Received: 19 April 2024 | Accepted: 1 August 2024

DOI: 10.1002/mas.21905

REVIEW ARTICLE

WILEY

# Exploring the dynamic landscape of immunopeptidomics: Unravelling posttranslational modifications and navigating bioinformatics terrain

**Daniel Flender**[1,2] | **Frédérique Vilenne**[2,3] | **Charlotte Adams**[4] | **Kurt Boonen**[1,5] | **Dirk Valkenborg**[3] | **Geert Baggerman**[4,5]

[1]Centre for Proteomics, University of Antwerp, Antwerpen, Belgium

[2]Health Unit, VITO, Mol, Belgium

[3]Data Science Institute, University of Hasselt, Hasselt, Belgium

[4]Department of Computer Science, University of Antwerp, Antwerp, Belgium

[5]ImmuneSpec, Niel, Belgium

**Correspondence**
Daniel Flender, Centre for Proteomics, University of Antwerp, 2020 Antwerpen, Belgium.
Email: daniel.flender@vito.be and frederique.vilenne@uhasselt.be

**Funding information**
Vlaamse Instelling voor Technologisch Onderzoek; Fonds Wetenschappelijk Onderzoek, Grant/Award Number: G070722N

**Abstract**

Immunopeptidomics is becoming an increasingly important field of study. The capability to identify immunopeptides with pivotal roles in the human immune system is essential to shift the current curative medicine towards personalized medicine. Throughout the years, the field has matured, giving insight into the current pitfalls. Nowadays, it is commonly accepted that generalizing shotgun proteomics workflows is malpractice because immunopeptidomics faces numerous challenges. While many of these difficulties have been addressed, the road towards the ideal workflow remains complicated. Although the presence of Post-translational modifications (PTMs) in the immunopeptidome has been demonstrated, their identification remains highly challenging despite their significance for immunotherapies. The large number of unpredictable modifications in the immunopeptidome plays a pivotal role in the functionality and these challenges. This review provides a comprehensive overview of the current advancements in immunopeptidomics. We delve into the challenges associated with identifying PTMs within the immunopeptidome, aiming to address the current state of the field.

**KEYWORDS**

bioinformatics, immunopeptidomics, neoantigens, PTMs

## 1 | INTRODUCTION

Our immune system has the capability to eliminate potential threats by engaging in intricate interactions between T cells and immunopeptides. Immunopeptides undergo processing and are bound to the major histocompatibility complexes (MHC), also referred to as human leukocyte antigen (HLA) complex in humans and will be presented on the surface of the cell. Through the presentation of immunopeptides, cells can communicate their internal condition to immune cells. When aberrant (none-self) proteins are being processed inside the cell, T-cell receptors on T-cells can recognize the immunopeptide-MHC complexes and will trigger an immune response. This process is essential for the detection and elimination of infections, and the removal of aberrant or cancerous cells.

In Immunopeptidomics, mass spectrometry is used for the analysis of immunopeptides. The comprehensive

Daniel Flender and Frédérique Vilenne contributed equally to this work.

overview of this peptide repertoire, commonly known as the immunopeptidome, serves as a valuable tool for unraveling immune recognition mechanisms and pathways. Beyond this primary application, it also provides insights into the cell's proteome, contributing to a deeper understanding of its physiological state, pathological conditions, and various cellular alterations (Shapiro & Bassani-Sternberg, 2023). As the field of immunopeptidomics advances, it significantly enhances our understanding of the intricate relationship between the immune system and different disease states. Immunopeptidomics, or the study of antigen presentation has shown its usefulness in many applications including the fields of personalized medicine, cancer immunotherapy, vaccinology, and related sciences (Arnaud et al., 2021; Barbier et al., 2022; Chu et al., 2022; Yamamoto et al., 2019).

In recent decades, significant advancements have been made in the field of immunopeptidomics, largely propelled by developments in sample preparation techniques, the enhanced sensitivity of mass spectrometry, and progress in bioinformatics. These strides build upon the pioneering work of Donald Hunt and Hans-Georg Rammensee in the early 1990s (Falk et al., 1991; Rammensee et al., 1993; Rötzschke et al., 1990). The aforementioned advancements have facilitated the examination and characterization of the immunopeptidome of cells or tissues in various states, such as healthy, stressed, and infected, while requiring ever smaller input material. This has, therefore, contributed to the creation of extensive databases (Marcu et al., 2021; Shao et al., 2018) and initiatives aimed at comprehensively mapping the human immunopeptidome (Vizcaíno et al., 2020). Identifying immunopeptides in general presents obstacles in both experimental and bioinformatic aspects, even more so when considering peptides that carry a posttranslational modification (PTM) (Faridi et al., 2018; Kacen et al., 2022; Smith & Rogowska-Wrzesinska, 2020). The abundance of immunopeptides can range from a minimum of a single copy to over 10,000 copies per cell (Hassan et al., 2014). This emphasizes the necessity for a high sensitivity and a large dynamic range of the mass spectrometry instrumentation (MS) typically used in immunopeptidomics. Furthermore, in contrast to the tryptic peptides generally analysed in proteomics, peptide antigens (immunopeptides) bound to MHC complexes are shorter (especially MHC I peptides) and often lack the typical basic amino acid residue at the C-terminus found in tryptic peptides. These factors result in unfavorable ionization properties and less predictable fragmentation patterns, ultimately posing challenges for interpretation of the spectra and identifying the peptides (Dudek et al., 2016).

The domain of cancer immunotherapy represents probably one of the most intriguing field of application for immunopeptidomics. For decades, researchers have concentrated on identifying tumor immunopeptides suitable for integration into epitope-specific cancer immunotherapies, aiming to elicit T-cell-mediated removal of tumors (Abbott & Ustoyev, 2019). These immunopeptides, known as tumor-specific antigens (TSAs) or neoantigens, result from the accumulation of generic alterations in cancers and must be absent in healthy tissue to minimize the potential for on-target off-tumor effects. Besides generic alterations and dysregulated transcription, neoantigens can also derive from dysregulated RNA splicing (Apavaloaei et al., 2020) or carry a PTM. Evidence of neoantigens derived from mutated proteins (Hogan et al., 1998) and PTM (Andersen et al., 1999; Haurum et al., 1999) has been documented for decades, both eliciting strong T-cell responses. However, despite their potential as leads for therapy, modified peptides are very often ignored in immunopeptidomics studies. This is likely due in part to the challenges that come with their identification which will be discussed later in this manuscript (Kacen et al., 2022). The peptides derived from dysregulated posttranslational modified proteins and presented by the MHC present an additional source of tumor-specific antigens that are interesting leads for cancer immunotherapies, potentially being highly specific but less dependent on a patient's personal genetic mutation profile. By decoding the distinct immunopeptides exhibited on malignant cells, scientists can customize individualized therapeutic interventions that effectively utilize the immune system's capabilities to target and eradicate these deviant cells specifically (Chakraborty et al., 2024).

This review delves into the field of immunopeptidomics, emphasizing the significance of PTMs on peptides bound to the MHC molecule as potential indicators of tumor-specific antigens or neoantigens. It provides an in-depth exploration of PTMs within the immunopeptidome, shedding light on the challenges associated with their identification. The comprehensive outline of the immunopeptidomics workflow is accompanied by a specific focus on detecting and characterizing PTMs. The review outlines various PTMs discovered in immunopeptidomics studies, highlighting the most common ones. Amidst the inherent complexity of identifying PTMs, the narrative extends to an overview of bioinformatic tools crucial for PTM identification, particularly those relevant to the intricate landscape of the immunopeptidome.

