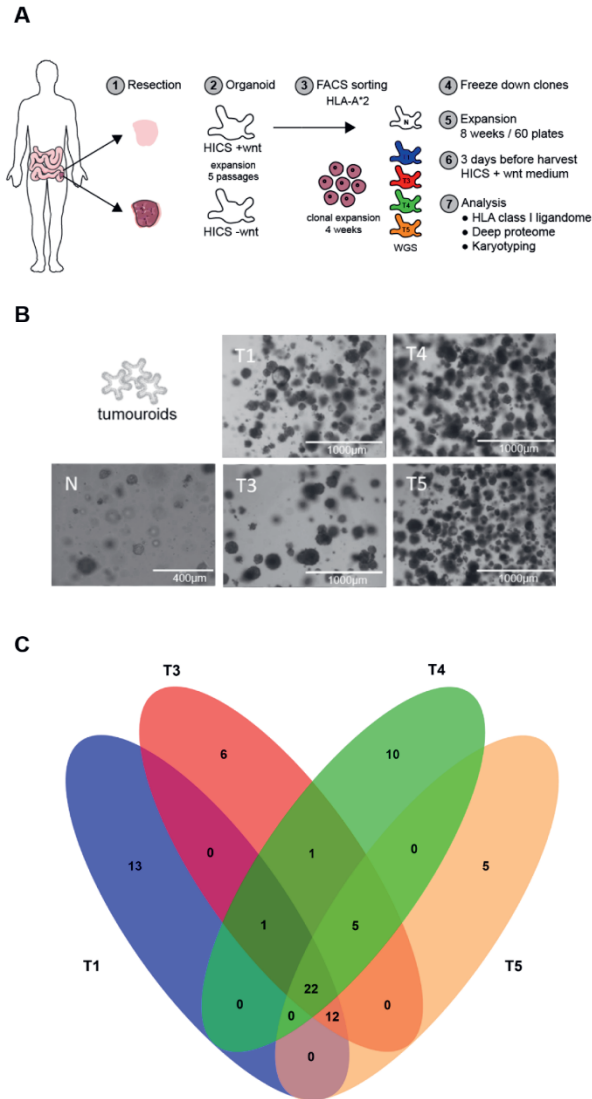
Figure 1: Preparation and characterization of single-cell derived clonal colorectal organoids. A) Organoids were made from colorectal cancer tissue and healthy tissue from the same CRC patient and grown in human intestinal stem cell medium (HICS) with or without additional wnt. After five passages, the cells were FACS sorted for HLA-A*2 positive cells and were then clonally expanded for four weeks. The clones were frozen and expanded again for approximately eight weeks until 60 plates. Three days before harvest, the tumor clones were cultured in HICS +wnt medium. **B)** Representative images of organoid morphology from 3 different passages. Four CRC tumor organoid lines (T1, T3, T4 and T5) were used together with a normal CRC organoid line (N). Magnification indicated by respective scale bars. **C)** Venn diagram showing the shared and unique somatic mutations (SBS and INDELs) of each CRC organoid line.
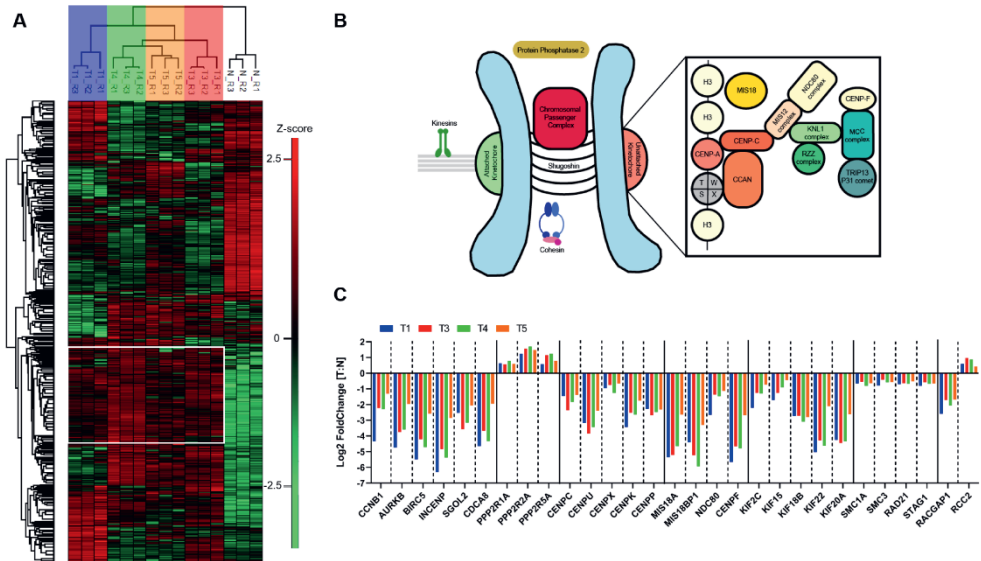
Figure 2: CRC organoids recapitulate tumor proteome characteristics and retain clonal heterogeneity. A) Clustering of ANOVA significant proteins (p<0.05). A conserved tumor proteome characteristic signature is indicated by the white box, whereas pockets of clonal heterogeneity are also observed in each single-cell derived tumor organoid line. **B)** Schematic representation of the kinetochore and associated protein complexes upon unattached microtubules. Numerous proteins in the kinetochore regulatory complex were significantly reduced in steady-state abundance. **C)** Proteome level fold change analysis revealed consistent loss of chromosomal regulatory proteins in every tumor clone compared to normal colon organoids. Plotting data tabulated in Source Data.

In addition, components in mTOR signaling also appeared to be consistently regulated in all tumor clones with respect to the normal organoid clone (Figure 3A). Uncontrolled mTOR signaling for proliferation is a well-documented functionality required to support CRC oncogenesis (35), and our proteome level observations reflect the known activation of mTOR pathway in CRC pathogenesis. Based on these evidence, the CRC tumor clones isolated exemplify the hallmarks of chromosomally instable CRC, and are good single-cell models to study intrinsic micro-heterogeneity in (i) proteome regulation and (ii) HLA class I peptide ligand presentation.

Despite substantial convergence, individual tumor clones were nonetheless not identical, as pockets of clone-specific proteome signatures were prominently detected. These clone-specific differences in general represent redundant hits in the same functional pathway that could modulate, replace or rescue one other. For instance, some clones (notably T1, T3 and T4) were hallmarked by the collective loss of DNA-level regulation through overall reduction in the capacity for methylation and acetylation turnover, but mosaic suppression on KMT2A/KMT2D/KDM2A and KAT5/KAT6A/KAT7/HDAC2 (Figure 3C). Substantial dysregulation of chromatin remodeling complexes and epigenetic regulation was found to be also a striking

signature for some of the clones (notably T3 and T4), and has also been documented as a hallmark of CRC and numerous other cancer types (36-39). Taken together, these findings demonstrate that although all four tumor clones were isolated from the same patient, substantial heterogeneity still exists between these clones, especially where there is functional redundancy. Since such heterogeneity still exists after in vitro organoid amplification in the same controlled environment, these differences are most likely to be intrinsic or imprinted in vivo, in agreement with prior report that clone-specific differences are largely preserved with organoid amplification (21).
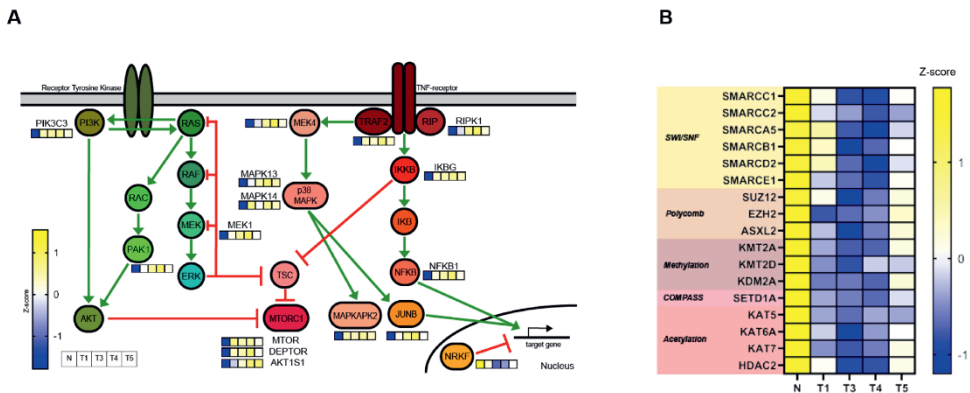
Figure 3: Proteome characteristics retained in clonal heterogeneity A) Schematic map of the mTOR pathway featuring proteins detected in the deep proteome analysis. Z-scored protein intensities are indicated in small squares. Collectively, the tumor organoid clones have higher steady-state abundance of mTOR proteins when compared to the normal organoid line. **B)** Heatmap of Z-scored protein abundances of chromatin remodeling complex components and various transferases, demethylases and deacetylases. Compensatory differences in protein abundance within the same functional processes illustrate heterogeneity between CRC organoid lines. Plotting data tabulated in Source Data.

HLA class I peptide ligand analysis from clonal organoids

To address if the observed heterogeneity also exists in antigen presentation between tumor organoid clones of the same individual, we isolated HLA class I peptide ligands from each tumor clone by immuno-affinity purification with a pan-HLA class I antibody (W6/32). Peptide LC-MS/MS analyses with complementary HCD and EThcD fragmentation modes (experiment 1 and 2 respectively) enabled high-sensitivity detection of a large number of peptide ligands (approximately 7000) from each tumor clone (Supplementary Figure 3A). In addition, at least 85% of all peptides identified were predicted to bind to the patient's HLA type (HLA-A*02:01, HLA-B*15:01/57:01, HLA-C*03:04/06:02) (Supplementary Figure 3B). The organoid-derived peptide ligands were also predominantly 9 amino acids long (Supplementary Figure 3C), and in good concordance with the theoretical HLA-type specific peptide

consensus motifs (Supplementary Figure 3D). Collectively, these data confirm the high quality and specificity in our ligandome sample preparation and data analysis.

Since the CRC clonal proteomes were easily markedly distinguishable from the normal organoids by various conserved signatures (Figure 2A), we hypothesized that corroborating differences in the ligands presented should be detectable, as HLA peptide ligands are known to originate as byproducts of protein turnover. To our surprise, only about 3% of all HLA peptide ligands detected were unique to each CRC tumor clone, and never presented by the normal organoids (Figure 4A). By ranking the peptide intensities (label-free quantification) and binding affinities of all ligands detected from each tumor clones, we found that clone-specific ligands were not necessarily always low abundant (Figure 4B, Supplementary Figure 4) or low in in peptide loading affinity to the patients HLA type (Figure 4C). Rather, these provide indications that abundant and high-affinity tumor-specific HLA class I peptide ligands could still be present in the small proportion of tumor-unique presented peptides, despite the low mutation load. In addition, we also observed a slight re-distribution in HLA class I peptide ligand length away from 9-mers (Figure 4D), potentially hinting at modulations in the HLA peptide trimming mechanism in the tumor cells.
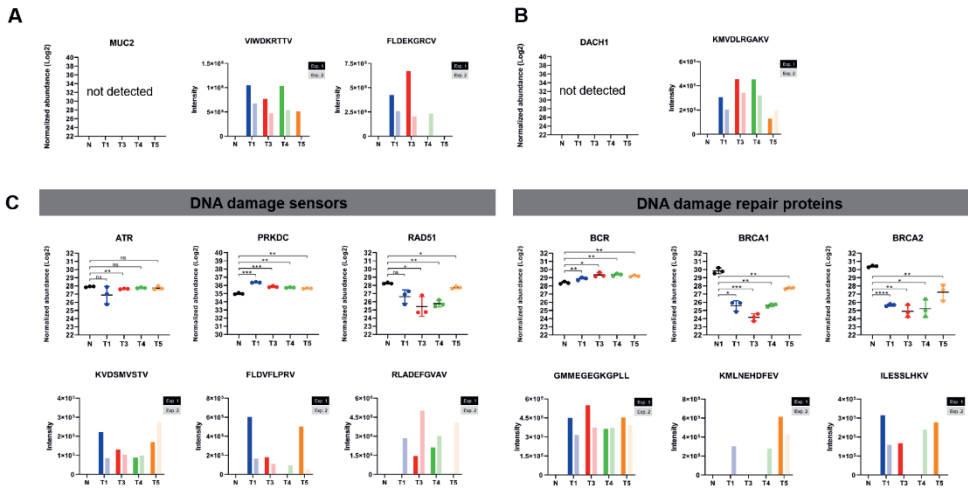
Clonal HLA class I ligandomes display over-representation of DNA repair source proteins

To rationalize the basis for tumor cells to present tumor-unique peptides, we looked into the source proteins of these ligands from each tumor clones, and compiled a master list of proteins that were reliably detected, in all six replicate mass spectrometry based ligandome measurements per clone. In this stringently curated list of 356 source proteins, we observed several CRC prognostic markers and an over-representation in proteins involved in DNA damage sensing and repair. We hypothesized that proteins which are disadvantageous for propagation are more likely to be degraded in the tumor clones, and that in turn more peptides from these proteins may be presented on the cell surface by HLA class I molecules. In concurrence with this notion, HLA peptide ligands from MUC2 and DACH1 were reliably detected from all the CRC tumor clones analyzed (Figure 5A-B), whereas we did not detect the proteins in the quantitative proteomics screen. This loss of MUC2 and DACH1 protein expression us not unique to our experimental model, but also broadly observed in CRC (40, 41, 42). Along similar lines, HLA peptide ligands from numerous DNA damage sensing proteins (ATR, PRKDC, RAD51) and repair proteins (BCR, BRCA1, BRCA2) were also detected only in CRC organoid lines (Figure 4C, Supplementary Figure 5). We believe these peptides are likely products of active degradation by tumor cells, to prevent the induction of DNA damage response and hamper with DNA repair. Together, this would also support the acquisition of further genome instability, a documented hallmark of cancer (43); since

