Seminars in Immunology 66 (2023) 101708

Contents lists available at ScienceDirect

Seminars in Immunology

journal homepage: www.elsevier.com/locate/ysmim

Contemplating immunopeptidomes to better predict them

David Gfeller [a,b,c,d,*], Yan Liu [a,b,c,d], Julien Racle [a,b,c,d]

[a] Department of Oncology, Ludwig Institute for Cancer Research Lausanne, University of Lausanne, Switzerland
[b] Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland
[c] Agora Cancer Research Centre, 1011 Lausanne, Switzerland
[d] Swiss Cancer Center Leman (SCCL), 1011 Lausanne, Switzerland

ARTICLE INFO

Keywords:
Immunopeptidomics
Antigen presentation
Epitope predictions
Machine learning

ABSTRACT

The identification of T-cell epitopes is key for a complete molecular understanding of immune recognition mechanisms in infectious diseases, autoimmunity and cancer. T-cell epitopes further provide targets for personalized vaccines and T-cell therapy, with several therapeutic applications in cancer immunotherapy and elsewhere. T-cell epitopes consist of short peptides displayed on Major Histocompatibility Complex (MHC) molecules. The recent advances in mass spectrometry (MS) based technologies to profile the ensemble of peptides displayed on MHC molecules – the so-called immunopeptidome – had a major impact on our understanding of antigen presentation and MHC ligands. On the one hand, these techniques enabled researchers to directly identify hundreds of thousands of peptides presented on MHC molecules, including some that elicited T-cell recognition. On the other hand, the data collected in these experiments revealed fundamental properties of antigen presentation pathways and significantly improved our ability to predict naturally presented MHC ligands and T-cell epitopes across the wide spectrum of MHC alleles found in human and other organisms. Here we review recent computational developments to analyze experimentally determined immunopeptidomes and harness these data to improve our understanding of antigen presentation and MHC binding specificities, as well as our ability to predict MHC ligands. We further discuss the strengths and limitations of the latest approaches to move beyond predictions of antigen presentation and tackle the challenges of predicting TCR recognition and immunogenicity.

1. Introduction

The immunopeptidome is defined as the ensemble of peptides displayed on Major Histocompatibility Complex (MHC) molecules [1]. It is also referred to as the ligandome, the MHC peptidome, or the human leukocyte antigen (HLA) peptidome when considering only peptides displayed on human MHC molecules. Peptides found in the immunopeptidome play a central role in thymic selection of T cells as well as T-cell recognition of infected and malignant cells. In particular, non-self peptides originating from pathogens or cancer specific alterations and displayed on MHC molecules can in theory be recognized by T cells. This recognition is triggered by the binding of the T-Cell Receptor (TCR) to peptide-MHC complexes, which initiates formation of the immunological synapse and T-cell activation. For this reason, a detailed understanding of the immunopeptidome is powerful to understand T-cell recognition processes and guide the design of T-cell based therapies, like personalized vaccines in cancer immunotherapy.

The antigen processing and presentation pathways, and especially the MHC molecules, play a central role in determining the immunopeptidome found in different individuals or species [2,3]. Two classes of MHC molecules have been identified: MHC-I molecules, which are targeted by $CD8^+$ T cells, and MHC-II molecules, which are targeted by $CD4^+$ T cells. MHC-I molecules are expressed in most nucleated cells. They bind peptides originating mainly from intracellular proteins that have undergone degradation in the proteasome (Fig. 1A). Some of the peptides leaving the proteasome are then transported to the endoplasmic reticulum (ER). This transport takes place via the transporter associated with antigen processing (TAP) complex, which consists of the TAP-1 and TAP-2 proteins, and is further regulated by different chaperones [3]. Multiple evidences have shown that antigen transport is especially efficient for peptides of 8–14 residues [4]. Once in the ER, peptides can be loaded on MHC-I molecules. These molecules form heterodimers with the β2m and the complex folds stably only upon binding to a peptide. Stable peptide–MHC-I complexes are then

* Corresponding author at: Department of Oncology, Ludwig Institute for Cancer Research Lausanne, University of Lausanne, Switzerland.
E-mail address: david.gfeller@unil.ch (D. Gfeller).

https://doi.org/10.1016/j.smim.2022.101708
Received 18 October 2022; Received in revised form 16 December 2022; Accepted 20 December 2022
Available online 6 January 2023

translocated to the cellular membrane, with the MHC-I peptide binding site in the extracellular milieu (Fig. 1A). The binding site of MHC-I molecules displays several pockets (Fig. 1B) and is highly specific both in terms of the amino acid sequence and the length of the peptides. This specificity can be described and visualized with sequence motifs and peptide length distributions (Fig. 1C, see also resources like the MHC Motif Atlas [5] or the Motif Viewer of NetMHCpan [6]). As a consequence, only a small fraction of peptides available for loading in the ER can bind to a given MHC-I molecule.

Unlike MHC-I, MHC-II molecules are primarily expressed in antigen presenting cells, such as B cells, macrophages or dendritic cells. Peptides displayed on MHC-II molecules can come from both exogenous and endogenous proteins (Fig. 1D). These peptides are processed through the endocytosis or phagocytosis pathways [7]. MHC-II molecules form dimers, with an alpha and a beta chain. The MHC-II dimers usually bind first the invariant chain, which is further cleaved into the CLIP peptide [8]. Upon encounter with other peptides in the late endosome, the CLIP peptide is replaced by some of these peptides and the peptide-MHC-II complexes are translocated to the cell surface (Fig. 1D). This process is controlled by several chaperones, including HLA-DM and HLA-DO in human [2,9]. The binding site of MHC-II molecules shows similarity with the one of MHC-I molecules, but is more open at both ends. As a result, MHC-II ligands are characterized by a binding core with flanking residues extending on both sides of the binding core (Fig. 1E). Akin to MHC-I, the binding between MHC-II and their ligands shows specificity both in terms of the amino acid sequence and the length of the peptides (see later for a few additional features in MHC-II specificity). This specificity can be described and visualized with sequence motifs and peptide length distributions (Fig. 1F) [5].

MHC molecules are encoded by different genes, which are among the most polymorphic ones. In human, MHC-I molecules are encoded by three widely expressed genes (HLA-A, HLA-B and HLA-C) as well as three other genes mainly expressed in specialized cell types (HLA-E, HLA-F and HLA-G). Tens of thousands of MHC-I alleles have been

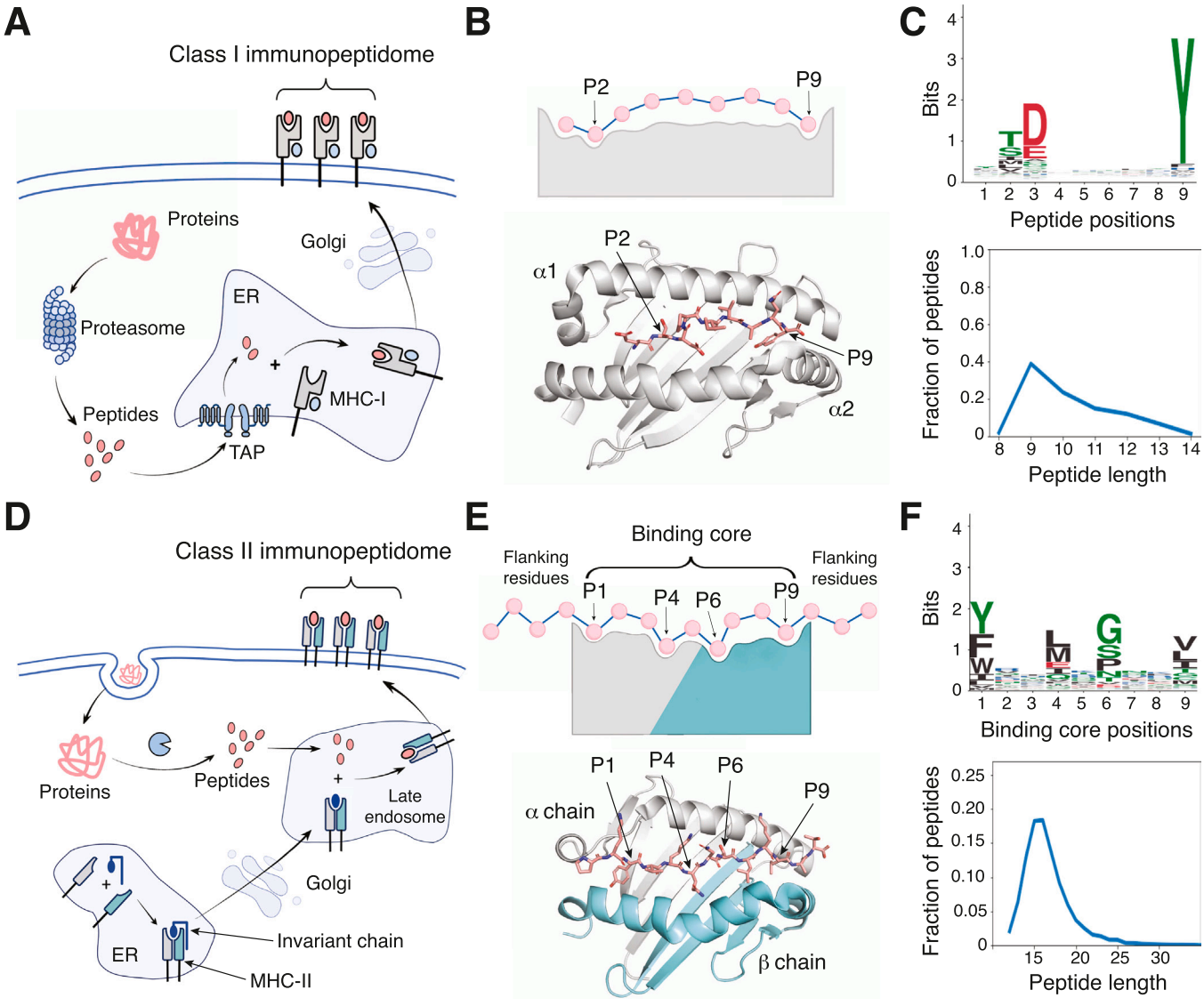

Fig. 1. Antigen presentation pathways and peptide-MHC interactions. (A) Schematic view of the main steps of the class I antigen presentation pathway. (B) Schematic and structural view of the binding of a 9-mer peptide to MHC-I. The structure represents the binding of a peptide to HLA-A*01:01 (PDB:5BS0) [144]. Typical anchor residues at P2 and P9 are indicated by their position, though additional residues can also determine the specificity in some alleles. (C) Binding motif and peptide length distribution for HLA-A*01:01. (D) Schematic view of the main steps of the class II antigen presentation pathway. (E) Schematic and structural view of the binding of peptides to MHC-II. The numbering P1 to P9 indicates anchor positions in the binding core, from N- to C-terminus. Binding core flanking residues extend on both sides of the binding core. The structure represents the binding of a peptide to HLA-DRB1*01:01 (PDB:3S5L) [145]. (F) Binding motif and peptide length distribution for HLA-DRB1*01:01.

identified in human [10]. MHC-II molecules are encoded by nine genes (HLA-DRA1, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1), with extensive genetic linkage between several of these genes. Except for HLA-DRA1, genes encoding MHC-II molecules also display a very high polymorphism. Within each type of genes (HLA-DR, -DP and -DQ), MHC-II molecules form dimers (e.g., HLA-DRA1 – HLA-DRB3). At the protein level, MHC molecules are named with two sets of digits (e.g., HLA-A*01:01 or HLA-DRB1*01:01) that unambiguously distinguish each protein sequence.

The high polymorphism of MHC molecules has a profound impact on the immunopeptidome found in different individuals. Most polymorphic residues are directly located in the ligand binding site. As a result, MHC molecules encoded by different alleles can have very different binding specificities and bind to different repertoires of peptides. For MHC-I molecules, the specificity dictates both the amino acid sequences and the lengths of their ligands (Fig. 1C). Amino acids most important for the binding (referred to as anchor residues) are typically located at the second and last positions of the ligands (see Fig. 1B), although some alleles can display strong specificity at other positions. For instance, the ligands of HLA-A*02:01 are highly enriched in hydrophobic residues at the second and last position, while ligands of HLA-A*01:01 prefer small polar residues at the second position, tyrosine at the last position, and have further preference for negatively charged residues at the third position (Fig. 1C). Most MHC-I molecules have preference for 9-mers, but some molecules bind almost exclusively 9-mers while others accommodate a substantial fraction of longer (e.g., 10- to 14-mers) or shorter (e.g., 8-mers) peptides (Fig. 1C) [11,12]. MHC-II molecules bind longer peptides (roughly 12–25 residues), with a 9-mer binding core located preferentially close to the middle of the ligands and flanking residues extending on both sides of the binding core [13] (Fig. 1E). Within the binding core, anchor residues are often observed at positions P1, P4, P6 and P9, although variability is observed across different genes and different alleles. As for MHC-I, different MHC-II encoded by different alleles have different binding motifs, giving rise to different peptides present in the class II immunopeptidome of different individuals.

A complete view of the immunopeptidome, and especially of non-self peptides originating from pathogens or cancer specific alterations, could be very powerful to map the potential targets of T cells in infectious diseases [14] and cancer [15]. For this reason, large efforts have been devoted to either experimentally determine the immunopeptidome of different individuals and different tissues, or computationally predict MHC ligands, with a focus on non-self peptides. In the next chapters, we discuss the evolution of immunopeptidomics data as well as the impact of these data on our ability to understand antigen presentation and predict MHC ligands and T-cell epitopes.

2. Immunopeptidomes – from a few hundreds to more than a million naturally presented MHC ligands

Soon after the discovery of the role of MHC molecules in antigen presentation, scientists became interested in characterizing the diversity of peptides displayed on MHC molecules and potentially recognized by T cells. Early on, pool peptide sequencing or mass spectrometry (MS) were used to identify MHC ligands and these approaches enabled researchers to find some of the first ligands and derive rough binding motifs for the most common MHC alleles in human and mouse [16–18]. Briefly, MS-based identification of MHC ligands consists of pulling down peptide-MHC complexes with antibodies recognizing MHC-I or MHC-II molecules, followed by the elution of the peptides from the MHC and MS analysis of the isolated peptides. Most of the antibodies used in such experiments are cross-reactive among either MHC-I or MHC-II molecules, although allele-specific or gene-specific antibodies have also been developed to profile the ligands of specific MHC alleles (e.g., HLA-A*02:01) or MHC genes (e.g., HLA-DR).