## 2 | IMMUNOPEPTIDE PROCESSING AND PRESENTATION

T-lymphocytes play a crucial role in the adaptive immune system by scanning the organism's tissues for infected and aberrant cells and eliminating them.

Conventional T-cells recognize repertoires of endogenously processed peptides bound to the MHC of cells. These immunopeptides can arise from intracellular pathogens, pathogen products absorbed from the extracellular fluid, or the natural cycling of regular (self-) proteins (Istrail et al., 2004).

There are two classes of MHC molecules: MHC I, recognized by CD8+ effector T cells, and MHC II, recognized by $CD4^+$ T helper cells. While MHC I is present in all somatic cells, except red blood cells, MHC II is mainly expressed by professional antigen-presenting cells. The peptides linked to MHC I and MHC II differ in length, intracellular pathways through which they are derived, and immune system functions, as extensively reviewed in (Cresswell et al., 2005) and (Roche & Furuta, 2015). Briefly, MHC I immunopeptides are shorter, predominantly 8–12 amino acids long, while MHC II peptides have more significant length variability, ranging from 13 to 25 amino acids (Wieczorek et al., 2017).

Figure 1 provides an overview of the antigen processing and presentation pathway of MHC I and II. Intracellular degradation of proteins into peptides through the proteasome plays a central role in processing MHC-presented peptides (Kloetzel, 2001). MHC I-bound peptides primarily result from the proteolysis of endogenous cytosolic proteins (Blum et al., 2013). Ubiquitin-labeled intracellular proteins, including those associated with pathogens, undergo degradation by the cytosolic proteasome. The products are then transported into the endoplasmic reticulum (ER) via the transmembrane protein TAP (transporter associated with antigen processing), where a variety of peptides are preselected and evaluated for binding to the MHC I molecule by the peptide loading complex, composed of the chaperones tapasin, calreticulin, and ERp57 (Blees et al., 2017). Peptides are N-terminally truncated by the aminopeptidase ERAP before and after loading onto the MHC complex. Finally, the entire peptide-MHC complex is transported to the cell surface via the Golgi complex and

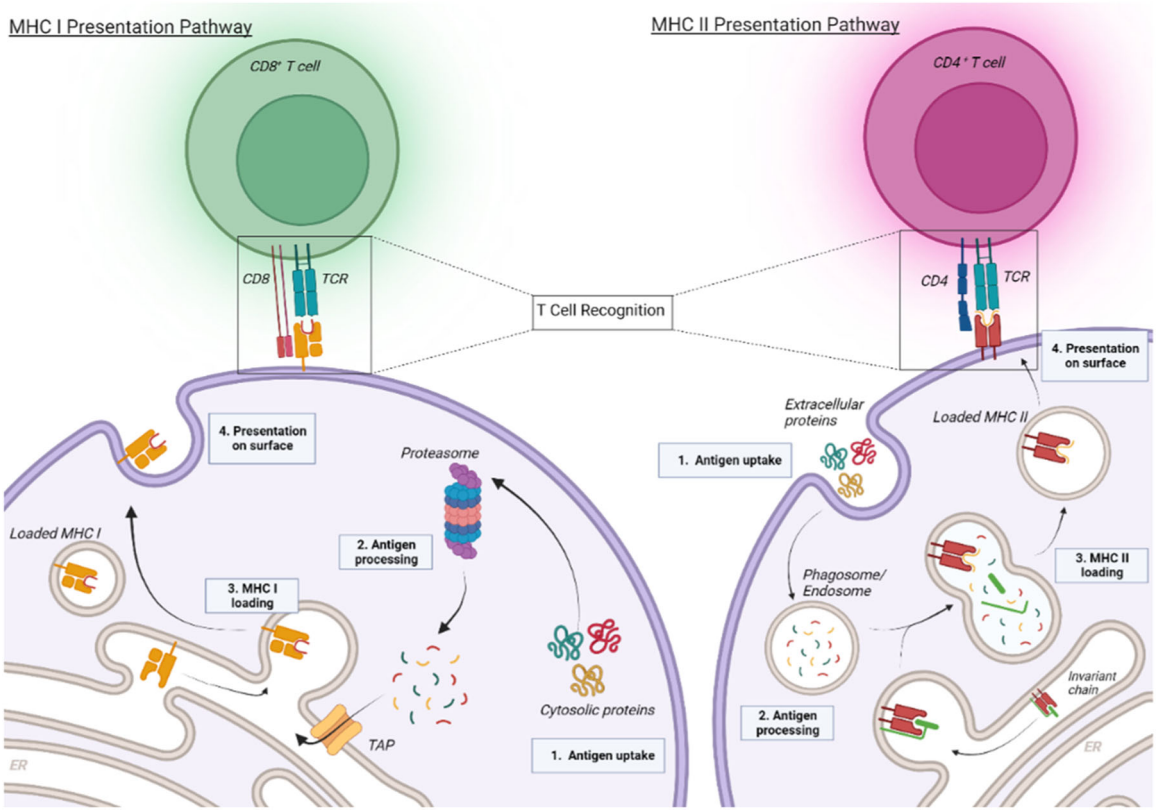


**FIGURE 1** Overview of major histocompatibility complex (MHC) Class I and Class II antigen presentation pathways: This schematic illustrates the key steps in MHC Class I and Class II antigen presentation pathways. In MHC Class I pathway (left), endogenous antigens are processed within the cytoplasm and presented on the cell surface to cytotoxic $CD8^+$ T cells. In MHC Class II pathway (right), exogenous antigens are engulfed, processed in endosomes, and presented to $CD4^+$ helper T cells. Both pathways play crucial roles in immune surveillance and orchestrate adaptive immune responses. Created with BioRender.com. [Color figure can be viewed at wileyonlinelibrary.com]

presented to $CD8^+$ effector cells. When the peptide is recognized and bound by a T-cell receptor on the $CD8^+$ T cell, it triggers an immune response. T-cell receptors on $CD8^+$ T cells typically recognize "nonself" peptides, such as those derived from pathogens or aberrant proteins.

In contrast to MHC I-presented peptides, the MHC II pathway presents peptides derived from extracellular proteins, degraded through the endosomal pathway (Suri et al., 2006). Mass spectrometry studies of MHC II-bound peptides indicate that 25%–55% of presented peptides originate from cytosolic proteins, suggesting that MHC II antigen processing and presentation are not restricted to exogenous antigens (Mommen et al., 2016; Stern & Santambrogio, 2016).

Autophagy is the primary process delivering intracellular proteins to phagosomes. In this process, protein aggregates or entire organelles in the cell are encased in a double membrane and fuse with lysosomes, ensuring the degradation of proteins and presentation of intracellular and extracellular protein-derived peptides via the MHC II complex (Stern & Santambrogio, 2016). Synthesized MHC class II molecules form complexes with an invariant chain (Ii), a nonpolymorphic protein. This chain includes targeting motifs guiding the Ii–MHC class II complex to antigen-processing compartments within endosomal–lysosomal pathways (Cresswell, 1996).