DNA damage remains 'silent' despite severe chromosomal aberrations (Figure 2B-C).
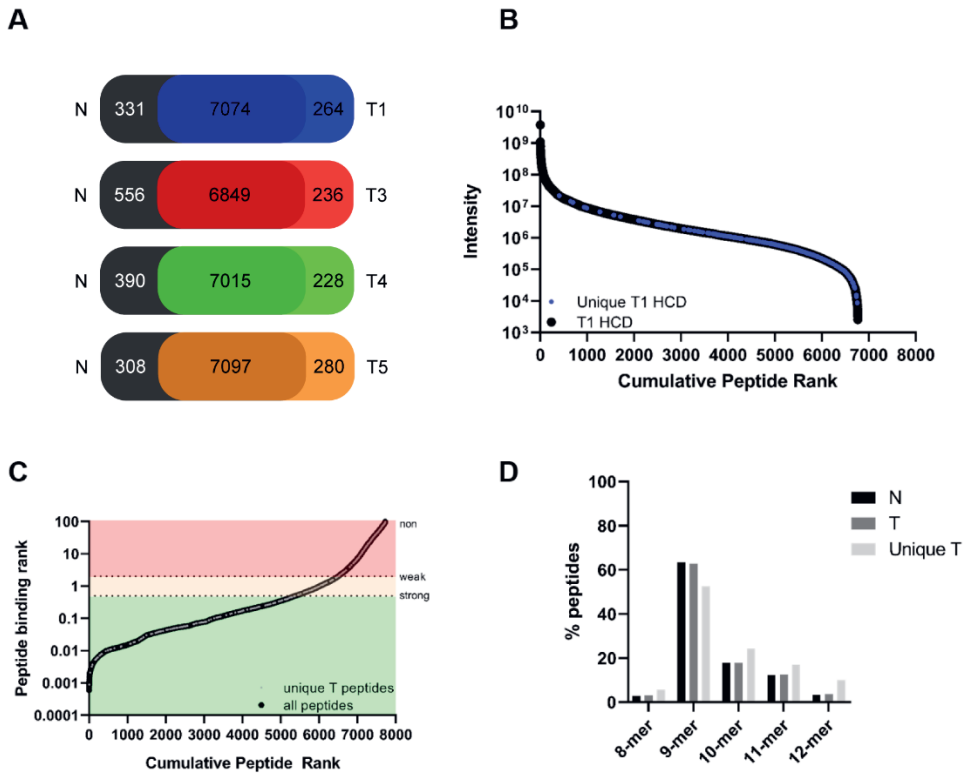
Figure 4: Tumor specific HLA class I peptide ligand characteristics. A) Identification overlap in unique HLA peptide ligands between normal CRC organoids and each tumor organoid line. In all comparisons, the majority of HLA peptide ligands (>6800) were presented by both normal colon organoids and tumor organoid clones. **B)** Peptide intensities plotted against the cumulative peptide rank. HLA peptide ligands identified from T1 were ranked by decreasing intensity. On the representative trace of all T1 – HCD peptides (black), peptides that are unique to T1 and never detected from normal colon organoids are annotated (blue). The spread of these T1 unique peptides over ranked intensity indicated that several T1 unique peptides are highly abundant. **C)** Peptide binding rank plotted against cumulative peptide rank. Peptide binding ranks were predicted using the NetMHC 4.0 pan algorithm. Peptides with a binding rank below 0.5 are considered strong binder. Peptides with a binding rank between 0.5 and 2.0 are considered weak binders and all peptides with a binding rank over 2.0 are considered non-binders. On the trace of all peptides identified (black), tumor specific peptides that are never detected from the normal colon organoids (in 6 MS measurements by 2 different fragmentation methods) are annotated (grey). Distribution of peptides presented uniquely by tumor organoid clones reveals many tumor specific ligands with high affinity. **D)** Length distribution of peptide ligands from normal (N), tumor (T) and unique tumor (T unique). Peptide ligands unique to CRC tumor organoids appear to be spread over a broader length distribution. Ligand characteristics for each individual tumor organoid clone are shown in Figure S1 and Figure S2. Plotting data tabulated in Source Data.

Figure 5: Clonal tumor HLA class I ligandomes over-represent peptides from DNA repair source proteins. A) The MUC2 proteins was not detected in the deep proteome analysis (left), but two HLA peptide ligands derived from MUC2 degradation were predominantly presented by all CRC tumor organoid clones (right). For each peptide, normalized peptide intensities were summed from three replicated measurements by HCD (Exp1), or EThcD (Exp2), respectively. **B)** The DACH1 protein was similarly not detected in the deep proteome analysis (left), but consistently presented on all tumor organoid clones, and consistently detected in both HCD and EThcD MS fragmentation methods. **C)** Peptides derived from degradation of DNA damage sensors and repair proteins were prominently presented by CRC organoid clones as HLA peptides. Significant changes in protein abundance with respect to normal colon organoids were determined based on three technical replicates with a two-sided student's t-test: p<0.05(*); p<0.01(**). Data presented as mean values ± standard deviation. Plotting data tabulated in Source Data.

To strengthen our hypothesis in a reciprocal manner, we also performed source protein analysis on ligands that were presented only by normal colon organoids (Supplementary Figure 6). ALDOA is a glycolytic enzyme observed to accumulate in CRC (44), COPS6 is a subunit of the COP9 signalosome specifically required to drive CRC (45-47), MCM2 functions as a DNA replication licensing factor needed to override once-per-cycle DNA replication in S-phase (48, 49) and SMARCA4 (BRG1) is a SWI/SNF component activating WNT and VEGF signaling to drive CRC (37, 50, 51). While HLA peptides from these four proteins were reliably detected on normal colon organoids, these were consistently absent from the ligandome of all four CRC organoid lines, suggesting that these proteins may be functionally important in CRC and therefore preserved from degradation (Supplementary Figure 6). Therefore, unique HLA peptide presentation on tumor organoids appears to be promoted by protein degradation events that favor oncogenesis and deter DNA damage remediation.

Clone-specific ligandome characteristics and quantitative variations

To further investigate ligandome heterogeneity between tumor clones, we focused on unique presentation by the CRC organoid clone T5. By plotting the overlap between tumor-specific HLA peptide ligands (from Figure 4A), we observed that only T5 presented 17 clone-specific ligands that are not shared with the other three tumor organoid lines (Figure 6A). HLA peptide ligands from MST1R, TP53 and TRAF2 were clearly differentially presented by T5, although proteome level trends of these proteins in all tumor clones were rather consistent with each other (Figure 6B). To rule out differences in the peptide presentation pathway between the tumor clones, we also verified that components of the HLA presentation pathway (10) are not prominently regulated in T5 compared to the other tumor clones (Supplementary Figure 7). Therefore, unique degradation and presentation of these functionally relevant proteins hints at differential signaling and functional regulation specifically in T5, but not in the HLA processing machinery per se.
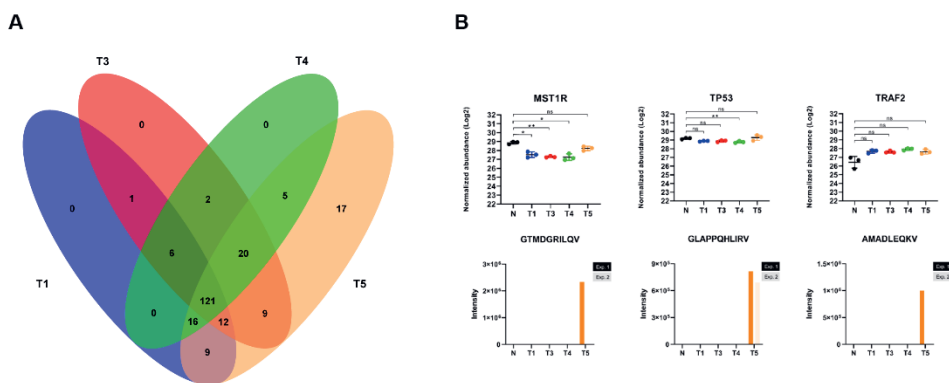

Figure 6: Clone specific ligandome characteristics. A) Overlap of all predicted HLA ligands unique to at least one CRC organoid clone. Ligands were detected by either HCD (Exp 1) or EThcD (Exp 2) fragmentation. Most ligands not detected on normal colon organoids were shared by at least three tumor clones, except for 17 ligands unique to T5 only. **B)** Protein abundance and peptide ligand abundance for MST1R, TP53 and TRAF2 from all organoid lines. Even though protein levels of MST1R, TRAF2 and TP53 were not significantly different between the four CRC tumor organoid lines, presentation of peptide ligands appear to be specific to only T5. Significant changes in protein abundance were determined with respect to normal colon organoids based on three technical replicates with a two-sided student's t-test: $p < 0.05$(*); $p < 0.01$(**). Data presented as mean values ± standard deviation. Plotting data tabulated in Source Data.

Even though the HLA class I ligandome of the tumor clones consisted of largely shared peptide sequences (97% between CRC clones) significant quantitative variations in peptide abundance could still be observed. By pairwise comparison of HLA peptide intensities between different tumor clones, we observed that up to 15-25% of all the peptides detected could vary in intensity by more than 2-fold (Figure

7). For instance, a SLINVGLISV peptide was 400 times more abundantly presented on T3 compared to T1. Therefore, this implies that one CRC organoid clone may present an HLA peptide ligand strongly, whereas another clone only weakly. We further verified that these intensity differences were not of technical nature, given that biological replicates of HLA ligandomes measured months apart did not vary by more than 1%. We believe such clone-to-clone variation in HLA peptide ligand presentation should be critically considered, especially in targeting residual disease after immunotherapy.
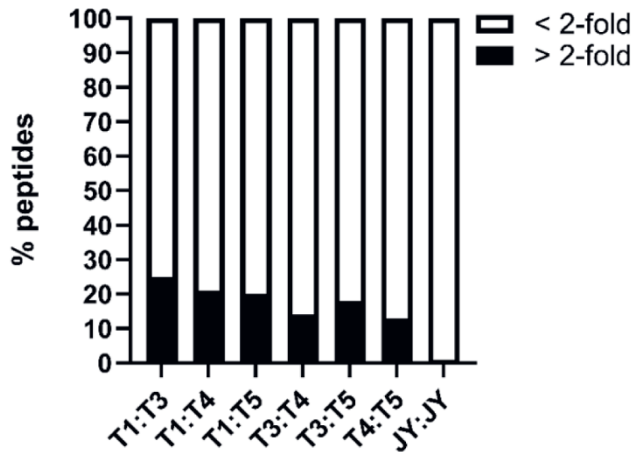
Figure 7: Clone-specific variations in peptide ligand presentation. Peptide ligand intensities from each tumor organoid clone were compared by retention time alignment and label-free quantification in a pair-wise manner. Between 15-25% of all ligands varied in intensity by more than 2-fold in each pairwise comparison (black). Quantitative variations by more than 2-fold was extremely few at <1%, between two independent preparations of JY cell line peptide ligands analyzed months apart with the same LC-MS settings, validating these quantitative ligand variations observed between tumor organoid clones were not of technical nature. Plotting data tabulated in Source Data.

Discussion

HLA class I peptide ligands presented abundantly, exclusively and uniformly on the tumor surface provide ideal starting points to design personalized immunotherapy. In this work, we modeled single-cell level heterogeneity in CRC from single cells using patient-derived clonal organoids, and correlated clonal tumor proteomes with their respective HLA ligandomes.

The strategy presented here has various advantages. The three-dimensional spatial signaling in organoid culture recapitulates the gut (patho-)physiology more accurately than flat in vitro expansions. Parallel analysis of normal colon organoids from the same patient allowed us to directly assess specificity of proteome and ligandome signatures to CRC cells, in largely the same patient genetic background. In the background of low mutation-load, we could subsequently survey the proteome and ligandome heterogeneity between individual clones largely without considering clone-specific genetic and protein-coding changes, to focus on studying the fundamental logic in cell-surface presentation. Since the proteome and ligandome are paired to each clone, steady-state protein abundance and HLA class I peptide ligand presentation could easily be correlated. Amplifying single cells using organoid technology also enables clone-specific proteome and ligandome signatures to be detected without signal averaging, unlike most of the tissue-based analyses reported to date. We believe our approach also adds sensitivity to identify a ligandome much larger than previously possible from patient-derived biopsies (52, 53).

Comparing between clonal proteomes of organoid-expanded tumor cells against normal colon organoids of the same patient, we observed that established cancer hallmarks were well-retained in every tumor organoid line, whereas these specific features were not present and also not acquired by the normal colon organoids during expansion, further verifying the validity of our experimental model. Whilst functional ontology of many proteome differences was conserved between tumor clones (e.g. mTOR), heterogeneity still exist in redundant pathways of epigenetic regulation and chromatin remodeling (e.g. KMTs and KATs). These observations also mirror the selective pressure *in vivo* that evolves single cancer cells via multiple routes towards the same signaling outcome (54).

