



Predicting Antigen Presentation— What Could We Learn From a Million Peptides?

David Gfeller^{1,2*} and Michal Bassani-Sternberg^{1,3*}

¹Department of Oncology, Ludwig Institute for Cancer Research, University of Lausanne, Lausanne, Switzerland,

²Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland, ³Department of Oncology, Ludwig Institute for Cancer Research, University Hospital of Lausanne, Lausanne, Switzerland

OPEN ACCESS

Edited by:

Victor Greiff,
University of Oslo, Norway

Reviewed by:

Nicola Ternette,
University of Oxford,
United Kingdom
Scheherazade Sadegh-Nasseri,
Johns Hopkins University,
United States
Pouya Faridi,
Monash University, Australia

*Correspondence:

David Gfeller
david.gfeller@unil.ch;
Michal Bassani-Sternberg
michal.bassani@chuv.ch

Specialty section:

This article was submitted
to B Cell Biology,
a section of the journal
Frontiers in Immunology

Received: 27 April 2018

Accepted: 12 July 2018

Published: 25 July 2018

Citation:

Gfeller D and Bassani-Sternberg M
(2018) Predicting Antigen
Presentation—What Could We Learn
From a Million Peptides?
Front. Immunol. 9:1716.
doi: 10.3389/fimmu.2018.01716

Antigen presentation lies at the heart of immune recognition of infected or malignant cells. For this reason, important efforts have been made to predict which peptides are more likely to bind and be presented by the human leukocyte antigen (HLA) complex at the surface of cells. These predictions have become even more important with the advent of next-generation sequencing technologies that enable researchers and clinicians to rapidly determine the sequences of pathogens (and their multiple variants) or identify non-synonymous genetic alterations in cancer cells. Here, we review recent advances in predicting HLA binding and antigen presentation in human cells. We argue that the very large amount of high-quality mass spectrometry data of eluted (mainly self) HLA ligands generated in the last few years provides unprecedented opportunities to improve our ability to predict antigen presentation and learn new properties of HLA molecules, as demonstrated in many recent studies of naturally presented HLA-I ligands. Although major challenges still lie on the road toward the ultimate goal of predicting immunogenicity, these experimental and computational developments will facilitate screening of putative epitopes, which may eventually help decipher the rules governing T cell recognition.

Keywords: human leukocyte antigen peptidomics, human leukocyte antigen ligand prediction, antigen presentation, T cell epitope, computational immunology

INTRODUCTION

Recognition of infected or malignant cells by T cells relies on the presentation of immunogenic self and non-self peptides at the cell surface. Two main pathways have been identified for antigen presentation and processing (1–3).

In the class I pathway, intracellular proteins are degraded into small peptides by the proteasome. These peptides are transported into the endoplasmic reticulum by the transporter associated with antigen processing (TAP) protein complex. There, they can bind to human leukocyte antigen class I (HLA-I) molecules in complex with beta2-microglobulin (β 2m). After trafficking to the cell surface, the complexes may be recognized by CD8 T cells. HLA-I proteins are primarily encoded by three genes (HLA-A, HLA-B, and HLA-C), which are widely expressed in most cell types in human. In addition, specialized cell types can express HLA-E, HLA-F, or HLA-G genes. HLA-A, -B, and -C genes (hereafter referred to as HLA-I) are the most polymorphic genes in the human genome and over 12,000 distinct alleles are documented in the human population (4). Humans have in general different combinations of HLA-I alleles and, therefore, express up to six different HLA-I proteins

(two for each gene). HLA-I molecules bind short peptides, mainly 9–11 amino acids, and different HLA-I alleles have distinct binding specificities, which implies that a broad spectrum of peptides can be displayed across different individuals.

In the class II pathway, peptides coming from the degradation of phagocytosed extracellular proteins are presented on HLA-II molecules for recognition by CD4 T cells (5). In addition, endogenous proteins can be presented on HLA-II molecules when degraded through autophagy (6). HLA-II proteins are encoded by several genes (HLA-DRA, HLA-DRB1,3,4,5, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1) and also show a very high level of polymorphism in the humans (except for HLA-DRA). HLA-II form heterodimers (HLA-DRA/HLA-DRB1,3,4,5; HLA-DPA1/HLA-DPB1 and HLA-DQA1/HLA-DQB1). These dimers bind longer peptides (12–20 amino acids) within an open-ended peptide-binding site. Several other steps are involved in presentation of class II epitopes, such as loading on HLA-II molecules catalyzed by HLA-DM, peptide exchange catalyzed by HLA-DO, the presence of other enzymes such as cathepsins or pH gradients (7–10). Unlike HLA-I, HLA-II molecules are mainly expressed on specific professional antigen-presenting cells (pAPCs) such as dendritic cells or B cells (1), and rarely also by cancer cells such as melanoma (11). pAPCs can also uptake exogenous antigens and present them on HLA-I (12). This process is called cross-presentation, and it is crucial for priming of naïve T cells (13, 14). Altogether, the cellular antigen processing and presentation machinery ensures that the restrictive loading of either intracellular (class I) or extracellular (class II) peptides of the right length will take place in specialized cellular compartments.

The set of peptides presented on HLA molecules is called the HLA peptidome, also referred to as immunopeptidome or HLA ligandome. The HLA peptidome is a rich and complex repertoire of peptides that inform T cells about abnormalities in the genome, transcriptome, and proteome of infected or malignant cells (15–17). It is constantly modulated by HLA or peptides' source protein expression levels, by posttranslational modifications and by the many enzymes, chaperones, and transporters that comprise the antigen processing and presentation machinery (7, 18–20). In particular, the catalytic subunits of the constitutive proteasome, the immunoproteasome, and the thymic proteasome are tightly regulated, leading to the production of distinct repertoires of presented peptides in different cell types and under different conditions (21–24).

Historically, the study and predictions of class I and class II T cell epitopes have mainly developed in the field of infectious diseases, and large datasets of peptides displayed at the surface of infected cells and recognized by T cells are available from HIV, dengue, or influenza (25, 26). In the field of cancer immunology, tumor-associated antigens (defined here as genes expressed in cancer cells and not, or very poorly, in normal cells) have received much attention for almost 30 years (27). For instance, T cell recognizing specific epitopes of NY-ESO or MAGE-1 proteins can be found in melanoma patients, indicating that the immune system can mount a response against tumor-specific antigens (27–29). More recently, many evidences have been accumulated indicating that cancer cells express unique mutated antigens, the so-called neoantigens, which can be recognized by the patients'

own (autologous) T cells (15, 30–35). The total number of somatic mutations in some tumors has been shown to correlate with the therapeutic efficacy of checkpoint blockade antibodies (36–39), suggesting that neoantigens could play an important role in tumor immune recognition. Moreover, several studies demonstrated clinical benefit mediated by the administration of highly enriched populations of neoantigen-reactive CD4⁺ and CD8⁺ T cells (34, 40) and by neoantigen-based vaccines (41, 42). Potential neoantigens are typically predicted first by identifying non-synonymous alterations from next generation sequencing data and second by predicting the binding to HLA molecules of peptides encompassing these non-synonymous genetic alterations (43). For these reasons, predictions of peptides presented on HLA-I and HLA-II molecules have gained renewed interest in the field of tumor immunology. Predicted neoantigens need to be then experimentally validated for HLA binding and immune recognition *in vitro* (44–47).

Here, we review approaches developed for predicting antigen presentation in human cells, with a focus on the latest experimental and computational developments to take advantage of in-depth and accurate mass spectrometry (MS) data of HLA peptidomics. Our aim is to describe the main steps of antigen presentation that proved to be successful in making quantitative predictions of antigens. The more biological aspects of antigen presentation and processing are covered in many other reviews (1–3, 8).

MAIN SOURCES OF HLA LIGAND DATA

A cornerstone in our ability to understand and predict antigen presentation has been the experimental identification of specific peptides interacting with HLA molecules. First, from an experimental point of view, HLA-I molecules do not fold stably in the absence of a ligand and, therefore, all biochemical, structural, and functional studies of HLA-I molecules rely on the availability of known HLA-I ligands. Second, all computational methods to predict HLA ligands at a large-scale use data-driven approaches based on sequence patterns identified within known ligands.

Two main classes of experimental assays have been developed to identify HLA ligands. The first class of assays consists of *in vitro* assays. For HLA-I molecules, refolding assays use conformational pan HLA-I antibodies to test whether the HLA-I complex is properly folded in the presence of a peptide (48–52). Peptide-rescuing assays consist of a photo-cleavable peptide that is stripped by UV radiation in the presence of another peptide (53–55). Competitive assays with radiolabeled peptides have been used to determine relative affinity (IC₅₀) (56). Dissociation assays based on radiolabeled β₂m have been used to probe the stability of peptide–HLA-I complexes (57, 58). Surface plasmon resonance techniques can be used to measure actual K_d values (59). *In vitro* binding assays have also been used for HLA-II ligands (60–62). Compared to class I ligands, screening of class II ligands at high throughput is facilitated since HLA-II molecules have an open-ended peptide-binding site. Therefore, peptides can be fixed on plates, which allow for the use of peptide microarrays (63), or directly encoded in different display systems such as phage or yeast display (64, 65).

In vitro binding assays play a central role in our ability to identify T cell epitopes from viral or cancer-specific antigens (66, 67). When used in combination with state-of-the-art prediction tools, they enable rapid validation of predicted targets and are currently key to most neoantigen discovery approaches in cancer immunotherapy (30, 31, 68, 69). The main caveat of *in vitro* assays for HLA-I ligands is that the peptides have to be determined *a priori* and chemically synthesized, since both the C- and N-terminus of most HLA-I ligands need to be free in most cases. This limits the use of high-throughput and unbiased peptide screening technologies. Furthermore, the involvement of the components of the antigen-loading complex is missing in *in vitro* binding assays and, therefore, signals related to antigen loading *in vivo* cannot be captured.

The second type of experimental assays for HLA ligand identification is based on MS measurement of eluted HLA-binding peptides. This approach is the only methodology to comprehensively interrogate the repertoire of HLA ligands presented naturally *in vivo* (16, 18, 70, 71). The best-established HLA peptidomics methodology is based on immunoaffinity purification (IP) of HLA complexes from detergent solubilized lysates, followed by extraction and purification of the peptides. Typically, either anti-pan-HLA class I, anti-HLA-DR, or anti-pan-HLA class II monoclonal antibodies are used. The extracted peptides are then separated by high-pressure liquid chromatography and directly injected into a mass spectrometer. The resulting spectra obtained from the fragmentation of the peptides are compared with *in silico* generated spectra of peptides from protein sequence databases with MS search tools. Therefore, this search is limited to the available databases, usually the annotated human proteome. Moreover, peptides that have features that make them incompatible with ionization, those that are too hydrophobic or too hydrophilic, might not be detected with standard methods. With the new generation of mass spectrometers, thousands of HLA ligands can be identified per sample (15, 18, 72, 73). Cell lines, including human cancer cell lines, tumors, healthy tissues, and body fluids such as plasma have been subjected to HLA peptidomics analyses (18, 70–84). However, MS-based HLA peptidomics approaches have limited sensitivity and require a relatively large amount of biological sample (~1 cm³ of tissue or 1 × 10⁸ cells) (21). Furthermore, despite major improvement in the quality of HLA peptidomics data, one can never exclude small residual contaminations from co-eluted peptides or wrong annotation of spectra depending on the false discovery rate threshold used in spectral searches.