As mentioned, MHC II is expressed primarily by professional antigen-presenting cells, such as B cells, monocytes, macrophages, and dendritic cells. The presentation of MHC II in dendritic cells activates naïve $CD4^+$ T cells, which, in turn, aids the activation of effector $CD8^+$ T cells, triggering the adaptive immune response. Additionally, MHC II presentation is a crucial step in the activation of B cells and macrophages (Unanue et al., 2016). The discovery of MHC II-restricted cancer- or other disease-associated epitopes, along with a thorough understanding of the molecular mechanisms underlying MHC II antigen processing and presentation (Alspach et al., 2019; Linnemann et al., 2015), significantly improves the capacity to design novel and improved $CD8^+$ and $CD4^+$ T cell-based therapies through specific modifications with the potential to lead to personalized treatment (Chen & Jensen, 2008; Duru et al., 2020; van Stipdonk et al., 2009).

## 3 | IMMUNOPEPTIDOMICS WORKFLOW

### 3.1 | Sample preparation

The immunopeptidomics workflow consists of multiple steps and a standardized protocol for this workflow has been published by Purcell et al. (2019). In this review, we aim to provide a concise overview of the crucial steps involved in immunopeptidome analysis. This analysis differs significantly from classical proteomic approaches, presenting unique challenges throughout the workflow, as extensively reviewed by Faridi et al. (2018). Many hampering factors have been circumvented by improvements in the efficiency of isolating immunopeptides, the rapid development of instrument sensitivity, and improved bioinformatic pipelines, resulting in the ongoing project of mapping of the human immune-peptidome project (HIPP) (Caron et al., 2017). The starting material in an immunopeptidomics experiment can consist of cell lines, primary cells, and fresh or fresh-frozen tissue samples. Although the recent advances in the sensitivity of mass spectrometry (e.g., the increased focus on single-cell proteomics) have reduced the amount of material needed for an immunopeptidome analysis, the amount of sample material available remains a crucial limiting factor. An interesting development in that aspect is the use of patient-derived organoids (PDO). Hereby, a small patient biopsy can be multiplied to a larger sample, with the intent of mimicking the patient's initial biopsy condition (Wang et al., 2022).

Figure 2 provides an overview of the current immunopeptidomics workflow. Two well-established techniques for isolating HLA-bound peptides from tissue or cells are mild acid elution (MAE) (Sugawara et al., 1987) and Immunoprecipitation (Subramanian, 2002). The commencement of the immunopeptidomics workflow with immunoprecipitation involves the lysis of tissue or cells. This critical step employs a mild detergent to ensure the solubilization of MHC, a transmembrane protein, for subsequent processing. Monoclonal antibodies (mAbs) are then used for the immunoaffinity capture of MHC-peptide complexes. The selection of mAbs is contingent upon the research objective and the HLA allomorph of the investigated cells. A comprehensive list of commonly used mAbs and their references can be found in the protocol published by Purcell et al. (2019). Depending on the research question at hand, using specific antibodies for a particular HLA allele may offer higher specificity. Most commonly used for HLA class I isolation is the mouse mAb W6/32, which is specific to the heavy chain of HLA-A, -B, and -C (Brodsky & Parham, 1982). This allows for the monomorphic isolation of HLA class I peptides. Polymorphic antibodies, such as ME1 against HLA-B07, BB7.2 against HLA-A02, and GAP.A3 against HLA-A*03, target and recognize determinants carried by specific alleles (Hilton & Parham, 2013). For HLA II on the other hand, the pan-HLA class II antibody supposedly pull down the HLA molecules DR, DQ, and DP. However,

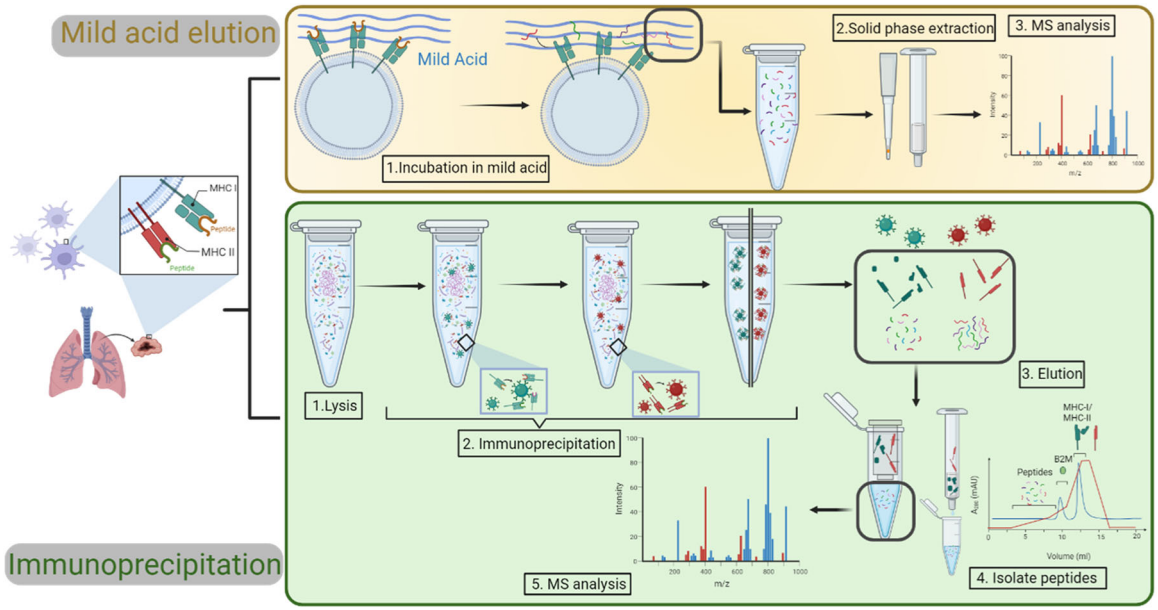


**FIGURE 2** Workflow of mild acid elution and immunoprecipitation in immunopeptidomics: This overview illustrates the alternative approaches of mild acid elution (top, in yellow) and immunoprecipitation (bottom, in green) in immunopeptidomics. The yellow pathway represents mild acid elution, where peptides are gently released from major histocompatibility complex (MHC) molecules, while the green pathway depicts immunoprecipitation, a technique involving the selective capture of MHC-peptide complexes using specific antibodies. Researchers can choose between these two methods based on their experimental goals, providing flexibility in the analysis of MHC-bound peptides and contributing to a comprehensive understanding of the immunopeptidome. Created with BioRender.com. [Color figure can be viewed at wileyonlinelibrary.com]

immunopeptidomics studies have demonstrated poor specificity towards DQ and DP molecules, hence resulting in low yields of peptides bound to these loci and underrepresentation of DQ and DP in datasets(Nilsson et al., 2023; Racle et al., 2019). Additionally, even though a lot of studies focus on DRB1 alleles, it has been shown that DRB3, 4, and 5 play an important role in forming the HLA II immunopeptidome, underlying the importance of full HLA typing (Kaabinejadian et al., 2022). Following immunoprecipitation, the nonspecific lysate components are thoroughly washed away, leaving only the MHC-peptide complexes. The MHC-peptide complex is then eluted from the mAbs using a mild acid, dissociating the peptide from the MHC. Eluted immunopeptides undergo additional enrichment processes to reduce complexity and eliminate contaminants. Standard techniques for enrichment are solid-phase extraction (SPE) (Bassani-Sternberg et al., 2010) and ultrafiltration using a molecular cutoff (MWCO) filter (Sturm et al., 2021). These techniques enhance the sensitivity and selectivity of subsequent analyses. During the enrichment procedure, it is crucial to remove the intact proteins that are part of the MHC complex, such as beta 2 microglobulin and the MHC alpha chain, as their presence interferes with downstream analysis. Utilizing restricted access material (RAM) into solid-phase extraction enhances the recovery of immunopeptides, particularly favoring hydrophobic ones. This improvement is attributed to the selective non-retention of contaminating proteins by the RAM material, allowing for the application of a higher concentration of organic solvent during the elution process (Bernhardt et al., 2022). Kuznetsov et al. (2020) published a thorough review emphasizing various techniques for the isolation of the immunopeptidome.