2.1. Evolution of immunopeptidomics data

Over the years, the throughput and sensitivity of MS have dramatically improved, making it today's most powerful technology to unravel immunopeptidomes [19–21]. This had a profound impact on our ability to profile immunopeptidomes in different tissues, different contexts (e.g., treatment versus control) or different species [22–25]. As of 2022, more than a million naturally presented MHC ligands have been identified and this number is increasing every year (see Fig. 2A). This is the result of both the large number of peptides that can be detected in individual samples (Fig. 2B), and the increasing number of large immunopeptidomics studies, often profiling dozens of samples and identifying tens of thousands of peptides across these samples (Fig. 2C).

2.2. Going beyond standard peptides

Most immunopeptidomics workflows rely on databases of peptides to annotate the spectra obtained by MS. Typically, databases consisting of all possible peptides from the proteome of the organism which the samples came from have been used. This explains why the majority of known naturally presented MHC ligands consist of self peptides coming from human proteins. For epitope discovery in infectious diseases, these databases can be augmented with the sequences of the peptides from the pathogens under investigation [26–28]. Similarly, for cancer neo-antigen discovery, these databases are augmented with peptides encompassing non-synonymous somatic alterations [29–31].

Several studies have also explored the presence of post-translational modification (PTMs) in MHC ligands, by allowing such modifications during the annotation of spectra [30,32–38]. Another type of post-translational modifications comes from proteasomal splicing [39]. The extent of proteasomal splicing in immunopeptidomes has been subject to intense controversy and we refer the readers to existing literature where this topic has been argued extensively [40–44].

Multiple studies have demonstrated that peptides originating from different non-canonical sources, including aberrant transcription and translation of genomic regions that were assumed to be non-coding, untranslated regions or altered frames, can be found in immunopeptidomes [45–50]. Although the fraction of non-self peptides among such non-canonical MHC ligands remains unknown, one cannot exclude that a large number of potential epitopes lie in the non-canonical immunopeptidome. Identification of such non-canonical MHC ligands is best accomplished by expanding the canonical proteome used to search MS raw data based on predictions from RNA-Seq and Ribo-Seq experiments [45,48,49].

Overall, these different examples illustrate how MS enabled us not only to identify many MHC ligands from human proteins, but also to expand the immunopeptidomes beyond standard unmodified peptides, including peptides from pathogens or specific to cancer cells, post-translationally modified peptides as well as non-canonical peptides. Provided false-discovery rates are carefully adapted to the size of the reference databases used to annotate the spectra, these different approaches are powerful to capture the diversity of peptides naturally displayed on MHC molecules.

Unfortunately, even with today's best MS instruments and analysis pipelines, it is widely recognized that only a fraction of the naturally presented peptides can be detected. For this reason, alternative approaches based on predictions of epitopes are likely to remain important in antigen discovery. In the next two sections, we discuss how our ability to identify large datasets of naturally presented MHC ligands by MS had a major impact on (i) our understanding of the specificity of antigen presentation and especially the binding specificity of MHC molecules (Section 3), and (ii) our ability to predict naturally presented MHC ligands (Section 4).

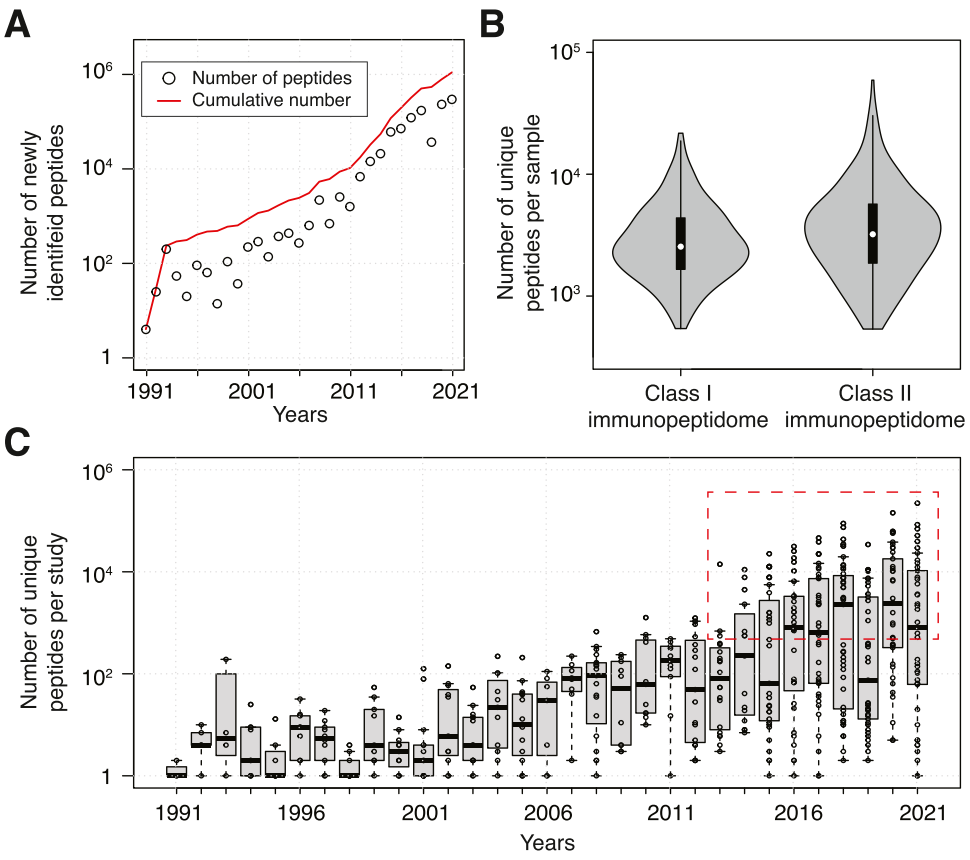

Fig. 2. Evolution of the size of immunopeptidomes. (A) Number of newly identified peptides by MS-based immunopeptidomics per year. The red curve shows the cumulative number of unique peptides. (B) Number of unique peptides reported in class I and class II immunopeptidomics samples with at least 500 peptides. (C) Number of unique peptides reported in immunopeptidomics studies for each year (of note, many studies include multiple samples). The red box shows the studies of origin for the samples considered in panel B. Data in panels A and C were retrieved from the IEDB database [146]. Data in panel B were retrieved from individual immunopeptidomics studies to get information about the samples of origin of each peptide.

3. The specificity of antigen presentation revealed by analyzes of immunopeptidomes

3.1. MHC binding motifs inferred from immunopeptidomes

Of the different steps that shape the immunopeptidome, from protein cleavage to display on cell surface by MHC molecules (Fig. 1A and D), the binding of peptides to MHC is the most selective. For this reason, MHC molecules have a major influence on which peptides can be displayed on the cell surface in a specific individual, and footprints of MHC binding motifs are expected to be visible in immunopeptidomes. This observation has been exploited to infer MHC binding motifs and train predictors of naturally presented MHC ligands based on immunopeptidomics data. Provided that MS based immunopeptidomics data have not been filtered by MHC ligand prediction tools, MHC binding motifs inferred from such data currently provide the most unbiased view of the specificity of MHC molecules [5].

Approaches to learn MHC binding motifs from immunopeptidomes can be broadly classified in two categories. The first category uses samples in which the isolated peptides come from a single allele (i.e., mono-allelic data). This can be achieved by using allele specific antibodies in immunopeptidomics experiments, or cell lines naturally expressing a single allele. Alternatively, cell lines have been transfected or genetically modified to express only one MHC molecule. The immunopeptidome of such samples is therefore directly reflecting the specificity of a single MHC allele. The second category starts with immunopeptidomes obtained from any kind of samples (multi-allelic cell lines, tissues, etc.) with antibodies cross-reactive within MHC-I or MHC-II alleles, and uses machine learning tools to infer models of MHC binding specificity for the different MHC alleles present in the sample. This approach is sometimes referred to as 'motif deconvolution', since it has to deal with the fact that multiple MHCs are expressed in most samples [51–54].

In the two sub-sections below, we review applications of these different approaches to infer MHC-I and MHC-II binding motifs.

3.2. MHC-I binding motifs

Immunopeptidomics profiling of mono-allelic samples has been used for several years to study the properties of naturally presented MHC-I ligands [55,56]. Different strategies have been developed. For instance, several studies used cell lines transfected with soluble MHC molecules of one allele (e.g., HLA-A*02:01, HLA-B*07:02) [12,57,58]. As an alternative to soluble MHCs, single MHC molecules were either expressed in cell lines genetically engineered to express only this MHC-I molecule [59,60] or transfected into cell lines with very low expression of the endogenous MHC-I molecules [61]. These two last approaches are more scalable to profiling a large number of alleles. Mono-allelic immunopeptidomics data are currently available for more than one hundred MHC-I molecules in human [41,59–62]. In parallel, different motif deconvolution approaches have been developed to infer MHC-I binding motifs from multi-allelic class I immunopeptidomics data. For instance, tools like MixMHCp [52] or GibbsCluster [63] allow users to directly identify multiple motifs and annotate peptides to their respective motifs in class I immunopeptidomics data from multi-allelic samples without any prior information about the binding specificity of MHC molecules. Other approaches based on machine learning have been proposed to train models on multi-allelic samples [54,62,64,65]. Although there is no guarantee that correct models can be built with these methods for every allele in a multi-allelic sample, this limitation can often be mitigated by considering multiple samples and capitalizing on the fact that many alleles appear in more than one sample [53]. Of particular interest is the NNalign-MA model, which combines an initial phase of learning a model on mono-allelic samples, and then uses multi-allelic samples to refine the model [54] (see also Pyke et al. [62] for a related framework). This semi-supervised approach has been used

to establish the training data of the widely popular NetMHCpan tool [6] (see Section 4). We also note that fully supervised motif deconvolution approaches have been developed to assign allelic restriction to peptides identified in immunopeptidome based on existing predictors of MHC-I ligands [66].

Overall, MHC-I binding motifs obtained from mono-allelic samples or from deconvolution of multi-allelic samples show very high similarity [53,54,66,67]. This indicates that both approaches can be used to derive MHC-I binding motifs, especially when integrating multiple samples. Moreover, motif deconvolution is a very useful quality control step for both mono- and multi-allelic immunopeptidomics data, by allowing users to verify that the predicted motif(s) correspond to those expected from HLA-I typing, and by filtering out peptides that cannot be assigned to any motif (i.e., trash cluster in GibbCluster or flat motif in MixMHCp) [11,66,68]. Given the different sources of noise or contaminants both in mono-allelic and multi-allelic immunopeptidomics data, we recommend to always apply motif deconvolution, together with other quality-control approaches [69,70], when analyzing immunopeptidomics samples.

3.3. MHC-II binding motifs

As for MHC-I molecules, the identification of MHC-II motifs in class II immunopeptidomes was performed using either mono-allelic samples or motif deconvolution. Historically, several studies used cell lines expressing only one MHC-II allele [18], or purified separately peptides displayed on different alleles [71]. Recently, large libraries of MHC-II mono-allelic cell lines were built to profile MHC-II binding motifs of multiple alleles [72,73]. In parallel, evidences that MHC-II motifs could be retrieved from the class II immunopeptidomes of unmodified multi-allelic samples were provided in several studies [17,74–77] using MoDec [51], GibbsCluster [63] or other tools [78]. As of today, several methods combine mono- and multi-allelic samples to reach the highest allelic coverage and reliable binding motifs are available for more than 80 MHC-II alleles [6,54,66,79]. Irrespective of the use of mono- or multi-allelic samples, an important challenge when learning MHC-II binding specificity models from class II immunopeptidomics data comes from the fact that the binding cores of MHC-II ligands are not naturally aligned, unlike for most MHC-I ligands. Moreover, signals related to antigen processing (see below) can complicate the alignment of MHC-II ligands. Failures to correctly align the peptides or learn the actual binding cores can be an important source of inaccuracies when training MHC-II ligand predictors. This alignment/binding core identification step is included in motif deconvolution tools like MoDec [51] or GibbsCluster [63]. In addition, class II immunopeptidomics samples can contain a substantial fraction of contaminants or wrongly identified peptides [51,66,69]. As for class I immunopeptidomics data, we advise to run unsupervised or supervised motif deconvolution [51,66] as a quality control for class II immunopeptidomics data, since a lack of motifs corresponding to those expected based on the HLA-II typing, or the presence of unexpected motifs, can be an indication of noise in the data (e.g., contaminants, wrongly identified peptides, issues with HLA typing, etc.).

3.4. Structural insights into MHC ligand binding modes

Unsupervised analysis of MHC binding motifs inferred from immunopeptidomes was also useful to gain structural insights into different binding modes of MHC ligands. For instance, the observation that the majority of MHC-I ligands with more than 9 residues have conserved anchors at the second and last positions is a strong indication that these MHC-I ligands bulge out in the middle of the MHC-I binding site, in agreement with most crystal structures [80]. Over the years, multiple studies have observed that MHC-I ligands can also extend at the N- and C-terminus (Fig. 3A). C-terminal extensions were first observed in 1994 [81] and were confirmed in several subsequent studies [82–88]. Analysis of class I immunopeptidomes covering more than fifty HLA-I alleles demonstrated that C-terminal extensions preferentially occur in specific alleles, which are often characterized by higher flexibility in the α1 helix as a result of a glycine at position 79, larger distance between the α1 and α2 helices around the F pocket, and preference for positively charged residues filling the F pocket [82]. N-terminal extensions were identified in several studies [89–93]. Analysis of class I immunopeptidomes indicated that they occur less frequently than C-terminal extensions in

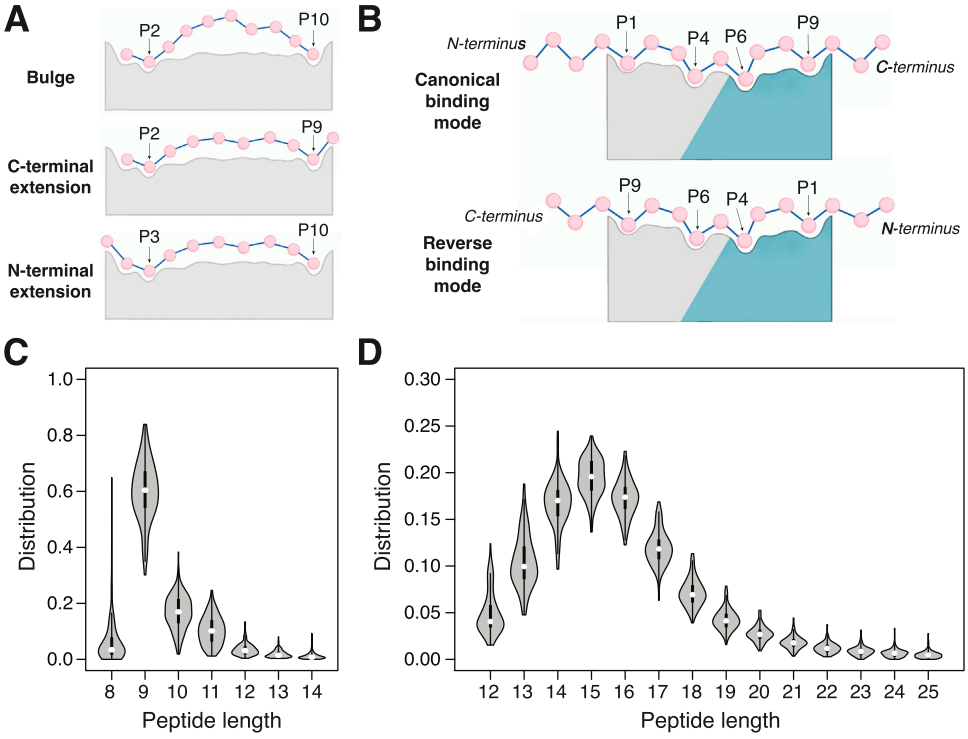

Fig. 3. Binding modes and peptide length distributions inferred from immunopeptidomes. (A) Representation of different binding modes for MHC-I ligands of more than 9 residues (10-mer in this example), including bulge, C-terminal extensions and N-terminal extensions. (B) Representation of different binding modes for MHC-II ligands, including canonical and reverse binders. The numbering P1 to P9 indicates the anchor positions in the binding core, from N- to C-terminus. (C) Variability in peptide length distributions across class I immunopeptidomics samples. (D) Variability in peptide length distributions across human class II immunopeptidomics samples. Data in panel C come from the collection class I immunopeptidomics studies considered in Gfeller et al. [67]. Data in panel D come from the collection class II immunopeptidomics studies considered in Racle et al. [79] (see also red box in Fig. 2C).

human [82].