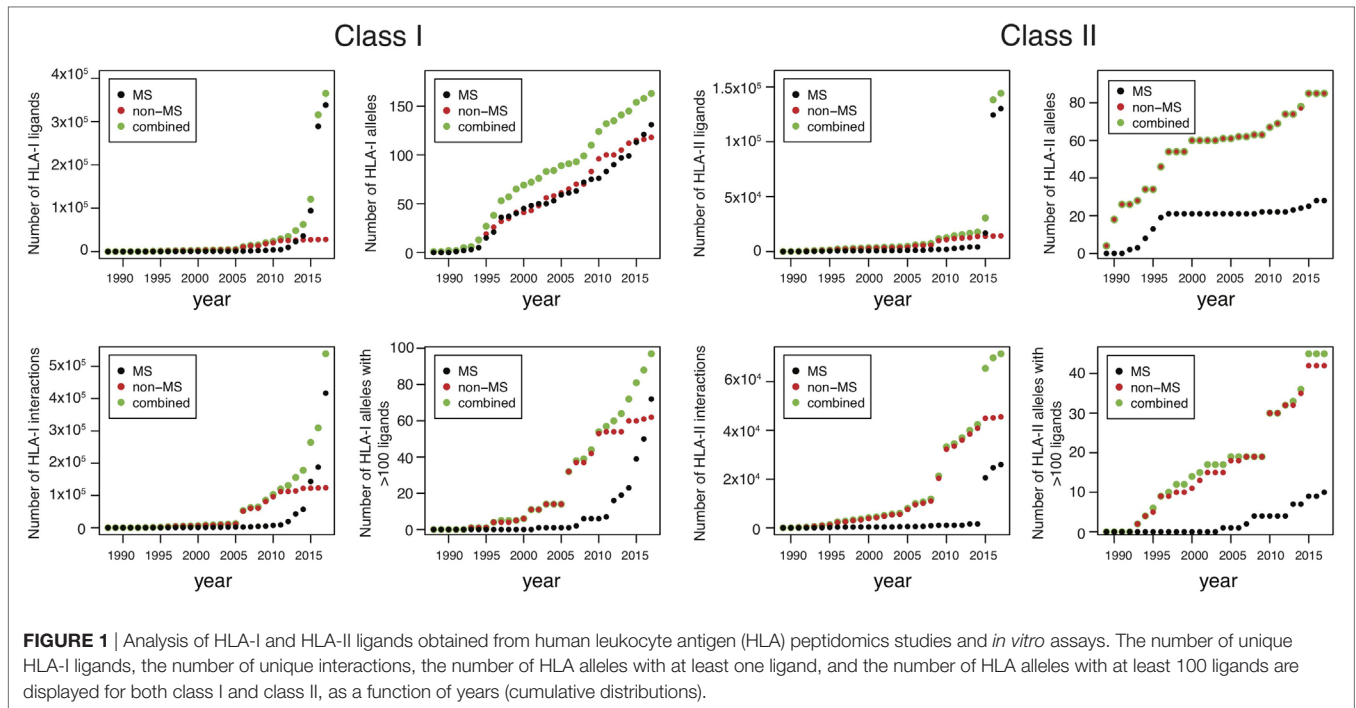
Dedicated proteogenomics computational pipelines for customized reference databases have been developed to expand the search space beyond the canonical human proteome. Customizing references to include somatic alterations observed in tumors have been used for direct identification of neoantigens by MS in murine and human cancer cell line models (31, 35, 80, 85), B cell lymphomas (86), and melanoma tissues (15). Similar approaches were also used for other cryptic peptides resulting from unconventional coding sequences in the genome (87) and new open reading frames (88) (see Non-Canonical HLA-I Ligands).

Historically, the first HLA-I motifs (e.g., HLA-A02:01) were found by looking at peptide sequences of eluted ligands identified

by MS (89, 90). To overcome the fact that eluted peptides come from up to six HLA-I alleles in unmodified cell lines or tissue samples, two experimental approaches have been developed. The first approach consists of transfecting a soluble HLA allele into a cell line and pulling down only the soluble HLA-I molecules in complex with their ligands (91, 92). While it has been shown that the repertoire of peptides presented on transfected soluble HLA-I and the endogenous membranous HLA-I molecules are highly similar (93), the non-physiological expression level of the soluble HLA-I molecules and the potential different environment in the loading compartment could affect the overall peptide repertoire. Furthermore, endogenous HLA-I alleles can be shaded or naturally secreted from cells in culture (94) and could contaminate the secreted peptidome (75). Nevertheless, this approach proved very powerful to identify HLA-I motifs (77, 78, 95–97). Of particular interest is the study by Di Marco and co-authors where the motifs of 15 HLA-C alleles could be determined, together with motif for HLA-G01:01 (75). This detailed view of HLA-C alleles binding specificities enabled the authors of this study to identify for the first time specificity determinant residues in the HLA-C-binding site that provide likely molecular mechanisms explaining the differences observed between HLA-C binding motifs. The second experimental approach consists of using genetically modified cell lines that express only one allele (98, 99) and was used to study binding motifs of highly similar alleles, like HLA-B27:02 to HLA-B27:09 (100). This approach was also recently used to screen 16 HLA-A and HLA-B alleles, and this work confirmed that predictors trained on MS data could improve predictions of naturally presented HLA-I ligands (70). One advantage of this approach is that theoretically all peptides come from one single allele (see above for potential sources of contaminations). In parallel, we and others introduced computational techniques based on motif deconvolution (72, 101) and peptide clustering (102, 103) to accurately determine HLA-I restriction of eluted ligands from pooled samples without requiring to experimentally isolate each HLA-I allele and without relying on HLA-I ligand predictors (see below for a detailed description of these approaches).

Comparison of MS and *In Vitro* Data

Until 2012, the number of MS datasets was significantly lower than *in vitro* data (Figure 1), which partly explains why *in vitro* binding data were mainly used for training HLA-I ligand predictors. However, the situation has changed quite dramatically over the last 4 years. Combining data from IEDB (25) together with recent HLA peptidomics studies (see Supplementary Material), we can observe that roughly 10 times more unique HLA-I ligands and three times more unique HLA-I-peptide interactions are currently available from MS studies (Figure 1, the lower number of interactions than peptides for MS data comes from the fact that several MS samples did not have HLA typing information or allele restriction could not be determined with motif deconvolution). The coverage of HLA-I alleles is also larger in HLA peptidomics samples compared to *in vitro* binding data (Figure 1). Moreover, all curves for MS data do not show signs of saturation, suggesting that these numbers are likely to further increase in the coming years, especially with the growing interest in HLA peptidomics profiling of cancer samples from patients with diverse ethnic



backgrounds for neoantigen discovery (15). Similar observations hold for HLA-II ligands, where the number of unique peptides identified by MS largely exceeds the number of peptides identified in *in vitro* assays. However, the number of HLA-II alleles with documented ligands is still larger for *in vitro* binding data. This likely reflects the fact that HLA-II ligands are easier to screen in a high-throughput way using peptide microarrays, and that allele restriction in HLA-II peptidomics data is still more difficult to determine with motif deconvolution or peptide clustering than for HLA-I peptidomics data.

MODELING HLA-I BINDING SPECIFICITY Allele-Specific Predictors

Modeling HLA-I-binding specificity has been carried out for almost 30 years since the first evidence of HLA-I motifs. Early studies used simple sequence motifs [e.g., xLxxxxx(L/V) for HLA-A02:01]. However, as more data started to accumulate, it became clear that simple motifs were too restrictive and not quantitative enough. To overcome these limitations, position weight matrices (PWM) (equally referred to as Position Specific Scoring Matrices or simply scoring matrices) were introduced (104–107). The basic idea is to compute the frequency of each amino acid at each position in a set of (pre-aligned) peptides. The score of a new peptide can then be computed by multiplying the PWM entries corresponding to the sequence of the new peptide (see Supplementary Material). Although the idea of computing amino acid frequencies is relatively simple to understand, several steps are important when building a predictor based on PWMs. First, one has to consider the amino acid background distribution and use this distribution to renormalize the scores (see Supplementary Material). In most existing approaches, amino acid frequencies of

the human proteome have been used. However, this approach may not be fully justified when using viral epitopes to train predictors. Similarly, eluted HLA-I ligands do not show the same amino acid distribution as human proteins and much lower frequency of cysteine has been reported by ourselves and others (70, 72). As such, the optimal choice of background distribution may depend on the origin (both biological and technical) of the data. Second, in most cases, estimating the frequency of amino acids occurring only a few times (or never) at a given position is highly susceptible to statistical noise. To address this issue, pseudo-counts are often used. A widely used approach is based on the BLOSUM62 matrix (see Supplementary Material) (105, 108, 109). Third, biases due to the design of specific experiments can be found in many *in vitro* datasets. For instance, if a mutagenesis was carried out at a fairly non-specific position in a given epitope, many sequences will have identical amino acids at all positions except the one used in the mutagenesis. One way to correct for such biases is to add a weight to all peptides that is inversely proportional to the number of highly similar sequences in the dataset (see Supplementary Material).

Since the last decade, most allele-specific HLA-I ligand predictors use machine learning frameworks such as neural networks, hidden Markov Models, support vector machines, or convolutional neural networks (110–114). One attractive aspect of these models is the ability to consider potential correlations between different positions within HLA-I ligands. For instance, we recently observed in HLA-B07:02 ligands that arginine is preferred at P3 or at P6, but not at both positions at the same time (101). This type of correlation is not captured by simple PWMs. However, it is still unclear how frequent these correlations are for HLA-I ligands. In particular, although many studies reported improved predictions of HLA-I ligands using machine learning algorithms (112, 115),

one has to be careful before concluding that correlation patterns are prevalent, since improvement in prediction accuracy may also result from more robust regularization frameworks. Finally, machine learning approaches are also susceptible to overfitting and correcting for potential biases in training sets can be more challenging than with simple PWMs.

Pan-Allele Predictors

Enough experimental ligands are available for roughly 100 HLA-I alleles, which represents only a small fraction of the >12,000 HLA-I alleles observed in the human population. To address this issue, pan-allele predictors have been introduced, where the input of the algorithm consists of both the sequence of the ligand and the sequence of the HLA-I allele (or of its binding site) (107, 116–118). These algorithms are powerful at capturing correlations between amino acids in the HLA-I-binding site and in the ligand. The most widely used and likely the most elaborate pan-specific algorithm is the NetMHCpan tool (117), which includes several features specific for HLA-I molecules, such as combining peptides of different lengths in the training and incorporating peptide length preferences.

Table 1 summarizes some of the most common predictors, together with information about the algorithm that is used, the type of training data and the output.

Choosing the Right Training Set

While extensive work has been performed to optimize the algorithms used in HLA-I predictors, less attention has been devoted to the choice of the training set. Prior to 2016, most approaches aimed at predicting binding affinity values (i.e., IC50) and, therefore, were trained on *in vitro* data mainly obtained from IEDB (25). Although high accuracy could be reached for many common alleles, several potential biases suggest that such data can be suboptimal for training predictors. In particular, it is important to remember that most HLA-I ligands tested *in vitro* for binding were first predicted with older versions of HLA-I ligand predictors [some exceptions that used random peptide libraries include Ref.

(58)]. Unfortunately, this can induce circularity when using these data to retrain predictors, and such biases are difficult to detect and correct for. Of note, the same circularity issue can also affect several published MS datasets when HLA-I ligand predictors or motifs were used to assign allele restriction and filter noise. Here, we argue that high-quality MS data not filtered with existing predictors provide a powerful solution toward overcoming the potential circularity inherent to many *in vitro* binding data.

Using MS Data for Identifying HLA-I Motifs and Training Predictors

Mono-allelic samples or transfected soluble HLA-I alleles have been used since many years to study the binding motifs of specific HLA-I molecules (91, 92). However, due the experimental work implied by such approaches, they were never applied to a large panel of HLA-I alleles [the largest studies consist of 16 alleles for mono-allelic cell lines (70) and 17 alleles for transfected soluble HLA-I alleles (75)]. For pooled HLA peptidomics dataset, the impossibility to experimentally assign allelic restriction was often considered as an important hurdle to use such data toward studying HLA-I-binding motifs.

However, in the last few years, it became clear that pooled HLA peptidomics data can be used to study HLA-I motifs and improve predictions, thereby overcoming the need of genetically modifying cell lines or transfecting soluble HLA-I alleles. The first attempt to determine HLA-I-binding motifs from pooled HLA peptidomics data was published in 2015 (18). A year later, we published the first evidence that such data can be used to improve predictions of HLA-I ligands (101). Since then, many studies have confirmed these results both for the identification of new motifs (72, 81, 102, 103, 119) and for improving predictions of HLA-I ligands by integrating MS data in the training of predictors (70, 72, 117, 120).

As of today, two algorithms have been used for motif deconvolution and peptide clustering of pooled HLA peptidomics data. One of them (MixMHCp) is based on mixture models and was initially developed for multiple specificity analysis in large PDZ

TABLE 1 | Summary of some of the most recent or most widely used human leukocyte antigen (HLA)-I predictors with available web interface or code repository.