MAE, in contrast to immunoprecipitation, utilizes viable cells to isolate peptides presented by the MHC on cell surfaces. In this method, cells are treated with a mild acid, causing peptides to separate from the MHC. Importantly, it's highlighted that peptides associated with MHCII are believed to remain unaffected by this process (Sugawara et al., 1987). When evaluating the two methods using B-cell lymphoblasts, the Immunoprecipitation technique results in a significant 6.4-fold increase in the detection immunopeptides (Lanoix et al., 2018). Furthermore, a different study comparing an enhanced MAE technique with immunoprecipitation revealed a nearly 50% overlap in the number of identifications between the two approaches (Sturm et al., 2021). There is a crucial need for a high-throughput analysis methodology to enable the application of immunopeptidomics in mid to large-sized cohorts and for future comparative studies. As Sian et al. suggested, the semi-automated workflow

revolves around the extraction of MHC-bound peptides through for instance magnetic beads for immunoprecipitation, performed on an automated platform employing 96-well plates. According to their protocol, it is possible to simultaneously process 12 samples, identifying roughly 400 to 13,000 distinct peptides, derived from cell quantities ranging from 0.5 to 50 million (Lim Kam Sian et al., 2023). However, it should be noted that in a high throughput set-up the MS analysis and the subsequent data analysis are very likely to become the bottleneck as both are time-consuming steps in the analysis. Considering the significance of both high throughput and sensitivity, Li et al.'s research team developed a workflow utilizing a microfluidics platform. This platform facilitates automated liquid handling and minimizes sample transfers. Through this streamlined process, the team successfully identified more than 4000 and 5000 MHC-I-restricted immunopeptides from 0.2 million RA957 cells (human B-cell line) and 5 mg of melanoma tissue, respectively (Li et al., 2023).

## 3.2 | Mass spectrometry analysis of isolated immunopeptides

The enriched and cleaned-up peptides are then subjected to liquid chromatography (LC) tandem mass spectrometry (MS) analysis. Selecting the proper instrumentation is crucial for achieving comprehensive immunopeptidome coverage. To acquire purified immunopeptides effectively, it is advisable to employ a high-end, state-of-the-art LC-MS/MS system. Given the biochemical similarities of immunopeptides, a high-resolution LC system is essential. Nano LCs are favored over microflow LCs in immunopeptidomics approaches, primarily due to the usual low sample quantities and the need for the highest possible sensitivity. Some studies suggest the inclusion of charge-enhancing additives such as dimethyl sulfoxide or m-nitrobenzyl alcohol to enhance performance (Nielsen & Abaye, 2013; Van Wanseele et al., 2019). However, the effectiveness of improving electrospray ionization and sensitivity may vary, depending on factors such as the emitter source type and specific parameters like source voltage, temperature, and gas flow. For the mass spectrometer, opting for a high-mass accuracy analyzer is a requirement as it enhances the reliability of peptide sequence identification. It is important to note that depending on the MHC allele of the sample, a substantial number of singly charged peptides may be present. Strategies for increasing the charge state of these peptides using chemical derivatization and isobaric labeling have been shown (Chen et al., 2018; Pfammatter et al., 2020).

The ongoing improvement of mass spectrometry instruments has advanced the field of immunopeptidomics immensely. For example, cutting-edge mass spectrometers coupled with ion-mobility separation and/or trapping devices (IMS) offer increased identifications and improved signal-to-noise ratios, further enhancing the immunopeptidome analysis process (Meier et al., 2018). Moreover, a recent study has revealed that MS/MS spectra from timsTOF instruments exhibit more reproducibility at low abundances compared to MS/MS spectra from Orbitrap instruments (Hoenisch Gravel et al., 2023).

In the realm of immunopeptidomics, three primary methods are employed for identifying and quantifying immunopeptides. Most studies in this field rely on data-dependent acquisition (DDA) to maximize the amount of information acquired within an experiment (Caron, Kowalewski, et al., 2015). In DDA, the ionized immunopeptides are first detected in a survey scan (MS1 scan). The most abundant precursor ions detected are further isolated for fragmentation, generating individual peptide spectra (fragmentation spectra) and maximizing the information extracted from the experiment (Gatlin et al., 2000). However, in DDA, precursor ion selection follows a straightforward intensity-dependent rule, resulting in the (semi-)random isolation of precursor ions for fragmentation. This randomness can lead to inconsistent peptide identification when analyzing the same sample repeatedly, compromising reproducibility. Consequently, DDA is less suitable for accurately quantifying immunopeptides across different samples. On average, approximately 20% of the selected MHC I immunopeptides have been observed to vary between replicate analyses of the same sample (Caron, Espona, et al., 2015). On the contrary, data-independent acquisition (DIA) adopts a different approach by acquiring fragmentation data for all ions within predefined precursor isolation windows, providing a comprehensive map of a given sample. Since its introduction in 2004, several DIA strategies have been described and reviewed in detail (Chapman et al., 2014; Distler et al., 2014; Sajic et al., 2015). DIA captures comprehensive data by simultaneously fragmenting and analyzing all precursor ions in a sample, creating a detailed digital map of the sample's molecular components. However, since chimeric spectra are produced from fragments of multiple peptide precursors rather than individual fragmentation spectra, successfully implementing a DIA strategy requires the development of a comprehensive spectral library. A specific constraint in DIA experiments is the need for specialized approaches to identify individual peptides, which require prior knowledge of the peptides' fragmentation and

chromatographic behavior. This information is typically obtained through the construction of spectral libraries based on previous DDA analysis of similar samples, or through the use of retention time and fragmentation spectra modeling based on peptide sequences. These approaches are particularly challenging for immunopeptidomics, as immunopeptides can vary significantly between samples depending on the HLA genotypes present (Lou & Shui, 2024). This can be challenging, especially when investigating non-canonical peptides and neoantigens. In proteomics, a spectral library is typically generated from experimental data using multiple data-dependent data sets. Since the immunopeptidome is by definition patient-specific, one cannot easily resort to publicly available datasets to construct a library. In essence, ideally a specific library would have to be generated per individual to be comprehensive enough to be of use for potentially actionable immunopeptides. Limited availability of clinical biopsies makes it difficult to generate DDA data for a spectral library while still having enough material for another DIA analysis. Recent prediction models, however, allow in silico generation of spectral libraries, offering a solution to the challenge of relying on clinical biopsies for spectral library creation. A recent study implementing a DIA immunopeptidomics workflow using spectral libraries of increasing complexity reported compelling results. Utilizing a complex multi-MHC spectral library of previously measured DDA results resulted in a twofold increase in peptide identification compared to a sample-specific spectral library and a threefold increase compared to a DDA approach. Additionally, the study demonstrated the successful application of DIA for neoantigen discovery by analyzing DIA data with predicted MS/MS spectra (generated by a trained model) of clinically relevant immunopeptides (Gessulat et al., 2019; Pak et al., 2021). Nevertheless, the utility of these in silico prediction models is constrained when it comes to identifying PTMs. In the study mentioned earlier, peptides containing cysteine were deliberately omitted. This decision was driven by the fact the trained model for spectral library prediction assumes the carbamidomethylation of all cysteines, reflecting the characteristics of the training data upon which the models rely, derived from proteomics data.