More intriguingly, despite clear clustering of CRC organoid proteomes away from normal colon organoids, we observed that these protein level changes are qualitatively buffered at the level of HLA peptide ligand presentation, such that tumor clones share up to 97% of ligandome overlap even with normal colon organoids. This enlightens on the perennial challenges in neo-antigen discovery (55, 56), where even in the context of a mutation-rich cancer, a large proportion of the tumor surface HLA molecules is still occupied by peptides derived from routine protein turnover. More recently, neo-antigen discovery from non-coding regions have been attempted involving proteogenomics approaches and large computational efforts (57, 58). While these pipelines are streamlined to pick up the rare tumor-specific mutated antigen events, these approaches do not focus enough, during data analysis, on wild type ligands that are still over-presented from altered proteasome degradation and functional processing.

We propose, on the other hand, that tumor-specific HLA peptide ligand presentation could also arise as a byproduct, from the rational need to degrade tumor suppressors and proteins needed for DNA damage sensing and repair. Evidence supporting this

'Achilles heel' is presented here in the paired whole-proteome and HLA class I peptide ligand analyses from multiple tumor clones, and is also logically concordant with the need to accumulate chromosomal lesions, which we observed in karyotypes. Even within the small tumor-specific ligand repertoire, and given the normally low abundance of BRCA proteins, we could still pick up tumor specific BRCA peptides consistently across all the CRC clones analyzed, suggesting that this is likely reflecting an important molecular alteration in CRC. Since suppression of DNA damage sensing and repair is a common trait of many cancers, we believe the findings and rationale we distill here could also have strong impact on a variety of cancer and a range of immuno-therapeutic routes. For instance, potentially also in breast cancer with HER2 amplification but rapid turnover of HER2 by degradation (59), where targeting HLA peptides derived from HER2 degradation may also be logical.

From a therapeutic point-of-view, HLA peptides with high tumor specificity and homogenously high tumor surface loading are best candidates for further testing. While intense efforts have been channeled into predicting (57) and detecting mutated and spliced ligands to boost tumor specificity (19, 52, 60, 61), relatively little has been studied regarding the presentation heterogeneity between clonal tumor populations. We believe the latter would be a strong determinant of therapeutic efficacy and residual disease, drawing upon lessons learned from chemo-resistance. We show here that by using single-cell amplified organoids from the same patient, that clone-specific ligandome signatures exist, and quantitative variations in clonal presentation are prevalent. In this respect, immunization with multiple peptides highly conserved in presentation, for instance BRCA peptides, may minimize the risk of immune escape.

5

Acknowledgements

the pan-HLA antibody W6/32, and thank the USEQ for sequencing support and the UBEC for bioinformatical support.

Supporting Information

The Supporting Information is available free of charge on the Nature Publications website at DOI: https://doi.org/10.1038/s41467-020-19142-9

Data availability

The mass spectrometry proteomics and peptidomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (62) partner repository with the data set identifier PXD016582. The sequencing data of the CRC organoid lines have been deposited at the European Genome-phenome Archive (https://www.ebi.ac.uk/ega/studies/) under accession number EGAS00001003366.

Author Contributions

L.C.D., H.C., A.J.R.H. and W.W conceptualized the project, designed the experiments, and wrote the manuscript. L.C.D. and W.W. designed and performed the proteome and HLA ligandome experiments and data analysis. H.W.P.vd.T. assisted in bioinformatics analysis. A.V.H and E.C. performed the genomic analysis on the CRC organoids. Under the supervision of H.C., K.K. and Y.E.B.E. generated and characterized the tumor and normal CRC organoids with assistance of M.K. and G.v.S. and. J.v.G., A.P. and N.S. provided primary CRC patient material.

Competing interests

H.C., K.K. and Y.E.B.E. are named inventors on several patents or patent applications related to organoid technology. All other authors declare no conflicts of interest.

References

1. Dagogo-Jack, I.; Shaw, A. T., Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol 2018, 15, (2), 81-94.

2. McGranahan, N.; Swanton, C., Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell 2017, 168, (4), 613-628.

3. Qiao, J.; Liu, Z.; Fu, Y. X., Adapting conventional cancer treatment for immunotherapy. J Mol Med (Berl) 2016, 94, (5), 489-95.

4. Darragh, L. B.; Oweida, A. J.; Karam, S. D., Overcoming Resistance to Combination Radiation-Immunotherapy: A Focus on Contributing Pathways Within the Tumor Microenvironment. Front Immunol 2018, 9, 3154.

5. Clarke, P. A.; Roe, T.; Swabey, K.; Hobbs, S. M.; McAndrew, C.; Tomlin, K.; Westwood, I.; Burke, R.; van Montfort, R.; Workman, P., Dissecting mechanisms of resistance to targeted drug combination therapy in human colorectal cancer. Oncogene 2019, 38, (25), 5076-5090.

6. Van der Jeught, K.; Xu, H. C.; Li, Y. J.; Lu, X. B.; Ji, G., Drug resistance and new therapies in colorectal cancer. World J Gastroenterol 2018, 24, (34), 3834-3848.

7. Pech, M. F.; Fong, L. E.; Villalta, J. E.; Chan, L. J.; Kharbanda, S.; O'Brien, J. J.; McAllister, F. E.; Firestone, A. J.; Jan, C. H.; Settleman, J., Systematic identification of cancer cell vulnerabilities to natural killer cell-mediated immune surveillance. Elife 2019, 8.

8. Hammond, W. A.; Swaika, A.; Mody, K., Pharmacologic resistance in colorectal cancer: a review. Ther Adv Med Oncol 2016, 8, (1), 57-84.

9. Marin-Acevedo, J. A.; Dholaria, B.; Soyano, A. E.; Knutson, K. L.; Chumsri, S.; Lou, Y., Next generation of immune checkpoint therapy in cancer: new developments and challenges. J Hematol Oncol 2018, 11, (1), 39.

10. Rock, K. L.; Reits, E.; Neefjes, J., Present Yourself! By MHC Class I and MHC Class II Molecules. Trends Immunol 2016, 37, (11), 724-737.

11. Slingluff, C. L., Jr., The present and future of peptide vaccines for cancer: single or multiple, long or short, alone or in combination? Cancer J 2011, 17, (5), 343-50.

12. Sakamoto, S.; Matsueda, S.; Takamori, S.; Toh, U.; Noguchi, M.; Yutani, S.; Yamada, A.; Shichijo, S.; Yamada, T.; Suekane, S.; Kawano, K.; Naitou, M.; Sasada, T.; Hattori, N.; Kohno, N.; Itoh, K., Immunological evaluation of

peptide vaccination for cancer patients with the HLA -A11(+) or -A33(+) allele. Cancer Sci 2017, 108, (4), 598-603.

13. Wei, X.; Chen, F.; Xin, K.; Wang, Q.; Yu, L.; Liu, B.; Liu, Q., Cancer-Testis Antigen Peptide Vaccine for Cancer Immunotherapy: Progress and Prospects. Transl Oncol 2019, 12, (5), 733-738.

14. Mahdavi, M.; Moreau, V., In silico designing breast cancer peptide vaccine for binding to MHC class I and II: A molecular docking study. Comput Biol Chem 2016, 65, 110-116.

15. Schappert, A.; Schneck, J. P.; Suarez, L.; Oelke, M.; Schutz, C., Soluble MHC class I complexes for targeted immunotherapy. Life Sci 2018, 209, 255-258.

16. Demmers, L. C.; Heck, A. J. R.; Wu, W., Pre-fractionation Extends but also Creates a Bias in the Detectable HLA Class Iota Ligandome. J Proteome Res 2019, 18, (4), 1634-1643.

17. Purcell, A. W.; Ramarathinam, S. H.; Ternette, N., Mass spectrometry-based identification of MHC-bound peptides for immunopeptidomics. Nat Protoc 2019, 14, (6), 1687-1707.

18. Mommen, G. P.; Frese, C. K.; Meiring, H. D.; van Gaans-van den Brink, J.; de Jong, A. P.; van Els, C. A.; Heck, A. J., Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc Natl Acad Sci U S A 2014, 111, (12), 4507-12.

19. Liepe, J.; Sidney, J.; Lorenz, F. K. M.; Sette, A.; Mishto, M., Mapping the MHC Class I-Spliced Immunopeptidome of Cancer Cells. Cancer Immunol Res 2019, 7, (1), 62-76.

20. Schumacher, T. N.; Schreiber, R. D., Neoantigens in cancer immunotherapy. Science 2015, 348, (6230), 69-74.

21. Huch, M.; Gehart, H.; van Boxtel, R.; Hamer, K.; Blokzijl, F.; Verstegen, M. M.; Ellis, E.; van Wenum, M.; Fuchs, S. A.; de Ligt, J.; van de Wetering, M.; Sasaki, N.; Boers, S. J.; Kemperman, H.; de Jonge, J.; Ijzermans, J. N.; Nieuwenhuis, E. E.; Hoekstra, R.; Strom, S.; Vries, R. R.; van der Laan, L. J.; Cuppen, E.; Clevers, H., Long-term culture of genome-stable bipotent stem cells from adult human liver. Cell 2015, 160, (1-2), 299-312.

22. Roerink, S. F.; Sasaki, N.; Lee-Six, H.; Young, M. D.; Alexandrov, L. B.; Behjati, S.; Mitchell, T. J.; Grossmann, S.; Lightfoot, H.; Egan, D. A.; Pronk, A.; Smakman, N.; van Gorp, J.; Anderson, E.; Gamble, S. J.; Alder, C.; van de Wetering, M.; Campbell, P. J.; Stratton, M. R.; Clevers, H., Intra-tumour

diversification in colorectal cancer at the single-cell level. Nature 2018, 556, (7702), 457-462.

23. Kopper, O.; de Witte, C. J.; Lohmussaar, K.; Valle-Inclan, J. E.; Hami, N.; Kester, L.; Balgobind, A. V.; Korving, J.; Proost, N.; Begthel, H.; van Wijk, L. M.; Revilla, S. A.; Theeuwsen, R.; van de Ven, M.; van Roosmalen, M. J.; Ponsioen, B.; Ho, V. W. H.; Neel, B. G.; Bosse, T.; Gaarenstroom, K. N.; Vrieling, H.; Vreeswijk, M. P. G.; van Diest, P. J.; Witteveen, P. O.; Jonges, T.; Bos, J. L.; van Oudenaarden, A.; Zweemer, R. P.; Snippert, H. J. G.; Kloosterman, W. P.; Clevers, H., An organoid platform for ovarian cancer captures intra- and interpatient heterogeneity. Nat Med 2019, 25, (5), 838-849.

24. Sato, T.; Stange, D. E.; Ferrante, M.; Vries, R. G.; Van Es, J. H.; Van den Brink, S.; Van Houdt, W. J.; Pronk, A.; Van Gorp, J.; Siersema, P. D.; Clevers, H., Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. Gastroenterology 2011, 141, (5), 1762-72.

25. van de Wetering, M.; Francies, H. E.; Francis, J. M.; Bounova, G.; Iorio, F.; Pronk, A.; van Houdt, W.; van Gorp, J.; Taylor-Weiner, A.; Kester, L.; McLaren-Douglas, A.; Blokker, J.; Jaksani, S.; Bartfeld, S.; Volckman, R.; van Sluis, P.; Li, V. S.; Seepo, S.; Sekhar Pedamallu, C.; Cibulskis, K.; Carter, S. L.; McKenna, A.; Lawrence, M. S.; Lichtenstein, L.; Stewart, C.; Koster, J.; Versteeg, R.; van Oudenaarden, A.; Saez-Rodriguez, J.; Vries, R. G.; Getz, G.; Wessels, L.; Stratton, M. R.; McDermott, U.; Meyerson, M.; Garnett, M. J.; Clevers, H., Prospective derivation of a living organoid biobank of colorectal cancer patients. Cell 2015, 161, (4), 933-45.

26. Priestley, P.; Baber, J.; Lolkema, M. P.; Steeghs, N.; de Bruijn, E.; Shale, C.; Duyvesteyn, K.; Haidari, S.; van Hoeck, A.; Onstenk, W.; Roepman, P.; Voda, M.; Bloemendal, H. J.; Tjan-Heijnen, V. C. G.; van Herpen, C. M. L.; Labots, M.; Witteveen, P. O.; Smit, E. F.; Sleijfer, S.; Voest, E. E.; Cuppen, E., Pan-cancer whole-genome analyses of metastatic solid tumours. Nature 2019, 575, (7781), 210-216.

27. Barnstable, C. J.; Bodmer, W. F.; Brown, G.; Galfre, G.; Milstein, C.; Williams, A. F.; 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.

28. Zheng, X.; Baker, H.; Hancock, W. S.; Fawaz, F.; McCaman, M.; Pungor, E., Jr., Proteomic analysis for the assessment of different lots of fetal bovine serum as a raw material for cell culture. Part IV. Application of proteomics to

the manufacture of biological drugs. Biotechnol Prog 2006, 22, (5), 1294-300.