Name	Training data	Output	Algorithm	Allele coverage	Access	Reference
NetMHC4.0	BA	BA	NN	S	http://www.cbs.dtu.dk/services/NetMHC/	(110)
NetMHCpan4.0	BA + MS	R (BA)	NN	Pan	http://www.cbs.dtu.dk/services/NetMHCpan-4.0/	(117)
MixMHCpred	MS	R	PWM	S	https://github.com/GfellerLab/MixMHCpred	(72, 120)
MHCflurry	BA	BA	NN	S	https://github.com/openvax/mhcfurry	(113)
PickPocket	BA	BA	PWM	Pan	http://www.cbs.dtu.dk/services/PickPocket/	(107)
NetMHCstabpan	BS	BS	NN	Pan	http://www.cbs.dtu.dk/services/NetMHCstabpan/	(118)
NetMHCstab	BS	BS	NN	S	http://www.cbs.dtu.dk/services/NetMHCstab/	(111)
NetMHCcons	BA	BA	C	S	http://www.cbs.dtu.dk/services/NetMHCcons/	(181)
IEDB consensus	BA	R	C	S	http://tools.iedb.org/mhci/	(182)
SMMPMBEC	BA	R	PWM	S	https://github.com/ykimbiology/smmpmbec	(104)
MHCnuggets	BA	BA	NN	S	https://github.com/KarchinLab/mhcnuggets-2.0	(183)
ConvMHC	BA	R	NN	Pan	http://jumong.kaist.ac.kr:8080/convmhc	(116)
HLA-CNN	BA	R	NN	S	https://github.com/uci-cbcl/HLA-bind	(114)
SYFPEITHI	BA + MS	R	PWM	S	http://syfpeithi.de/0-Home.htm	(106)
PSSMHCpan	BA	BA	PWM	Pan	https://github.com/BGI2016/PSSMHCpan	(184)

Column 2, BA, binding affinity; BS, binding stability; MS, HLA peptidomics data; column 3, BA, binding affinity; R, ranking; column 4, NN, Neural network (including deep networks); PWM, position weight matrices; C, consensus; column 5, S, allele specific; Pan, pan-class I.

or SH3 ligand datasets obtained by phage display (121–123). In this framework, the idea is to let the algorithm infer K distinct PWMs that optimally model the eluted peptides (101). Since peptides identified by MS come from K different HLA-I alleles ($K \leq 6$), it is not surprising that the motifs that optimally describe the data correspond precisely to the specificity of these alleles. The other algorithm (GibbsCluster) is based on simulated annealing to group the peptides into different clusters optimizing a global cost function that models how well each peptide fits into its respective cluster (103, 124). Somehow unexpectedly, both algorithms were initially developed for other purposes (i.e., multiple specificity analysis for MixMHCp and simultaneous clustering and alignments of short peptides for GibbsCluster) and their use for motif identification in HLA peptidomics data was realized only later (18, 101, 102). The two approaches have many conceptual similarities since the likelihood function optimized in MixMHCp differs only slightly from the cost function optimized in GibbsCluster. In practice, the two algorithms lead most of the time to very similar results for HLA-I peptidomics data (101) and nearly identical motifs as those obtained from mono-allelic samples or transfected soluble alleles (72) (see also examples in Figure S1 in Supplementary Material). In some cases, as we have reported, the mixture model tends to be slightly more sensitive to identify motifs supported by few peptides, such as those describing HLA-C alleles (101). Conversely, the GibbsCluster has several advantages, such as the ability to combine peptides of different lengths and the simultaneous clustering and alignment of the peptides (which is critical for HLA-II ligands) (102, 103). Both methods can be used as command line or through web servers (see <http://www.mixmhcp.org> and <http://www.cbs.dtu.dk/services/GibbsCluster-2.0/>). The availability of these algorithms strongly supports the notion that allele assignment in MS data should not be done based on HLA-I ligand predictors, since this may remove all peptides that are not well modeled with existing predictors, and hence bias determination of motifs and prevent improving the predictors. It is also important to emphasize that accurate motif deconvolution requires a large number of peptides, and ideally, many samples to test the robustness of the motifs (72). For this reason, it is likely the combination of higher accuracy and throughput of MS instruments (18) together with these novel algorithms that enabled accurate HLA-I motifs identification in pooled HLA peptidomics data.

Annotation of the motifs deconvolved from pooled HLA peptidomics data can be done in different ways. For alleles for which a reasonable description of the motifs is known, one can simply compare the motifs found in MS data to the known references (18). Using Euclidean distance to quantify the similarity between the PWMs appears to provide stable results and most of the time the mapping is quite obvious (72, 101). If the motifs are not known, two approaches have been developed. One fully unsupervised approach was proposed by ourselves based on co-occurrence of HLA-I alleles across different samples (72). In this way, we could identify and annotate HLA-I motifs for more than 40 alleles, including 7 alleles that had no experimental ligands at the time of this study. Another semi-supervised approach that works well in most cases consists of comparing with motifs predicted from pan-allele predictors such as NetMHCpan (119).

An important limitation of motif deconvolution approaches comes from the fact that motifs for some alleles (especially HLA-C alleles) are more difficult to detect in many samples. Also, in the presence of highly similar motifs (e.g., HLA-A23:01 and HLA-A24:02, or HLA-C07:01 and HLA-C07:02), the two motifs often cannot be split (72). Because of this, not all HLA peptidomics datasets are appropriate for training predictors for each allele expressed in the corresponding sample. This limitation can be alleviated by considering large collections of HLA peptidomics studies and focusing on cases where the motifs are clearly visible and can be unambiguously annotated (72). Finally, it is sometimes useful to consider more motifs than the number of alleles in order to identify motifs for each allele (Figure S2 in Supplementary Material).

Figures 2–4 summarize the HLA-A, HLA-B, and HLA-C motifs currently available by combining motifs deconvolved from recent MS studies together with IEDB data (see Supplementary Material). As expected, the clustering based on the similarity between the motifs (see Supplementary Material) broadly recapitulates the supertype assignment for HLA-A and HLA-B alleles and helps highlighting differences among alleles classified within the same superotypes.

Biases in MS Data

While MS data are not suffering from the potential circularity present in many *in vitro* binding data, they are not free from any biases. First, as already mentioned, only peptides that are part of the database used for spectral searches can be detected in HLA peptidomics data, or else, the less accurate *de novo* method may be applied. This has direct implication for cysteine-containing peptides. Since this amino acid can be chemically modified by oxidation and as such modifications are typically not included in standard MS searches, cysteine occurs at very low frequency in HLA peptidomics datasets. Attempts to correct for this bias when training predictors tried to renormalize PWMs based on observed amino acid frequencies at non-anchor positions (72) or expand the MS spectral search to include modified cysteines (70). Second, peptides that are too hydrophobic or too hydrophilic might be missed applying the common purification methods that rely on retaining peptides through hydrophobic interactions with the solid phase. Furthermore, some peptides have features that make them incompatible with ionization or lead to poor fragmentation. Combining fragmentation methods, such as higher-energy collision-induced dissociation and electron-transfer dissociation, have been shown to improve spectra annotation of HLA peptides (73). Despite these limitations, inspection of HLA peptidomics data and comparison with motifs obtained from *in vitro* data did not reveal major differences, except for the low frequency of cysteine [slightly higher frequency of charged amino acids at some positions has been reported in some studies (101, 102)]. Third, immuno-purification based MS data cannot distinguish between HLA-I ligands presented on the cell surface from those resident in the ER. This can be achieved by purifying HLA-I peptides from the cell surface by mild acid elution (125, 126). However, in a head-to-head comparison, the IP method outperformed the mild acid elution in terms of peptide recovery (127). Last, when considering MS

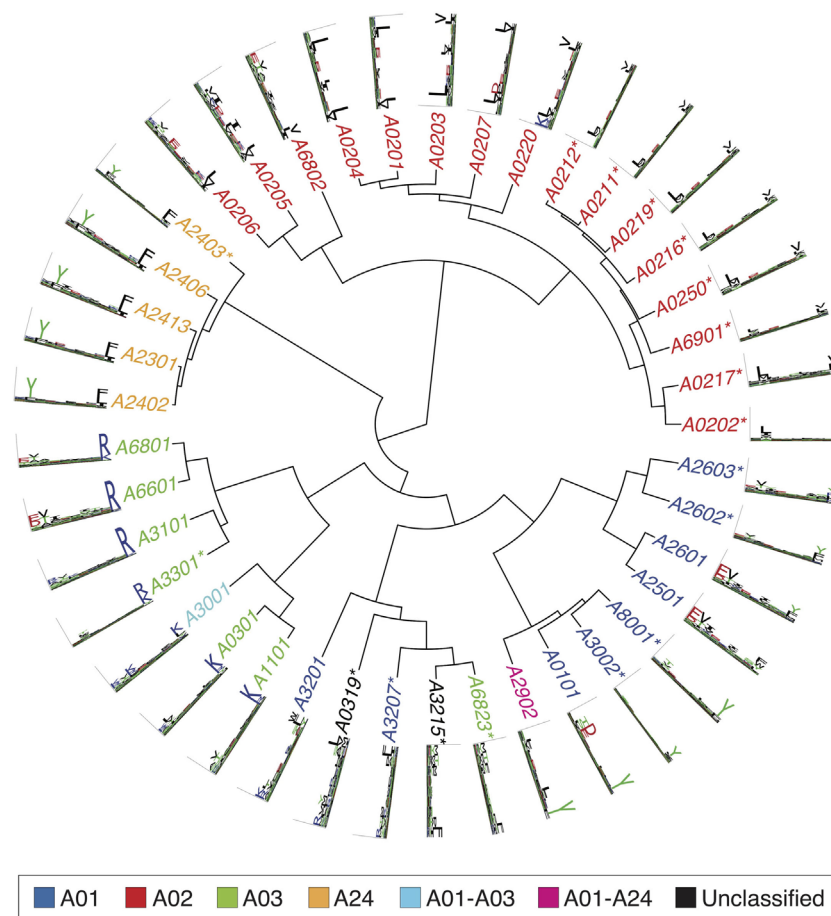


FIGURE 2 | Hierarchical clustering of HLA-A alleles based on their binding specificity. Stars indicate cases where only *in vitro* binding data were available to generate the motifs. In all other cases, only mass spectrometry data were used. Name colors and their descriptions in the legend indicate supertypes as defined in Ref. (185).

data, it is important to remember that these peptides come from human proteins and that proteins or domains within proteins can display significant homology (especially for class II ligands where in addition many peptides can originate from the same core region). This can artificially enhance the frequency of some amino acids. This issue is especially important when building random models of MS data to infer whether amino acid frequencies (either within a motif or at flanking regions) differ from what is expected by chance.

MODELING HLA-II-BINDING SPECIFICITY

Predictions of HLA-I ligands, especially with the recent incorporation of high-quality MS data in the training of predictors, have reached a high level of accuracy (70, 72, 117, 120). The situation is unfortunately not the same for HLA-II ligands, which are still much more difficult to predict despite the large amount of experimental data acquired over the years (Figure 1). Several challenges arise when modeling the binding specificity of HLA-II alleles. First, HLA-II alleles tend to have more degenerate and less specific motifs. Second, all current approaches rely on first aligning

peptides with tools such as NN-align (128). Although these tools have been optimized to handle HLA-II ligands, automated alignment of small peptides is known to be a difficult computational problem. Finally, the fact that HLA-II molecules form dimers further increases the diversity for HLA-DP and HLA-DQ alleles where both members of the dimers are polymorphic. Allele-specific HLA-II ligand predictors include NetMHCII (129), ProPred Singh (130), MHCpred (131), TEPITOPE (132), and consensus methods (133). Pan-specific class II predictors mainly consist of NetMHCIIpan (129). While all these predictors show better than random performances, their accuracy is lower than for HLA-I ligand predictors. This may be due to the challenges of determining class II motifs, as well as to the complex machinery of class II presentation, whose specificity is still poorly understood from a quantitative and predictive point of view [see Ref. (7–10) for a detailed discussion of the more biological aspects of this process and the importance of HLA-DM and other enzymes]. In particular, it appears that properties such as conformational flexibility play a role in loading onto HLA-II molecules (134), and these properties are difficult to predict directly from peptide sequences.

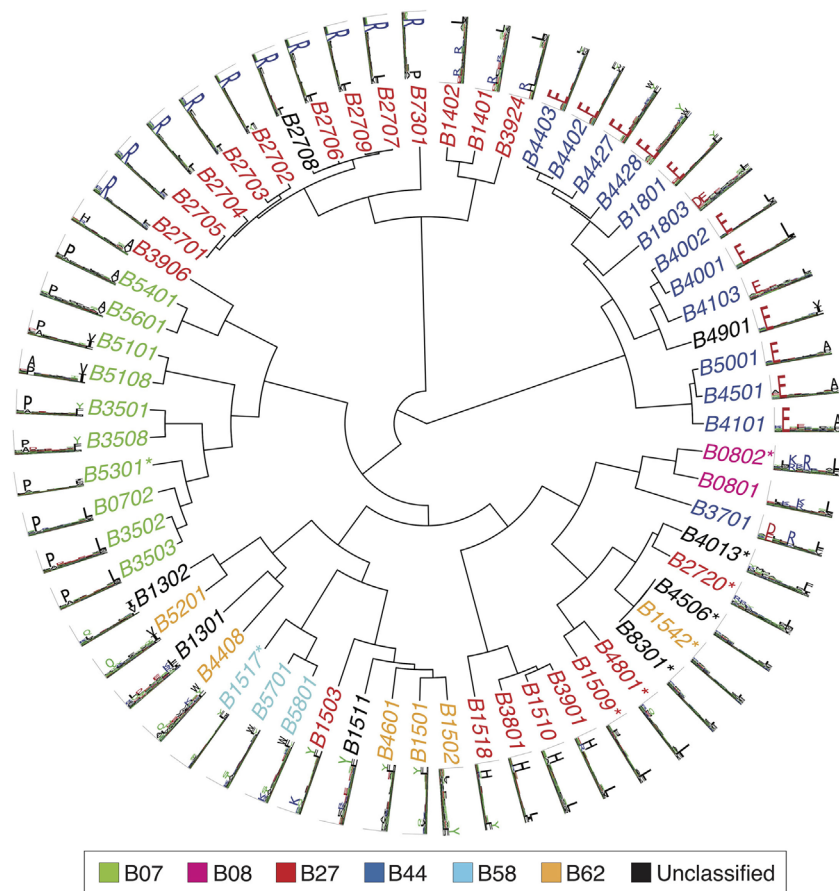


FIGURE 3 | Hierarchical clustering of HLA-B alleles based on their binding specificity. Stars indicate cases where only *in vitro* binding data were available to generate the motifs. In all other cases, only mass spectrometry data were used. Name colors and their description in the legend indicate supertypes as defined in Ref. (185).