An alternative method for the identification and quantification of immunopeptides is the targeted data acquisition mode. In contrast to DDA and DIA, a targeted approach offers high specificity, sensitivity, reproducibility, quantitative accuracy, and a wide dynamic range (Gallien et al., 2014; Peterson et al., 2012; Picotti & Aebersold, 2012). Typically conducted in Selected/Multiple Reaction Monitoring (S/MRM) and Parallel Reaction Monitoring (PRM) modes, this method is commonly employed on triple quadrupole and quadrupole Orbitrap instruments (Peterson et al., 2012; Picotti & Aebersold, 2012). It is important to note that a targeted approach necessitates prior knowledge of the specific target of interest (Croft et al., 2013; Gubin et al., 2014; Hassan et al., 2014; Hogan et al., 2005; Tan et al., 2011). In certain instances, studies employing Selected/Multiple Reaction Monitoring (S/MRM) for both relative and absolute quantification have been conducted. For instance, in a study by Hassan et al. (2014), heavy-labeled peptides loaded on MHC class I were utilized for accurate quantitation of peptide yield isolated using immunoprecipitation (IP) (Hassan et al., 2014). Utilizing isobaric labeling for targeted measurements enables the simultaneous analysis of multiple samples. A recent technique known as TOMAHAQ employs synchronous precursor selection (SPS-MS3) to enhance quantitative accuracy. This approach uses a dedicated MS3 fragmentation event for quantification, effectively eliminating interference from co-isolated or co-fragmented peptides (Pollock et al., 2021; Rose et al., 2019). Applying this methodology in immunopeptidomics extends quantitative sensitivity to the low amol/μL level (Pollock et al., 2021). Despite the advantages of targeted approaches, the drawback of requiring prior knowledge about the investigated targets limits their standalone applicability. Consequently, the implementation of targeted approaches appears to be more complementary to other acquisition methods in immunopeptidomics and its use is largely restricted to validation studies.

Another crucial factor influencing immunopeptidome coverage is the choice of fragmentation technique. In a study by Mommen et al. (2014), the combination of electron-transfer/higher-energy collision dissociation (EThcD) led to a threefold increase in the identification of MHC-I immunopeptides compared to traditional fragmentation techniques like collision-induced dissociation (CID), higher-energy collisional dissociation (HCD), or a combination thereof. EThcD outperforms conventional methods in sequencing peptides bound to HLA-B7 and -B27 molecules, which are challenging due to the preferred presence of proline or arginine as an anchor residue as the second amino acid in immunopeptide's sequence. ETD cannot cleave proline's N-Cα bond, and CID/HCD faces difficulties with internal arginine residues. EThcD's dual-ion series overcomes these limitations, leading to an almost twofold improvement in identifying peptides associated with HLA-B7 and –B27 peptides.

## 4 | THE ROLE AND CHALLENGES OF PTMS IN IMMUNOPEPTIDOMICS

### 4.1 | PTMs—A key role in healthy and diseased cells

The regulation of cellular activities, including the biological activity of proteins, cellular localization, and the assembly of protein complexes, often relies on PTMs that occur naturally on proteins. Enzymes like kinases, phosphatases, and glycosyl- or acyltransferases, among others, catalyze these modifications, which are reversible in most cases. The common naturally occurring PTMs, such as phosphorylation, glycosylation, methylation, acetylation, ubiquitination, and citrullination, are essential for cellular processes, including for proper immune responses (Mann & Jensen, 2003; Seo & Lee, 2004; Zavala-Cerna et al., 2014). An overview of the common PTMs in eukaryotic cells can be seen in Figure 3.

As previously noted, the modifications of proteins can result in modified immunopeptides, which are then presented by MHC I/II molecules, possibly triggering T-cell responses. Despite their potential significance, PTMs on immunopeptides remain understudied. This primarily stems from the limitation of the methods employed. In general, detection of PTMs requires highly sensitive mass spectrometers, dedicated fragmentation techniques, specific enrichment strategies, and/or dedicated data analysis. These methods have been applied extensively in proteomics. However, it is only with the recent advancements in mass spectrometry instrumentation, particularly in terms of enhanced sensitivity, that the extension of these analyses to immunopeptides has become feasible.

### 4.2 | Experimental challenges in mass spectrometry analysis of PTMs

MS/MS offers a range of essential analytical capabilities for determining the amino acid sequence of a peptide and, in most cases, not only identifying a PTM but also pinpointing the specific amino acid residue it affects. However, analyzing PTMs using mass spectrometry can be challenging due to several factors. Numerous reviews discuss the challenges and approaches for identifying PTMs at the protein and peptide levels (Doll & Burlingame, 2015; Kim et al., 2016; Smith & Rogowska-Wrzesinska, 2020). While many challenges in detection and localization from a proteomics perspective also apply to immunopeptidomics, there are key challenges that significantly impact PTM detection in immunopeptidomics.

To analyze immunopeptides with PTMs using mass spectrometry, they must be ionized. However, certain PTMs can affect a peptide's ionization efficiency, making

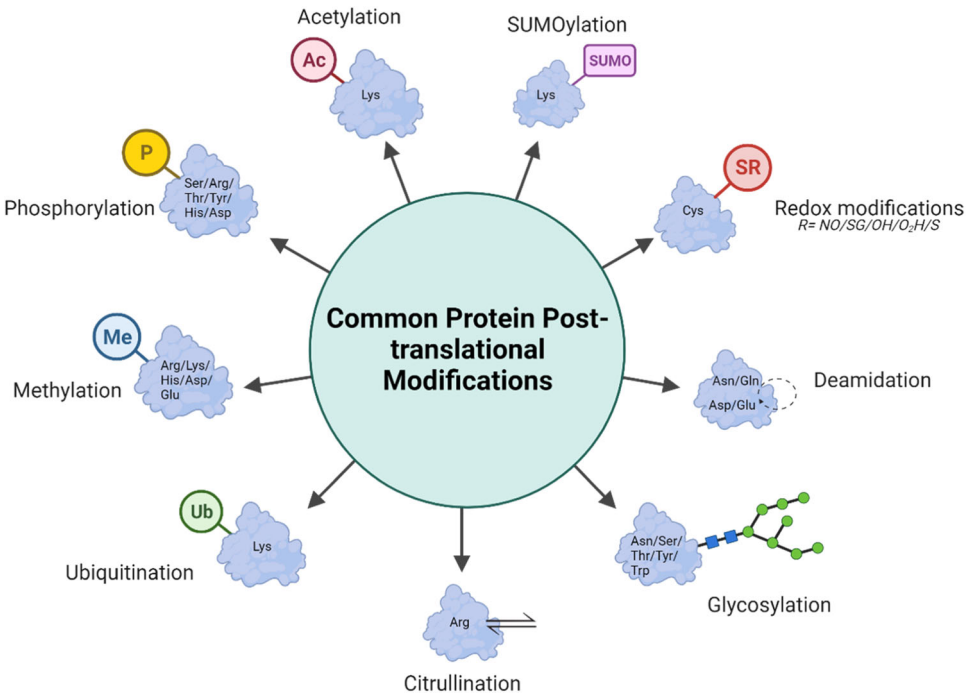


**FIGURE 3** Overview of common posttranslational modifications (PTMs) in eukaryotic cells. The schematic representation illustrates various PTMs occurring on proteins in eukaryotic cells. [Color figure can be viewed at wileyonlinelibrary.com]