29.   Andreatta, M.; Nielsen, M., Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics 2016, 32, (4), 511-7.

30.   Taylor, A. M.; Shih, J.; Ha, G.; Gao, G. F.; Zhang, X.; Berger, A. C.; Schumacher, S. E.; Wang, C.; Hu, H.; Liu, J.; Lazar, A. J.; Cancer Genome Atlas Research, N.; Cherniack, A. D.; Beroukhim, R.; Meyerson, M., Genomic and Functional Approaches to Understanding Cancer Aneuploidy. Cancer Cell 2018, 33, (4), 676-689 e3.

31.   Musacchio, A.; Salmon, E. D., The spindle-assembly checkpoint in space and time. Nat Rev Mol Cell Biol 2007, 8, (5), 379-93.

32.   Makrantoni, V.; Marston, A. L., Cohesin and chromosome segregation. Curr Biol 2018, 28, (12), R688-R693.

33.   Carmena, M.; Wheelock, M.; Funabiki, H.; Earnshaw, W. C., The chromosomal passenger complex (CPC): from easy rider to the godfather of mitosis. Nat Rev Mol Cell Biol 2012, 13, (12), 789-803.

34.   Duijf, P. H.; Schultz, N.; Benezra, R., Cancer cells preferentially lose small chromosomes. Int J Cancer 2013, 132, (10), 2316-26.

35.   Saxton, R. A.; Sabatini, D. M., mTOR Signaling in Growth, Metabolism, and Disease. Cell 2017, 168, (6), 960-976.

36.   Nair, S. S.; Kumar, R., Chromatin remodeling in cancer: a gateway to regulate gene transcription. Mol Oncol 2012, 6, (6), 611-9.

37.   Lin, S.; Jiang, T.; Ye, L.; Han, Z.; Liu, Y.; Liu, C.; Yuan, C.; Zhao, S.; Chen, J.; Wang, J.; Tang, H.; Lu, S.; Yang, L.; Wang, X.; Yan, D.; Peng, Z.; Fan, J., The chromatin-remodeling enzyme BRG1 promotes colon cancer progression via positive regulation of WNT3A. Oncotarget 2016, 7, (52), 86051-86063.

38.   Kumar, R.; Li, D. Q.; Muller, S.; Knapp, S., Epigenomic regulation of oncogenesis by chromatin remodeling. Oncogene 2016, 35, (34), 4423-36.

39.   Kadoch, C.; Crabtree, G. R., Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. Sci Adv 2015, 1, (5), e1500447.

40. Hsu, H. P.; Lai, M. D.; Lee, J. C.; Yen, M. C.; Weng, T. Y.; Chen, W. C.; Fang, J. H.; Chen, Y. L., Mucin 2 silencing promotes colon cancer metastasis through interleukin-6 signaling. Sci Rep 2017, 7, (1), 5823.

41. Wang, P., Suppression of DACH1 promotes migration and invasion of colorectal cancer via activating TGF-beta-mediated epithelial-mesenchymal transition. Biochem Biophys Res Commun 2015, 460, (2), 314-9.

42. Yan, W.; Wu, K.; Herman, J. G.; Brock, M. V.; Fuks, F.; Yang, L.; Zhu, H.; Li, Y.; Yang, Y.; Guo, M., Epigenetic regulation of DACH1, a novel Wnt signaling component in colorectal cancer. Epigenetics 2013, 8, (12), 1373-83.

43. Hanahan, D.; Weinberg, R. A., Hallmarks of cancer: the next generation. Cell 2011, 144, (5), 646-74.

44. Dai, L.; Pan, G.; Liu, X.; Huang, J.; Jiang, Z.; Zhu, X.; Gan, X.; Xu, Q.; Tan, N., High expression of ALDOA and DDX5 are associated with poor prognosis in human colorectal cancer. Cancer Manag Res 2018, 10, 1799-1806.

45. Watanabe, K.; Yokoyama, S.; Kaneto, N.; Hori, T.; Iwakami, Y.; Kato, S.; Hayakawa, Y.; Sakurai, H.; Fukuoka, J.; Saiki, I., COP9 signalosome subunit 5 regulates cancer metastasis by deubiquitinating SNAIL. Oncotarget 2018, 9, (29), 20670-20680.

46. Richardson, K. S.; Zundel, W., The emerging role of the COP9 signalosome in cancer. Mol Cancer Res 2005, 3, (12), 645-53.

47. Schlierf, A.; Altmann, E.; Quancard, J.; Jefferson, A. B.; Assenberg, R.; Renatus, M.; Jones, M.; Hassiepen, U.; Schaefer, M.; Kiffe, M.; Weiss, A.; Wiesmann, C.; Sedrani, R.; Eder, J.; Martoglio, B., Targeted inhibition of the COP9 signalosome for treatment of cancer. Nat Commun 2016, 7, 13166.

48. Giaginis, C.; Georgiadou, M.; Dimakopoulou, K.; Tsourouflis, G.; Gatzidou, E.; Kouraklis, G.; Theocharis, S., Clinical significance of MCM-2 and MCM-5 expression in colon cancer: association with clinicopathological parameters and tumor proliferative capacity. Dig Dis Sci 2009, 54, (2), 282-91.

49. Burger, M., MCM2 and MCM5 as prognostic markers in colon cancer: a worthwhile approach. Dig Dis Sci 2009, 54, (2), 197-8.

50. Pyo, J. S.; Son, B. K.; Oh, D.; Kim, E. K., BRG1 is correlated with poor prognosis in colorectal cancer. Hum Pathol 2018, 73, 66-73.

5

51.   Zhu, X.; Sun, L.; Lan, J.; Xu, L.; Zhang, M.; Luo, X.; Gong, J.; Wang, G.; Yuan, X.; Hu, J.; Wang, J., BRG1 targeting STAT3/VEGFC signaling regulates lymphangiogenesis in colorectal cancer. Oncotarget 2016, 7, (24), 36501-36509.

52.   Bassani-Sternberg, M.; Braunlein, E.; Klar, R.; Engleitner, T.; Sinitcyn, P.; Audehm, S.; Straub, M.; Weber, J.; Slotta-Huspenina, J.; Specht, K.; Martignoni, M. E.; Werner, A.; Hein, R.; D, H. B.; Peschel, C.; Rad, R.; Cox, J.; Mann, M.; Krackhardt, A. M., Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. Nat Commun 2016, 7, 13404.

53.   Loffler, M. W.; Kowalewski, D. J.; Backert, L.; Bernhardt, J.; Adam, P.; Schuster, H.; Dengler, F.; Backes, D.; Kopp, H. G.; Beckert, S.; Wagner, S.; Konigsrainer, I.; Kohlbacher, O.; Kanz, L.; Konigsrainer, A.; Rammensee, H. G.; Stevanovic, S.; Haen, S. P., Mapping the HLA Ligandome of Colorectal Cancer Reveals an Imprint of Malignant Cell Transformation. Cancer Res 2018, 78, (16), 4627-4641.

54.   Ovens, K.; Naugler, C., Preliminary evidence of different selection pressures on cancer cells as compared to normal tissues. Theor Biol Med Model 2012, 9, 44.

55.   Garcia-Garijo, A.; Fajardo, C. A.; Gros, A., Determinants for Neoantigen Identification. Front Immunol 2019, 10, 1392.

56.   Mardis, E. R., Neoantigen Discovery in Human Cancers. Cancer J 2017, 23, (2), 97-101.

57.   Weinzierl, A. O.; Lemmel, C.; Schoor, O.; Muller, M.; Kruger, T.; Wernet, D.; Hennenlotter, J.; Stenzl, A.; Klingel, K.; Rammensee, H. G.; Stevanovic, S., Distorted relation between mRNA copy number and corresponding major histocompatibility complex ligand density on the cell surface. Mol Cell Proteomics 2007, 6, (1), 102-13.

58.   Laumont, C. M.; Vincent, K.; Hesnard, L.; Audemard, E.; Bonneil, E.; Laverdure, J. P.; Gendron, P.; Courcelles, M.; Hardy, M. P.; Cote, 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).

59.   Luoh, S. W.; Ramsey, B.; Hanlon Newell, A.; Troxell, M.; Hu, Z.; Chin, K.; Spellman, P.; Olson, S.; Keenan, E., HER-2 gene amplification in human breast cancer without concurrent HER-2 over-expression. Springerplus 2013, 2, 386.

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

61.   Faridi, P.; Li, C.; Ramarathinam, S. H.; Vivian, J. P.; Illing, P. T.; Mifsud, N. A.; Ayala, R.; Song, J.; Gearing, L. J.; Hertzog, P. J.; Ternette, N.; Rossjohn, J.; Croft, N. P.; Purcell, A. W., A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands. Sci Immunol 2018, 3, (28).

62.   Vizcaino, J. A.; Csordas, A.; del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; Xu, Q. W.; Wang, R.; Hermjakob, H., 2016 update of the PRIDE database and its related tools. Nucleic Acids Res 2016, 44, (D1), D447-56.

5

Supplementary Figure 1
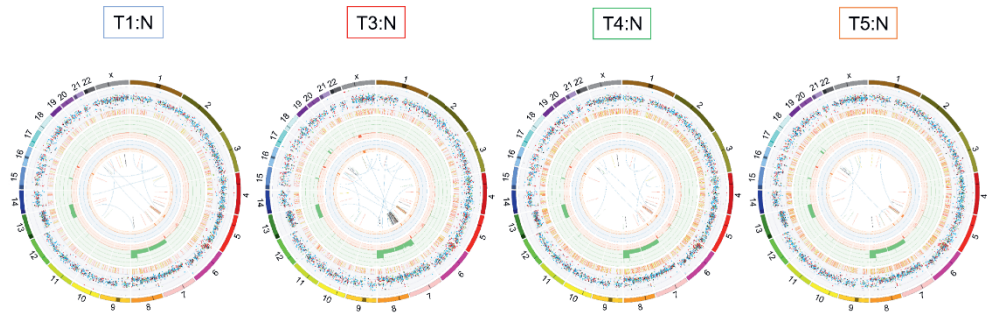
Supplementary Figure 1: Circoplots of genetic alterations detected in the single-cell derived clonal colorectal organoids. The outermost circle demarcate chromosomes. The next concentric ring shows the somatic variants (including exon, intron and intergenic regions) with respect to N. Somatic variants are further divided into an outer ring of substitutions and an inner ring of INDELs. Each dot represents a single somatic variant scaled to 100% by its allele frequency score. Point mutations are colored to the type of base change (e.g. C>T/G>A in red) and are in concordance with colors used previously to describe mutational signatures63. INDELs are colored yellow and red for insertions and deletions respectively. The third ring shows copy number changes adjusted against observed tumor purity. Copy number loss and gain are indicated in red and green respectively. The scale ranges from 0 (complete loss) to 6 (high level gains). The fourth ring represents the observed 'minor allele copy numbers' across the chromosome. The range of this chart is from 0 to 3. The expected normal minor allele copy number is 1 and anything below 1 is shown as a loss (orange) and represents a LOH event. Minor allele copy numbers above 1 (blue) indicate amplification events of both A and B alleles at the indicated locations. The innermost circle displays the observed structural variants within or between the chromosomes. Translocations are indicated in blue, deletions in red, insertions in yellow, tandem duplications in green and inversions in black.

Supplementary Figure 2
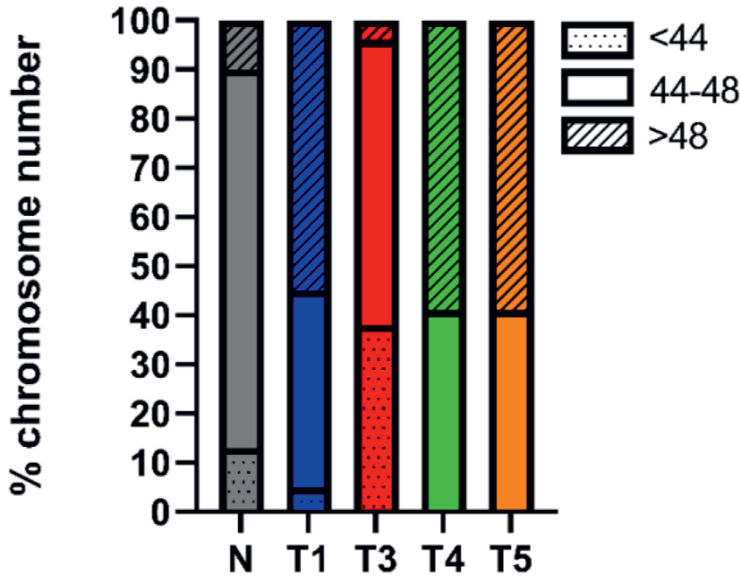
Supplementary Figure 2: Single-cell derived clonal colorectal organoids karyotyping. Percentage of cells with aberrant number of chromosomes. CRC tumor organoids were evidently enriched for aneuploid cells. Karyotyping was performed as described in Drost et al, 2015 on organoid lines cryopreserved at the time of collection for proteomics measurements, subsequently, recovered and expanded for karyotyping. Per organoid line, between 30 and 46 metaphase spreads were counted. Plotting data tabulated in Source Data.