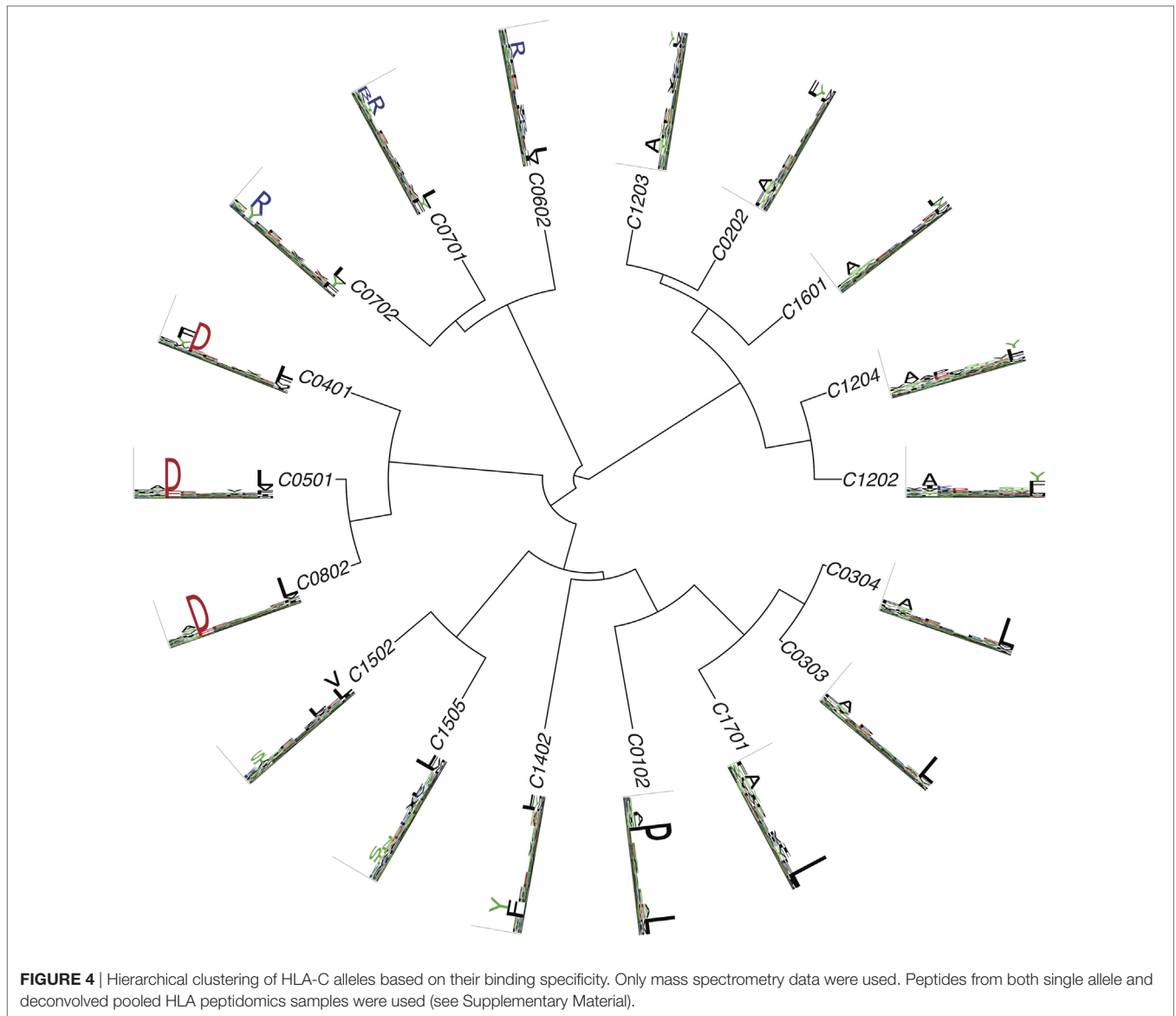
Whether similar improvement for class II predictions as for class I will be reached by incorporating class II peptidome data in the training of algorithms has not been investigated at a large scale. Nevertheless, it has been recognized already long ago that eluted ligands could provide important information about HLA-II-binding motifs (135). More recently, HLA-II peptidomics was performed in BALB/c and C57BL/6 mice and demonstrated that clear motifs for H-2 I-Ad and H-2 I-Ab could be obtained (136). A subsequent study suggested that predictors trained on these data perform better than NetMHCIIpan when repredicting the MS data (137). A similar strategy was carried out in transgenic DR1+ and DR15+ mice to identify the motifs of these two alleles (138). Recent studies also indicate that motif deconvolution with the GibbsCluster algorithm may work in pooled HLA-II peptidome datasets (21, 139), which could lead to refinement of HLA-II motifs and improved predictions in the coming years, as suggested in a recent preprint (140). However, the results are still more challenging to interpret and some motifs predicted by GibbsCluster are difficult to annotate, while the motifs for some alleles are sometimes not detected (21, 139, 141).

INVESTIGATING OTHER PROPERTIES OF HLA-PEPTIDE INTERACTIONS

Many other important properties of HLA-I molecules beyond the 9-mer-binding motifs themselves can be studied through the analysis of HLA peptidomics data.

Peptide Length Distribution

Arguably, the most important information beyond the binding motifs that can be extracted from MS data is the characterization of peptide length distributions. Many studies have demonstrated high heterogeneity of peptide length distributions between different alleles, with alleles such as HLA-B51:01 displaying high frequency of 8-mers (only slightly smaller than 9-mers) and very few longer peptide, while others such as HLA-A01:01 show high frequency of longer (≥ 12 amino acids) peptides, which can still be recognized by T cells (15, 70, 97, 103, 142). Structurally, most longer peptides are known to form bulges, with anchor residues conserved at the second and last positions of the peptides. Some patterns emerged from analysis of peptide length distributions. For instance, HLA-I alleles with anchor residues at middle



positions (e.g., HLA-B08:01, HLA-B14:01, HLA-B14:02, HLA-B37:01) displayed peptide length distributions peaked at 9-mers, which is consistent with the fact that the middle anchor residue needs to be structurally conserved in the presence of an anchor at such positions (101). The study by Trolle and co-authors (97) demonstrated that peptide length distributions observed in MS data for five alleles could not be simply explained by differences in binding affinity, suggesting that the pool of peptides available for loading in the ER is skewed toward 9-mers. This likely implies that predictors trained on MS data will differ from those predicting binding affinity when comparing peptides of different lengths. In a recent preprint (143), we performed a large-scale analysis of peptide length distributions across 85 HLA-I alleles and could identify clusters of HLA-I molecules based on the similarity of their peptide length distributions. Peptide length distribution has been incorporated into the latest versions of NetMHC and NetMHCpan, by adding one additional input node encoding

for peptide length in the neural networks (110, 117), and into MixMHCpred by directly fitting distributions observed in MS data (143).

As observed in our recent paper (21), peptide length distribution can also be affected by different treatments such as INF γ likely due to modulating the activity of catalytic subunits of the proteasome, and these aspects are not captured by existing predictors.

C- And N-Terminal Extensions

Human leukocyte antigen peptidomics data have been instrumental in exploring non-canonical binding modes in HLA-I ligands. In particular, several recent studies have used MS data to study C- and N-terminal extensions in HLA-I ligands. Although such extensions had been identified long ago [first crystal structure in 1994 (144), PDB:2CLR, followed by another structure in 2009 (145), PDB:3GIV], their prevalence had remained unclear.

In 2016, HLA peptidomics profiling and X-ray crystallography were combined to explore C-terminal extensions in HLA-A02:01 and demonstrated that such extensions were especially common among peptides originating from pathogens (146). This was followed by additional work that better described the structural mechanisms and cellular origin of such extensions (147). N-terminal extensions have been identified in HLA-B57:01 (148) and HLA-B58:01 (149). More recently, we have demonstrated that C-terminal extension occur in a substantial fraction of HLA-I molecules and can be recognized by CD8 T cells (120). Our work further enabled us to identify both sequence and structural features predictive of such extensions. In particular, it appeared that C-terminal extensions are especially frequent in alleles displaying specificity for positively charged residues at the last anchor position (e.g., HLA-A03:01, HLA-A31:01, HLA-A68:01). While MS data potentially provide a rich source of information about C- and N-terminal extensions, identifying these extensions by looking at the sequence of the peptides can be challenging, especially when the residue at the extension has similar specificity as the anchor residue (i.e., same residues at P9 and P10 for putative C-terminal extensions, same residues a P2 and P3 for putative N-terminal extensions). Our work suggests that many ambiguous cases may actually follow the bulging conformation (120).

Posttranslationally Modified HLA-I Ligands

Posttranslationally modified peptides have been identified by MS analysis of eluted ligands (15, 150–152). These include mainly phosphorylated peptides, which can be recognized by T cells (153–155). Phosphorylation was observed to occur mainly at position 4 (15), suggesting that it does not impact too much the binding to the HLA-I molecules. Existing HLA-I ligand predictors do not include phosphorylated peptides, although the increasingly larger MS datasets of phosphorylated HLA-I ligands suggest that predictions of phosphorylated HLA-I ligands may soon become feasible. As for now, one approach is to treat the phosphorylated residue as its unmodified counterpart and use available predictors to predict such ligands.

HLA-II Molecules

Fewer studies used MS data to investigate properties of HLA-II molecules other than the actual-binding motifs. Studies reported broad peptide length distributions peaked around 15-mers (15, 21, 139, 156, 157), but it is still unclear to what extent distinct alleles show distinct peptide length distributions. Other properties of HLA-II molecules that could be studied based on MS data include the cellular origin of class II peptides (156, 158, 159) and the impact of different biological processes such as autophagy (160). MS studies also indicated preference for proline at the second and second to last position of peptides degraded in the endolysosomal pathway (156, 161), and preference for lysine at the C-terminus and for aspartate at the N-terminally flanking residue of class II epitopes degraded in the cytosolic pathway (156). Along these lines, many studies support the idea that presentation of class II peptides is not only driven by the binding specificity to the HLA-II molecules but also involves some (still uncharacterized)

specificity in the processing machinery, flanking regions (162), or presentation hotspots in the human proteome (159).

Considering the increasingly higher quality and throughput of class II HLA peptidomics data (15, 21, 86, 138, 139), we anticipate that analysis of HLA-II peptidomes will further enable researchers to investigate new properties of HLA-II molecules. For instance, it will be interesting to see whether the presence of bulging class II ligands, as recently reported from an analysis of *in vitro* binding data (163), can be confirmed in large-scale unbiased MS data.

ANTIGEN PRESENTATION—BEYOND BINDING TO HLA

Integrating Cleavage Site and TAP Transport Predictions, Signals from Flanking Regions and Other Proteomic Information

Mass spectrometry-based HLA peptidomics analysis can reveal crucial information about the rules underlying the biogenesis of the HLA peptidome, including signatures of cleavage site specificity, influence of source protein expression or other patterns characterizing naturally presented HLA ligands. Predictions of cleavage sites have been available since many years and have been used to narrow-down the list of predicted HLA-I ligands (164). Although some improvement has been observed, cleavage site predictions have only a limited effect on prediction accuracy of naturally presented HLA ligands. For this reason, it is not widely used in many existing pipelines for neoantigen predictions from exome sequencing data, for instance. Predictions of TAP transport has also been integrated with affinity and cleavage site predictions to model antigen presentation (165–167). Interrogation of properties of thousands of HLA-I ligands source-proteins has revealed that the proteome is not randomly sampled. Several biological determinants correlate with presentation, such as level of translation (71), expression, and turnover rate (18) and selective regions of the human proteome (71). Specific amino acid signals in flanking regions of naturally presented HLA-I ligands, like lower frequency of proline, have also been demonstrated (70). While binding to HLA still appears to be the most selective step of class I antigen presentation, integrating these additional features into a single predictor further improves the accuracy of predictions of naturally presented peptides (70, 71).

Presentation Hotspots

After deep interrogation of HLA peptidomics large scale data, we and others have recently suggested that HLA ligands are not randomly distributed along the protein sequences but are located within “hotspots” (15, 71), which fit proteasomal cleavage, peptide processing, and HLA-binding rules (168). Recently, we envisioned that these hotspots reflect regions of proteins with enhanced proteasomal or endosomal peptide production prior to HLA loading and may, therefore, provide complementary information to HLA-binding predictions (159). To this end, we collected a large dataset of MS detected HLA class I and class

II ligands from different cancer and healthy tissues and variety of cell lines. We used this dataset to score potential neoantigens based on how well their un-mutated source proteins are naturally presented. In a proof of concept study, we tested this hypothesis with published data (33) and could show that MS-based features improved the prioritization of confirmed neoepitopes (159). Large scale databases of HLA peptidomics data capture the global nature of the *in vivo* peptidome averaged over many HLA alleles and, therefore, reflect the propensity of peptides to be presented, which can complement binding-affinity predictions.