them more difficult to detect than their unmodified counterparts (Gao & Wang, 2007). Once a modified immunopeptide is ionized, the PTM can alter the peptide's fragmentation pattern, complicating the accurate interpretation of the mass spectra (Kim et al., 2016). Additionally, post-translationally modified peptides are often less abundant than unmodified peptides, further complicating their detection (Larsen et al., 2006).

Moreover, identifying immunopeptides with PTMs does not necessarily attribute the modification to a biological origin. For example, methionine oxidation can occur during sample preparation, and asparagine deamidation is sensitive to pH, temperature, and buffer type, underscoring the importance of selecting appropriate sample preparation methods and materials. These sample preparation artifacts can ultimately complicate data interpretation (Liu et al., 2013; Morand et al., 1993; Pace et al., 2013).

Confidently identifying PTMs on a peptide requires the PTM to persist throughout the sample preparation process and downstream MS/MS analysis. Chemical stability is crucial for efficient PTM detection in MS/MS, with stable PTMs like acetyl-lysine maintaining integrity and causing a 42 Da mass increment for the intact peptide. Less stable PTMs, such as phosphoserine and phosphothreonine, often undergo neutral loss ($-H_3PO_4$) during MS/MS, resulting in a mass deficit of 98 Da in fragment ions (Annan et al., 2000). These Δm values prove valuable for annotating PTM peptide MS/MS spectra, aiding in the identification of modified peptides. However assignment of modified residues requires often a specific approach. In electron transfer dissociation (ETD), the protein or peptide is fragmented using low-energy anions produced by the chemical ionization of fluoranthene, ETD typically causes fragmentation along the N-Cα bonds, preserving the modified amino acids (Yu & Veenstra, 2021). Beyond sample preparation and MS/MS analysis, challenges persist in the data analysis phase of identifying PTMs. These challenges will be further explored in the forthcoming data analysis chapter.

## 4.3 | Significance and applications of modified immunopeptides

The elusive question concerning modified immunopeptides is, whether they do exist in vivo, and what their relevance is. Can they trigger a specific T-cell response? As mentioned earlier, PTMs play a crucial role in the function and structure of proteins. Since the immunopeptidome mirrors the proteome, various PTMs should be present on MHC I and MHC II (Admon & Bassani-Sternberg, 2011). In a couple of studies, the presentation and recognition by cytotoxic T-lymphocytes of modified peptides have been demonstrated for phosphorylated and glycosylated peptides in particular (Andersen et al., 1999; Haurum et al., 1999). More recent MS and bioinformatics analysis suggest that modified peptides may account for 12%–25% of the total immunopeptidome (Prus et al., 2019). Additionally, research has demonstrated that the physiochemical alterations caused by PTMs can influence the binding affinity of peptides to specific allotypes, further indicating that PTMs may have an impact on antigen presentation (Sidney et al., 2018).

Aberrant PTMs have been confirmed to result in antigenic peptides and the cognate T-cell receptor (TCR) recognition in various instances (Apostolopoulos et al., 2003; Vlad et al., 2002). Notably, these immunogenic peptides, arising from dysregulated PTMs in cancer cells, represent an understudied class of potentially tumor-specific (neo)antigens. These neoantigens hold great potential as lead candidates to be used in anticancer immunotherapy. Moreover, because they do not involve patient-specific or tumor-specific mutations, they have the potential of being shared across individuals independent of the specific tumor-associated mutations, making them promising 'off-the-shelf' targets for immunotherapy.

Accumulating evidence suggests that alternative sources of cancer neoantigens, such as gene fusions, alternative splicing variants, translated noncoding sequences -and PTMs, hold promise as novel targets for immunotherapy. Particularly, neoantigens derived from gene fusions, recurrent mutations in cancer driver genes, noncoding regions, and aberrant PTMs are more likely to be shared among patients, providing a readily available resource for immunotherapy (Capietto et al., 2022; Smith et al., 2019; Zhou et al., 2019). In the upcoming chapters, we will explore the PTMs that exhibit the highest potential as neoantigens, spotlight the modified immunopeptides currently identified and briefly describe the challenges in PTM detection. Advancements and Applications in Glycosylation Analysis Glycosylation is essential for physiological and pathological cellular activities; advancements in analytical techniques have fueled progress in the study of glycobiology during the last decade. For a more detailed overview of the function of glycosylation in health and disease, we refer to the review written by Reily et al. (2019).

A significant function of glycans is to contribute to protein stability and solubility, including appropriate molecular orientation and reduction of nonspecific protein–protein interactions. However, glycosylation can also be present on peptides presented by the MHC and

form tumor-specific neoantigens (Malaker, Penny, et al., 2017; Marino et al., 2015). While glycopeptides have been isolated from MHC-II molecules for some time, the evidence for the existence of MHC-I glycopeptides was initially limited (Chicz et al., 1993). The first documentation of glycosylated MHCI immunopeptides was reported by Haurum et al. (1999). Their groundbreaking work demonstrated that naturally presented MHC-I peptides include a small subset of glycopeptides, primarily featuring O-β-N-acetylglucosamine (GlcNAc) substitutions on serine and threonine residues. Using synthesized glycopeptides, they further established that these peptides were immunogenic in mice and could be transported by the TAP transporter for antigen presentation (Haurum et al., 1999).

Characterizing and quantifying intact glycopeptides from complex datasets remains challenging due to their inherent glycan heterogeneity, unique ionization and separation characteristics, and relatively low abundance compared to non-modified peptide counterparts. In a study conducted by Mukherjee, ion mobility-coupled mass spectrometry devices were used to separate glycosylated peptides by their collisional cross-section (CCS). Ion mobility emerged as a promising tool for characterizing glycopeptides due to their distinct physical properties. By isolating glycopeptides and optimizing glycan-dedicated stepped collision energy, the extraction of information is improved, resulting in less ambiguous and more complete fragmentation spectra. Although this study was limited to trypsin-digested peptides, the analysis should also be applicable to immunopeptidomics samples (Mukherjee et al., 2023).