For MHC-II ligands, analyzes of class II immunopeptidomes have confirmed the fact that the vast majority of MHC-II ligands have a binding core of 9 residues (see one exception in chicken [94]). It appears therefore that the bulging mechanism frequently observed in MHC-I is quite rare in MHC-II alleles, possibly reflecting the higher number and middle position of anchor residues of MHC-II ligands (e.g., P1, P4, P6 and P9 for many HLA-DR and HLA-DP alleles). The vast majority of studies reported a conserved orientation of the MHC-II ligands in the binding site, with the only exceptions of the CLIP peptide [95] and two other peptides covalently linked to MHC-II [96], which were shown to bind in the reverse orientation (Fig. 3B). However, this assumption of conserved orientation has been recently challenged, based on the analysis of class II immunopeptidomics data. Reverse binders were first predicted in van Balen et al. [73] for HLA-DPA1*02:02 – HLA-DPB1*05:01. In a recent preprint from June 2022, we have observed that roughly half of the HLA-DP alleles accommodate ligands binding in the reverse orientation and that these peptides represent from 5 % to 60 % of the immunopeptidome for these alleles [79]. These observations were independently confirmed in [97]. Predictions of ligands following the newly characterized binding mode are now included in the latest version of MixMHC2pred (v2.0) [79]. We anticipate that the ability to predict this new type of HLA-DP ligands will expand the repertoire of class II epitopes in infectious diseases and cancer.

These examples demonstrate how unsupervised profiling of MHC ligands helps generate new hypotheses about MHC ligand binding modes, which can then be experimentally validated.

3.5. Peptide length distribution in immunopeptidomes

Beyond binding motifs, another important determinant of the immunopeptidome comes from the length specificity of naturally presented MHC ligands. For class I immunopeptidomes in human, the typical length of peptides ranges from 8 to 14 residues, with a peak at 9-mers. However, the fraction of peptides of different lengths changes considerably depending on which alleles are present in a sample (Fig. 3C). In a remarkable study, a comparison was performed between the length preference observed in class I immunopeptidomes and the one predicted from binding affinity measurements for four different alleles [12]. The results revealed a clear discrepancy between the two observations, demonstrating that the peptide length distributions observed in class I immunopeptidomes are not only explained by the binding preferences of MHC-I molecules but represent a convolution between binding to MHC-I and skewed distribution of peptide lengths in the pool of peptides available for loading on MHC-I molecules in the ER. Clustering of MHC-I alleles based on their length distribution further provided evidences of groups of MHC-I molecules that share similarity in their peptide length distributions and possible structural mechanisms to explain some of the peptide length preferences [11]. Most peptides in class I immunopeptidomes have length between 8 and 14, but shorter or (much) longer peptides are also observed. A large fraction of contaminants is expected among these peptides [11,67], but it is likely that several real naturally presented MHC-I ligands can have such lengths and may even be immunologically relevant [98]. For instance, a 20-mer peptide recently observed by MS was shown to bind HLA-A*02:01 and elicit CD8+ T-cell recognition [99].

For class II immunopeptidomes in human, the most frequent peptide length is 15, and the majority of peptides have length 12–25. Compared to MHC-I, less variability in peptide length distributions is observed across different samples (Fig. 3D) and different alleles [51,66,72]. This is consistent with the fact that MHC-II ligands have a binding core of conserved length (i.e., 9 residues), with binding core flanking residues extending on both sides and making less interactions with the MHC-II molecules. Differences between MHC-II alleles have therefore less impact on the peptide length distributions, which are presumably more influenced by the class II antigen processing pathways.

It is also important to note that some differences in length distribution have been reported between mono-allelic and multi-allelic samples, possibly reflecting technical biases in protocols used for some mono-allelic samples [67,79].

3.6. Footprints of proteasomal cleavage and transport of antigen

Peptides found in immunopeptidomes have been processed by the class I or class II antigen processing machinery, which includes cleavage, antigen transport and loading on MHC, and is regulated by several chaperones. This suggests that footprints of antigen processing could be found in the sequence of these peptides. For naturally presented MHC-I ligands, this was confirmed in multiple studies and signals related to antigen processing have been found both within the peptide and in the peptide flanking regions [86,100–102]. Within the peptides, the strongest influence of the class I antigen presentation pathway can be seen in the peptide length distribution, as previously discussed. In addition, an over-representation of proline (i.e., a residue that tends to decrease cleavage propensity) was observed in the middle of MHC-I ligands, and a depletion of proline after the C-terminus [59]. Over-representation of arginine/lysine (i.e., residues that enhance cleavage propensity) upstream and downstream of MHC-I ligands was also reported [59,100]. These observations are consistent with the fact that MHC-I ligands need to be cleaved at both their N- and C-termini, but not in their middle.

For MHC-II ligands, signals of antigen processing have also been observed both within MHC-II ligands and in peptide flanking regions. For instance, a clear enrichment of proline at the second position of MHC-II ligands was observed in multiple studies [17,74,103]. Mutation of this proline had no influence on the in vitro binding of MHC-II ligands, demonstrating that this preference for proline comes from antigen processing and not binding to MHC-II [51].

3.7. Impact of source protein expression and localization

Another important determinant of the peptides found in the immunopeptidome comes from the expression of the source proteins of these peptides. For instance, it is expected that peptides encoded by genes that are not expressed in a given tissue will not be found in the immunopeptidome of this tissue [23]. The correlation of source transcript and/or source protein expression with MHC presentation was observed in multiple studies of both class I and class II immunopeptidomes [59,72,75,78,104]. Preferences for specific sub-cellular localization were also observed [25,105–107]. For instance, MHC-II ligands were shown to preferentially originate from secreted proteins, which is consistent with what is known about the class II antigen presentation pathway [72].

4. Predicting immunopeptidomes

As mentioned earlier, an important limitation of direct identification of naturally presented MHC ligands by MS comes from the limited sensitivity of MS-based immunopeptidomics. In addition, immunopeptidomics experiments are not trivial to conduct properly, rely on expensive and dedicated MS equipment and require a minimal amount of cells ($\sim 10^7$) [19,75]. For these reasons this approach is not always applicable to any kind of patients' samples anywhere in the world.

To compensate for this limitation, a widely used alternative consists of predicting putative MHC ligands and T-cell epitopes based on their sequence. The most powerful techniques for predicting naturally presented MHC ligands are based on machine learning algorithms trained on large datasets of naturally presented MHC ligands. In such a framework, the key to reliable predictions is both the algorithm itself and the depth and quality of the training data. Owing to their size, their allelic coverage and their biological relevance, data from immunopeptidomes determined by MS play a central role in the latest versions of predictors of naturally presented MHC ligands.

4.1. Predictions of naturally presented MHC-I ligands

Predictions of MHC-I ligands have been carried out for more than thirty years. These predictions are based on the idea of identifying patterns in MHC-I ligands that distinguish them from 'random' peptides. Historically, the first attempts used simple sequence patterns that can be represented as regular expression, like x[L/M/V]xxxxxx[L/V] for 9-mers HLA-A*02:01 ligands [16]. These regular expressions already capture some of the binding specificity and are reasonable when only a handful of ligands are known. However, they fail to capture specificity at positions designed with x, and fail to capture differences in the relative importance of each amino acid at each positions. These limitations can be addressed by the use of position weight matrices (also referred to as position specific scoring matrices), where the frequency of each amino acid at each position is learned from the ligands of one specific MHC-I molecule. These matrices include renormalization by background amino acid frequency (typically taken from the human proteome) and can be used as a predictor, with the additional advantage that they can be easily visualized as sequence logos [108,109]. Over the years other machine learning tools have been used to predict MHC-I ligands [6,60,62,65,110–115]. These include mixture models, hidden Markov models, artificial neural networks, deep neural networks, etc. Previous benchmarks have not shown superiority of one of these algorithms versus others, indicating that the information present in MHC-I ligands can be accurately captured by different tools. Some of these predictors can only make predictions for alleles for which ligands are available (i.e., allele-specific predictors), while others can be applied to any MHC-I molecule (i.e., pan-allele predictors, see discussion below). In many cases, the peptides in the training set need to be encoded as numerical values, which can be done by one-hot encoding (i.e., treating each amino acid at each position as an independent variable) or BLOSUM based encoding (i.e., including information about biophysical similarity between amino acids). For training or renormalization, many of these approaches need negative data. The standard approach is to use random peptides from the human proteome.

A specific challenge for MHC-I molecules comes from the different lengths of their ligands. This has been addressed in different ways. Some methods, like MixMHCpred [67], train distinct predictors for each peptide length and then renormalize the scores of each predictor to fit peptide length distributions observed experimentally in immunopeptidomes. Methods like NetMHC [116] or NetMHCpan [6] first align the peptides (allowing for gaps or insertions) and then use the aligned peptides as input to a neural network, together with additional input nodes encoding for the length of the peptides. Other approaches like MHCflurry [112] capitalize on methods like convolutional neural networks, and include padding in the encoding of the peptides. When using methods that require negative data for training (e.g., neural networks), it is important to use data with a uniform peptide length distribution (e.g., random peptides from the human proteome), so that the model can learn the length distribution of the actual MHC-I ligands.

Although nearly 150 MHC-I molecules have reliable experimental ligands, this falls short of the tens of thousands of alleles observed in human, not to mention the broadly unknown allelic diversity in other species. To expand predictions beyond alleles with available ligands, several machine learning tools have been developed. The underlying idea is to capture correlations between the amino acid sequences of the MHC-I (and especially the binding site) and the amino acid preferences in their ligands. Such correlations can reflect direct contacts between pairs of residues in the MHC-I binding site and the ligand, or indirect correlations arising from the presence of phylogenetically related groups of MHC-I alleles (sometimes referred to HLA supertypes). These correlations can be efficiently captured by neural networks. The first pan-allele predictor, called NetMHCpan, was developed in 2009 and it remains today's most widely used MHC-I ligand predictor [6,117]. In general, these pan-allele tools are very powerful to extrapolate predictions to alleles relatively similar to those in their training set (e.g., other human HLA-I alleles). The accuracy tends to decrease when considering MHC-I molecules with lower sequence similarity to those with known ligands, such as alleles in evolutionary more distant species.

Different approaches have been developed to incorporate additional features in predictions of naturally presented MHC-I ligands, including specificity from proteasomal cleavage, antigen transport or source protein expression. The first two aspects have been shown to improve predictions of naturally presented MHC-I ligands in several studies [100,101,112,118]. However, the selectivity of cleavage and antigen transport is much lower than the one of MHC-I molecules, and may depend on the biological context (e.g., expression of the immunoproteasome). Regarding source protein expression, different tools allow to input such data in MHC-I ligand predictors [60,119]. There is no doubt that expression (either at the RNA or even better at the protein level) can significantly improve predictions [59,120], for instance by excluding peptides coming from source proteins which are not expressed in a given sample. The main challenge when integrating source protein expression or abundance comes from the ability to accurately measure these values in different contexts (e.g., viral infections, cancer). In the case of neo-antigens, another important point to consider is whether the expression of the source protein corresponds to the mutated allele or to other sources (e.g., non-mutated allele, cancer cells from a clone without the mutation, non-cancer cells).

The natural output of machine learning predictors of MHC ligands trained on immunopeptidomics data consists of a numerical score. To improve interpretability and selection of potential thresholds, a useful approach, first introduced in the NetMHC/NetMHCpan tools is to use % ranks. %ranks are computed by comparing the score of a peptide with the scores obtained on hundreds of thousands of random peptides typically taken from the human proteome. A %rank of 1 % means that 99 % of the random peptides had a score worse than the peptide for which predictions are made. %rank values tend to be more stable than predicted affinity (i.e., IC50 [nM]) or raw scores from machine learning algorithms, and more comparable across tools. They are therefore recommended for practical applications.

4.2. Predictions of naturally presented MHC-II ligands

Over the last few years, predictions of naturally presented MHC-II ligands have also witnessed major improvements by incorporating class II immunopeptidomics data in the training of the predictors [6,51,72,74,78,79]. In terms of computational framework, predictions of MHC-II ligands share many features with predictions of MHC-I ligands (motifs, length preferences). Predictors of MHC-II ligands can also be classified into allele-specific predictors, where a model is trained for each allele individually and predictions can only be done for the alleles in the training set, and pan-allele models where predictions can be done for any allele. Still, a few notable differences need to be highlighted. The main one is that MHC-II ligands have a 9-mer binding core with flanking residues extending on both sides (Fig. 1F). As such multiple overlapping peptides with the same binding core can bind to a given allele. For practical applications, people often focus on 15-mers, assuming that the binding core flanking residues may not impact too much binding to MHC-II and $CD4^+$ T-cell recognition. Another difference is that the position of the binding core shows positional specificity, with a binding core offset close to the middle but slightly shifted towards the C-terminus of the peptide [51].