Supplementary Figure 3
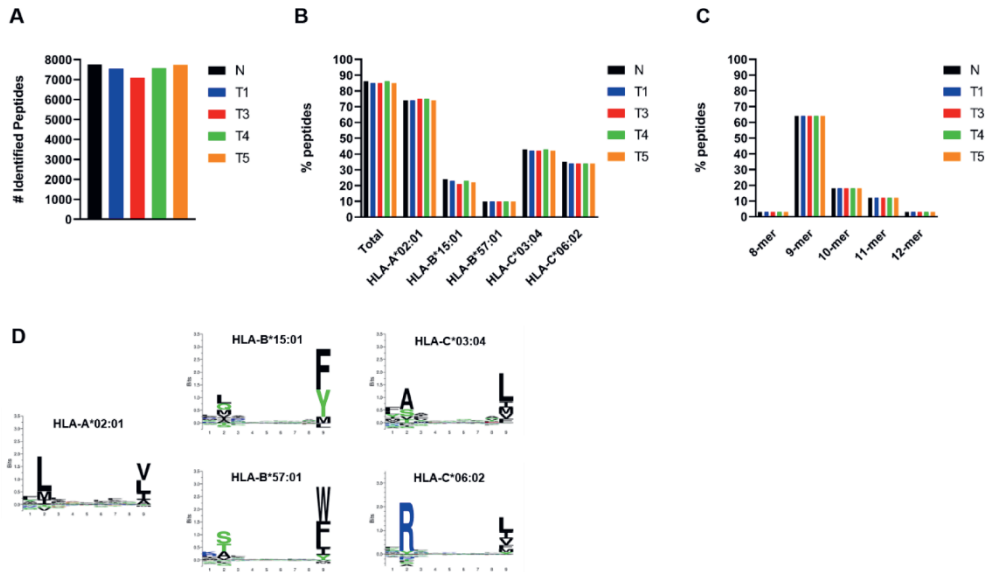
Supplementary Figure 3: HLA class I peptide ligand characteristics. A) Number of unique peptides identified per organoid clone (N = 7750, T1 = 7541, T3 = 7087, T4 = 7571, T5 = 7741). **B)** Assignment of all identified peptides to the patient's HLA alleles. **C)** Length distribution of identified HLA class I peptide ligands. **D)** Theoretical binding motifs of the HLA class I peptide ligands from the patient's HLA alleles. Plotting data tabulated in Source Data.

Supplementary Figure 4
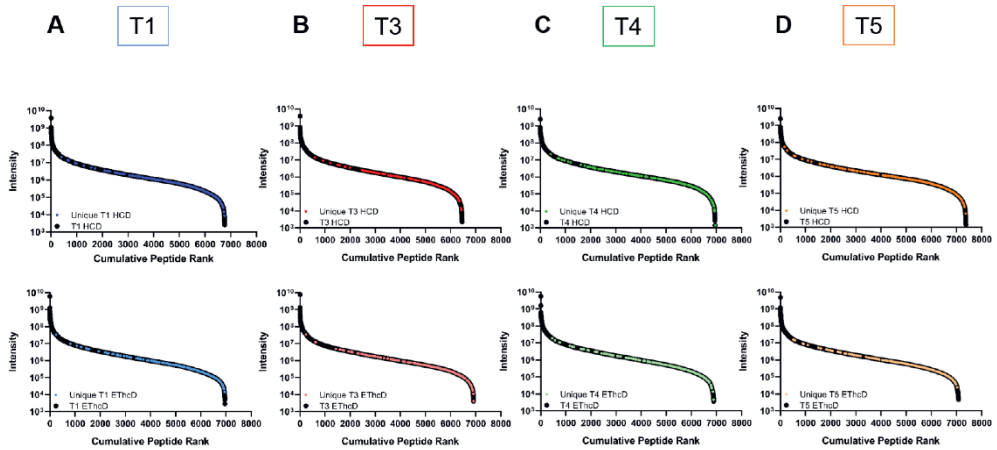
Supplementary Figure 4: Unique peptide ligand intensity distribution per tumor organoid. A) Peptide intensity plotted against the cumulative peptide rank for all T1 peptides (black) and T1 unique peptides (HCD in blue, EThcD in light blue). **B)** Peptide intensity plotted against the cumulative peptide rank for all T3 peptides (black) and T3 unique peptides (HCD in red, EThcD in light red). **C)** Peptide intensity plotted against the cumulative peptide rank for all T4 peptides (black) and T4 unique peptides (HCD in green, EThcD in light green). **D)** Peptide intensity plotted against the cumulative peptide rank for all T5 peptides (black) and T5 unique peptides (HCD in orange, EThcD in light orange). Plotting data tabulated in Source Data.

Supplementary Figure 5
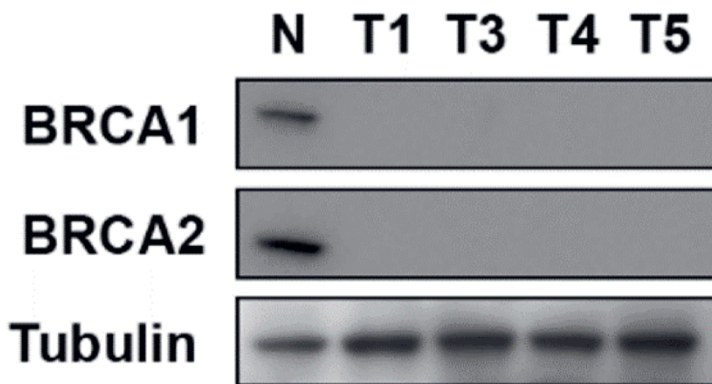
Supplementary Figure 5: BRCA1 and BRCA2 western blot. Both BRCA1 (SC-6954, dilution 1:500) and BRCA2 (SC-293185, dilution 1:500) were not detectable in all four tumor organoid clones, in agreement with the down-regulation observed in fractionated organoid proteomes (Figure 5C). Tubulin was used as loading control. Raw images in Source Data.

Supplementary Figure 6
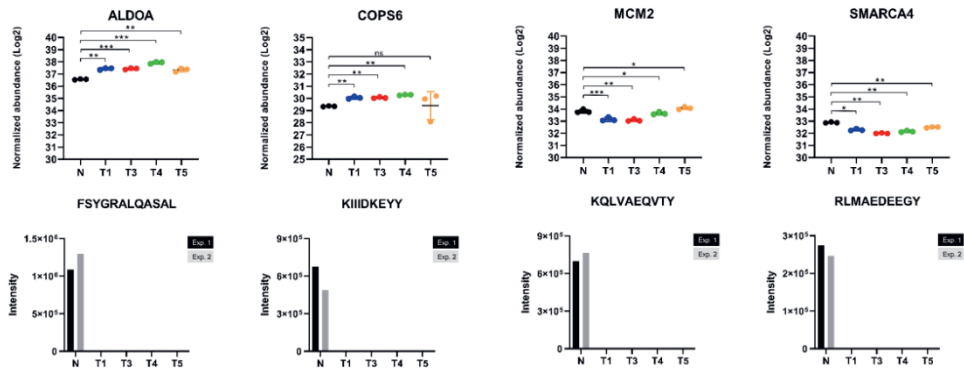
Supplementary Figure 6: Protein abundance and HLA peptide ligand abundance for ALDOA, COPS6, MCM2 and SMARCA4 in all analyzed organoid clones. Top panels: Normalized abundance of the proteins ALDOA, COPS6, MCM2 and SMARCA4, as measured by label-free quantitative proteomics. Bottom panels: HLA peptides originating from ALDOA, COPS6, MCM2 and SMARCA4 were reproducibly detected on normal colon organoids by both HCD (Exp 1) and EThcD (Exp 2) MS fragmentation modes, but not detected in any of the four CRC tumor clones by either fragmentation mode. Significant changes in protein abundance were determined with respect to normal colon organoids based on three technical replicates with a two-sided student's t-test: $p < 0.05$ (*); $p < 0.01$(**). Data presented as mean values ± standard deviation. Plotting data tabulated in Source Data.

Supplementary Figure 7
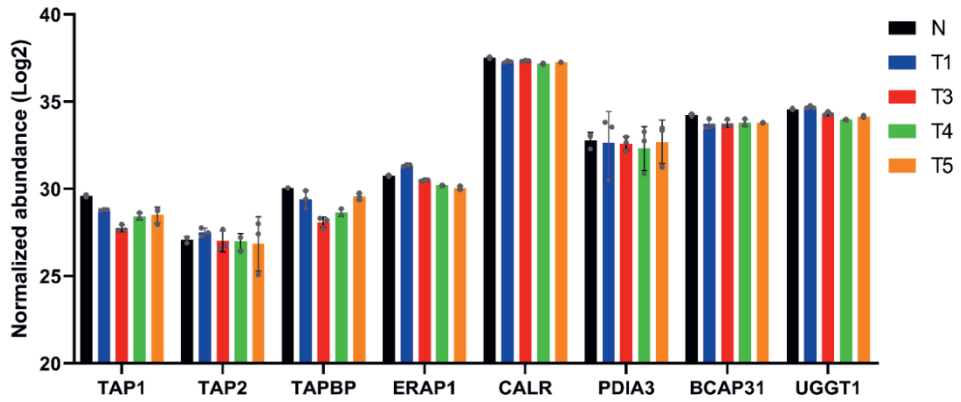
Supplementary Figure 7: Protein abundance of protein involved in HLA class I processing in all analyzed organoid clones. n = 3 technical replicates. Data presented as mean values ± standard deviation. Plotting data tabulated in Source Data.

Chapter 6

Synopsis

Summary and Future Outlook

HLA peptide ligands which are presented abundantly, exclusively and uniformly on the tumor cell surface provide an essential starting point in the design of personalized immunotherapy. In this thesis, we have attempted to tackle several analytical challenges in the field of identifying HLA antigens, also termed immunopeptidomics. Based on these advances we have applied novel strategies to the translational context of colorectal cancer.

In HLA peptide isolation and measurement techniques applied, we have incorporated high-pH fractionation, which is complementary to the widely used strong-cation exchange (SCX) approach. This allowed us to analyze the highly complex HLA peptide ligandome to greater depth. By using a pulldown flowthrough 'recycling' approach, in which HLA ligands are purified from the same sample in multiple rounds of purification, we could further maximize the amount of HLA peptide ligands isolated from a single sample. This may be applied critically to increase the yield of isolated HLA peptides for analysis, especially from clinical samples, where the amount of material is often a limiting factor.

These developments, used in combination in a different study, allowed us to obtain new mechanistic insights into HLA loading in an artificial 'fever-induced' state. We first challenged a cell line system with high temperature to mimic a fever, and consequently detected specific changes in the HLA class II ligandome, mediated through alteration of the CD74 HLA class II chaperone. In this simple context of temperature effects on HLA peptide presentation, we could also focus on the presentation of peptides derived from HLA proteins themselves, taking on a frequently neglected angle than regular mass spectrometry-based ligandomics. By further applying glycoproteomics to characterize HLA protein molecules, we have also characterized their distinct glycan compositions, and shown that different HLA molecules have different glycosylation patterns, in strong correlation with their physical location inside the cell.

Besides using mass spectrometry to understand the fundamental rules in HLA peptide processing and loading in different physiological states, as well as glycosylation-dependent translocation of peptide-bound HLA molecules, the tools established also made it possible to explore cancer antigen discovery in the patient context. In addition to the fundamental research questions, we also delved deep to look for clinically relevant tumor specific antigens and neo-antigens, using the organoid system as means to amplify limited patient material. Despite gains in ligand identification brought about by advances in instrumentation and sample preparation, the routine discovery of neo-antigens from patient derived biopsies remains challenging.

To ensure that neo-antigen based immunotherapy can be more routinely used in the near future, significant steps still need to be made to circumvent the current limitations. One such limitation is the enormous number of *in silico* predicted neo-antigens from the whole patient genome that cannot be experimentally verified one by one. Even with the most stringent prediction methods to date, the large number of potential neo-antigen candidates still cannot be tested one by one, and the chance of clinical utility from these predictions remain very low. Another limitation is the limited sensitivity of shotgun ligandomics in detecting low abundant peptide species. Due to the stochastic nature of shotgun analysis, it is almost impossible to rule out whether a non-identified antigen is really absent, or is presented but not identified due to its inherent low abundance. This may in part be solved with a targeted mass spectrometry approach. Here, we will discuss four main challenges in routine neo-antigen identification that still remain.

Patient derived organoids in immunopeptidomics research

Sample input from patient derived biopsies will always be a limiting factor for subsequent immunopeptidome analysis due to the inherent low abundance of antigens. In the research described in this thesis, we have used an organoid based system to amplify patient derived material. In this way, the organoids serve as a "PCR equivalent" for the amplification of antigens and proteins. Considering that organoids do not accumulate significant mutations in vitro and can retain features of the (tumor) cells and tissue they originate from, make them a very suitable platform for this particular purpose.

To date, patient derived organoids have been only very sporadically used in the immunopeptidomics field. A recent relevant study focused on neo-antigen discovery in microsatellite stable colorectal cancer organoids (1).The major difference with our study, concerning organoids, is how the organoid lines were established. While we clonally expand tumor organoid lines, the other study made use of 'bulk' organoid lines, where one organoid line is established from multiple tumor cells. In our approach, heterogeneity of the original tumor is retained and can thus also be analyzed.

Capturing tumor heterogeneity is very important in cancer research, as it is a known driver for cancer progression. Such heterogeneity is known to results in diversely adapted to survive chemotoxic stress. Using the bulk organoid lines, these traits will be averaged, making it nearly impossible to capture enough tumor vulnerabilities for a potential treatment.