FUTURE PERSPECTIVES

Expanding the Description of HLA Motifs

Accurate and unbiased binding motifs are available for a bit more than 100 HLA-I alleles (Figures 1–4). This is only a tiny fraction of the >12,000 HLA-I alleles listed in IMGT/HLA database (4). For this reason, much has still to be learned about the specificity of HLA-I molecules. We anticipate that the ability to deconvolve HLA-I motifs from pooled HLA peptidomics data will play an important role to expand our understanding of HLA-I-binding specificities. This is especially promising in light of the current interest in using MS to identify neoantigens in cancer patients. However, even with the current efforts in HLA peptidomics, extrapolation of the curves in Figure 1 suggests that experimentally determined HLA-I ligands will remain available for only a small fractions of HLA-I alleles in the coming years. For this reason, development of pan-specific HLA-I ligand predictors leveraging high-quality MS data available for a few (~100) alleles to model the binding specificity of other alleles are expected to play an important role in broadening the scope of HLA-I ligand predictions to rarer alleles without document ligands (117). Accurate and in-depth HLA peptidomics data will also likely play an important role in improving our understanding and description of HLA-II motifs. Use of HLA-II gene-specific antibodies (i.e., pan-DR, pan-DP, or pan-DQ) may facilitate accurate motif deconvolution in such datasets.

Better Understanding of Antigen Presentation

While binding to HLA molecules is the most specific and best quantitatively characterized step of the antigen presentation process, it is likely that some additional filtering comes from cleavage in the proteasome, transport with TAP, and loading in the ER. As mentioned earlier, several recent studies suggest that including these additional parameters further improves prediction accuracy (70, 71, 159, 166). One of the challenges there is to disentangle real biological signals from potential technical biases in MS data. Despite this caveat, it is likely that accumulating very large datasets of naturally presented HLA-I ligands is the only way to improve the accuracy of models of antigen presentation that go beyond the binding to HLA molecules. In addition, it could provide new information about how the HLA peptidome can be remodeled in response to extracellular signals, such as IFN γ stimulation (19, 21). We, therefore, envision that screening how inhibition or activation of components of the antigen processing

and presentation affect the nature of naturally presented HLA ligands on a large scale may reveal their role in shaping the HLA peptidome.

Non-Canonical HLA-I Ligands

Increasing evidences also suggest that non-canonical and cryptic peptides contribute to the HLA peptidome and expand the range of putative T cell epitopes. Laumont et al. have constructed a reference database of stop-to-stop translation products of six open reading frames of expressed RNAs and revealed that about 10% of the peptidome derive from presumably noncoding genomic sequences or exonic out-of-frame translation (87). Liepe et al. have reported that around 30% of the peptidome is derived from non-contiguous peptides spliced by the proteasome (169). Unexpectedly, spliced peptides displayed significantly lower predicted affinity than the normal peptides identified in the same samples (169) and did not show the expected HLA-I motifs. A very large database that is about two orders of magnitude larger than the typical protein-coding database was used to incorporate theoretical spliced products (169). Searching such large databases, especially in order to identify HLA peptides that have no enzymatic restrictions, may lead to improper control of false positives (170). In a recent preprint (171), we proposed an alternative, more conservative, approach to identify spliced peptides among HLA-I ligands based on *de novo* interpretation of high-quality spectra, suggesting that the number of such peptides may have been overestimated in the original study. The exact amount of spliced HLA-I ligands is still a matter of debate, and further studies will be needed to precisely estimate the fraction of spliced peptides actually displayed on HLA-I molecules. However, these potential issues suggest that putative spliced peptides may not all be appropriate for training HLA-I ligand predictors. Exploring non-canonical HLA ligands derived from translation of non-conventional regions in our genome or posttranslation events such as splicing is like finding a needle in a haystack. *In silico* predictions of such potential HLA ligands with existing tools may, therefore, lead to in-controlled numbers of false-positives, since the non-canonical space is theoretically orders of magnitude larger than the current canonical protein space. Hence, intensive proteogenomics based investigation of acquired HLA peptidomics data will likely play a central role in this endeavor and will require advanced computational tools and statistics to properly control for false positives.

Toward Predictions of Immunogenicity

Recent years have witnessed an unprecedented growth of in-depth and accurate MS data (Figure 1) that significantly enhanced our ability to predict antigen presentation. Unfortunately, these data cannot inform us about the most critical step in immune recognition, namely, the recognition of presented antigens by T cells. Much less is known there, and it is for instance, a disappointing fact that most predicted neoantigens from mutations found by exome sequencing of tumors are not recognized by T cells, although many resulting peptides do bind to HLA-I molecules. While direct identification of mutated peptides presented on the surface of cancer cells will likely improve the fraction of truly immunogenic epitopes (101), it is likely that many mutated

peptides seen by MS will still not be immunogenic. Moreover, although binding affinity has been demonstrated to be useful for enriching pools of peptide in immunogenic epitopes (especially for class I), many known immunogenic epitopes display low-binding affinity, suggesting that they would be missed by approaches based on affinity predictions only. This is especially true for class II epitopes, where clear evidences indicate that different enzymes, peptide exchange mediated by HLA-DM or HLA-DO, pH gradients and peptide conformational flexibility play a role in selecting immunodominant epitopes (8–10, 134). Unfortunately, currently, very little of this biological knowledge about class II antigen presentation could be used to improve predictions of class II epitopes.

Work by Calis et al. (172) suggested that some amino acids at non-anchor positions confer increased immunogenicity to HLA-I ligands. More recently, it has been observed that dissimilarity to self among mutated peptides predicted to have similar binding affinity as their wild-type counterpart can further help predicting immunogenic epitopes (173). Differences between the affinity of the wild-type and the mutated peptide, as well as stability of the MHC-I peptide interaction were also suggested to narrow down the list of immunogenic epitopes (174). Unfortunately, datasets of true immunogenic peptides from cancer or infectious diseases are still restricted to a few 100 peptides, limiting the power of machine learning approaches to infer properties of immunogenic epitopes (175, 176). This is likely the main bottleneck toward our understanding of the determinants of immunogenicity. Therefore, recent high-throughput methods for screening T cells using for instance DNA barcoded multimers have the potential to provide critical information about the differences between immunogenic and non-immunogenic peptides (46). Importantly, most of these approaches require to select *a priori* the HLA ligands to be screened [with the exception of a recent phage display system (177)]. Therefore, improved prediction of HLA ligands and antigen presentation will likely play an important role in optimizing the set of ligands currently tested for immunogenicity.

CONCLUSION

The first HLA-I motifs were described almost 30 years ago by looking at sequences obtained from MS analysis of eluted MHC-I ligands (89, 90). Since then, much has been learned about HLA-I and HLA-II molecules through the analysis of their ligands. In human, this has resulted in a detailed description of HLA-I alleles

binding specificities for the most common alleles and culminated with the development of pan-allele predictors. Recent years have witnessed an explosion of new high-quality data generated by MS about HLA-I ligands. Combined with advances in algorithms to analyze such data, this has led to refinement of known HLA-I motifs, discovery of new HLA-I motifs, characterization of peptide length distributions, analysis of N- and C-terminal extensions, characterization of antigen processing signals in flanking regions, analysis of the interplay between gene/protein expression, protein localization and peptide presentation, and evidences for presentation hotspots in the human proteome. For HLA-II ligands, MS studies have been recently used to study HLA-II motifs, suggesting that similar improvements may be observed there as well (21, 138–140). Moreover, the current interest in neoantigen discovery will likely result in many more HLA peptidomics datasets from donors with diverse HLA backgrounds and different pathogeneses. This will provide unique opportunities to further improve our understanding of the rules of antigen presentation. To this end, it will be crucial that raw MS data are made publicly available, and that the reporting of HLA peptidomics data will comply with the recent minimal information about an Immuno-Peptidomics Experiment (MIAIPE) guidelines (178). Databases such as IEDB (25), PRIDE (179), or the SysteMHC Atlas (180) play a key role in this process, and it is our hope that soon all journals publishing HLA peptidomics studies will require deposition of the raw MS data in PRIDE and unfiltered lists of peptides in appropriate databases, or at least accessible in supplementary datasets.

AUTHOR CONTRIBUTIONS

DG and MB-S designed the review and wrote the manuscript. DG analyzed the data and prepared the figures.

ACKNOWLEDGMENTS

We thank Julien Racle for insightful comments about the manuscript. DG is supported by the Swiss Cancer League (KFS-4104-02-2017-R).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <https://www.frontiersin.org/articles/10.3389/fimmu.2018.01716/full#supplementary-material>.

REFERENCES

1. Neeftjes J, Jongsma MLM, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol* (2011) 11:823–36. doi:10.1038/nri3084
2. Blum JS, Wearsch PA, Cresswell P. Pathways of antigen processing. *Annu Rev Immunol* (2013) 31:443–73. doi:10.1146/annurev-immunol-032712-095910
3. Vyas JM, Van der Veen AG, Ploegh HL. The known unknowns of antigen processing and presentation. *Nat Rev Immunol* (2008) 8:607–18. doi:10.1038/nri2368
4. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SGE. The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res* (2015) 43:D423–31. doi:10.1093/nar/gku1161
5. Roche PA, Furuta K. The ins and outs of MHC class II-mediated antigen processing and presentation. *Nat Rev Immunol* (2015) 15:203–16. doi:10.1038/nri3818
6. Crotzer VL, Blum JS. Autophagy and its role in MHC-mediated antigen presentation. *J Immunol* (2009) 182:3335–41. doi:10.4049/jimmunol.0803458
7. Yin L, Maben ZJ, Becerra A, Stern LJ. Evaluating the role of HLA-DM in MHC class II-peptide association reactions. *J Immunol* (2015) 195:706–16. doi:10.4049/jimmunol.1403190
8. Kim A, Sadegh-Nasseri S. Determinants of immunodominance for CD4 T cells. *Curr Opin Immunol* (2015) 34:9–15. doi:10.1016/j.coi.2014.12.005
9. Sadegh-Nasseri S, Kim A. MHC class II auto-antigen presentation is unconventional. *Front Immunol* (2015) 6:372. doi:10.3389/fimmu.2015.00372