Given the higher impact of cleavage and antigen processing in shaping class II immunopeptidomes compared to class I immunopeptidomes, many prediction tools are not restricted to MHC-II binding motifs, but also consider peptide residues flanking the binding core, as well as peptide flanking regions. Source transcript and/or protein expression and sub-cellular localization were also shown to influence presentation of peptides on MHC-II molecules, and this aspect has been considered in some tools [72,78]. Some of the challenges when using gene/protein expression in MHC-II ligand predictions come from the fact that MHC-II

molecules are expressed mainly in specialized cell types and bind peptides resulting from endocytosis or phagocytosis of extracellular proteins. As such, one cannot exclude important differences between the source gene/protein expression measurements done on bulk samples and the abundance of peptides available for loading on MHC-II, especially in the case of pathogen or cancer specific peptides. Despite these limitations, it is expected that accurate information about source gene/protein expression and subcellular localization can help narrow down the list of peptides most likely to be displayed on MHC-II molecules.

4.3. Predictions of post-translationally modified MHC ligands

Most predictions of MHC ligands are performed on peptides consisting of the standard twenty amino acids. As mentioned previously, several studies have demonstrated that immunopeptidomes also contain ligands bearing PTMs, such as phosphorylation, cysteinylation, dimethylation or deamination [30,32–38]. The presence of such modifications could significantly alter the binding of a peptide to MHC molecules. Currently, tools incorporating PTMs only exist for the prediction of phosphorylated MHC-I and MHC-II ligands [36,121,122]. We anticipate the (re-)analysis of MS-based immunopeptidomics data will help fill this gap and that MHC ligand predictions tools will be expanded to consider additional PTMs in the future.

4.4. Potential limitations of immunopeptidomics data to train predictors of MHC ligands

The immunopeptidomes identified by MS in multiple samples have dramatically increased the number of known MHC-I and MHC-II ligands. This has led to much more accurate definitions of MHC binding motifs, as well as understanding of properties related to antigen processing (e.g., cleavage motifs, peptide length distributions,…) that could not be derived from the results of binding assays. For this reason, almost all recent machine learning tools to predict MHC ligands rely heavily on MS-based immunopeptidomics data for their training. As with all data used to train machine learning models, an important question to ask is to what extent immunopeptidomes identified by MS may have biases or limitations. The main bias in immunopeptidomes, which has been known since several years, is the under-representation of cysteine-containing peptides [53,59,72,123]. This is likely because chemical modifications of cysteines that can occur during the immunopeptidomics experimental workflow (e.g., dimerization of cysteine-containing peptides) are typically not included in the annotation of MS spectra. It was also observed that methionine and tryptophan tend to be depleted in MHC ligands, possibly reflecting modifications that can occur on these residues in immunopeptidomics experiments [53,72].

Precise identification of amino acid biases arising from MS-based immunopeptidomics experimental pipeline is complicated by the fact that naturally presented MHC ligands display footprints of antigen processing pathways and MHC binding motifs, which impact amino acid frequencies in these peptides. To shed additional light on these issues, we used our collection of MHC-II ligands [79] and computed the amino acid frequency at positions outside of the binding core and outside of the first and last three residues of the peptides. The underlying idea is to use positions that are least influenced by MHC binding motifs or antigen processing, although we cannot totally exclude any impact of these factors. The amino acid frequencies were compared to those in the human proteome (excluding the first and last three residues on each protein since these cannot occur in the positions considered in our MHC-II ligands). The results confirmed the observations of previous studies, with an under-representation of cysteine, methionine and tryptophan. We further identified under-representation of some hydrophobic amino acids (i.e., L, F, I), and an over-representation of polar or charged sidechains (i.e., Q, E, D, K) (Fig. 4).

The impact of any technical bias in MS-based immunopeptidomics data has important consequences on MHC ligand predictors trained on such data. For instance, machine learning predictors trained without correcting for the cysteine bias will invariably give much lower scores to cysteine-containing peptides, since the latter will be under-represented in the positives compared to the negatives (typically peptides randomly selected in the human proteome) used for training [53,123]. Several approaches have been proposed to correct for such biases, either by expanding the search of MS spectra to include more chemical modifications [59], or by fine tuning amino acid background frequencies [53]. Despite these attempts, one cannot exclude that MHC ligand prediction tools are still influenced by (possibly uncharacterized) biases in immunopeptidomics data. Although they do not capture the role of antigen processing, display technologies (e.g., phage or yeast display) to screen MHC ligands may also help revealing MS-related biases in large datasets of MHC ligands [124,125].

5. Future challenges in epitope predictions

5.1. Predictions of TCR recognition

Most epitope prediction tools are trained on naturally presented MHC ligands. As discussed earlier, this includes modeling MHC binding motifs and peptides length distributions, signal of proteasomal cleavage and antigen processing, and source gene/protein expression (Fig. 5A). As such, these approaches do not consider TCR recognition [126], even if multiple evidences exist that TCR recognition propensity is not perfectly correlated with MHC binding. For example, T-cell epitopes displaying very weak experimental binding to MHC molecules have been identified [127]. The focus on antigen presentation in epitope prediction tools comes from the fact that (i) we know orders of magnitudes more MHC ligands than T-cell epitopes, and (ii) TCR recognition

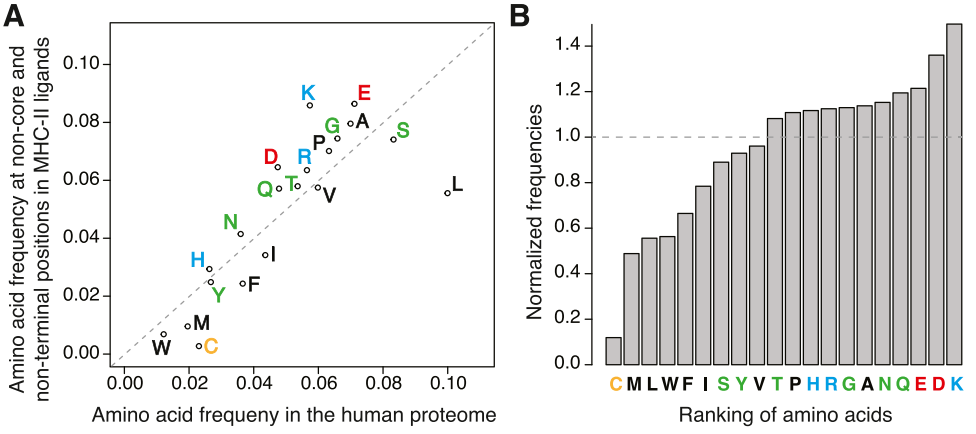

Fig. 4. Potential amino acid biases in MS-based immunopeptidomics data. (A) Comparison between amino acid frequencies in the human proteome (excluding the first and last three residues of each protein) and amino acid frequencies in MHC-II ligands collected in Racle et al. [79] at positions outside of the binding core and outside of the three N- and C-terminal residues (i.e. positions that are expected to show the lowest influence of the binding to MHC-II and class II antigen processing pathways). (B) Ordering of amino acids based on the ratio between the two frequencies computed in panel A. Positively charged residues are shown in blue, negatively charged residues are shown in red, polar residues are shown in green, hydrophobic residues are show in black and cysteine is shown in orange.

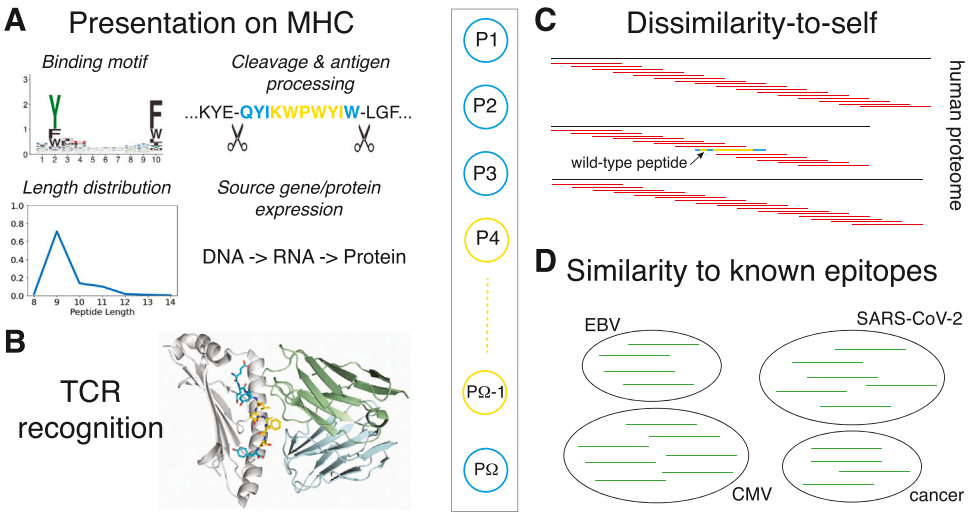

Fig. 5. Summary of different approaches to predict T-cell epitopes. (A) Predictions of presentation on MHC: this approach relies on MHC binding motifs (primarily but not exclusively determined by the residues at the beginning and the end of the peptides for MHC-I – shown in blue in the Figure), peptide length distributions, cleavage and antigen processing signals, and source gene/protein expression. (B) TCR recognition predictions: this approach aims at capturing properties of epitopes that enhance the propensity for TCRs to bind to them. Different studies converged to the observation that aromatic and hydrophobic residues in the middle of the peptide (yellow residues in the Figure) tend to increase TCR recognition [67,128–130]. (C) Dissimilarity-to-self: this approach aims at capturing the impact of negative selection by excluding putative epitopes that show high similarity to some self-peptide. In general peptides arising from single nucleotide variants in cancer show low dissimilarity-to-self, while most pathogen peptides show high dissimilarity-to-self. In the case of cancer mutations, one related approach aims at prioritizing cases where the mutant peptide has a much higher predicted binding to MHC compared to the wild-type peptide. This is typically the case if a mutation modifies an unfavorable anchor residue (e.g., at P2) into a favorable one. (D) Similarity to known epitopes: this approach is based on the hypothesis that peptides showing high similarity to a known epitope are more likely to be immunogenic. In cancer, most putative neo-antigens are quite distinct from any known epitope. Out of these different approaches, only the prediction of presentation on MHC has truly passed the test of time for epitope predictions.

has proved much more challenging to model as a consequence of the huge diversity of TCRs that can potentially recognize any epitope. Still, several attempts to go beyond antigen presentation have been proposed, partly driven by the interest in cancer neo-antigen discovery in recent years and mainly focusing on $CD8^+$ T-cell epitopes. These approaches can be categorized into two different classes:

1) Attempts to capture biochemical properties of MHC-I ligands that enhance TCR recognition (Fig. 5B).
2) Attempts to capture negative selection of $CD8^+$ T cells in the thymus by considering dissimilarity to self peptides (Fig. 5C) or foreignness of potential epitopes (Fig. 5D).

Approaches falling in the first category have investigated differences in amino acid frequencies between immunogenic and non-immunogenic MHC-I ligands. The first evidence that some amino acids (mainly W, F and I) may help for TCR recognition was provided by Calis and coworkers [128]. Later on, a study by Chowell indicated that hydrophobic residues are enriched in immunogenic versus non-immunogenic MHC-I ligands [129]. In our own work, we proposed an algorithmic framework (PRIME) to combine affinity to MHC with propensity for TCR recognition [67,130]. This tool revealed enrichment of aromatic and hydrophobic residues, and depletion of charged residues, in immunogenic MHC-I ligands, and more precisely in the region that is recognized by the TCRs (Fig. 5B). These results are consistent with the fact that hydrophobic and aromatic residues tend to be enriched in protein structural cores and protein-protein interfaces. However, many known immunodominant epitopes do not have hydrophobic or aromatic residues in positions interacting with the TCRs. This indicates that the pattern observed in the aforementioned studies is at best a propensity, and not a strict rule.

In the second category, different approaches have been proposed to capture dissimilarity-to-self or foreignness properties of epitopes. Some approaches attempted to quantify dissimilarity-to-self by computing the sequence similarity between the peptides for which predictions are made and all possible peptides in the human proteome (Fig. 5C) [131,132]. In case of neo-antigens, it is expected that T cells recognizing peptides highly similar to self peptides (e.g., peptides with a mutation that preserves the biochemical properties of the original amino acid, or at a position with minimal contact with the TCR) could have been eliminated in the thymus. However, it is also important to realize that most neo-antigens fall into a similar range of dissimilarity-to-self values, since many of them are just one amino acid away from their wild-type counterpart. Moreover, even subtle changes at the sequence level can have important impact on TCR recognition [133]. Other approaches in the field of neo-antigen discovery have proposed to compute the ratio between the predicted binding affinity of the wild-type and the mutant peptides, a measure sometimes referred to as agretopicity [134,135]. These approaches are based on the hypothesis that if a wild-type peptide has much lower binding affinity than the mutant (and possibly does not even bind to the MHC), T cells recognizing the mutated peptide will be more likely to survive thymic negative selection. Similarity to known epitopes has also been proposed as a way to prioritize potential epitopes and was included in a model of antigen fitness that correlated with survival in cancer patients (Fig. 5D) [136–138]. Although high homology to a known epitope could be a strong indication that individuals can mount a response against a given peptide, few mutated peptides in cancer show such a high similarity to known epitopes. Moreover, the space of actual epitopes is likely very sparsely understood, and T cells have evolved a remarkable ability to recognize a large diversity of antigens.

Altogether, the approaches described in this section are in general based on reasonable assumptions or models. Many of them have been shown to increase predictions of TCR recognition in several studies [130,132,134,139]. However these studies are often based on a relatively limited number of epitopes (from a few tens to a few hundreds). Others have used indirect evidences, such as correlation with survival or response to therapy, with the caveat that the conclusions may be influenced by confounding factors not directly linked to improvement in epitope predictions. For these reasons, there is still no global consensus on how big the improvement in epitope predictions is compared to approaches based only on antigen presentation predictions [140]. It is also important to realize that T cells have evolved to maximize their ability to recognize a large spectrum of antigens, which makes it challenging to accurately model TCR recognition processes.

5.2. Predictions of immunogenicity

Many of the aforementioned predictors that attempted to capture TCR recognition use results of T-cell assays in their training set. T-cell assays can be quite diverse and include IFNγ ELISpot, multimer staining, killing assays, immunization, vaccination, etc. They are also performed on different cells, which comprise bulk T cells, naive T cells, memory T cells or effector T cells, and different individuals (e.g., patients or healthy donors). In addition, multiple rounds of stimulation with the peptides are often performed to enhance a weak initial recognition that could be difficult to distinguish from noise in some assays. These different approaches have an impact on the interpretation of the results of these assays. For instance, stimulation of naive T cells from healthy donors reflects the existence of some TCRs that can recognize a specific epitope, but not necessarily an ongoing or past immune response. Unfortunately, experimental data capturing physiologically relevant natural or vaccine-induced immune responses are much less frequent than those capturing merely TCR recognition. Recent technologies to analyze TCR repertoires and identify clusters of TCRs enriched across patients can reveal TCRs which are more likely to recognize immunodominant epitopes. This information is powerful to guide immunodominant epitope discovery by using these TCRs in epitope screening pipelines [141,142]. However, despite these promising approaches, the set of known immunodominant epitopes is still limited, and prediction of immunogenicity or immunodominance remains an unsolved problem as of today.