Glycosylation plays a role in many autoimmune diseases and cancer, therefore providing an interesting target to study in the context of immunopeptidomics. Rheumatoid arthritis (RA) is linked to specific MHC class II alleles and is characterized by a recurrent autoimmune response in the joints. Collagen type II (CII), the primary component of hyaline cartilage, has been hypothesized as a potential autoantigen in rheumatoid arthritis because CII-specific antibodies are commonly identified in RA patients. In a study using transgenic mice expressing human DR4 and human CD4, T cells predominantly recognized the immunodominant type II collagen antigen in its glycosylated form (Bäcklund et al., 2002; Corthay et al., 1998). In tuberculosis, $CD8^+$ T cells recognize the glycosylated peptide presented by HLA-E, a nonclassical MHC I molecule (Harriff et al., 2017). Additionally, the alteration in protein glycosylation within tumor cells has a decisive impact in all stages of the disease. This could impact the HLA-epitopes and lead to the overexpression of standard glycoproteins and tumor-specific glycoproteins (Peixoto et al., 2019). Several studies have shown T-cell reactivation towards glycosylated epitopes presented by MHC I and II, underlining the importance of these epitopes (Kastrup et al., 2001; Unanue et al., 2016).

In a study profiling the prevalence of PTMs among a subset of HLA allotypes, about 17%–38% of the identified HLA I-bound peptides were found to be post-translationally modified. Deamidated peptides were the second most common PTM found in the immunopeptidome, accounting for 2.5%–7% of the detected modified peptides. A large proportion, nearly 59%, of the deamidated peptides, has the common N-glycosylation motif NX (S/T), suggesting that most deamidations originate from glycosylated proteins that have undergone deglycosylation via the ER-associated protein degradation pathway (ERAD) (Mei et al., 2020). In a study conducted by Malaker, Penny, et al. (2017), investigating melanoma cell lines, a total of 93 MHC II-bound glyco-peptides were identified (Malaker, Ferracane, et al., 2017). Out of these peptides, a vast majority carried the glycosylation on flanking residues, not interfering with the peptide binding domain of the MHC. Recent research has identified glycosylated HLA-DR-bound peptides in dendritic cells pulsed with the SARS-CoV-2 spike protein. The glycosylation profile of the HLA-II-bound S-protein-derived peptides shows substantial trimming of glycan residues compared to the S protein, likely due to antigen processing (Parker, Partridge, et al., 2021).

### 4.3.1 | Phosphorylation in immunopeptidomics

Phosphorylation, a common PTM of proteins, involves the addition of a phosphate group ($-PO_4$) to specific amino acid residues, predominantly serine, threonine, and tyrosine. This modification is a crucial cellular signaling pathway that regulates various physiological processes, including cell division, growth, differentiation, and apoptosis. Dysregulation or overexpression of protein phosphorylation is implicated in several disorders, such as cancer, diabetes, and Alzheimer's disease.

Functioning as a vital physiological regulatory mechanism, phosphorylation is orchestrated by protein kinases and phosphatases. These enzymes govern diverse cellular activities, encompassing protein synthesis, cell division, signal transduction, growth, development, and aging. Protein kinases, key players in cellular transduction signaling, are often associated with pathologies, particularly cancer, when overexpressed or dysfunctional (Ardito et al., 2017). Exploiting specific modifications, notably phosphorylation, has become a focal point in

novel cancer therapies (Mazhar et al., 2019). In the review by Zeneyedpur et al. (2020), the emphasis is on utilizing proteomics and next-generation sequencing to identify phosphopeptide neoantigens for potential immunotherapeutic and diagnostic applications in cancer and autoimmune diseases (Zeneyedpour et al., 2020).

When conducting phosphorylation analysis, immunopurified and phosphorylated peptides can undergo different enrichment techniques, like immobilized metal-affinity chromatography (IMAC). IMAC selectively captures and enriches phosphorylated peptides for subsequent mass spectrometry analysis (Ficarro et al., 2002; Zarling et al., 2006). This technique has been subsequently modified for enriching phosphorylated immunopeptides (Abelin et al., 2015). In 1998, a mass spectrometry HLA-I peptide analysis revealed the first phosphorylated MHC-bound peptide capable of triggering a cytotoxic T-cell response (Hogan et al., 1998). Shortly thereafter, a study substantiated the effective transportation of phosphorylated peptides from the cytosol to the ER by the Transporter Associated with Antigen Processing (TAP), facilitating their subsequent loading onto MHC-I molecules (Andersen et al., 1999). Overall, the antigen processing and presentation pathway exhibit general consistency for both phosphoproteins and unmodified proteins, involving both MHC-I and MHC-II presentation. At present, in 15 different types of cancer, over 2500 phosphorylated peptides have been identified, with about 1000 of these peptides being part of cancer-associated pathways in multiple cancer types. Nearly 80% of the studied phosphorylated peptides could indeed generate memory T cells from monocytes of healthy donors demonstrating their antigenic properties (Mahoney et al., 2021). In the case of phosphorylation, modified peptides that can induce a T-cell response in healthy donors are promising immunotherapeutic targets for treatment of cancers and other diseases (Cobbold et al., 2013). Briefly summarized, there are multiple examples in which MHC I-bound phosphorylated peptides are uniquely expressed on malignant cells in breast cancer (Hunt et al., 2022), melanoma (Topalian et al., 2016), colorectal cancer (Hunt et al., 2019), leukemia (Cobbold et al., 2013) and other cancer types (Hunt et al., 2013; Mohammed et al., 2008, 2017; Zarling et al., 2006). Notable are the results of a recent preclinical trial utilizing two class I MHC phosphopeptides to treat high-risk melanomas (Engelhard et al., 2020). While adverse effects in patients were minimal, T-cell responses were observed for one phosphopeptide in 5 out of 12 patients and 2 out of 12 for the second phosphopeptide, demonstrating the potential of specific protein phosphorylation as a genuine target for T-cell-based immune therapy.

In general, phosphopeptides comprise up to 1% of its composition and approximately 3%–4% of all PTMs within the immunopeptidome, underscoring that while understudied they may provide valuable leads for immune therapy development (Zarling et al., 2000).

### 4.3.2 | Citrullination in immunopeptidomics

In general, the detection of citrullinated peptides presents a challenging task, primarily because of the subtle net increase in the molecular mass of the peptide, amounting to only 0.984016 Da. Peptides are not visible in a spectrum as a single peak, but rather as an isotopic envelope. This means that the selection of the wrong peaks in this envelope can result in a mass shift of + 1 Da, potentially falsely identifying the peptide as being modified. To ensure accurate assignment of the modification, it is imperative that the fragment ion MS/MS comprehensively covers the citrullination site and that isotopic distribution is taken into account (Hensen & Pruijn, 2014).

An analysis of C1R cell lines expressing HLA-A*01:01, HLA-A*02:01, or HLA-A*24:02 revealed that approximately 1% of PTM peptides presented by each of these prevalent MHC-I allotypes were citrullinated (Mei et al., 2020). Contrary to other PTMs, citrullination does not change the structure of the peptide but instead modifies the electrostatic potential of the side chain (Sandalova et al., 2022). A study has demonstrated that the citrullination of an MHC II peptide induces $CD4^+$ T-cell response, while the amino acid sequence without the PTM does not elicit this T-cell response. This demonstrates the potential of TCR-interacting citrullinated peptides (Ting et al., 2018).