10. Sadegh-Nasseri S. A step-by-step overview of the dynamic process of epitope selection by major histocompatibility complex class II for presentation to helper T cells. *F1000Res* (2016) 5:1305. doi:10.12688/f1000research.7664.1
11. Johnson DB, Estrada MV, Salgado R, Sanchez V, Doxie DB, Opalenik SR, et al. Melanoma-specific MHC-II expression represents a tumour-autonomous phenotype and predicts response to anti-PD-1/PD-L1 therapy. *Nat Commun* (2016) 7:10582. doi:10.1038/ncomms10582
12. West MA, Wallin RPA, Matthews SP, Svensson HG, Zaru R, Ljunggren H-G, et al. Enhanced dendritic cell antigen capture via toll-like receptor-induced actin remodeling. *Science* (2004) 305:1153–7. doi:10.1126/science.1099153
13. Ackerman AL, Cresswell P. Cellular mechanisms governing cross-presentation of exogenous antigens. *Nat Immunol* (2004) 5:678–84. doi:10.1038/ni1082
14. Cruz FM, Colbert JD, Merino E, Kriegsman BA, Rock KL. The Biology and underlying mechanisms of cross-presentation of exogenous antigens on MHC-I molecules. *Annu Rev Immunol* (2017) 35:149–76. doi:10.1146/annurev-immunol-041015-055254
15. Bassani-Sternberg M, Bräunlein E, Klar R, Engleitner T, Sinitcyn P, Audehm S, et al. Direct identification of clinically relevant neopeptides presented on native human melanoma tissue by mass spectrometry. *Nat Commun* (2016) 7:13404. doi:10.1038/ncomms13404
16. Caron E, Kowalewski DJ, Chiek Koh C, Sturm T, Schuster H, Aebersold R. Analysis of major histocompatibility complex (MHC) immunopeptidomes using mass spectrometry. *Mol Cell Proteomics* (2015) 14:3105–17. doi:10.1074/mcp.M115.052431
17. Vaughan K, Xu X, Caron E, Peters B, Sette A. Deciphering the MHC-associated peptidome: a review of naturally processed ligand data. *Expert Rev Proteomics* (2017) 14:729–36. doi:10.1080/14789450.2017.1361825
18. Bassani-Sternberg M, Pletscher-Frankild S, Jensen LJ, 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) 14:658–73. doi:10.1074/mcp.M114.042812
19. Caron E, Vincent K, Fortier M-H, Laverdure J-P, Bramoullé A, Hardy M-P, et al. The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation. *Mol Syst Biol* (2011) 7:533. doi:10.1038/msb.2011.68
20. Fortier M-H, Caron E, Hardy M-P, Voisin G, Lemieux S, Perreault C, et al. The MHC class I peptide repertoire is molded by the transcriptome. *J Exp Med* (2008) 205:595–610. doi:10.1084/jem.20071985
21. Chong C, Marino F, Pak H, Racle J, Daniel RT, Müller M, et al. High-throughput and sensitive immunopeptidomics platform reveals profound interferon-mediated remodeling of the human leukocyte antigen (HLA) ligandome. *Mol Cell Proteomics* (2018) 17:533–48. doi:10.1074/mcp.TIR117.000383
22. Ferrington DA, Gregerson DS. Immunoproteasomes: structure, function, and antigen presentation. *Prog Mol Biol Transl Sci* (2012) 109:75–112. doi:10.1016/B978-0-12-397863-9.00003-1
23. Kincaid EZ, Murata S, Tanaka K, Rock KL. Specialized proteasome subunits have an essential role in the thymic selection of CD8(+) T cells. *Nat Immunol* (2016) 17:938–45. doi:10.1038/ni.3480
24. Robek MD, Garcia ML, Boyd BS, Chisari FV. Role of immunoproteasome catalytic subunits in the immune response to hepatitis B virus. *J Virol* (2007) 81:483–91. doi:10.1128/JVI.01779-06
25. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res* (2015) 43:D405–12. doi:10.1093/nar/gku938
26. Weiskopf D, Yauch LE, Angelo MA, John DV, Greenbaum JA, Sidney J, et al. Insights into HLA-restricted T cell responses in a novel mouse model of dengue virus infection point toward new implications for vaccine design. *J Immunol* (2011) 187:4268–79. doi:10.4049/jimmunol.1101970
27. van der Bruggen P, Traversari C, Chomez P, Lurquin C, De Plaen E, Van den Eynde B, et al. A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. *Science* (1991) 254:1643–7. doi:10.1126/science.1840703
28. Reuschenbach M, von Knebel DM, Wentzensen N. A systematic review of humoral immune responses against tumor antigens. *Cancer Immunol Immunother* (2009) 58:1535–44. doi:10.1007/s00262-009-0733-4
29. Simpson AJG, Caballero OL, Jungbluth A, Chen Y-T, Old LJ. Cancer/testis antigens, gametogenesis and cancer. *Nat Rev Cancer* (2005) 5:615–25. doi:10.1038/nrc1669
30. Bobisse S, Genolet R, Roberti A, Tanyi JL, Racle J, Stevenson BJ, et al. Sensitive and frequent identification of high avidity neo-epitope specific CD8 + T cells in immunotherapy-naïve ovarian cancer. *Nat Commun* (2018) 9:1092. doi:10.1038/s41467-018-03301-0
31. Carreno BM, Magrini V, Becker-Hapak M, Kaabinejadian S, Hundal J, Petti AA, et al. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* (2015) 348:803–8. doi:10.1126/science.aaa3828
32. Gros A, Parkhurst MR, Tran E, Pasetto A, Robbins PF, Ilyas S, et al. Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nat Med* (2016) 22:433–8. doi:10.1038/nm.4051
33. Strønen E, Toebes M, Kelderman S, van Buuren MM, Yang W, van Rooij N, et al. Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science* (2016) 352:1337–41. doi:10.1126/science.aaf2288
34. Tran E, Turcotte S, Gros A, Robbins PF, Lu Y-C, Dudley ME, et al. Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science* (2014) 344:641–5. doi:10.1126/science.1251102
35. Yadav M, Jhunjhunwala S, Phung QT, Lupardus P, Tanguay J, Bumbaca S, et al. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* (2014) 515:572–6. doi:10.1038/nature14001
36. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med* (2015) 372:2509–20. doi:10.1056/NEJMoa1500596
37. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* (2015) 348:124–8. doi:10.1126/science.aaa1348
38. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med* (2014) 371:2189–99. doi:10.1056/NEJMoa1406498
39. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* (2015) 350:207–11. doi:10.1126/science.aad0095
40. Tran E, Robbins PF, Lu Y-C, Prickett TD, Gartner JJ, Jia L, et al. T-cell transfer therapy targeting mutant KRAS in cancer. *N Engl J Med* (2016) 375:2255–62. doi:10.1056/NEJMoa1609279
41. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* (2017) 547:217–21. doi:10.1038/nature22991
42. Sahin U, Derhovanessian E, Miller M, Kloke B-P, Simon P, Löwer M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* (2017) 547:222–6. doi:10.1038/nature23003
43. Gfeller D, Bassani-Sternberg M, Schmidt J, Luescher IF. Current tools for predicting cancer-specific T cell immunity. *Oncoimmunology* (2016) 5:e1177691. doi:10.1080/2162402X.2016.1177691
44. Andersen RS, Thru CA, Junker N, Lyngaa R, Donia M, Ellebæk E, et al. Dissection of T-cell antigen specificity in human melanoma. *Cancer Res* (2012) 72:1642–50. doi:10.1158/0008-5472.CAN-11-2614
45. Bentzen AK, Hadrup SR. Evolution of MHC-based technologies used for detection of antigen-responsive T cells. *Cancer Immunol Immunother* (2017) 66:657–66. doi:10.1007/s00262-017-1971-5
46. Bentzen AK, Marquard AM, Lyngaa R, Saini SK, Ramskov S, Donia M, et al. Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nat Biotechnol* (2016) 34:1037–45. doi:10.1038/nbt.3662
47. Linnemann C, van Buuren MM, Bies L, Verdegaal EME, Schotte R, Calis JJA, et al. High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4+ T cells in human melanoma. *Nat Med* (2015) 21:81–5. doi:10.1038/nm.3773
48. Elvin J, Cerundolo V, Elliott T, Townsend A. A quantitative assay of peptide-dependent class I assembly. *Eur J Immunol* (1991) 21:2025–31. doi:10.1002/eji.1830210909
49. Harndahl M, Rasmussen M, Røder G, Dalgaard Pedersen I, Sørensen M, Nielsen M, et al. Peptide-MHC class I stability is a better predictor than

- peptide affinity of CTL immunogenicity. *Eur J Immunol* (2012) 42:1405–16. doi:10.1002/eji.201141774
50. Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, et al. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. *ImmunoRes* (2008) 4:2. doi:10.1186/1745-7580-4-2
 51. Townsend A, Elliott T, Cerundolo V, Foster L, Barber B, Tse A. Assembly of MHC class I molecules analyzed in vitro. *Cell* (1990) 62:285–95. doi:10.1016/0092-8674(90)90366-M
 52. Wulf M, Hoehn P, Trinder P. Identification of human MHC class I binding peptides using the iTOPIA- epitope discovery system. *Methods Mol Biol* (2009) 524:361–7. doi:10.1007/978-1-59745-450-6_26
 53. Bakker AH, Hoppes R, Linnemann C, Toebes M, Rodenko B, Berkers CR, et al. Conditional MHC class I ligands and peptide exchange technology for the human MHC gene products HLA-A1, -A3, -A11, and -B7. *Proc Natl Acad Sci U S A* (2008) 105:3825–30. doi:10.1073/pnas.0709717105
 54. Hadrup SR, Toebes M, Rodenko B, Bakker AH, Egan DA, Ovaa H, et al. High-throughput T-cell epitope discovery through MHC peptide exchange. *Methods Mol Biol* (2009) 524:383–405. doi:10.1007/978-1-59745-450-6_28
 55. Rodenko B, Toebes M, Hadrup SR, van Esch WJE, Molenaar AM, Schumacher TNM, et al. Generation of peptide-MHC class I complexes through UV-mediated ligand exchange. *Nat Protoc* (2006) 1:1120–32. doi:10.1038/nprot.2006.121
 56. Sidney J, Southwood S, Moore C, Oseroff C, Pinilla C, Grey HM, et al. Measurement of MHC/peptide interactions by gel filtration or monoclonal antibody capture. *Curr Protoc Immunol*. (2013) 100:18.3.1–36. doi:10.1002/0471142735.im1803s100
 57. Harndahl M, Rasmussen M, Røder G, Buus S. Real-time, high-throughput measurements of peptide-MHC-I dissociation using a scintillation proximity assay. *J Immunol Methods* (2011) 374:5–12. doi:10.1016/j.jim.2010.10.012
 58. Rasmussen M, Harndahl M, Stryhn A, Boucherma R, Nielsen LL, Lemonnier FA, et al. Uncovering the peptide-binding specificities of HLA-C: a general strategy to determine the specificity of any MHC class I molecule. *J Immunol* (2014) 193:4790–802. doi:10.4049/jimmunol.1401689
 59. Miles KM, Miles JJ, Madura E, Sewell AK, Cole DK. Real time detection of peptide-MHC dissociation reveals that improvement of primary MHC-binding residues can have a minimal, or no, effect on stability. *Mol Immunol* (2011) 48:728–32. doi:10.1016/j.molimm.2010.11.004
 60. Justesen S, Harndahl M, Lamberth K, Nielsen L-LB, Buus S. Functional recombinant MHC class II molecules and high-throughput peptide-binding assays. *Immunome Res* (2009) 5:2. doi:10.1186/1745-7580-5-2
 61. Salvat R, Moise L, Bailey-Kellogg C, Griswold KE. A high throughput MHC II binding assay for quantitative analysis of peptide epitopes. *J Vis Exp* (2014) (85):e51308. doi:10.3791/51308
 62. Yin L, Stern LJ. Measurement of peptide binding to MHC class II molecules by fluorescence polarization. *Curr Protoc Immunol*. (2014) 106:5.10.1–12. doi:10.1002/0471142735.im0510s106
 63. Gaseitsiwe S, Valentini D, MahdaviFar S, Reilly M, Ehrnst A, Maeurer M. Peptide microarray-based identification of *Mycobacterium tuberculosis* epitope binding to HLA-DRB1*0101, DRB1*1501, and DRB1*0401. *Clin Vaccine Immunol* (2010) 17:168–75. doi:10.1128/CVI.00208-09
 64. Hammer J, Takacs B, Sinigaglia F. Identification of a motif for HLA-DR1 binding peptides using M13 display libraries. *J Exp Med* (1992) 176:1007–13. doi:10.1084/jem.176.4.1007
 65. Jiang W, Boder ET. High-throughput engineering and analysis of peptide binding to class II MHC. *Proc Natl Acad Sci U S A* (2010) 107:13258–63. doi:10.1073/pnas.1006344107
 66. Rajasagi M, Shukla SA, Fritsch EF, Keskin DB, DeLuca D, Carmona E, et al. Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood* (2014) 124:453–62. doi:10.1182/blood-2014-04-567933
 67. Robbins PF, Lu Y-C, El-Gamil M, Li YF, Gross C, Gartner J, et al. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat Med* (2013) 19:747–52. doi:10.1038/nm.3161
 68. Castle JC, Kreiter S, Diekmann J, Löwer M, van de Roemer N, de Graaf J, et al. Exploiting the mutanome for tumor vaccination. *Cancer Res* (2012) 72:1081–91. doi:10.1158/0008-5472.CAN-11-3722
 69. Kreiter S, Castle JC, Türeci Ö, Sahin U. Targeting the tumor mutanome for personalized vaccination therapy. *Oncoimmunology* (2012) 1:768–9. doi:10.4161/onci.19727
 70. Abelin JG, Keskin DB, Sarkizova S, Hartigan CR, Zhang W, Sidney J, et al. Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity* (2017) 46:315–26. doi:10.1016/j.immuni.2017.02.007
 71. Pearson H, Daouda T, Granados DP, Durette C, Bonneil E, Courcelles M, et al. MHC class I-associated peptides derive from selective regions of the human genome. *J Clin Invest* (2016) 126:4690–701. doi:10.1172/JCI88590
 72. Bassani-Sternberg M, Chong C, Guillaume P, Solleder M, Pak H, Gannon PO, et al. Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allosterically regulating HLA specificity. *PLoS Comput Biol* (2017) 13:e1005725. doi:10.1371/journal.pcbi.1005725
 73. Mommen GPM, Frese CK, Meiring HD, van Gaans-van den Brink J, de Jong APJM, van Els CACM, et al. Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proc Natl Acad Sci U S A* (2014) 111:4507–12. doi:10.1073/pnas.1321458111
 74. Dargel C, Bassani-Sternberg M, Hasreiter J, Zani F, Bockmann J-H, Thiele F, et al. T cells engineered to express a T-cell receptor specific for glypican-3 to recognize and kill hepatoma cells in vitro and in mice. *Gastroenterology* (2015) 149:1042–52. doi:10.1053/j.gastro.2015.05.055
 75. Di Marco M, Schuster H, Backert L, Ghosh M, Rammensee H-G, Stevanović S. Unveiling the peptide motifs of HLA-C and HLA-G from naturally presented peptides and generation of binding prediction matrices. *J Immunol* (2017) 199:2639–51. doi:10.4049/jimmunol.1700938
 76. Gloger A, Ritz D, Fugmann T, Neri D. Mass spectrometric analysis of the HLA class I peptidome of melanoma cell lines as a promising tool for the identification of putative tumor-associated HLA epitopes. *Cancer Immunol Immunother* (2016) 65:1377–93. doi:10.1007/s00262-016-1897-3
 77. Guasp P, Alvarez-Navarro C, Gomez-Molina P, Martín-Esteban A, Marcilla M, Barnea E, et al. The peptidome of Behçet's disease-associated HLA-B*51:01 includes two subpeptidomes differentially shaped by endoplasmic reticulum aminopeptidase 1. *Arthritis Rheumatol* (2016) 68:505–15. doi:10.1002/art.39430
 78. Hilton HG, McMurtrey CP, Han AS, Djaoud Z, Guethlein LA, Blokhuis JH, et al. The intergenic recombinant HLA-B*46:01 has a distinctive peptidome that includes KIR2DL3 ligands. *Cell Rep* (2017) 19:1394–405. doi:10.1016/j.celrep.2017.04.059
 79. Jarmalavicius S, Welte Y, Walden P. High immunogenicity of the human leukocyte antigen peptidomes of melanoma tumor cells. *J Biol Chem* (2012) 287:33401–11. doi:10.1074/jbc.M112.358903
 80. Kalaora S, Barnea E, Merhavi-Shoham E, Qutob N, Teer JK, Shimony N, et al. Use of HLA peptidomics and whole exome sequencing to identify human immunogenic neo-antigens. *Oncotarget* (2016) 7:5110–7. doi:10.18632/oncotarget.6960
 81. Ritz D, Gloger A, Weide B, Garbe C, Neri D, Fugmann T. High-sensitivity HLA class I peptidome analysis enables a precise definition of peptide motifs and the identification of peptides from cell lines and patients' sera. *Proteomics* (2016) 16:1570–80. doi:10.1002/pmic.201500445
 82. Shraibman B, Kadosh DM, Barnea E, Admon A. Human leukocyte antigen (HLA) peptides derived from tumor antigens induced by inhibition of DNA methylation for development of drug-facilitated immunotherapy. *Mol Cell Proteomics* (2016) 15:3058–70. doi:10.1074/mcp.M116.060350
 83. Singh-Jasuja H, Emmerich NPN, Rammensee H-G. The Tübingen approach: identification, selection, and validation of tumor-associated HLA peptides for cancer therapy. *Cancer Immunol Immunother* (2004) 53:187–95. doi:10.1007/s00262-003-0480-x
 84. Weinschenk T, Gouttefangeas C, Schirle M, Obermayr F, Walter S, Schoor O, et al. Integrated functional genomics approach for the design of patient-individual antitumor vaccines. *Cancer Res* (2002) 62:5818–27.
 85. Gubin MM, Zhang X, Schuster H, Caron E, Ward JP, Noguchi T, et al. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* (2014) 515:577–81. doi:10.1038/nature13988
 86. Khodadoust MS, Olsson N, Wagar LE, Haabeth OAW, Chen B, Swaminathan K, et al. Antigen presentation profiling reveals recognition of lymphoma immunoglobulin neoantigens. *Nature* (2017) 543:723–7. doi:10.1038/nature21433