6. Conclusion

TCR recognition of peptides displayed on MHC molecules is necessary to initiate the formation of the immunological synapse and elicit cellular immune responses. The ability to determine and predict the specificity of the class I and class II antigen presentation pathways from antigen processing to display on MHCs is therefore a cornerstone towards molecular understanding of T-cell recognition of infected or malignant cells. As of today, the richest source of information about the specificity of antigen presentation has come from MS-based immunopeptidome profiling. In particular, analysis of immunopeptidomes has enabled researchers to (i) define robust MHC binding motifs and peptide length distributions for hundreds of MHC-I and MHC-II molecules [5], (ii) identify signals related to cleavage and antigen processing both within MHC ligands and in peptide flanking regions, (iii) explore the role of source transcript/protein expression and (iv) expand the universe of MHC ligands to pathogen or cancer specific peptides, post-translationally modified peptides and non-canonical peptides. Several of these features have been included in powerful predictors of naturally presented MHC ligands. Owing to the large number (more than a million) of MHC ligands used to train these predictors, a high level of accuracy has been reached for many MHC molecules, including the most common HLA-I or HLA-DR alleles in human. Further improvements are likely possible for less common MHC-I alleles, as well as HLA-DP and HLA-DQ alleles whose motifs have been less studied, although the situation is rapidly changing [79,143]. In addition, a better understanding of the universe of peptides available for loading on MHC molecules, including post-translationally modified or non-canonical ligands, will further expand our ability to predict such MHC ligands.

A much more challenging task is to incorporate TCR recognition and immunogenicity in T-cell epitope prediction tools. The difficulty comes from both the limited number of known T-cell epitopes (especially immunodominant epitopes) and the challenges related to the enormous diversity of TCR sequences. As such, it will be interesting to see whether similar improvements as those observed in predictions of naturally presented MHC ligands can be reached for modeling TCR recognition simply by expanding the number of known epitopes (i.e., the size of the training data), or whether new frameworks and other types of biological information will be needed to overcome the current limitations in immunogenicity predictions.

CRediT authorship contribution statement

DG wrote the manuscript, analyzed the data and made the figures. YL contributed to the figures. JR provided feedback on the manuscript.

Acknowledgment

We thank Daniel M Tadros for useful comments on the manuscript. The project was supported by the Swiss Cancer Research Foundation (KFS-4961-02-2020).

References

[1] J.W. Yewdell, MHC class I immunopeptidome: past, present, and future, Mol. Cell. Proteom. 21 (2022), 100230, https://doi.org/10.1016/j.mcpro.2022.100230.
[2] J. Neefjes, M.L.M. Jongsma, P. Paul, O. Bakke, Towards a systems understanding of MHC class I and MHC class II antigen presentation, Nat. Rev. Immunol. 11 (2011) 823–836.
[3] J.M. Vyas, A.G. Van der Veen, H.L. Ploegh, The known unknowns of antigen processing and presentation, Nat. Rev. Immunol. 8 (2008) 607–618.
[4] U. Ritz, B. Seliger, The transporter associated with antigen processing (TAP): structural integrity, expression, function, and its clinical relevance, Mol. Med. 7 (2001) 149–158.
[5] D.M. Tadros, S. Eggenschwiler, J. Racle, D. Gfeller, The MHC Motif Atlas: a database of MHC binding specificities and ligands, Nucleic Acids Res. (2022), gkac965, https://doi.org/10.1093/nar/gkac965.
[6] B. Reynisson, B. Alvarez, S. Paul, B. Peters, M. Nielsen, NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data, Nucleic Acids Res. 48 (2020) W449–W454, https://doi.org/10.1093/nar/gkaa379.
[7] E.R. Unanue, V. Turk, J. Neefjes, Variations in MHC class II antigen processing and presentation in health and disease, Annu. Rev. Immunol. 34 (2016) 265–297, https://doi.org/10.1146/annurev-immunol-041015-055420.
[8] P.A. Roche, P. Cresswell, Invariant chain association with HLA-DR molecules inhibits immunogenic peptide binding, Nature 345 (1990) 615–618, https://doi.org/10.1038/345615a0.
[9] P.A. Roche, K. Furuta, The ins and outs of MHC class II-mediated antigen processing and presentation, Nat. Rev. Immunol. 15 (2015) 203–216.
[10] J. Robinson, D.J. Barker, X. Georgiou, M.A. Cooper, P. Flicek, S.G.E. Marsh, IPD-IMGT/HLA database, Nucleic Acids Res. 48 (2020) D948–D955, https://doi.org/10.1093/nar/gkz950.
[11] D. Gfeller, P. Guillaume, J. Michaux, H.-S. Pak, R.T. Daniel, J. Racle, G. Coukos, M. Bassani-Sternberg, The length distribution and multiple specificity of naturally presented HLA-I ligands, J. Immunol. 201 (2018) 3705–3716, https://doi.org/10.4049/jimmunol.1800914.
[12] T. Trolle, C.P. McMurtrey, J. Sidney, W. Bardet, S.C. Osborn, T. Kaever, A. Sette, W.H. Hildebrand, M. Nielsen, B. Peters, The length distribution of class I-restricted T cell epitopes is determined by both peptide supply and MHC allele-specific binding preference, J. Immunol. 196 (2016) 1480–1487, https://doi.org/10.4049/jimmunol.1501721.
[13] E.Y. Jones, L. Fugger, J.L. Strominger, C. Siebold, MHC class II proteins and disease: a structural perspective, Nat. Rev. Immunol. 6 (2006) 271–282, https://doi.org/10.1038/nri1805.
[14] S.K. Dhanda, S. Mahajan, S. Paul, Z. Yan, H. Kim, M.C. Jespersen, V. Jurtz, M. Andreatta, J.A. Greenbaum, P. Marcatili, A. Sette, M. Nielsen, B. Peters, IEDB-AR: immune epitope database-analysis resource in 2019, Nucleic Acids Res. 47 (2019) W502–W506, https://doi.org/10.1093/nar/gkz452.
[15] Z. Koşaloğlu-Yalçın, N. Blazeska, R. Vita, H. Carter, M. Nielsen, S. Schoenberger, A. Sette, B. Peters, The cancer epitope database and analysis resource (CEDAR), Nucleic Acids Res. (2022), gkac902, https://doi.org/10.1093/nar/gkac902.
[16] K. Falk, O. Rötzschke, S. Stevanović, G. Jung, H.-G. Rammensee, Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules, Nature 351 (1991) 290–296.
[17] K. Falk, O. R tzschke, S. Stevanov c, G. nther Jung, H.-G. Rammensee, Pool sequencing of natural HLA-DR, DQ, and DP ligands reveals detailed peptide motifs, constraints of processing, and general rules, Immunogenetics 39 (1994) 230–242.
[18] D.F. Hunt, H. Michel, T.A. Dickinson, J. Shabanowitz, A.L. Cox, K. Sakaguchi, E. Appella, H.M. Grey, A. Sette, Peptides presented to the immune system by the murine class II major histocompatibility complex molecule I-Ad, Science 256 (1992) 1817–1820, https://doi.org/10.1126/science.1319610.
[19] L.E. Stopfer, A.D. D'Souza, F.M. White, 1,2,3, MHC: a review of mass-spectrometry-based immunopeptidomics methods for relative and absolute quantification of pMHCs, Immuno-Oncol. Technol. 11 (2021), 100042, https://doi.org/10.1016/j.iotech.2021.100042.
[20] E. Caron, R. Aebersold, A. Banaei-Esfahani, C. Chong, M. Bassani-Sternberg, A case for a human immuno-peptidome project consortium, Immunity 47 (2017) 203–208, https://doi.org/10.1016/j.immuni.2017.07.010.

[21] I. Sirois, M. Isabelle, J.D. Duquette, F. Saab, E. Caron, Immunopeptidomics: isolation of mouse and human MHC class I- and II-associated peptides for mass spectrometry analysis, J. Vis. Exp. (2021), https://doi.org/10.3791/63052.

[22] S.M. Barth, C.M. Schreitmüller, F. Proehl, K. Oehl, L.M. Lumpp, D.J. Kowalewski, M. Di Marco, T. Sturm, L. Backert, H. Schuster, S. Stevanović, H.-G. Rammensee, O. Planz, Characterization of the canine MHC class I DLA-88*50101 peptide binding motif as a prerequisite for canine T cell immunotherapy, PLoS One 11 (2016), e0167017, https://doi.org/10.1371/journal.pone.0167017.

[23] P. Kubiniok, A. Marcu, L. Bichmann, L. Kuchenbecker, H. Schuster, D.J. Hamelin, J.D. Duquette, K.A. Kovalchik, L. Wessling, O. Kohlbacher, H.-G. Rammensee, M. C. Neidert, I. Sirois, E. Caron, Understanding the constitutive presentation of MHC class I immunopeptidomes in primary tissues, IScience 25 (2022), 103768, https://doi.org/10.1016/j.isci.2022.103768.

[24] H. Schuster, W. Shao, T. Weiss, P.G.A. Pedrioli, P. Roth, M. Weller, D.S. Campbell, E.W. Deutsch, R.L. Moritz, O. Planz, H.-G. Rammensee, R. Aebersold, E. Caron, A tissue-based draft map of the murine MHC class I immunopeptidome, Sci. Data 5 (2018), 180157, https://doi.org/10.1038/sdata.2018.157.

[25] A. Marcu, L. Bichmann, L. Kuchenbecker, D.J. Kowalewski, L.K. Freudenmann, L. Backert, L. Mühlenbruch, A. Szolek, M. Lübke, P. Wagner, T. Engler, S. Matovina, J. Wang, M. Hauri-Hohl, R. Martin, K. Kapolou, J.S. Walz, J. Velz, H. Moch, L. Regli, M. Silginer, M. Weller, M.W. Löffler, F. Erhard, A. Schlosser, O. Kohlbacher, S. Stevanović, H.-G. Rammensee, M.C. Neidert, HLA ligand atlas: a benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy, J. Immunother. Cancer 9 (2021), e002071, https://doi.org/10.1136/jitc-2020-002071.

[26] B. Nicholas, A. Bailey, K.J. Staples, T. Wilkinson, T. Elliott, P. Skipp, Immunopeptidomic analysis of influenza A virus infected human tissues identifies internal proteins as a rich source of HLA ligands, PLoS Pathog. 18 (2022), e1009894, https://doi.org/10.1371/journal.ppat.1009894.

[27] N.P. Croft, S.A. Smith, J. Pickering, J. Sidney, B. Peters, P. Faridi, M.J. Witney, P. Sebastian, I.E.A. Flesch, S.L. Heading, A. Sette, N.L. La Gruta, A.W. Purcell, D. C. Tscharke, Most viral peptides displayed by class I MHC on infected cells are immunogenic, Proc. Natl. Acad. Sci. USA 116 (2019) 3112–3117, https://doi.org/10.1073/pnas.1815239116.

[28] A. Nagler, S. Kalaora, C. Barbolin, A. Gangaev, S.L.C. Ketelaars, M. Alon, J. Pai, G. Benedek, Y. Yahalom-Ronen, N. Erez, P. Greenberg, G. Yagel, A. Peri, Y. Levin, A.T. Satpathy, E. Bar-Haim, N. Paran, P. Kvistborg, Y. Samuels, Identification of presented SARS-CoV-2 HLA class I and HLA class II peptides using HLA peptidomics, Cell Rep. 35 (2021), 109305, https://doi.org/10.1016/j.celrep.2021.109305.

[29] C. Chong, G. Coukos, M. Bassani-Sternberg, Identification of tumor antigens with immunopeptidomics, Nat. Biotechnol. 40 (2022) 175–188, https://doi.org/10.1038/s41587-021-01038-8.

[30] M. Bassani-Sternberg, E. Bräunlein, R. Klar, T. Engleitner, P. Sinitcyn, S. Audehm, M. Straub, J. Weber, J. Slotta-Huspenina, K. Specht, M.E. Martignoni, A. Werner, R. Hein, D.H. Busch, C. Peschel, R. Rad, J. Cox, M. Mann, A.M. Krackhardt, Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry, Nat. Commun. 7 (2016) 13404, https://doi.org/10.1038/ncomms13404.

[31] S. Kalaora, E. Barnea, E. Merhavi-Shoham, N. Qutob, J.K. Teer, N. Shimony, J. Schachter, S.A. Rosenberg, M.J. Besser, A. Admon, Y. Samuels, Use of HLA peptidomics and whole exome sequencing to identify human immunogenic neo-antigens, Oncotarget 7 (2016) 5110–5117, https://doi.org/10.18632/oncotarget.6960.

[32] M. Cobbold, H. De La Peña, A. Norris, J.M. Polefrone, J. Qian, A.M. English, K. L. Cummings, S. Penny, J.E. Turner, J. Cottine, J.G. Abelin, S.A. Malaker, A. L. Zarling, H.-W. Huang, O. Goodyear, S.D. Freeman, J. Shabanowitz, G. Pratt, C. Craddock, M.E. Williams, D.F. Hunt, V.H. Engelhard, MHC class I-associated phosphopeptides are the targets of memory-like immunity in leukemia, Sci. Transl. Med. 5 (2013) 203ra125, https://doi.org/10.1126/scitranslmed.3006061.

[33] F. Marino, G.P.M. Mommen, A. Jeko, H.D. Meiring, J.A.M. van Gaans-van den Brink, R.A. Scheltema, C.A.C.M. van Els, A.J.R. Heck, Arginine (di)methylated human leukocyte antigen class I peptides are favorably presented by HLA-B*07, J. Proteome Res. 16 (2017) 34–44.

[34] S. Mei, R. Ayala, S.H. Ramarathinam, P.T. Illing, P. Faridi, J. Song, A.W. Purcell, N.P. Croft, Immunopeptidomic analysis reveals that deamidated HLA-bound peptides arise predominantly from deglycosylated precursors, Mol. Cell. Proteom. 19 (2020) 1236–1247, https://doi.org/10.1074/mcp.RA119.001846.

[35] G.P.M. Mommen, C.K. Frese, H.D. Meiring, J. van Gaans-van den Brink, A.P.J. M. de Jong, C.A.C.M. van Els, A.J.R. Heck, Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD), Proc. Natl. Acad. Sci. USA 111 (2014) 4507–4512, https://doi.org/10.1073/pnas.1321458111.