The question that remains is if the antigens presented by organoids are representative for antigens presented by the originating tumor. Antigen presentation is by definition a dynamic process with constant turnover of HLA molecules and HLA

peptide ligands. Therefore, we can only take a snapshot of which antigens are presented at a certain point in time. The major determining factors of antigen presentation are shared between the originating tumor and the organoids, therefore, most antigens presented by organoids are expected to be shared by the originating tumor. Currently, organoids are the closest possible *in vitro* model to an *in vivo* tumor and therefore, organoids are the most feasible model systems to study patient-specific antigens, until we have detection methods which are sensitive enough for the direct identification of antigens from patient-derived biopsies.

Patient derived sources of neo-antigens

The holy grail of antigen discovery is a personalized vaccine that can target patient-specific neo-antigens. These neo-antigens arise from tumor specific mutations. Although most tumors accumulate mutations as the default state, it remains challenging to identify neo-antigens. Most immunopeptidomics workflows are rather blind to patient-specific mutations, unless the search space is refined with personalized sequence information. Using whole genome sequencing, it should be possible to detect all patient mutations, among which single nucleotide variants (SNVs), frameshifts and genomic rearrangements which can all potentially lead to mutated proteins and thus neo-antigens.

Genomic rearrangements can have a tremendous impact on protein function, but are not the most frequent candidates for neo-antigens. Such potential neo-antigens from chromosomal rearrangements could likely only arise around the breakpoint/fusion region of the rearrangement, leading to only few neo-antigen candidates, and only in patients with the specific chromosomal rearrangements.

On the other hand, single nucleotide variants can lead to very promising neo-antigen candidates, even there is only one amino acid change. HLA peptides are only very short peptide fragments ranging from 9 to 12 amino acids, therefore, a single amino acid change can drastically alter the physical, biochemical and loading properties, resulting in presentation of a neo-antigen.

Even more promising than single nucleotide variants are frameshift mutations. These mutations can truncate the subsequent protein-coding sequence, potentially generating a stretch of new protein, up to hundreds of amino acids long. Depending on the stability of these neo-proteins, they can be trimmed and loaded as peptide antigens for which the patient has not developed peripheral tolerance for. Such frameshifts have the potential to generate tremendous numbers of neo-antigens, which may greatly increase the spectrum of loading on HLA types. The approach has great potential for immunotherapy for patients harboring the same conserved frameshift mutations, and a generic long mRNA vaccine covering the frameshifted

sequence may be broadly used for all patients with the same frameshift, almost regardless of patient HLA type.

Lastly, whole genome sequencing can provide information about the non-coding parts of the genome. Tumors often over-express non-coding genes, such as pseudogenes, lncRNAs and alternative reading frames (2, 3), which also have the potential to become a new source of neo-antigens, by some termed the 'dark antigens' (4).

Improving neo-antigen prediction

Improving neo-antigen prediction is key for fast and reliable neo-antigen identification in the future. The current prediction pipelines are almost exclusively focused on the basis of sequencing. Though this can be performed with very high throughput, generating up to tens of thousands of putative neo-antigens for a single patient, it is still currently not possible to measure all these predicted antigens with even the most sensitive mass spectrometers available. Therefore, it is crucial to let the mass spectrometer focus only on prioritized antigens. A streamlined method to prioritize neo-antigens is still not complete, although the tools for antigen prioritization are clearly under development.

As a first filtering step for prioritization, transcriptomic (mRNA) data is often used as a source for neo-antigens. Unfortunately, the correlation between the transcriptome, proteome and ligandome is often rather poor, and therefore neo-antigen predictions based on mRNA seem not be adequate (5-7). In this respect, neo-antigen prediction may be improved if it can be rather based on the patient's proteome. By doing so, the translated proteome is taken into consideration, since translated proteins always function with a limited lifespan, will be degraded eventually and will inevitable enter the pathway of proteasome degradation of HLA presentation. By prioritizing mutated proteins that are detectable in the proteome, the number of predicted neo-antigens could become more manageable for targeted detection. Neo-antigen prediction may also be appended with neo-antigens derived from proteins detected in degradasomes, since the transit of degraded proteins through the proteasome also provides raw material peptides for presentation. Since antigen peptides are by-products of protein degradation, degraded mutant proteins might also be a good source for potential neo-antigens.

Neo-antigen identification by targeted mass spectrometry

In most immunopeptidomics experiments to date, shot-gun profiling mass spectrometry approaches are used to identify neo-antigens. This approach works well when the goal is to analyze a large repertoire of ligands and when there is no

specific interest in particular antigens. Untargeted mass spectrometry experiments rely on intensity-based precursor prioritization, which favors sequencing of high abundant peptides. This biases against detection of low abundant neo-antigens, which may require longer accumulation times to reach the detection limit.

In this respect, parallel reaction monitoring (PRM) may be the method of choice to specifically detect prioritized and predicted neo-antigens. PRM depends on an inclusion list of peptide precursors, which can be composed only of predicted and high-priority neo-antigens. Using this inclusion list, the mass spectrometer spends only time on accumulating and fragmenting the pre-determined precursors, as opposed to intensity-based precursor selection in profiling mode. With this method, it is possible to reliably sample the low abundant antigens and quantitatively detect neo-antigens.

It is not completely certain yet if a low abundant neo-antigen is sufficient for an immune response, although, it has been shown previously that one single copy of a neo-antigen on the cell surface is sufficient for recognition by a TCR and subsequent immune response (8). Furthermore, abundance is not the only determining factor for immunogenicity. Amino acid side chains of mutated residues pointing outside towards the TCR (9) was also proposed as a critical factor for immunogenicity, but also substitutions of antigen anchor positions (10, 11) and antigen hydrophobicity (12). Therefore, the current consensus is that neo-antigens low abundance of peptide presentation seems not to hamper with immunogenicity in a correlative manner.

Taken together, in the work presented in this thesis, we have contributed to better measurement and understanding of HLA peptides and translated these knowledge to improve HLA peptide antigen discovery from human material. The knowledge and roadmaps presented here may contribute to develop new neo-antigen based immunotherapies in the near future.

References

1.   Newey, A.; Griffiths, B.; Michaux, J.; Pak, H. S.; Stevenson, B. J.; Woolston, A.; Semiannikova, M.; Spain, G.; Barber, L. J.; Matthews, N.; Rao, S.; Watkins, D.; Chau, I.; Coukos, G.; Racle, J.; Gfeller, D.; Starling, N.; Cunningham, D.; Bassani-Sternberg, M.; Gerlinger, M., Immunopeptidomics of colorectal cancer organoids reveals a sparse HLA class I neoantigen landscape and no increase in neoantigens with interferon or MEK-inhibitor treatment. J Immunother Cancer 2019, 7, (1), 309.

2.   Kloor, M.; Reuschenbach, M.; Pauligk, C.; Karbach, J.; Rafiyan, M. R.; Al-Batran, S. E.; Tariverdian, M.; Jager, E.; von Knebel Doeberitz, M., A Frameshift Peptide Neoantigen-Based Vaccine for Mismatch Repair-Deficient Cancers: A Phase I/IIa Clinical Trial. Clin Cancer Res 2020, 26, (17), 4503-4510.

3.   Ballhausen, A.; Przybilla, M. J.; Jendrusch, M.; Haupt, S.; Pfaffendorf, E.; Seidler, F.; Witt, J.; Hernandez Sanchez, A.; Urban, K.; Draxlbauer, M.; Krausert, S.; Ahadova, A.; Kalteis, M. S.; Pfuderer, P. L.; Heid, D.; Stichel, D.; Gebert, J.; Bonsack, M.; Schott, S.; Blaker, H.; Seppala, T.; Mecklin, J. P.; Ten Broeke, S.; Nielsen, M.; Heuveline, V.; Krzykalla, J.; Benner, A.; Riemer, A. B.; von Knebel Doeberitz, M.; Kloor, M., The shared frameshift mutation landscape of microsatellite-unstable cancers suggests immunoediting during tumor evolution. Nat Commun 2020, 11, (1), 4740.

4.   Slavoff, S. A.; Mitchell, A. J.; Schwaid, A. G.; Cabili, M. N.; Ma, J.; Levin, J. Z.; Karger, A. D.; Budnik, B. A.; Rinn, J. L.; Saghatelian, A., Peptidomic discovery of short open reading frame-encoded peptides in human cells. Nat Chem Biol 2013, 9, (1), 59-64.

5.   Branca, R. M.; Orre, L. M.; Johansson, H. J.; Granholm, V.; Huss, M.; Perez-Bercoff, A.; Forshed, J.; Kall, L.; Lehtio, J., HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. Nat Methods 2014, 11, (1), 59-62.

6.   Xiang, R.; Ma, L.; Yang, M.; Zheng, Z.; Chen, X.; Jia, F.; Xie, F.; Zhou, Y.; Li, F.; Wu, K.; Zhu, Y., Increased expression of peptides from non-coding genes in cancer proteomics datasets suggests potential tumor neoantigens. Commun Biol 2021, 4, (1), 496.

7.   Wang, X.; Liu, Q.; Zhang, B., Leveraging the complementary nature of RNA-Seq and shotgun proteomics data. Proteomics 2014, 14, (23-24), 2676-87.

8.   Muller, 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.

9.   Ramarathinam, S. H.; Ho, B. K.; Dudek, N. L.; Purcell, A. W., HLA class II immunopeptidomics reveals that co-inherited HLA-allotypes within an extended haplotype can improve proteome coverage for immunosurveillance. Proteomics 2021, e2000160.

10.   Sykulev, Y.; Joo, M.; Vturina, I.; Tsomides, T. J.; Eisen, H. N., Evidence that a single peptide-MHC complex on a target cell can elicit a cytolytic T cell response. Immunity 1996, 4, (6), 565-71.

11.   Fritsch, E. F.; Rajasagi, M.; Ott, P. A.; Brusic, V.; Hacohen, N.; Wu, C. J., HLA-binding properties of tumor neoepitopes in humans. Cancer Immunol Res 2014, 2, (6), 522-9.

12.   Yadav, M.; Jhunjhunwala, S.; Phung, Q. T.; Lupardus, P.; Tanguay, J.; Bumbaca, S.; Franci, C.; Cheung, T. K.; Fritsche, J.; Weinschenk, T.; Modrusan, Z.; Mellman, I.; Lill, J. R.; Delamarre, L., Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. Nature 2014, 515, (7528), 572-6.

13.   Duan, F.; Duitama, J.; Al Seesi, S.; Ayres, C. M.; Corcelli, S. A.; Pawashe, A. P.; Blanchard, T.; McMahon, D.; Sidney, J.; Sette, A.; Baker, B. M.; Mandoiu, II; Srivastava, P. K., Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. J Exp Med 2014, 211, (11), 2231-48.

14.   Chowell, D.; Krishna, S.; Becker, P. D.; Cocita, C.; Shu, J.; Tan, X.; Greenberg, P. D.; Klavinskis, L. S.; Blattman, J. N.; Anderson, K. S., TCR contact residue hydrophobicity is a hallmark of immunogenic CD8+ T cell epitopes. Proc Natl Acad Sci U S A 2015, 112, (14), E1754-62.

Lay summary of this thesis

Our body consists of approximately 37 trillion cells. Every cell can be compared to a small city. There are many different factories, shops and inhabitants. There are factories to generate energy (mitochondria), garbage trucks that continuously collect and dispose of garbage (lysozymes) and trucks that drive around packages (Golgi-system and endosomes). We call the factory workers and other city inhabitant's **proteins**.

As most cities, our cells also have an army to protect us against invading intruders, which are often pathogens such as bacteria and viruses. We call this army the **immune system**. The immune system is extremely diverse. In this thesis, we focus on the **Human Leukocyte Antigen system (HLA)**. You can compare these proteins with merciless soldiers, who rip apart pieces of bacteria and viruses and show them to whole army, so they can be recognized and destroyed easily. These pieces are called **antigens** and are very small pieces of protein.

We can study these antigens by **mass spectrometry**. This is a machine that can extremely accurately measure the mass of a protein. A protein is made out of 20 different building blocks, amino acids, which together form a very long chain. Almost every amino acid has a different mass, therefore, the mass spectrometer can determine which they are based on their respective masses. Though the mass spectrometer has difficulties in measuring full proteins. Therefore, we first need to cut them into tiny pieces, peptides. Antigens are already peptides, and therefore do not need to be cleaved before we analyze them by mass spectrometry. Besides measuring the mass of a peptide, the mass spectrometer can also determine in which order the amino acids are placed in a peptide. With this information, we know the amino acid sequence of a peptide and this will help us subsequently to identify the protein it originated from.

In this thesis, we use mass spectrometry to determine which antigens are presented by HLA molecules. By understanding this process, we can use this knowledge for e.g. vaccine development against diseases such as SARS-CoV-2 or non-infectious diseases such as cancer.

In **chapter 1** of this thesis, I describe the basic principles of antigen presentation by HLA molecules. First, we describe that there are two different classes of HLA molecules: HLA class I and HLA class II. Whereas HLA class I is more important for clearance of potential cancer cells, HLA class II is more important in immunity against bacteria and viruses. We describe in detail how antigens are presented for both classes. Next, we focus on the role of HLA in different autoimmune diseases, but also its critical role in clearance of infections. Then, I describe all current methods that can be used to purify these antigens and HLA proteins from cells and which are the best way to measure them by mass spectrometry. In the end of chapter 1, we

discuss how HLA research has led to improved immunotherapy for cancer patients and how this will revolutionize future immunotherapy.