87. Laumont CM, Daouda T, Laverdure J-P, Bonneil E, Caron-Lizotte O, Hardy M-P, et al. Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames. *Nat Commun* (2016) 7:10238. doi:10.1038/ncomms10238
88. Erhard F, Halenius A, Zimmermann C, L'Hernault A, Kowalewski DJ, Weekes MP, et al. Improved Ribo-seq enables identification of cryptic translation events. *Nat Methods* (2018) 4:e08890. doi:10.1038/nmeth.4631
89. Falk K, Rötzschke O, Stevanović S, Jung G, Rammensee H-G. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* (1991) 351:290–6. doi:10.1038/351290a0
90. Hunt DF, Henderson RA, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, et al. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* (1992) 255:1261–3. doi:10.1126/science.1546328
91. Barnea E, Beer I, Patoka R, Ziv T, Kessler O, Tzevoval E, et al. Analysis of endogenous peptides bound by soluble MHC class I molecules: a novel approach for identifying tumor-specific antigens. *Eur J Immunol* (2002) 32:213–22. doi:10.1002/1521-4141(200201)32:1<213::AID-IMMU213>3.0.CO;2-8
92. Prilliman K, Lindsey M, Zuo Y, Jackson KW, Zhang Y, Hildebrand W. Large-scale production of class I bound peptides: assigning a signature to HLA-B*1501. *Immunogenetics* (1997) 45:379–85. doi:10.1007/s002510050219
93. Scull KE, Dudek NL, Corbett AJ, Ramarathinam SH, Gorasia DG, Williamson NA, et al. Secreted HLA recapitulates the immunopeptidome and allows in-depth coverage of HLA A*02:01 ligands. *Mol Immunol* (2012) 51:136–42. doi:10.1016/j.molimm.2012.02.117
94. Bassani-Sternberg M, Barnea E, Beer I, Avivi I, Katz T, Admon A. Soluble plasma HLA peptidome as a potential source for cancer biomarkers. *Proc Natl Acad Sci U S A* (2010) 107:18769–76. doi:10.1073/pnas.1008501107
95. Mobbs JJ, Illing PT, Dudek NL, Brooks AG, Baker DG, Purcell AW, et al. The molecular basis for peptide repertoire selection in the human leucocyte antigen (HLA) C*06:02 molecule. *J Biol Chem* (2017) 292:17203–15. doi:10.1074/jbc.M117.806976
96. Schittenhelm RB, Dudek NL, Croft NP, Ramarathinam SH, Purcell AW. A comprehensive analysis of constitutive naturally processed and presented HLA-C*04:01 (Cw4)-specific peptides. *Tissue Antigens* (2014) 83:174–9. doi:10.1111/tan.12282
97. Trolle T, McMurtrey CP, Sidney J, Bardet W, Osborn SC, Kaever T, et al. The length distribution of class I-restricted T cell epitopes is determined by both peptide supply and MHC allele-specific binding preference. *J Immunol* (2016) 196:1480–7. doi:10.4049/jimmunol.1501721
98. Giam K, Ayala-Perez R, Illing PT, Schittenhelm RB, Croft NP, Purcell AW, et al. A comprehensive analysis of peptides presented by HLA-A1. *Tissue Antigens* (2015) 85:492–6. doi:10.1111/tan.12565
99. Yair-Sabag S, Tedeschi V, Vitulano C, Barnea E, Glaser F, Melamed Kadosh D, et al. The peptide repertoire of HLA-B27 may include ligands with lysine at P2 anchor position. *Proteomics* (2018) 18:e1700249. doi:10.1002/pmic.201700249
100. Schittenhelm RB, Sian TCCLK, Wilmann PG, Dudek NL, Purcell AW. Revisiting the arthritogenic peptide theory: quantitative not qualitative changes in the peptide repertoire of HLA-B27 allotypes. *Arthritis Rheumatol* (2015) 67:702–13. doi:10.1002/art.38963
101. Bassani-Sternberg M, Gfeller D. Unsupervised HLA peptidome deconvolution improves ligand prediction accuracy and predicts cooperative effects in peptide-HLA interactions. *J Immunol* (2016) 197:2492–9. doi:10.4049/jimmunol.1600808
102. Alvarez B, Barra C, Nielsen M, Andreatta M. Computational tools for the identification and interpretation of sequence motifs in immunopeptidomes. *Proteomics* (2018) 14:1700252. doi:10.1002/pmic.201700252
103. Andreatta M, Alvarez B, Nielsen M. GibbsCluster: unsupervised clustering and alignment of peptide sequences. *Nucleic Acids Res* (2017) 45:W458–63. doi:10.1093/nar/gkx248
104. Kim Y, Sidney J, Pinilla C, Sette A, Peters B. Derivation of an amino acid similarity matrix for peptide: MHC binding and its application as a Bayesian prior. *BMC Bioinformatics* (2009) 10:394. doi:10.1186/1471-2105-10-394
105. Lund O, Nielsen M, Lundegaard C, Keşmir C, Brunak S. *Immunological Bioinformatics*. MIT Press (2005).
106. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanović S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* (1999) 50:213–9. doi:10.1007/s002510050595
107. Zhang H, Lund O, Nielsen M. The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. *Bioinformatics* (2009) 25:1293–9. doi:10.1093/bioinformatics/btp137
108. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* (1997) 25:3389–402. doi:10.1093/nar/25.17.3389
109. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* (1992) 89:10915–9. doi:10.1073/pnas.89.22.10915
110. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* (2016) 32:511–7. doi:10.1093/bioinformatics/btv639
111. Jørgensen KW, Rasmussen M, Buus S, Nielsen M. NetMHCstab – predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. *Immunology* (2014) 141:18–26. doi:10.1111/imm.12160
112. Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lambeth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* (2003) 12:1007–17. doi:10.1110/ps.0239403
113. Rubinsteyn A, O'Donnell T, Damaraju N, Hammerbacher J. Predicting peptide-MHC binding affinities with imputed training data. *bioRxiv* (2016) 054775. doi:10.1101/054775
114. Vang YS, Xie X. HLA class I binding prediction via convolutional neural networks. *Bioinformatics* (2017) 33:2658–65. doi:10.1093/bioinformatics/btx264
115. Peters B, Tong W, Sidney J, Sette A, Weng Z. Examining the independent binding assumption for binding of peptide epitopes to MHC-I molecules. *Bioinformatics* (2003) 19:1765–72. doi:10.1093/bioinformatics/btg247
116. Han Y, Kim D. Deep convolutional neural networks for pan-specific peptide-MHC class I binding prediction. *BMC Bioinformatics* (2017) 18:585. doi:10.1186/s12859-017-1997-x
117. 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) 199:3360–8. doi:10.4049/jimmunol.1700893
118. Rasmussen M, Fenoy E, Harndahl M, Kristensen AB, Nielsen IK, Nielsen M, et al. Pan-specific prediction of peptide-MHC class I complex stability, a correlate of T cell immunogenicity. *J Immunol* (2016) 197:1517–24. doi:10.4049/jimmunol.1600582
119. Nielsen M, Connelley T, Ternette N. Improved prediction of Bovine Leucocyte Antigens (BoLA) presented ligands by use of mass spectrometry-determined ligand- and in-vitro binding data. *J Proteome Res* (2017) 17(1):559–67. doi:10.1021/acs.jproteome.7b00675
120. Guillaume P, Picaud S, Baumgaertner P, Montandon N, Schmidt J, Speiser DE, et al. The C-terminal extension landscape of naturally presented HLA-I ligands. *Proc Natl Acad Sci U S A* (2018) 115:5083–8. doi:10.1073/pnas.1717277115
121. Gfeller D. Uncovering new aspects of protein interactions through analysis of specificity landscapes in peptide recognition domains. *FEBS Lett* (2012) 586:2764–72. doi:10.1016/j.febslet.2012.03.054
122. Gfeller D, Butty F, Wierzbicka M, Verschueren E, Vanhee B, Huang H, et al. The multiple-specificity landscape of modular peptide recognition domains. *Mol Syst Biol* (2011) 7:484. doi:10.1038/msb.2011.18
123. Kim T, Tyndel MS, Huang H, Sidhu SS, Bader GD, Gfeller D, et al. MUSI: an integrated system for identifying multiple specificity from very large peptide or nucleic acid data sets. *Nucleic Acids Res* (2012) 40:e47. doi:10.1093/nar/gkr1294
124. Andreatta M, Lund O, Nielsen M. Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach. *Bioinformatics* (2013) 29:8–14. doi:10.1093/bioinformatics/bts621
125. Storkus WJ, Zeh HJ, Salter RD, Lotze MT. Identification of T-cell epitopes: rapid isolation of class I-presented peptides from viable cells by mild acid elution. *J Immunother Emphasis Tumor Immunol* (1993) 14:94–103. doi:10.1097/00002371-199308000-00003
126. 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) 100:83–90. doi:10.1016/0022-1759(87)90175-X