[36] M. Solleder, P. Guillaume, J. Racle, J. Michaux, H.-S. Pak, M. Müller, G. Coukos, M. Bassani-Sternberg, D. Gfeller, Mass spectrometry based immunopeptidomics leads to robust predictions of phosphorylated HLA class I ligands, Mol. Cell. Proteom. 19 (2020) 390–404, https://doi.org/10.1074/mcp.TIR119.001641.

[37] A.L. Zarling, J.M. Polefrone, A.M. Evans, L.M. Mikesh, J. Shabanowitz, S.T. Lewis, V.H. Engelhard, D.F. Hunt, Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy, Proc. Natl. Acad. Sci. USA 103 (2006) 14889–14894.

[38] A. Kacen, A. Javitt, M.P. Kramer, D. Morgenstern, T. Tsaban, M.D. Shmueli, G. C. Teo, F. da Veiga Leprevost, E. Barnea, F. Yu, A. Admon, L. Eisenbach, Y. Samuels, O. Schueler-Furman, Y. Levin, A.I. Nesvizhskii, Y. Merbl, Post-translational modifications reshape the antigenic landscape of the MHC I immunopeptidome in tumors, Nat. Biotechnol. (2022), https://doi.org/10.1038/s41587-022-01464-2.

[39] N. Vigneron, V. Stroobant, J. Chapiro, A. Ooms, G. Degiovanni, S. Morel, P. van der Bruggen, T. Boon, B.J. Van den Eynde, An antigenic peptide produced by peptide splicing in the proteasome, Science 304 (2004) 587–590, https://doi.org/10.1126/science.1095522.

[40] A. Admon, Are there indeed spliced peptides in the immunopeptidome, Mol. Cell. Proteom. 20 (2021), 100099, https://doi.org/10.1016/j.mcpro.2021.100099.

[41] P. Faridi, C. Li, S.H. Ramarathinam, J.P. Vivian, P.T. Illing, N.A. Mifsud, R. Ayala, J. Song, L.J. Gearing, P.J. Hertzog, N. Ternette, J. Rossjohn, N.P. Croft, A. W. Purcell, A subset of HLA-I peptides are not genomically templated: evidence for cis- and trans-spliced peptide ligands, Sci. Immunol. 3 (2018) eaar3947, https://doi.org/10.1126/sciimmunol.aar3947.

[42] C.F. Lichti, N. Vigneron, K.R. Clauser, B.J. Van den Eynde, M. Bassani-Sternberg, Navigating critical challenges associated with immunopeptidomics-based detection of proteasomal spliced peptide candidates, Cancer Immunol. Res. 10 (2022) 275–284, https://doi.org/10.1158/2326-6066.CIR-21-0727.

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

[44] R. Mylonas, I. Beer, C. Iseli, C. Chong, H. Pak, D. Gfeller, G. Coukos, I. Xenarios, M. Mueller, M. Bassani-Sternberg, Estimating the contribution of proteasomal spliced peptides to the HLA-I ligandome, BioRxiv (2018), 288209.

[45] C. Chong, M. Müller, H. Pak, D. Harnett, F. Huber, D. Grun, M. Leleu, A. Auger, M. Arnaud, B.J. Stevenson, J. Michaux, I. Bilic, A. Hirsekorn, L. Calviello, L. Simó-Riudalbas, E. Planet, J. Lubiński, M. Bryśkiewicz, M. Wiznerowicz, I. Xenarios, L. Zhang, D. Trono, A. Harari, U. Ohler, G. Coukos, M. Bassani-Sternberg, Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes, Nat. Commun. 11 (2020) 1293, https://doi.org/10.1038/s41467-020-14968-9.

[46] C.M. Laumont, T. Daouda, J.-P. Laverdure, E. Bonneil, O. Caron-Lizotte, M.-P. Hardy, D.P. Granados, C. Durette, S. Lemieux, P. Thibault, C. Perreault, Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames, Nat. Commun. 7 (2016) 10238.

[47] A. Pataskar, J. Champagne, R. Nagel, J. Kenski, M. Laos, J. Michaux, H.S. Pak, O. B. Bleijerveld, K. Mordente, J.M. Navarro, N. Blommaert, M.M. Nielsen, D. Lovecchio, E. Stone, G. Georgiou, M.C. de Gooijer, O. van Tellingen, M. Altelaar, R.P. Joosten, A. Perrakis, J. Olweus, M. Bassani-Sternberg, D. S. Peeper, R. Agami, Tryptophan depletion results in tryptophan-to-phenylalanine substitutants, Nature (2022), https://doi.org/10.1038/s41586-022-04499-2.

[48] M.V. Ruiz Cuevas, M.-P. Hardy, J. Hollý, É. Bonneil, C. Durette, M. Courcelles, J. Lanoix, C. Côté, L.M. Staudt, S. Lemieux, P. Thibault, C. Perreault, J. W. Yewdell, Most non-canonical proteins uniquely populate the proteome or immunopeptidome, Cell Rep. 34 (2021), 108815, https://doi.org/10.1016/j.celrep.2021.108815.

[49] T. Ouspenskaia, T. Law, K.R. Clauser, S. Klaeger, S. Sarkizova, F. Aguet, B. Li, E. Christian, B.A. Knisbacher, P.M. Le, C.R. Hartigan, H. Keshishian, A. Apffel, G. Oliveira, W. Zhang, S. Chen, Y.T. Chow, Z. Ji, I. Jungreis, S.A. Shukla, S. Justesen, P. Bachireddy, M. Kellis, G. Getz, N. Hacohen, D.B. Keskin, S.A. Carr, C.J. Wu, A. Regev, Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer, Nat. Biotechnol. 40 (2022) 209–217, https://doi.org/10.1038/s41587-021-01021-3.

[50] Y. Cai, D. Lv, D. Li, J. Yin, Y. Ma, Y. Luo, L. Fu, N. Ding, Y. Li, Z. Pan, X. Li, J. Xu, IEAtlas: an atlas of HLA-presented immune epitopes derived from non-coding regions, Nucleic Acids Res. (2022), gkac776, https://doi.org/10.1093/nar/gkac776.

[51] J. Racle, J. Michaux, G.A. Rockinger, M. Arnaud, S. Bobisse, C. Chong, P. Guillaume, G. Coukos, A. Harari, C. Jandus, M. Bassani-Sternberg, D. Gfeller, Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes, Nat. Biotechnol. 37 (2019) 1283–1286, https://doi.org/10.1038/s41587-019-0289-6.

[52] M. Bassani-Sternberg, D. Gfeller, Unsupervised HLA peptidome deconvolution improves ligand prediction accuracy and predicts cooperative effects in peptide-HLA interactions, J. Immunol. 197 (2016) 2492–2499, https://doi.org/10.4049/jimmunol.1600808.

[53] M. Bassani-Sternberg, C. Chong, P. Guillaume, M. Solleder, H. Pak, P.O. Gannon, L.E. Kandalaft, G. Coukos, D. Gfeller, Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allostery regulating HLA specificity, PLoS Comput. Biol. 13 (2017), e1005725, https://doi.org/10.1371/journal.pcbi.1005725.

[54] B. Alvarez, B. Reynisson, C. Barra, S. Buus, N. Ternette, T. Connelley, M. Andreatta, M. Nielsen, NNAlign_MA; MHC peptidome deconvolution for accurate MHC binding motif characterization and improved T-cell epitope predictions, Mol. Cell. Proteom. 18 (2019) 2459–2477, https://doi.org/10.1074/mcp.TIR119.001658.

[55] E. Barnea, I. Beer, R. Patoka, T. Ziv, O. Kessler, E. Tzehoval, L. Eisenbach, N. Zavazava, A. Admon, Analysis of endogenous peptides bound by soluble MHC class I molecules: a novel approach for identifying tumor-specific antigens, Eur. J. Immunol. 32 (2002) 213–222.

[56] K. Prilliman, M. Lindsey, Y. Zuo, K.W. Jackson, Y. Zhang, W. Hildebrand, Large-scale production of class I bound peptides: assigning a signature to HLA-B*1501, Immunogenetics 45 (1997) 379–385.

[57] T. Huyton, N. Ladas, H. Schumacher, R. Blasczyk, C. Bade-Doeding, Pocketcheck: updating the HLA class I peptide specificity roadmap, Tissue Antigens 80 (2012) 239–248, https://doi.org/10.1111/j.1399-0039.2012.01928.x.

[58] J.C. Yaciuk, M. Skaley, W. Bardet, F. Schafer, D. Mojsilovic, S. Cate, C.J. Stewart, C. McMurtrey, K.W. Jackson, R. Buchli, A. Olvera, S. Cedeño, M. Plana, B. Mothe, C. Brander, J.T. West, W.H. Hildebrand, Direct interrogation of viral peptides presented by the class I HLA of HIV-infected T cells, J. Virol. 88 (2014) 12992–13004.

[59] J.G. Abelin, D.B. Keskin, S. Sarkizova, C.R. Hartigan, W. Zhang, J. Sidney, J. Stevens, W. Lane, G.L. Zhang, T.M. Eisenhaure, K.R. Clauser, N. Hacohen, M. S. Rooney, S.A. Carr, C.J. Wu, Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction, Immunity 46 (2017) 315–326, https://doi.org/10.1016/j.immuni.2017.02.007.

[60] S. Sarkizova, S. Klaeger, P.M. Le, L.W. Li, G. Oliveira, H. Keshishian, C. R. Hartigan, W. Zhang, D.A. Braun, K.L. Ligon, P. Bachireddy, I.K. Zervantonakis, J.M. Rosenbluth, T. Ouspenskaia, T. Law, S. Justesen, J. Stevens, W.J. Lane, T. Eisenhaure, G. Lan Zhang, K.R. Clauser, N. Hacohen, S.A. Carr, C.J. Wu, D. B. Keskin, A large peptidome dataset improves HLA class I epitope prediction across most of the human population, Nat. Biotechnol. (2019), https://doi.org/10.1038/s41587-019-0322-9.

[61] M. Di Marco, H. Schuster, L. Backert, M. Ghosh, H.-G. Rammensee, S. Stevanović, Unveiling the peptide motifs of HLA-C and HLA-G from naturally presented peptides and generation of binding prediction matrices, J. Immunol. 199 (2017) 2639–2651, https://doi.org/10.4049/jimmunol.1700938.

[62] R.M. Pyke, D. Mellacheruvu, S. Dea, C.W. Abbott, S.V. Zhang, N.A. Phillips, J. Harris, G. Bartha, S. Desai, R. McClory, J. West, M.P. Snyder, R. Chen, S. M. Boyle, Precision neoantigen discovery using large-scale immunopeptidomes and composite modeling of MHC peptide presentation, Mol. Cell. Proteom. 20 (2021), 100111, https://doi.org/10.1016/j.mcpro.2021.100111.

[63] M. Andreatta, B. Alvarez, M. Nielsen, GibbsCluster: unsupervised clustering and alignment of peptide sequences, Nucleic Acids Res. 45 (2017) W458–W463, https://doi.org/10.1093/nar/gkx248.

[64] B. Bravi, J. Tubiana, S. Cocco, R. Monasson, T. Mora, A.M. Walczak, RBM-MHC: a semi-supervised machine-learning method for sample-specific prediction of antigen presentation by HLA-I alleles, Cell Syst. 12 (2021) 195–202, https://doi.org/10.1016/j.cels.2020.11.005 (.e9).

[65] B. Bulik-Sullivan, J. Busby, C.D. Palmer, M.J. Davis, T. Murphy, A. Clark, M. Busby, F. Duke, A. Yang, L. Young, N.C. Ojo, K. Caldwell, J. Abhyankar, T. Boucher, M.G. Hart, V. Makarov, V.T.D. Montpreville, O. Mercier, T.A. Chan, G. Scagliotti, P. Bironzo, S. Novello, N. Karachaliou, R. Rosell, I. Anderson, N. Gabrail, J. Hrom, C. Limvarapuss, K. Choquette, A. Spira, R. Rousseau, C. Voong, N.A. Rizvi, E. Fadel, M. Frattini, K. Jooss, M. Skoberne, J. Francis, R. Yelensky, Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification, Nat. Biotechnol. (2018), https://doi.org/10.1038/nbt.4313.

[66] S. Kaabinejadian, C. Barra, B. Alvarez, H. Yari, W.H. Hildebrand, M. Nielsen, Accurate MHC motif deconvolution of immunopeptidomics data reveals a significant contribution of DRB3, 4 and 5 to the total DR immunopeptidome, Front. Immunol. 13 (2022), 835454, https://doi.org/10.3389/fimmu.2022.835454.

[67] D. Gfeller, J. Schmidt, G. Croce, P. Guillaume, S. Bobisse, R. Genolet, L. Queiroz, J. Cesbron, J. Racle, A. Harari, Improved predictions of antigen presentation and TCR recognition with MixMHCpred2.2 and PRIME2.0 reveal potent SARS-CoV-2 CD8$^+$ T-cell epitopes, Cell Syst. 14 (2023) 1–12, https://doi.org/10.1016/j.cels.2022.12.002.

[68] C. Sricharoensuk, T. Boonchalermvichien, P. Muanwien, P. Somparn, T. Pisitkun, S. Sriswasdi, Unsupervised mining of HLA-I peptidomes reveals new binding motifs and potential false positives in the community database, Front. Immunol. 13 (2022), 847756, https://doi.org/10.3389/fimmu.2022.847756.

[69] J. Fritsche, D.J. Kowalewski, L. Backert, F. Gwinner, S. Dorner, M. Priemer, C.-C. Tsou, F. Hoffgaard, M. Römer, H. Schuster, O. Schoor, T. Weinschenk, Pitfalls in HLA ligandomics – how to catch a li(e)gand, Mol. Cell. Proteom. (2021), 100110, https://doi.org/10.1016/j.mcpro.2021.100110.

[70] K.A. Kovalchik, Q. Ma, L. Wessling, F. Saab, J.D. Duquette, P. Kubiniok, D. J. Hamelin, P. Faridi, C. Li, A.W. Purcell, A. Jang, E. Paramithiotis, M. Tognetti, L. Reiter, R. Bruderer, J. Lanoix, É. Bonneil, M. Courcelles, P. Thibault, E. Caron, I. Sirois, MhcVizPipe: a quality control software for rapid assessment of small- to large-scale immunopeptidome datasets, Mol. Cell. Proteom. 21 (2022), 100178, https://doi.org/10.1016/j.mcpro.2021.100178.

[71] A. Rudensky, P. Preston-Hurlburt, S.C. Hong, A. Barlow, C.A. Janeway, Sequence analysis of peptides bound to MHC class II molecules, Nature 353 (1991) 622–627, https://doi.org/10.1038/353622a0.