In **chapter 2** I focus on the difficulties of analyzing antigens by mass spectrometry. We apply methods often used in proteomics research, pre-fractionation methods high-pH reversed phase (HPH) and strong cation exchange (SCX). By applying these methods to antigens, we can detect >100% more antigens than in a workflow without pre-fractionation. But, by doing so, I introduce an identification bias which is determined by the chemical properties of antigens. Furthermore, we show that these methods are necessary to identify a subset of post-translational modifications.

In **chapter 3** I wanted to study the pathways of antigen presentation. Therefore, I grew cells at 40°C to simulate a fever. I saw that this was enough for our cells to start initiating an immune response. Also, I only saw changes in HLA class II and not HLA class I which indicates that the initial immune response is mediated through HLA class II. I saw specific changes in the HLA class II chaperone CLIP, which might be so it can easily exchange for a pathogenic peptide upon infection.

In **chapter 4** we describe a post-translational modification on HLA molecules. We identified multiple forms of this modification on the HLA molecules, dependent on the type of HLA molecule (A, B, C). Furthermore, we saw that the form of modification correlates with the physical place of an HLA molecule in the cell.

In **chapter 5** we wanted to study antigen presentation in a cancer context. We used colorectal cancer organoids (mini organs in a dish) as model system. First, we noticed that these organoids retain the characteristics of their originating cells. When we analyzed the antigen that are presented, we found that these overlap 97% between the tumor organoids and healthy colon organoids. The antigens that do not overlap come specifically from proteins that are detrimental for the tumor. These unique antigens could potentially be used in the future for anti-cancer vaccines for this subtype of colon cancers.

Finally, in **chapter 6** of this thesis, I share my view on the future of antigen research for cancer immunotherapy. Additionally, it includes a list of publications and my acknowledgements will be shared in here.

6

Lekensamenvatting van dit proefschrift

Ons lichaam bestaat uit ongeveer 37 biljoen cellen. Elke cel kun je vergelijken met een kleine stad. Er zijn veel verschillende fabrieken, winkels en inwoners. Zo zijn er fabrieken die de stad voorzien van energie (mitochondriën), vuilniswagens die continu het afval ophalen en wegbrengen (lysozymen), en pakketbezorgers (Golgi-systeem en endosomen). We noemen deze fabrieksarbeiders en alle andere inwoners **eiwitten**.

Net zoals de meeste middeleeuwse steden hebben ook onze cellen een leger dat ze beschermt tegen binnenvallende indringers. Deze indringers zijn vaak verschillende soorten bacteriën en virussen. We noemen het leger ook wel het **immuunsysteem**. Het immuun systeem bestaat uit veel verschillende soorten gespecialiseerde cellen en eiwitten. In dit proefschrift richt ik me op het **Humaan Leukocytenantigeensysteem (HLA)**. HLA-eiwitten kun je vergelijken met genadeloze soldaten die ledematen van vijanden afrukken en die laten zien aan de rest van het leger. Op deze manier weet het hele leger meteen welke vijanden het moet aanvallen. Deze ledematen noemen we ook wel **antigenen**, en in dit geval zijn het heel kleine stukjes eiwit.

We kunnen deze antigenen bestuderen met **massa spectrometrie**. Een massa spectrometer is een machine die zeer nauwkeuring de massa van een eiwit kan meten. Een eiwit is samengesteld uit twintig verschillende bouwstenen, aminozuren geheten, die samen een lange keten vormen. Bijna ieder aminozuur heeft een andere massa. Daardoor kan de massa spectrometer bepalen welke aminozuren er in een eiwit zitten. De massaspectrometer heeft echter moeite met het meten van hele eiwitten. Daarom knippen we ze eerst in stukjes. Een stukje eiwit noemen we een peptide. Antigenen zijn al peptiden en daarom hoeven we die niet eerst op te knippen voordat we ze kunnen meten met massa spectrometrie. Naast het meten van de massa van een peptide kan de massaspectrometer ook bepalen in welke volgorde de aminozuren voorkomen in een peptide. Met deze informatie kunnen we later, aan de hand van een database, weer bepalen uit welk eiwit we de peptiden hebben geknipt.

In dit proefschrift gebruiken we massaspectrometrie om te bepalen welke antigenen de HLA-moleculen aan het immuunsysteem laten zien. We noemen dit ook wel antigeen presentatie. Wanneer we dit process beter begrijpen, kunnen we het bijvoorbeeld inzetten voor het ontwikkelen van vaccins tegen SARS-CoV-2 of als therapie voor bepaalde soorten kanker.

In **hoofdstuk 1** van dit proefschrift beschrijven we de basisprincipes van antigeenpresentatie door HLA-moleculen. Eerst leggen we uit dat er twee verschillende soorten HLA- moleculen zijn: HLA-klasse I en HLA-klasse II. HLA-klasse I is het belangrijkst voor het tijdig opruimen van potentiële kankercellen. HLA-

klasse II zorgt voor immuniteit tegen bacteriën en virussen. In detail beschrijven we hoe HLA-moleculen antigenen aan het immuunsysteem laten zien. Verder richten we ons op de rol van HLA in verschillende auto-immuunziektes, maar ook op de cruciale rol bij het opruimen van infecties. Daarna beschrijven we welke methoden er op dit moment zijn om deze antigenen te isoleren uit cellen en welke massa spectrometrie methoden het beste zijn om ze daarna te analyseren. Aan het eind van hoofdstuk 1 bespreek ik hoe de HLA-onderzoeken die tot nu toe zijn gedaan hebben bijgedragen aan de verbetering van immuuntherapie voor kankerpatiënten, en welke grote verbeteringen het nog zal brengen voor de toekomst.

In **hoofdstuk 2** kijken we naar de moeilijkheden die er zijn bij het analyseren van antigenen met massa spectrometrie. We gebruiken daarvoor methoden die vaak worden gebruikt in standaard-proteomica-onderzoek. Het zijn de methoden reversed phase-fractionering onder hoge pH-condities en strong cation exchange, waarbij fractionering plaatsvindt op basis van de lading van een peptide. Wanneer we deze twee methoden toepassen, kunnen we 100% meer antigenen detecteren met massa spectrometrie.

In **hoofdstuk 3** laten we zien hoe we het mechanisme achter antigeenpresentatie bestudeerd hebben. Daarvoor hebben we cellen laten groeien bij 40 °C, wat koorts simuleert. We zijn erachter gekomen dat dat genoeg was om in onze cellen een immuunrespons te starten. Verder zien we vooral verschillen in antigenen gepresenteerd door HLA-klasse II en niet door HLA-klasse I. Dat laat zien dat de immuunreactie waarschijnlijk door middel van verloopt door HLA-klasse II gaat.

In **hoofdstuk 4** beschrijven we een post-translationele modificatie op HLA-moleculen. We hebben verschillende vormen van deze modificatie op HLA-moleculen geïdentificeerd die afhankelijk zijn van het type HLA-molecuul (A, B of C). Verder hebben we aangetoond dat de vorm van deze modificatie samenhangt met de fysieke plaats van het HLA-molecuul in de cel.

In **hoofdstuk 5** hebben we antigeen presentatie bestudeerd in de context van darmkanker. Daarvoor hebben we gebruikgemaakt van organoïden (miniorgaantjes) die worden gemaakt van darmkankercellen. We hebben opgemerkt dat deze organoïden dezelfde eigenschappen hebben als de cellen waaruit ze zijn gemaakt. Daarna hebben we de antigenen geanalyseerd en hebben we gezien dat ze voor 97% overlappen met antigenen van gezonde darmcellen. De antigenen die uniek zijn voor de darmkanker cellen, komen allemaal van eiwitten die erg schadelijk zijn voor de tumor. Deze antigenen komen niet voor op gezonde darmcellen en zijn daarom een goed doelwit voor immuuntherapie voor deze soort darmkanker.

Tot slot, in **hoofdstuk 6** van dit proefschrift, deel ik mijn visie op de toekomt van antigeenonderzoek voor immuuntherapie voor kankerpatiënten. Ook bevat dat hoofdstuk een lijst van publicaties en mijn dankwoord.

Leekesaomevatting van di proefschift

Oons lichaom bestòi uit ongeveer 37 biljoen celle. Elke cel kande verglijke meej 'n klèèn stad. D'r zen veul verschillende febrieke, winkels en inwoôners. Zoo zen d'r febrieke die de stad vurzien van eenergie (mitochondriën), vùlniswaoges die hil de tijd d'n afval ophaolen en wegbrenge (lysozymen), en pakketbezùrgers (Golgi-systeem en endosomen). We noeme dees febriekserbaaiers en alle aander inwoôners **aaiwitte**.

Net es de miste middeleêuwse steeje hebbe oons celle ok 'n leeger dè ze beschermt teege indringers die binnevalle. Dees indringers zen dikkels verschaaiene sorte bacteeriën en virusse. We noemen 't leeger ok wel 't **immuunsysteem**. 't Immuunsysteem bestòi uit veul verschillende sorte gespeesjelizeerde cellen en aaiwitte. In di proefschrift kijk ik nòr 't **Humaan Leukocytenantigeensysteem (HLA)**. HLA-aaiwitte kande verglijke meej genaodeloôze soldaote die lichaomsdeêle van vijaande afrukke en die laote zién òn de rest van 't leeger. Op die menier wit heêl 't leeger welke vijaanden 't moet aonvalle. Dees lichaomsdeêle noeme we ok wel **antigeene**, en in dees geval zen 't heêl klèèn stukskes aaiwit.

We kanne dees antigeene bestudeere meej **massaspectometrie**. 'ne Massaspectromeeter ies 'n mesjién dè heêl naauw de massa van 'n aaiwit kan meete. 'n Aaiwit bestòi uit twintig verschillende bouwsteêne, die aminozuure hiete, en die vùrme saomen 'n lange ketting. Bekaant ieder aminozuur heej 'n aander massa. Dè mòkt dè de massaspectromeeter kan bepaole welke aminizuure d'r in 'n aaiwit zitte. De massaspectromeeter heej alleên wel moeite meej 't meete van heêl aaiwitte. Daorom knippe we ze irst in stukskes. 'n Stukske aaiwit noeme we 'n peptide. Antigeene zen al peptide en daorom hoeve we die nie irst op te knippe vurdè we ze kanne meete meej massaspectomeetrie. Nòst 't meete van de massa van 'n peptide kan de massaspectromeeter ok bepaole in welke vòlgòrde de aminozuure vurkoome in 'n peptide. Meej dees infermaosie kanne we laoter, meej 'nen daotabees, wir bepaole uit welk aaiwit dè we peptide geknipt hebbe.

In di proefschrift gebruike we massaspectomeetrie om te bepaole welke antigeene de HLA-moolecuule òn 't immuunsysteem laote zién. Dè noeme we ok wel antigeenprizzentaosie. Es we beeter begrijpe hoe dè-tè gòi, kanne we't bevurbeeld gebruike vur 't maoke van vaccèèns teege SARS-CoV-2 of als thirrepie vur bepaolde sorte kaanker.

In **hoofdstuk 1** van di proefschrift beschrijve we de baozisprincipes van antigeenprizzentaosie dur HLA-moolecuule. Irst legge we uit dè-t'r tweej verschillende sorte HLA-moolecuule zen: HLA-klasse I en HLA-klasse II. HLA-klasse I ies 't belangrijkst vur 't op tijd opruime van moogelijke kaankercelle. HLA-klasse II zùrgt vur de immuniteit teege bacteeriën en virusse. We vertelle in deetail hoe HLA-moolecuule antigeene òn 't immuunsysteem laote zién. Verder kijke we nòr de ròl

van HLA in verschaaiene auto-immuunziektes, mer ok nòr de cruciale ròl bij 't opruime van infecsies. Dòrnao vertelle we welke methoodes d'r op di mement zen om dees antigeene te izoleere uit de cellen en welke massaspectometriemethoodes 't beste zen om ze daornao te analyzeere. Òn 't èènd van hoofdstuk 1 bespreek ik hoe de HLA-onderzoeke die tot nou toe gedaon zen, bijgedraoge hebbe òn de verbeetering van immuunthirrepie vur kaankerpesjente, en weffer groôte verbeeteringen 't nog zal hebbe vur de toekomst.

In **hoofdstuk 2** kijke we nòr de prebleeme die ge kant hebbe bij 't analyzeere van antigeene meej massaspectomeetrie. Daorvur gebruike we methoodes die dikkels wòrre gebrùkt in standaord-proteomica-onderzoek. 't Gòi om de methoodes reversed phase-fracsjoneering onder hoôge pH-condisies en strong cation exchange, wòrbij fracsjoneering plòtsvijnt op baozis van de laoding van 'n peptide. Es we dees tweej methoodes toepaase, kanne we 100% meêr antigeene dittecteere meej massaspectomeetrie.

In **hoofdstuk 3** laote we zién hoe dè we 't meechanisme aachter antigeenprizzentaosie bestudeerd hebbe. Daorvur hemme celle laote groeie bij 40 °C, duus daormeej doeme net ofdè-t'r korts ies. We zen d'r aachter gekoome dè-tè genoeg was om in oons celle 'n immuunrespons te starte. Verder zieme vural verschille in antigeene die geprizzenteerd wòrre dur HLA-klasse II en nie dur HLA-klasse II. Dè löt zién dè-te immuunreacsie wòrschijnlijk dur middel van HLA-klasse II gòi.