127. Lanoix J, Durette C, Courcelles M, Cossette É, Comtois-Marotte S, Hardy M-P, et al. Comparison of the MHC I immunopeptidome repertoire of B-cell lymphoblasts using two isolation methods. *Proteomics* (2018) 18(12): e1700251. doi:10.1002/ptm.201700251
128. Nielsen M, Lund O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics* (2009) 10:296. doi:10.1186/1471-2105-10-296
129. Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, et al. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* (2018) 154:394–406. doi:10.1111/imm.12889
130. Singh H, Raghava GP. ProPred: prediction of HLA-DR binding sites. *Bioinformatics* (2001) 17:1236–7. doi:10.1093/bioinformatics/17.12.1236
131. Guan P, Doytchinova IA, Zygouri C, Flower DR. MHCpred: a server for quantitative prediction of peptide-MHC binding. *Nucleic Acids Res* (2003) 31:3621–4. doi:10.1093/nar/gkg510
132. Sturniolo T, Bono E, Ding J, Radrizzani L, Tuercio O, Sahin U, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol* (1999) 17:555–61. doi:10.1038/9858
133. Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Comput Biol* (2008) 4:e1000048. doi:10.1371/journal.pcbi.1000048
134. Yin L, Stern LJ. HLA-DM focuses on conformational flexibility around P1 pocket to catalyze peptide exchange. *Front Immunol* (2013) 4:336. doi:10.3389/fimmu.2013.00336
135. Vogt AB, Kropshofer H, Kalbacher H, Kalbus M, Rammensee H-G, Coligan JE, et al. Ligand motifs of HLA-DRB5*0101 and DRB1*1501 molecules delineated from self-peptides. *J Immunol* (1994) 153:1665–73.
136. Sofron A, Ritz D, Neri D, Fugmann T. High-resolution analysis of the murine MHC class II immunopeptidome. *Eur J Immunol* (2015) 46:319–28. doi:10.1002/eji.201545930
137. Fugmann T, Sofron A, Ritz D, Bootz F, Neri D. The MHC class II immunopeptidome of lymph nodes in health and in chemically induced colitis. *Journal Immunol* (2016) 198:1357–64. doi:10.4049/jimmunol.1601157
138. Ooi JD, Petersen J, Tan YH, Huynh M, Willett ZJ, Ramarathinam SH, et al. Dominant protection from HLA-linked autoimmunity by antigen-specific regulatory T cells. *Nature* (2017) 545:243–7. doi:10.1038/nature22329
139. Ritz D, Sani E, Debiec H, Ronco P, Neri D, Fugmann T. Membranal and blood-soluble HLA class II peptidome analyses using data-dependent and independent acquisition. *Proteomics* (2018) 34:1700246. doi:10.1002/ptm.201700246
140. Barra CM, Alvarez B, Andreatta M, Buus S, Nielsen M. Footprints of antigen processing boost MHC class II natural ligand binding predictions. *bioRxiv* (2018):285767. doi:10.1101/285767
141. Mommen GPM, Marino F, Meiring HD, Poelen MCM, van Gaans-van den Brink JAM, Mohammed S, et al. Sampling from the proteome to the human leukocyte antigen-DR (HLA-DR) ligandome proceeds via high specificity. *Mol Cell Proteomics* (2016) 15:1412–23. doi:10.1074/mcp.M115.055780
142. Hassan C, Chabrol E, Jahn L, Kester MGD, de Ru AH, Drijfhout JW, et al. Naturally processed non-canonical HLA-A*02:01 presented peptides. *J Biol Chem* (2015) 290:2593–603. doi:10.1074/jbc.M114.607028
143. Gfeller D, Guillaume P, Michaux J, Pak H-S, Daniel RT, Racle J, et al. Peptide length distribution and multiple specificity in naturally presented HLA-I ligands. *bioRxiv* (2018):335661. doi:10.1101/335661
144. Collins EJ, Garboczi DN, Wiley DC. Three-dimensional structure of a peptide extending from one end of a class I MHC binding site. *Nature* (1994) 371:626–9. doi:10.1038/371626a0
145. Tenzer S, Wee E, Burgevin A, Stewart-Jones G, Friis L, Lamberth K, et al. Antigen processing influences HIV-specific cytotoxic T lymphocyte immunodominance. *Nat Immunol* (2009) 10:636–46. doi:10.1038/ni.1728
146. McMurtrey C, Trolle T, Sansom T, Remesh SG, Kaever T, Bardet W, et al. Toxoplasma gondii peptide ligands open the gate of the HLA class I binding groove. *Elife* (2016) 5:246. doi:10.7554/eLife.12556
147. Remesh SG, Andreatta M, Ying G, Kaever T, Nielsen M, McMurtrey C, et al. Unconventional peptide presentation by major histocompatibility complex (MHC) class I Allele HLA-A*02:01: BREAKING CONFINEMENT. *J Biol Chem* (2017) 292:5262–70. doi:10.1074/jbc.M117.776542
148. Pymm P, Illing PT, Ramarathinam SH, O'Connor GM, Hughes VA, Hitchen C, et al. MHC-I peptides get out of the groove and enable a novel mechanism of HIV-1 escape. *Nat Struct Mol Biol* (2017) 219:277. doi:10.1038/nsmb.3381
149. Li X, Lamothe PA, Walker BD, Wang J-H. Crystal structure of HLA-B*5801 with a TW10 HIV Gag epitope reveals a novel mode of peptide presentation. *Cell Mol Immunol* (2017) 14:631–4. doi:10.1038/cmi.2017.24
150. Alpizar A, Marino F, Ramos-Fernández A, Lombardía M, Jeko A, Pazos F, et al. A molecular basis for the presentation of phosphorylated peptides by HLA-B antigens. *Mol Cell Proteomics* (2016) 16(2):181–93. doi:10.1074/mcp.M116.063800
151. Marino F, Bern M, Mommen GPM, Leney AC, van Gaans-van den Brink JAM, Bonvin AMJJ, et al. Extended O-GlcNAc on HLA class-I-bound peptides. *J Am Chem Soc* (2015) 137:10922–5. doi:10.1021/jacs.5b06586
152. Marino F, Mommen GPM, Jeko A, Meiring HD, van Gaans-van den Brink JAM, Scheltema RA, et al. Arginine (Di)methylated human leukocyte antigen class I peptides are favorably presented by HLA-B*07. *J Proteome Res* (2017) 16:34–44. doi:10.1021/acs.jproteome.6b00528
153. Andersen MH, Bonfill JE, Neisig A, Arsequell G, Sondergaard I, Valencia G, et al. Phosphorylated peptides can be transported by TAP molecules, presented by class I MHC molecules, and recognized by phosphopeptide-specific CTL. *J Immunol* (1999) 163:3812–8.
154. Petersen J, Wurzbacher SJ, Williamson NA, Ramarathinam SH, Reid HH, Nair AKN, et al. 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) 106:2776–81. doi:10.1073/pnas.0812901106
155. Zarling AL, Polefrone JM, Evans AM, Mikesh LM, Shabanowitz J, Lewis ST, et al. Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. *Proc Natl Acad Sci U S A* (2006) 103:14889–94. doi:10.1073/pnas.0604045103
156. Ciudad MT, Sorvillo N, van Alphen FP, Catalán D, Meijer AB, Voorberg J, et al. Analysis of the HLA-DR peptidome from human dendritic cells reveals high affinity repertoires and nonconventional pathways of peptide generation. *J Leukoc Biol* (2017) 101:15–27. doi:10.1189/jlb.6HI0216-069R
157. Collado JA, Alvarez I, Ciudad MT, Espinosa G, Canals F, Pujol-Borrell R, et al. Composition of the HLA-DR-associated human thymus peptidome. *Eur J Immunol* (2013) 43:2273–82. doi:10.1002/eji.201243280
158. Clement CC, Becerra A, Yin L, Zolla V, Huang L, Merlin S, et al. The dendritic cell major histocompatibility complex II (MHC II) peptidome derives from a variety of processing pathways and includes peptides with a broad spectrum of HLA-DM sensitivity. *J Biol Chem* (2016) 291:5576–95. doi:10.1074/jbc.M115.655738
159. Müller M, Gfeller D, Coukos G, Bassani-Sternberg M. “Hotspots” of antigen presentation revealed by human leukocyte antigen ligandomics for neoantigen prioritization. *Front Immunol* (2017) 8:1367. doi:10.3389/fimmu.2017.01367
160. Dengjel J, Schoor O, Fischer R, Reich M, Kraus M, Müller M, et al. Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. *Proc Natl Acad Sci U S A* (2005) 102:7922–7. doi:10.1073/pnas.0501190102
161. Kropshofer H, Max H, Halder T, Kalbus M, Muller CA, Kalbacher H. Self-peptides from four HLA-DR alleles share hydrophobic anchor residues near the NH2-terminal including proline as a stop signal for trimming. *J Immunol* (1993) 151:4732–42.
162. Godkin AJ, Smith KJ, Willis A, Tejada-Simon MV, Zhang J, Elliott T, et al. Naturally processed HLA class II peptides reveal highly conserved immunogenic flanking region sequence preferences that reflect antigen processing rather than peptide-MHC interactions. *J Immunol* (2001) 166:6720–7. doi:10.4049/jimmunol.166.11.6720
163. Andreatta M, Jurtz VI, Kaever T, Sette A, Peters B, Nielsen M. Machine learning reveals a non-canonical mode of peptide binding to MHC class II molecules. *Immunology* (2017) 152:255–64. doi:10.1111/imm.12763
164. Nielsen M, Lundegaard C, Lund O, Keşmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* (2005) 57:33–41. doi:10.1007/s00251-005-0781-7
165. Doytchinova IA, Guan P, Flower DR. Epijen: a server for multistep T cell epitope prediction. *BMC Bioinformatics* (2006) 7:131. doi:10.1186/1471-2105-7-131

166. Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM, et al. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell Mol Life Sci* (2005) 62:1025–37. doi:10.1007/s00018-005-4528-2
167. Zhang GL, Petrovsky N, Kwok CK, August JT, Brusic V. PRED(TAP): a system for prediction of peptide binding to the human transporter associated with antigen processing. *Immunome Res* (2006) 2:3. doi:10.1186/1745-7580-2-3
168. Jappe EC, Kringelum J, Trolle T, Nielsen M. Predicted MHC peptide binding promiscuity explains MHC class I “hotspots” of antigen presentation defined by mass spectrometry eluted ligand data. *Immunology* (2018) 154:407–17. doi:10.1111/imm.12905
169. Liepe J, Marino F, Sidney J, Jeko A, Bunting DE, Sette A, et al. A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science* (2016) 354:354–8. doi:10.1126/science.aaf4384
170. Nesvizhskii AI. Proteogenomics: concepts, applications and computational strategies. *Nat Methods* (2014) 11:1114–25. doi:10.1038/nmeth.3144
171. Mylonas R, Beer I, Iseli C, Chong C, Pak H, Gfeller D, et al. Estimating the contribution of proteasomal spliced peptides to the HLA-I ligandome. *bioRxiv* (2018):288209. doi:10.1101/288209
172. Calis JJA, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput Biol* (2013) 9:e1003266. doi:10.1371/journal.pcbi.1003266
173. Bjerregaard A-M, Nielsen M, Jurtz V, Barra CM, Hadrup SR, Szallasi Z, et al. An analysis of natural T cell responses to predicted tumor neoepitopes. *Front Immunol* (2017) 8:1566. doi:10.3389/fimmu.2017.01566
174. Duan F, Duitama J, Al Seesi S, Ayres CM, Corcelli SA, Pawashe AP, et al. Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J Exp Med* (2014) 211:2231–48. doi:10.1084/jem.20141308
175. van Buuren MM, Calis JJ, Schumacher TN. High sensitivity of cancer exome-based CD8 T cell neo-antigen identification. *Oncoimmunology* (2014) 3:e28836. doi:10.4161/onci.28836
176. Fritsch EF, Rajasagi M, Ott PA, Brusic V, Hacohen N, Wu CJ. HLA-binding properties of tumor neoepitopes in humans. *Cancer Immunol Res* (2014) 2:522–9. doi:10.1158/2326-6066.CIR-13-0227
177. Gee MH, Han A, Lofgren SM, Beausang JF, Mendoza JL, Birnbaum ME, et al. Antigen identification for orphan t cell receptors expressed on tumor-infiltrating lymphocytes. *Cell* (2018) 172:549.e–63.e. doi:10.1016/j.cell.2017.11.043
178. Lill JR, van Veelen PA, Tenzer S, Admon A, Caron E, Elias J, et al. Minimal information about an immuno-peptidomics experiment (MIAIPE). *Proteomics* (2018) 18(12):e1800110. doi:10.1002/pmic.201800110
179. Vizcaino JA, Csordas A, Del-Toro N, Dianas JA, Griss J, Lavidas I, et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* (2016) 44:11033–11033. doi:10.1093/nar/gkv1145
180. Shao W, Pedrioli PGA, Wolski W, Scurtescu C, Schmid E, Vizcaino JA, et al. The SystemMHC Atlas project. *Nucleic Acids Res* (2018) 46:D1237–47. doi:10.1093/nar/gkx664
181. Karosiene E, Lundegaard C, Lund O, Nielsen M. NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics* (2012) 64:177–86. doi:10.1007/s00251-011-0579-8
182. Moutaftsi M, Peters B, Pasquetto V, Tschärke DC, Sidney J, Bui H-H, et al. A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. *Nat Biotechnol* (2006) 24:817–9. doi:10.1038/nbt1215
183. Bhattacharya R, Sivakumar A, Tokheim C, Guthrie VB, Anagnostou V, Velculescu VE, et al. Evaluation of machine learning methods to predict peptide binding to MHC Class I proteins. *bioRxiv* (2017) 154757. doi:10.1101/154757
184. Liu G, Li D, Li Z, Qiu S, Li W, Chao C-C, et al. PSSMHCPan: a novel PSSM-based software for predicting class I peptide-HLA binding affinity. *GigaScience* (2017) 6:1–11. doi:10.1093/gigascience/gix017
185. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I super-types: a revised and updated classification. *BMC Immunol* (2008) 9:1. doi:10.1186/1471-2172-9-1

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Gfeller and Bassani-Sternberg. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.