[72] J.G. Abelin, D. Harjanto, M. Malloy, P. Suri, T. Colson, S.P. Goulding, A.L. Creech, L.R. Serrano, G. Nasir, Y. Nasrullah, C.D. McGann, D. Velez, Y.S. Ting, A. Poran, D.A. Rothenberg, S. Chhangawala, A. Rubinsteyn, J. Hammerbacher, R. B. Gaynor, E.F. Fritsch, J. Greshock, R.C. Oslund, D. Barthelme, T.A. Addona, C. M. Arieta, M.S. Rooney, Defining HLA-II ligand processing and binding rules with mass spectrometry enhances cancer epitope prediction, Immunity 51 (2019) 766–779, https://doi.org/10.1016/j.immuni.2019.08.012 (.e17).

[73] P. van Balen, M.G.D. Kester, W. de Klerk, P. Crivello, E. Arrieta-Bolaños, A.H. de Ru, I. Jedema, Y. Mohammed, M.H.M. Heemskerk, K. Fleischhauer, P.A. van Veelen, J.H.F. Falkenburg, Immunopeptidome analysis of HLA-DPB1 allelic variants reveals new functional hierarchies, J. Immunol. 204 (2020) 3273–3282, https://doi.org/10.4049/jimmunol.2000192.

[74] C. Barra, B. Alvarez, S. Paul, A. Sette, B. Peters, M. Andreatta, S. Buus, M. Nielsen, Footprints of antigen processing boost MHC class II natural ligand predictions, Genome Med. 10 (2018) 84, https://doi.org/10.1186/s13073-018-0594-6.

[75] C. Chong, F. Marino, H. Pak, J. Racle, R.T. Daniel, M. Müller, D. Gfeller, G. Coukos, M. Bassani-Sternberg, High-throughput and sensitive immunopeptidomics platform reveals profound interferonγ-mediated remodeling of the human leukocyte antigen (HLA) ligandome, Mol. Cell. Proteom.: MCP 17 (2018) 533–548.

[76] G.P.M. Mommen, F. Marino, H.D. Meiring, M.C.M. Poelen, J.A.M. van Gaans-van den Brink, S. Mohammed, A.J.R. Heck, C.A.C.M. van Els, Sampling from the proteome to the human leukocyte antigen-DR (HLA-DR) ligandome proceeds via high specificity, Mol. Cell. Proteom.: MCP 15 (2016) 1412–1423.

[77] A. Sofron, D. Ritz, D. Neri, T. Fugmann, High-resolution analysis of the murine MHC class II immunopeptidome, Eur. J. Immunol. 46 (2015) 319–328.

[78] B. Chen, M.S. Khodadoust, N. Olsson, L.E. Wagar, E. Fast, C.L. Liu, Y. Muftuoglu, B.J. Sworder, M. Diehn, R. Levy, M.M. Davis, J.E. Elias, R.B. Altman, A. A. Alizadeh, Predicting HLA class II antigen presentation through integrated deep learning, Nat. Biotechnol. (2019), https://doi.org/10.1038/s41587-019-0280-2.

[79] J. Racle, P. Guillaume, J. Schmidt, J. Michaux, A. Larabi, K. Lau, M.A.S. Perez, G. Croce, R. Genolet, G. Coukos, V. Zoete, F. Pojer, M. Bassani-Sternberg, A. Harari, D. Gfeller, Machine learning predictions of MHC-II specificities reveal alternative binding mode of class II epitopes, BioRxiv (2022), 497561, https://doi.org/10.1101/2022.06.26.497561.

[80] H.C. Guo, T.S. Jardetzky, T.P. Garrett, W.S. Lane, J.L. Strominger, D.C. Wiley, Different length peptides bind to HLA-Aw68 similarly at their ends but bulge out in the middle, Nature 360 (1992) 364–366.

[81] E.J. Collins, D.N. Garboczi, D.C. Wiley, Three-dimensional structure of a peptide extending from one end of a class I MHC binding site, Nature 371 (1994) 626–629.

[82] P. Guillaume, S. Picaud, P. Baumgaertner, N. Montandon, J. Schmidt, D.E. Speiser, G. Coukos, M. Bassani-Sternberg, P. Filippakopoulos, D. Gfeller, The C-terminal extension landscape of naturally presented HLA-I ligands, Proc. Natl. Acad. Sci. USA, vol. 115, 2018, pp. 5083–8. ⟨https://doi.org/10.1073/pnas.1717277115⟩.

[83] C. McMurtrey, T. Trolle, T. Sansom, S.G. Remesh, T. Kaever, W. Bardet, K. Jackson, R. McLeod, A. Sette, M. Nielsen, D.M. Zajonc, I.J. Blader, B. Peters, W. Hildebrand, Toxoplasma gondii peptide ligands open the gate of the HLA class I binding groove, Elife 5 (2016), https://doi.org/10.7554/eLife.12556.

[84] S.G. Remesh, M. Andreatta, G. Ying, T. Kaever, M. Nielsen, C. McMurtrey, W. Hildebrand, B. Peters, D.M. Zajonc, Breaking confinement: unconventional peptide presentation by major histocompatibility (MHC) class I allele HLA-A*02:01, J. Biol. Chem. 292 (2017) 5262–5270.

[85] A. Stryhn, L.O. Pedersen, A. Holm, S. Buus, Longer peptide can be accommodated in the MHC class I binding site by a protrusion mechanism, Eur. J. Immunol. 30 (2000) 3089–3099.

[86] S. Tenzer, E. Wee, A. Burgevin, G. Stewart-Jones, L. Friis, K. Lamberth, C. Chang, M. Harndahl, M. Weimershaus, J. Gerstoft, N. Akkad, P. Klenerman, L. Fugger, E. Y. Jones, A.J. McMichael, S. Buus, H. Schild, P. van Endert, A.K.N. Iversen, Antigen processing influences HIV-specific cytotoxic T lymphocyte immunodominance, Nat. Immunol. 10 (2009) 636–646.

[87] J. Xiao, W. Xiang, Y. Zhang, W. Peng, M. Zhao, L. Niu, Y. Chai, J. Qi, F. Wang, P. Qi, C. Pan, L. Han, M. Wang, J. Kaufman, G.F. Gao, W.J. Liu, An invariant arginine in common with MHC class II allows extension at the C-terminal end of peptides bound to chicken MHC class I, J. Immunol. 201 (2018) 3084–3095, https://doi.org/10.4049/jimmunol.1800611.

[88] M. Yarmarkovich, Q.F. Marshall, J.M. Warrington, R. Premaratne, A. Farrel, D. Groff, W. Li, M. di Marco, E. Runbeck, H. Truong, J.S. Toor, S. Tripathi, S. Nguyen, H. Shen, T. Noel, N.L. Church, A. Weiner, N. Kendsersky, D. Martinez, R. Weisberg, M. Christie, L. Eisenlohr, K.R. Bosse, D.S. Dimitrov, S. Stevanovic, N. G. Sgourakis, B.R. Kiefel, J.M. Maris, Cross-HLA targeting of intracellular oncoproteins with peptide-centric CARs, Nature 599 (2021) 477–484, https://doi.org/10.1038/s41586-021-04061-6.

[89] L. Li, M. Batliwala, M. Bouvier, ERAP1 enzyme-mediated trimming and structural analyses of MHC I-bound precursor peptides yield novel insights into antigen processing and presentation, J. Biol. Chem. 294 (2019) 18534–18544, https://doi.org/10.1074/jbc.RA119.010102.

[90] X. Li, P.A. Lamothe, B.D. Walker, J.-H. Wang, Crystal structure of HLA-B*5801 with a TW10 HIV Gag epitope reveals a novel mode of peptide presentation, Cell. Mol. Immunol. 14 (2017) 631–634.

[91] P. Pymm, P.T. Illing, S.H. Ramarathinam, G.M. O'Connor, V.A. Hughes, C. Hitchen, D.A. Price, B.K. Ho, D.W. McVicar, A.G. Brooks, A.W. Purcell, J. Rossjohn, J.P. Vivian, MHC-I peptides get out of the groove and enable a novel mechanism of HIV-1 escape, Nat. Struct. Mol. Biol. 219 (2017) 277.

[92] X. Wei, S. Li, S. Wang, G. Feng, X. Xie, Z. Li, N. Zhang, Peptidomes and structures illustrate how SLA-I micropolymorphism influences the preference of binding peptide length, Front. Immunol. 13 (2022), 820881, https://doi.org/10.3389/fimmu.2022.820881.

[93] X. Wei, S. Wang, S. Wang, X. Xie, N. Zhang, Structure and peptidomes of swine MHC class I with long peptides reveal the cross-species characteristics of the novel N-terminal extension presentation mode, J. Immunol. 208 (2022) 480–491, https://doi.org/10.4049/jimmunol.2001207.

[94] S. Halabi, M. Ghosh, S. Stevanović, H.-G. Rammensee, L.D. Bertzbach, B. B. Kaufer, M.C. Moncrieffe, B. Kaspers, S. Härtle, J. Kaufman, The dominantly expressed class II molecule from a resistant MHC haplotype presents only a few Marek's disease virus peptides by using an unprecedented binding motif, PLoS Biol. 19 (2021), e3001057, https://doi.org/10.1371/journal.pbio.3001057.

[95] S. Günther, A. Schlundt, J. Sticht, Y. Roske, U. Heinemann, K.-H. Wiesmüller, G. Jung, K. Falk, O. Rötzschke, C. Freund, Bidirectional binding of invariant chain peptides to an MHC class II molecule, Proc. Natl. Acad. Sci. USA 107 (2010) 22219–22224, https://doi.org/10.1073/pnas.1014708107.