In **hoofdstuk 4** beschrijve we 'n post-translasjoneêle moodificaosie op HLA-moolecuule. We hebbe verschaaiene vùrme van dees moodificaosie op HLA-moolecuule geïdentificeerd die afhaankelijk zen van 't type HLA-moolecuul (A, B of C). We hebbe ok laote zién dèt de vùrm van dees moodificaosie te maoke heej meej de fyzieke plek van 't HLA-moolecuul in de cel.

In **hoofdstuk 5** hemme antigeenprizzentaosie bestudeerd es 't gòi om dèèrmkaanker. Daorvur hemme organooïde (miniòrgaontjes) gebrùkt die wòrre gemòkt van dèèrmkaankercelle. We hebbe gezién dè dees organooïde dezelfde èègeschappe hebbe es de celle waor ze uit zen gemòkt. Dòrnao hemme de antigeene geanalyzeerd en hemme gezién dè ze vur 97% 't zelfde zen es de antigeene van gezonde dèèrmcelle. De antigeene die uniek zen vur de dèèrmkaankercelle, koome allemaol van aaiwitte die heêl schaoilijk zen vur de tuumor. Dees antigeene koome nie vur op gezonde dèèrmcelle en daorom zen ze 'n goei doelwit vur immuunthirrepie vur di sort dèèrmkaanker.

Tot slòt, in **hoofdstuk 6** van di proefschrift, vertel ik hoe dè ik aonkijk teege de toekomst van antigeenonderzoek vur immuunthirrepie vur kaankerpesjente. Dè hoofdstuk bevat ok 'n lijst van puublicaosies en m'n daankwoord.

Curriculum Vitae

I studied the bachelor Biology and Applied Medical Laboratory Technology at Avans University of Applied Sciences in Breda, where I focused on (bio-) medical research. Afterwards, I followed the master Cancer, Stem Cells and Developmental Biology at Utrecht University. I conducted two research projects. In my first project, under the guidance of Dr. Eelco Tromer in the research lab of Prof. Dr. Geert Kops, I studied the functional domains of the mitotic kinase BUB1. At the end of my master I did an internship in the group of Prof. Dr. Albert Heck under supervision of Prof. Dr. Fan Liu. Here, we validated crosslinking mass spectrometry data of crosslinked nuclei. This internship was where my interest in mass spectrometry and proteomics was sparked.

After my studies, I started my PhD in the biomolecular mass spectrometry and proteomics group of Prof. Dr. Albert Heck. I performed my PhD under the guidance of Dr. Wei Wu. During my PhD, we focused on antigen presentation by the HLA system. We focused on technical advancement for the identification of antigens, but also on the mechanistic pathways underlying antigen presentation. Furthermore, we focused on antigen presentation in context of colorectal cancer. The resulting work of the last 4 years is presented in this thesis.

List of publications

Demmers, L.C., Heck, A.J.R., Wu, W. <u>Pre-fractionation extends, but also creates a bias in the detectable HLA class I ligandome</u>. *Journal of Proteome Research* 18, 4, 1634–1643 (2019)

Demmers, L.C., Kretzschmar K., Van Hoeck, A., Bar-Epraïm, Y.E., van den Toorn, H.W.P., Koomen, M., van Son, G., van Gorp, J., Pronk, A., Smakman, N., Cuppen, E., Clevers, H., Heck, A.J.R., Wu, W. <u>Single-cell derived tumor organoids display diversity in HLA class I peptide presentation</u>. *Nature Communications* 11, 5338 (2020)

Demmers, L.C., Wu, W., Heck, A.J.R. <u>HLA class II presentation is specifically altered at elevated temperatures in the B-lymphoblastic cell line JY</u>. *Molecular and Cellular Proteomics*, 20, 100089 (2021)

Hoek, M. [§], **Demmers, L.C.** [§], Wu, W., Heck, A.J.R. <u>Allotype-specific glycosylation and cellular localization of human leukocyte antigen class I proteins</u>. *Journal of Proteome Research*, epub ahead of print (2021)

Acknowledgements

This chapter marks the end of my PhD. When I graduated high school, I could have never imagined that I would ever obtain a PhD, but luckily, I proved myself wrong and did it! I have learned so much in the past four years. Not only about conducting scientific research, but also about our roles as scientist in society and also about our role as female scientists. I am sure I would have never been able to complete this journey without the help of all my amazing colleagues, friends and family!

First and foremost, I want to thank **Wei**. I want to thank you for all the guidance you have been giving me (and still are) over the past four years. You have thought me everything I needed to know to complete this PhD. From experimental design to data analysis and from writing manuscripts to making and giving presentations and posters. In the first year, we (you) managed to set up the whole tricky HLA IP protocol, and I'm proud to see how this matured and gave rise to some cool manuscripts. Besides all the work related things, I really enjoyed the dinners we had in Utrecht. Also, that you showed me proper Chinese cuisine (and not the Dutch variant). The jellyfish still haunts my dreams sometimes though. And, that you shared some amazing recipes with me for ramen and curry! During this PhD, you for sure made me a better scientist and helped me grow as a person, for which I am very grateful.

I also want to thank **Albert**, without whom, this PhD would have never started in the first place. I want to thank you for the opportunity to start my PhD in the lab without having any prior knowledge on mass spectrometry. I really enjoyed our monthly AIO meetings. Call me stubborn, but I really used to think that many of your ideas were (are) crazy and would definitely not work, but in the end they always worked out and led to some cool manuscripts. These past four years were an amazing experience and I consider myself lucky if I have inherited some of your talent.

I would like to thank all members of the yMHC subgroup: Julia, Max, Gadi and Álvaro. Thanks everyone for the great discussion we had about all our HLA related research. **Julia**, you are not just a colleague, but also a very good friend. I am honored that you will be standing next to me as my paranymph at the defense. We started our journey together at the HUPO-HIPP summer school in Madrid, which was basically the first time we actually met. It was nice to share that experience with you. One thing I will never forget, is that, during the first months, we thought about each other that we were idiots, all because of some stupid language barrier. I really enjoyed all our coffee and tea breaks, Friday afternoon beers and Thursday night dinners (thanks for introducing me to ramen!). We have seen each other's highs and lows, and through it all, you were always there for me no matter what! I have seen you growing into an amazing scientist, female rights fighter and aspirant artist. I know

that sometimes you really struggle with what you want to do after your PhD. For the future, I hope that you will find the strength to follow all your dreams and passions, regardless of what they are. Do not let anyone ever tell you that you are not good enough or cannot do something, because I know you can! Thanks for the amazing years we had together. I really want to visit you sometime in Ibiza when you start your own pottery business, so better make at least that one dream come true! **Max**, it was my pleasure to work together with you on the HLA glycosylation project. Also, thanks so much for helping me out with many annoying task for which I needed python (which I can't)! I really appreciate it. Also, it was very nice to be able to talk with someone about videogames. Always good to know that I was not the only nerd in the lab. I wish you all the best for the final stretches of your PhD and also your new job. Hope to catch you on Diablo every now and then! **Gadi**, although we did not work together that much, it was very nice to have you around in our little subgroup. I liked the conversations about work, but also about videogames, cheeses and all other things we have been talking about. **Álvaro**, I wish you all the best for your upcoming PhD journey, I'm sure it will be going great with all the experience you already had in the field (well, if you don't use Peaks of course, no offense).

Z604: Franziska, Nadine and Juan. It was really nice to have you as my officemates, especially because I had the feeling we could ask each other anything. Stupid questions about calculations upon brain malfunctioning, but also questions about data analysis, experiments and basically anything else. **Franziska**, it was nice that I could talk to someone about all the gardening/cooking/fermenting/cheese making adventures. Also, it was nice to have some around to give all my cheese to when I did not really like them myself. Guess I have to start looking for a new victim now. **Nadine (and Donna)**, I enjoyed discussing/ranting with you about all problems in the world (especially feministic issues). If we decide at some point that science is not where we want to spend our entire careers, I suggest we start some feministic action group. **Juan**, it's a bit of a shame we only recently found out we both enjoy gaming, but I'm happy we found out before it was too late! I wish you the best of luck with finishing your PhD and I wish you the best for the future.

Mirjam and Ceri, first of all, I want to thank you both for all the work you do for the lab. You have both thought me so much about protocols/proteomics/HPLCs and so much more. Also, I really appreciate the fact that you were always willing to help me out with some labwork, when I either forgot to add trypsin for example of when I was stuck in traffic for like the millionth time. Besides work, I liked going on Pokémon raids, coffee breaks and lunch breaks with you. Mirjam, thanks for all the coffee you have made for me over the past years, without it, I would have never finished this thesis. Also, it was very nice to go to creative life with you and to have someone around from Brabant. Ceri, it was nice to meet someone with basically the same personality as I have and the same love for the Efteling as I have. I'm glad you want

to be my paranymph, because I know that is not really your thing haha. I consider both of you my friends, and I hope we will keep seeing each other in the future!

Of course, I would also like to thank all other members of the Hecklab, but there are way too many to mention all of you here. I would like to address some of you in particular. All the **Technicians**, thanks a lot for keeping the whole lab up and running. It's a tremendous amount of work, which we sometimes take for granted. But without you, the lab for sure will fall apart. **Fleur, Danique, Tatiana, Wouter and Marie**, thanks a lot for the nice company at the Friday afternoon borrels, it was a very nice way to end the week! **Tim**, it was a pleasure to work with you on my last PhD project. You managed to learn me a lot about PRM, which I really appreciate. I'm happy I did not drive you crazy (I think) with all the times I was asking you to tweak some tiny little things in the figures you made. **Barbara**, it was so nice to have met you! You always helped me out with all my chemistry questions, which I appreciate. I hope we can see each other soon on the slopes for some skiing!

Naast veel steun van mijn collega's heb ik ook veel steun gehad aan mijn vrienden en familie. **Liesbeth, Lois en Yoïn**, ik wil jullie bedanken voor het feit dat ik echt altijd stoom kon afblazen bij jullie! Het is erg fijn om met mensen te kunnen praten die zich niet dagelijks bezig houden met experimenten, dat geeft vaak veel ruimte voor nieuwe inzichten. Ik ben blij dat ik jullie heb leren kennen in de eerste en tweede van de middelbare school. Ik hoop dat we snel weer welke maand met elkaar kunnen eten (ja, nu echt). En natuurlijk, speciale dank aan Yoïn, voor het vertalen van mijn lekensamenvatting in het Brabants!

Bedankt aan alle leden van "Bier": **Stijn, Ingrid, Jeroen, Stefan, Erik, Paul, Martin, Mark, Marcel en Jen.** Het was fijn om met regelmaat op vrijdag en/of zaterdagavond ene biertje te kunnen drinken en spelletjes te kunnen doen zonder over werk te hoeven nadenken. Verder wil ik Jeroen nog bedanken voor het financieren van mijn studie met zijn belastinggeld.

Esther, ik heb je leren kennen in Nijmegen toen we daar samen stage liepen. Ik denk dat we vrienden zijn geworden door onze liefde voor Muse. Het was leuk om met een PhD aan een andere universiteit te kunnen praten over al het gedoe rondom een PhD. Door jou ben ik nu wel verslaafd geraakt aan kaasmaken, maar toch bedankt daarvoor. Ik hoop dat als je ooit terugkomt uit Nieuw-Zeeland we hier een leuk boerderijtje op de kop kunnen tikken en daar ons boerderijwinkeltje kunnen beginnen.

Martie, bedankt dat je me enthousiast hebt gemaakt voor wetenschappelijk onderzoek en voor het feit dat je me het vertrouwen geeft om elk jaar een gastles te geven in de proteomics cursus op Avans. Ik hoop dat er ooit 1 student denkt dat het toch wel een leuk vakgebied is om in verder te gaan.

Daarnaast heb ik veel steun gehad van mijn familie. Natuurlijk van ons **pap en mam.** Bedankt voor de steun en support die jullie me hebben gegeven. Ik snap dat wat ik doe niet altijd te begrijpen is (en ook lastig uit te leggen), maar jullie waren wel altijd ontzettend geïnteresseerd en trots. Mede dankzij jullie heb ik dit tot een mooi einde kunnen brengen. Verder is er natuurlijk nog mijn bonus familie. **Peter, Monique, Ilse, Cecko en oma Joyce.** Ik vond het leuk dat jullie altijd interesse hadden in wat ik aan het doen was. Oma Joyce, ik vond het heel leuk dat u altijd zoveel moeite deed om mijn onderzoek te begrijpen door van alles op te zoeken! Ik vind dat bewonderingswaardig voor een vrouw van over de 90. Ik hoop dat ik op die leeftijd nog altijd zo leergierig ben.

Als laatste, maar zeker niet de minste, **Maarten**. Bedankt voor alle steun en mogelijkheden die je me hebt gegeven om al mijn studies te volgen en uiteindelijk mijn PhD. Ik weet vrij zeker dat ik net zo weinig weet van jouw werk als jij van dat van mij, maar dat neemt niet weg dat het fijn was om 's avonds altijd even over het werk te kunnen praten met je. Ik ben je dankbaar voor het feit dat je me altijd heb geholpen en me hebt laten groeien op persoonlijk vlak. Hierdoor heb je een beter mens van me gemaakt. Ik ben ook heel erg blij met alle kleine dingen, bijvoorbeeld dat je vast de groente voor het avondeten wilde snijden als ik weer eens laat thuis was, dat je niet klaagde over het bizar late tijdstip dat we vaak pas zaten te eten en dat je me de vrijheid geeft om al mijn gekke hobbies uit te voeren. Ik zie uit naar wat de toekomst brengt.

6