[96] A. Schlundt, S. Günther, J. Sticht, M. Wieczorek, Y. Roske, U. Heinemann, C. Freund, Peptide linkage to the α-subunit of MHCII creates a stably inverted antigen presentation complex, J. Mol. Biol. 423 (2012) 294–302, https://doi.org/10.1016/j.jmb.2012.07.008.
[97] S. Klobuch, J.J. Lim, P. van Balen, M.G.D. Kester, W. de Klerk, A.H. de Ru, C. R. Pothast, I. Jedema, J.W. Drijfhout, J. Rossjohn, H.H. Reid, P.A. van Veelen, J. H.F. Falkenburg, M.H.M. Heemskerk, Human T cells recognize HLA-DP-bound peptides in two orientations, Proc. Natl. Acad. Sci. USA 119 (2022), e2214331119, https://doi.org/10.1073/pnas.2214331119.
[98] C. Hassan, E. Chabrol, L. Jahn, M.G.D. Kester, A.H. de Ru, J.W. Drijfhout, J. Rossjohn, J.H.F. Falkenburg, M.H.M. Heemskerk, S. Gras, P.A. van Veelen, Naturally processed non-canonical HLA-A*02:01 presented peptides, J. Biol. Chem. 290 (2015) 2593–2603.
[99] M.H. Meeuwsen, A.K. Wouters, R.S. Hagedoorn, M.G.D. Kester, D.F.G. Remst, D. M. van der Steen, A. de Ru, P.A. van Veelen, J. Rossjohn, S. Gras, J.H. F. Falkenburg, M.H.M. Heemskerk, Cutting edge: unconventional CD8+ T cell recognition of a naturally occurring HLA-A*02:01-restricted 20mer epitope, J. Immunol. 208 (2022) 1851–1856, https://doi.org/10.4049/jimmunol.2101208.
[100] M. Nielsen, C. Lundegaard, O. Lund, C. Keşmir, The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage, Immunogenetics 57 (2005) 33–41.
[101] S. Tenzer, B. Peters, S. Bulik, O. Schoor, C. Lemmel, M.M. Schatz, P.-M. Kloetzel, H.-G. Rammensee, H. Schild, H.-G. Holzhütter, Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding, Cell. Mol. Life Sci. 62 (2005) 1025–1037, https://doi.org/10.1007/s00018-005-4528-2.
[102] D. de Verteuil, T.L. Muratore-Schroeder, D.P. Granados, M.-H. Fortier, M.-P. Hardy, A. Bramoullé, E. Caron, K. Vincent, S. Mader, S. Lemieux, P. Thibault, C. Perreault, Deletion of immunoproteasome subunits imprints on the transcriptome and has a broad impact on peptides presented by major histocompatibility complex I molecules, Mol. Cell. Proteom. 9 (2010) 2034–2047, https://doi.org/10.1074/mcp.M900566-MCP200.
[103] M.T. Ciudad, N. Sorvillo, F.P. van Alphen, D. Catalán, A.B. Meijer, J. Voorberg, D. Jaraquemada, Analysis of the HLA-DR peptidome from human dendritic cells reveals high affinity repertoires and nonconventional pathways of peptide generation, J. Leukoc. Biol. 101 (2017) 15–27.
[104] M. Bassani-Sternberg, S. Pletscher-Frankild, L.J. Jensen, M. Mann, Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation, Mol. Cell. Proteom. 14 (2015) 658–673, https://doi.org/10.1074/mcp.M114.042812.
[105] A.S. Juncker, M.V. Larsen, N. Weinhold, M. Nielsen, S. Brunak, O. Lund, Systematic characterisation of cellular localisation and expression profiles of proteins containing MHC ligands, PLoS One 4 (2009), e7448, https://doi.org/10.1371/journal.pone.0007448.
[106] J. Dengjel, O. Schoor, R. Fischer, M. Reich, M. Kraus, M. Müller, K. Kreymborg, F. Altenberend, J. Brandenburg, H. Kalbacher, R. Brock, C. Driessen, H.-G. Rammensee, S. Stevanovic, Autophagy promotes MHC class II presentation of peptides from intracellular source proteins, Proc. Natl. Acad. Sci. USA 102 (2005) 7922–7927.
[107] H.D. Hickman, A.D. Luis, R. Buchli, S.R. Few, M. Sathiamurthy, R.S. VanGundy, C.F. Giberson, W.H. Hildebrand, Toward a definition of self: proteomic evaluation of the class I peptide repertoire, J. Immunol. 172 (2004) 2944–2952, https://doi.org/10.4049/jimmunol.172.5.2944.
[108] D. Gfeller, M. Bassani-Sternberg, Predicting antigen presentation–what could we learn from a million peptides? Front. Immunol. 9 (2018) 1716, https://doi.org/10.3389/fimmu.2018.01716.
[109] J. Sidney, E. Assarsson, C. Moore, S. Ngo, C. Pinilla, A. Sette, B. Peters, Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries, Immun. Res. 4 (2008) 2.
[110] J. Hu, Z. Liu, DeepMHC: Deep Convolutional Neural Networks for High-performance peptide-MHC Binding Affinity Prediction, 2017.
[111] Y. Hu, Z. Wang, H. Hu, F. Wan, L. Chen, Y. Xiong, X. Wang, D. Zhao, W. Huang, J. Zeng, ACME: pan-specific peptide-MHC class I binding prediction through attention-based deep neural networks, Bioinformatics 35 (2019) 4946–4954, https://doi.org/10.1093/bioinformatics/btz427.
[112] T.J. O'Donnell, A. Rubinsteyn, U. Laserson, MHCflurry 2.0: improved pan-allele prediction of MHC class I-presented peptides by incorporating antigen processing, Cell Syst. 11 (2020) 42–48, https://doi.org/10.1016/j.cels.2020.06.010 (.e7).
[113] Y.S. Vang, X. Xie, HLA class I binding prediction via convolutional neural networks, Bioinformatics 33 (2017) 2658–2665.
[114] Y. Ye, J. Wang, Y. Xu, Y. Wang, Y. Pan, Q. Song, X. Liu, J. Wan, MATHLA: a robust framework for HLA-peptide binding prediction integrating bidirectional LSTM and multiple head attention mechanism, BMC Bioinform. 22 (2021) 7, https://doi.org/10.1186/s12859-020-03946-z.
[115] H. Zeng, D.K. Gifford, DeepLigand: accurate prediction of MHC class I ligands using peptide embedding, Bioinformatics 35 (2019) i278–i283, https://doi.org/10.1093/bioinformatics/btz330.
[116] M. Andreatta, M. Nielsen, Gapped sequence alignment using artificial neural networks: application to the MHC class I system, Bioinformatics 32 (2016) 511–517, https://doi.org/10.1093/bioinformatics/btv639.
[117] I. Hoof, B. Peters, J. Sidney, L.E. Pedersen, A. Sette, O. Lund, S. Buus, M. Nielsen, NetMHCpan, a method for MHC class I binding prediction beyond humans, Immunogenetics 61 (2009) 1–13, https://doi.org/10.1007/s00251-008-0341-z.
[118] M.V. Larsen, C. Lundegaard, K. Lamberth, S. Buus, S. Brunak, O. Lund, M. Nielsen, An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions, Eur. J. Immunol. 35 (2005) 2295–2303.
[119] H.M. Garcia Alvarez, Z. Koşaloğlu-Yalçın, B. Peters, M. Nielsen, The role of antigen expression in shaping the repertoire of HLA presented ligands, IScience 25 (2022), 104975, https://doi.org/10.1016/j.isci.2022.104975.
[120] Z. Koşaloğlu-Yalçın, J. Lee, J. Greenbaum, S.P. Schoenberger, A. Miller, Y.J. Kim, A. Sette, M. Nielsen, B. Peters, Combined assessment of MHC binding and antigen abundance improves T cell epitope predictions, IScience 25 (2022), 103850, https://doi.org/10.1016/j.isci.2022.103850.
[121] C.T. Refsgaard, C. Barra, X. Peng, N. Ternette, M. Nielsen, NetMHCphosPan – pan-specific prediction of MHC class I antigen presentation of phosphorylated ligands, ImmunoInformatics 1–2 (2021), 100005, https://doi.org/10.1016/j.immuno.2021.100005.
[122] M. Solleder, J. Racle, P. Guillaume, G. Coukos, M. Bassani-Sternberg, D. Gfeller, Deciphering the landscape of phosphorylated HLA-II ligands, IScience 25 (2022), 104215, https://doi.org/10.1016/j.isci.2022.104215.
[123] A. Sachs, E. Moore, Z. Kosaloglu-Yalcin, B. Peters, J. Sidney, S.A. Rosenberg, P. F. Robbins, A. Sette, Impact of cysteine residues on MHC binding predictions and recognition by tumor-reactive T cells, J. Immunol. 205 (2020) 539–549, https://doi.org/10.4049/jimmunol.1901173.
[124] C.G. Rappazzo, B.D. Huisman, M.E. Birnbaum, Repertoire-scale determination of class II MHC peptide binding via yeast display improves antigen prediction, Nat. Commun. 11 (2020) 4414, https://doi.org/10.1038/s41467-020-18204-2.
[125] B.D. Huisman, Z. Dai, D.K. Gifford, M.E. Birnbaum, A high-throughput yeast display approach to profile pathogen proteomes for MHC-II binding, Elife 11 (2022), e78589, https://doi.org/10.7554/eLife.78589.
[126] B. Peters, M. Nielsen, A. Sette, T cell epitope predictions, Annu. Rev. Immunol. 38 (2020) 123–145, https://doi.org/10.1146/annurev-immunol-082119-124838.
[127] H. Ebrahimi-Nik, M. Moussa, R.P. Englander, S. Singhaviranon, J. Michaux, H. Pak, H. Miyadera, W.L. Corwin, G.L.J. Keller, A.T. Hagymasi, T.V. Shcheglova, G. Coukos, B.M. Baker, I.I. Mandoiu, M. Bassani-Sternberg, P.K. Srivastava, Reversion analysis reveals the in vivo immunogenicity of a poorly MHC I-binding cancer neoepitope, Nat. Commun. 12 (2021) 6423, https://doi.org/10.1038/s41467-021-26646-5.
[128] J.J.A. Calis, M. Maybeno, J.A. Greenbaum, D. Weiskopf, A.D. De Silva, A. Sette, C. Keşmir, B. Peters, Properties of MHC class I presented peptides that enhance immunogenicity, PLoS Comput. Biol. 9 (2013), e1003266.
[129] D. Chowell, S. Krishna, P.D. Becker, C. Cocita, J. Shu, X. Tan, P.D. Greenberg, L.S. Klavinskis, J.N. Blattman, K.S. Anderson, TCR contact residue hydrophobicity is a hallmark of immunogenic CD8+ T cell epitopes, Proc. Natl. Acad. Sci. USA, vol. 112, 2015, pp. E1754–62. ⟨https://doi.org/10.1073/pnas.1500973112⟩.
[130] J. Schmidt, A.R. Smith, M. Magnin, J. Racle, J.R. Devlin, S. Bobisse, J. Cesbron, V. Bonnet, S.J. Carmona, F. Huber, G. Ciriello, D.E. Speiser, M. Bassani-Sternberg, G. Coukos, B.M. Baker, A. Harari, D. Gfeller, Prediction of neo-epitope immunogenicity reveals TCR recognition determinants and provides insight into immunoediting, Cell Rep. Med. 2 (2021), 100194, https://doi.org/10.1016/j.xcrm.2021.100194.
[131] A.-M. Bjerregaard, M. Nielsen, V. Jurtz, C.M. Barra, S.R. Hadrup, Z. Szallasi, A. C. Eklund, An analysis of natural T cell responses to predicted tumor neoepitopes, Front. Immunol. 8 (2017) 1566.
[132] L.P. Richman, R.H. Vonderheide, A.J. Rech, Neoantigen dissimilarity to the self-proteome predicts immunogenicity and response to immune checkpoint blockade, Cell Syst. 9 (2019) 375–382, https://doi.org/10.1016/j.cels.2019.08.009 (.e4).
[133] J.R. Devlin, J.A. Alonso, C.M. Ayres, G.L.J. Keller, S. Bobisse, C.W. Vander Kooi, G. Coukos, D. Gfeller, A. Harari, B.M. Baker, Structural dissimilarity from self drives neoepitope escape from immune tolerance, Nat. Chem. Biol. (2020), https://doi.org/10.1038/s41589-020-0610-1.
[134] A.-H. Capietto, S. Jhunjhunwala, S.B. Pollock, P. Lupardus, J. Wong, L. Hänsch, J. Cevallos, Y. Chestnut, A. Fernandez, N. Lounsbury, T. Nozawa, M. Singh, Z. Fan, C.C. de la Cruz, Q.T. Phung, L. Taraborrelli, B. Haley, J.R. Lill, I. Mellman, R. Bourgon, L. Delamarre, Mutation position is an important determinant for predicting cancer neoantigens, J. Exp. Med. 217 (2020), https://doi.org/10.1084/jem.20190179.
[135] F. Duan, J. Duitama, S. Al Seesi, C.M. Ayres, S.A. Corcelli, A.P. Pawashe, T. Blanchard, D. McMahon, J. Sidney, A. Sette, B.M. Baker, I.I. Mandoiu, P. K. Srivastava, Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity, J. Exp. Med. 211 (2014) 2231–2248.
[136] V.P. Balachandran, M. Łuksza, J.N. Zhao, V. Makarov, J.A. Moral, R. Remark, B. Herbst, G. Askan, U. Bhanot, Y. Senbabaoglu, D.K. Wells, C.I.O. Cary, O. Grbovic-Huezo, M. Attiyeh, B. Medina, J. Zhang, J. Loo, J. Saglimbeni, M. Abu-Akeel, R. Zappasodi, N. Riaz, M. Smoragiewicz, Z.L. Kelley, O. Basturk, Australian Pancreatic Cancer Genome Initiative, Garvan Institute of Medical Research, Prince of Wales Hospital, Royal North Shore Hospital, University of Glasgow, St Vincent's Hospital, QIMR Berghofer Medical Research Institute, University of Melbourne, Centre for Cancer Research, University of Queensland, Institute for Molecular Bioscience, Bankstown Hospital, Liverpool Hospital, Royal Prince Alfred Hospital, Chris O'Brien Lifehouse, Westmead Hospital, Fremantle Hospital, St John of God Healthcare, Royal Adelaide Hospital, Flinders Medical Centre, Envoi Pathology, Princess Alexandria Hospital, Austin Hospital, Johns Hopkins Medical Institutes, ARC-Net Centre for Applied Research on Cancer, M. Gönen, A. J. Levine, P.J. Allen, D.T. Fearon, M. Merad, S. Gnjatic, C.A. Iacobuzio-Donahue, J.D. Wolchok, R.P. DeMatteo, T.A. Chan, B.D. Greenbaum, T. Merghoub, S. D. Leach, Identification of unique neoantigen qualities in long-term survivors of

pancreatic cancer, Nature 551 (2017) 512–516, https://doi.org/10.1038/nature24462.
[137] M. Łuksza, N. Riaz, V. Makarov, V.P. Balachandran, M.D. Hellmann, A. Solovyov, N.A. Rizvi, T. Merghoub, A.J. Levine, T.A. Chan, J.D. Wolchok, B.D. Greenbaum, A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy, Nature 551 (2017) 517–520, https://doi.org/10.1038/nature24473.
[138] M. Łuksza, Z.M. Sethna, L.A. Rojas, J. Lihm, B. Bravi, Y. Elhanati, K. Soares, M. Amisaki, A. Dobrin, D. Hoyos, P. Guasp, A. Zebboudj, R. Yu, A.K. Chandra, T. Waters, Z. Odgerel, J. Leung, R. Kappagantula, A. Makohon-Moore, A. Johns, A. Gill, M. Gigoux, J. Wolchok, T. Merghoub, M. Sadelain, E. Patterson, R. Monasson, T. Mora, A.M. Walczak, S. Cocco, C. Iacobuzio-Donahue, B. D. Greenbaum, V.P. Balachandran, Neoantigen quality predicts immunoediting in survivors of pancreatic cancer, Nature 606 (2022) 389–395, https://doi.org/10.1038/s41586-022-04735-9.
[139] D.K. Wells, M.M. van Buuren, K.K. Dang, V.M. Hubbard-Lucey, K.C.F. Sheehan, K. M. Campbell, A. Lamb, J.P. Ward, J. Sidney, A.B. Blazquez, A.J. Rech, J. M. Zaretsky, B. Comin-Anduix, A.H.C. Ng, W. Chour, T.V. Yu, H. Rizvi, J.M. Chen, P. Manning, G.M. Steiner, X.C. Doan, T. Tumor Neoantigen Selection Alliance, J. Merghoub, A. Guinney, C. Kolom, A. Selinsky, M.D. Ribas, N. Hellmann, A. Hacohen, J.R. Sette, N. Heath, F. Bhardwaj, R.D. Ramsdell, T.N. Schreiber, P. Schumacher, N.A. Kvistborg, Defranoux, Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction, Cell 183 (2020) 818–834, https://doi.org/10.1016/j.cell.2020.09.015 (.e13).
[140] P.R. Buckley, C.H. Lee, R. Ma, I. Woodhouse, J. Woo, V.O. Tsvetkov, D. S. Shcherbinin, A. Antanaviciute, M. Shughay, M. Rei, A. Simmons, H. Koohy, Evaluating performance of existing computational models in predicting CD8+ T cell pathogenic epitopes and cancer neoantigens, Brief. Bioinform. 23 (2022), bbac141, https://doi.org/10.1093/bib/bbac141.
[141] H. Huang, C. Wang, F. Rubelt, T.J. Scriba, M.M. Davis, Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening, Nat. Biotechnol. 38 (2020) 1194–1202, https://doi.org/10.1038/s41587-020-0505-4.
[142] M.V. Pogorelyy, E. Rosati, A.A. Minervina, R.C. Mettelman, A. Scheffold, A. Franke, P. Bacher, P.G. Thomas, Resolving SARS-CoV-2 CD4+ T cell specificity via reverse epitope discovery, Cell Rep. Med. 3 (2022), 100697, https://doi.org/10.1016/j.xcrm.2022.100697.
[143] J.B. Nilsson, S. Kaabinejadian, H. Yari, B. Peters, C. Barra, L. Gragert, W. Hildebrand, M. Nielsen, Machine learning reveals limited contribution of trans-only encoded variants to the HLA-DQ immunopeptidome by accurate and comprehensive HLA-DQ antigen presentation prediction, BioRxiv (2022), 507934, https://doi.org/10.1101/2022.09.14.507934.
[144] M.C.C. Raman, P.J. Rizkallah, R. Simmons, Z. Donnellan, J. Dukes, G. Bossi, G. S. Le Provost, P. Todorov, E. Baston, E. Hickman, T. Mahon, N. Hassan, A. Vuidepot, M. Sami, D.K. Cole, B.K. Jakobsen, Direct molecular mimicry enables off-target cardiovascular toxicity by an enhanced affinity TCR designed for cancer immunotherapy, Sci. Rep. 6 (2016) 18851, https://doi.org/10.1038/srep18851.
[145] X.X. Wang, Y. Li, Y. Yin, M. Mo, Q. Wang, W. Gao, L. Wang, R.A. Mariuzza, Affinity maturation of human CD4 by yeast surface display and crystal structure of a CD4-HLA-DR1 complex, Proc. Natl. Acad. Sci. USA 108 (2011) 15960–15965, https://doi.org/10.1073/pnas.1109438108.
[146] R. Vita, S. Mahajan, J.A. Overton, S.K. Dhanda, S. Martini, J.R. Cantrell, D. K. Wheeler, A. Sette, B. Peters, The Immune Epitope Database (IEDB): 2018 update, Nucleic Acids Res. 47 (2019) D339–D343, https://doi.org/10.1093/nar/gky1006.