In a study comparing HLA subtypes HLA-B*27:05 and HLA-B*27:09, where the former is closely linked to ankylosing spondylitis, it was revealed that peptide citrullination can markedly change the binding configuration of the modified epitope (Beltrami et al., 2008). Crystallographic studies reveal that both HLA subtypes only differ in residue 116 within the peptide binding groove, yet they present the modified self-peptide in distinct formations (Beltrami et al., 2008). These structural disparities lead to varying responsiveness of $CD8^+$ T cells restricted by HLA-B27, indicating that the introduction of citrullinated peptides in presentation holds the capability to impact immune reactions (Beltrami et al., 2008).

## 5 | DATA ANALYSIS IN IMMUNOPEPTIDOMICS

As immunopeptidomics relies on the same technologies as proteomics, most data analysis concepts are shared. However, immunopeptidomics suffers from additional

challenges. Both Menschaert et al. (2010) and Maes et al. (2019) discussed several of these challenges in the field of peptidomics from a bioinformatics point of view. In 2018, Faridi et al. discussed several key challenges within immunopeptidomics (Faridi et al., 2018). Specifically, it discusses the drawbacks of generalizing shotgun proteomics workflows in the field of immunopeptidomics and suggests improvements.

Specific challenges in immunopeptidomics complicate the data analysis in comparison with proteomics. As mentioned previously, the absence of a specific cleavage enzyme, less favorable ionization characteristics, and as a result, less predictable fragmentation patterns all make identifying immunopeptides more difficult compared to tryptic peptides in a typical proteomics set-up. The absence of specific enzymatic cleavage is a substantial challenge in immunopeptidomics. One of the key elements here is that the underlying proteolytic processes are not completely known and complex since multiple proteases and peptidases are involved making the outcome of the protein processing unpredictable and may vary among individuals. Furthermore, the biochemical properties, precisely the length and sequence motif, between immunopeptides are dissimilar and dependent on the HLA genotype. This requires the algorithm to be able to perform searches with nonspecific enzymatic cleavage when relying on reference sequences or spectra, for example, database search engines or spectral library searching. Nonspecific cleavage of databases, either spectral or regular databases, generates a larger variety of peptides in comparison to specific enzymatic cleavage. This results in a larger pool of potential peptide sequences that must be considered during the database search. This larger search space has three major consequences. Firstly, the increased database size increases the computational complexity. Secondly, the peptide fragments resulting from a nonspecific cleavage may match multiple proteins in the database. This ambiguity makes it challenging to confidently assign a peptide to a specific protein, leading to a higher probability of incorrect identifications. Lastly, the larger search space increases the likelihood of finding peptide-spectrum matches by chance, leading to a higher number of false positive identifications. In addition, the search space in immunopeptidomics is frequently extended beyond the canonical human proteome, by incorporating somatic mutations (Tretter et al., 2023), pathogen genomes (Leddy et al., 2021), novel or unannotated open reading frames (Guilloy et al., 2023; Ouspenskaia et al., 2022), and PTMs (Kacen et al., 2022). In addition, especially for MHC-I, the relatively short length of the peptides (8–12 amino acids) can make the identification even more difficult. The analysis of short peptides is cumbersome due to several hindrances. Short peptides have a wide range of polarity, and very hydrophilic peptides may be lost during the initial chromatographic separation. When looking towards the bioinformatics perspective of short peptides, different issues arise. Typical search engines are limited towards detecting peptides of a certain length. Shorter peptides usually acquire lower search engine scores, and different short peptides can be isobaric. Lastly, short peptides have a harder time becoming at least doubly charged. This reduces the information obtained by MS/MS experiments as noise is increased and fragmentation of the backbone is reduced (Piovesana et al., 2019).

## 5.1 | Bioinformatic challenges in the identification of PTMs

In the identification of modified peptides, researchers typically have two options. They can either specify a predetermined list of potentially occurring PTMs in a closed search or perform an open modification search. During a closed search, a narrow precursor mass window is used to select candidate peptides to compare the experimental spectrum against. Only the specified PTMs will be considered, leaving any unspecified PTMs undetected. As each PTM that is added will increase the search space drastically, usually only a handful of PTMs are considered during a closed search. However, an open modification search approach allows unbiased detection of any PTM. During open modification searching, a wide precursor mass window is used, allowing the matching of a spectrum from a modified peptide to its unmodified counterpart in the database. Consequently, a mass shift, wherein the identified peptide-spectrum match (PSM) deviates in mass from the precursor ion, can be observed. These mass shifts are often attributable only to a single known PTM or, in some cases, a combination of PTMs, neutral losses, or chemical modifications. While an open search enables the handling of numerous unidentified PTMs, in immunopeptidomics, it is encumbered by two main challenges. First, the search space expands significantly, resulting in longer search times. Second, the search outcomes are prone to errors, often stemming from the misidentification of peptide sequences and the mis-localization of PTM sites, necessitating additional postprocessing steps or manual validation of the results compared to a conventional closed search approach. Recently, new tools have been emerging to resolve some of the main challenges in open modification searching such as reducing the search time, and the identification of localization of PTMs (An et al., 2019; Avtonomov et al., 2019; Han et al., 2011).

Not only is the presence of many PTMs challenging, but the nature of these PTMs can also make their detection difficult. Modified peptides are usually less abundant, making identification more challenging than their unmodified counterpart, especially in complex samples, where acquiring high-quality spectra is difficult. While easily overlooked, each PTM may also impose different challenges. Take some of the previously mentioned PTMs as an example. Deamidation and Citrullination cause a mass shift of 0.9840 Da, roughly corresponding to the mass of a single hydrogen atom (1.0078 Da). The minor mass shift of approximately 1 Da between the deamidated peptide and unmodified counterpart causes the deamidated peptide to overlap with the isotopic pattern of the unmodified peptide. Without employing proper separation techniques, distinguishing between the deamidated and unmodified peptides can be highly challenging (Badgett et al., 2017). Phosphorylation is a labile PTM, especially in the case of phosphothreonine and phosphoserine. These modified amino acids are prone to spontaneous dephosphorylation, causing a mass shift of −80 Da, making it troublesome to localize the modification. Glycosylation is also a challenging PTM to analyze because of the variability in the attached polysaccharides. Adding to this challenge is that glycosylation is often a set of modifications in a peptide with a high heterogeneity, resulting in not a single mass shift being associated with glycosylation, but rather a broad distribution. Also, the glycans are high molecular weight modifications, potentially shifting the molecular weight of the peptide out of the detection region of the MS. Lastly, conventional fragmentation of glycosylated peptides results into fragmentation of the glycan, yielding a spectrum containing information about the nature of the modification instead of the peptide sequence or the location of the modification. This can be circumvented through fragmentation by electron-capture dissociation. A popular approach to detect glycosylation in peptides is by removing the polysaccharides from the peptide and analyzing both separately (Mann & Jensen, 2003). Recently, a deep learning framework, called DeepGlyco, was developed. DeepGlyco is capable of predicting fragment spectra of glycopeptides. Leveraging these predictions makes it possible to create spectral libraries for processing steps (Yang & Fang, 2024).

A final bioinformatic challenge regarding PTMs is the localization of the modification within the peptide. Localization is a crucial aspect of MS-based proteomics in general to properly understand the protein function and regulation. In immunopeptidomics, this is of even greater importance given the relevance of immunopeptides and their localization of the modification in terms of T-cell recognition and MHC binding affinity. While many search engines can identify modified peptides, only a few algorithms have incorporated score statistics for modification site localization. In the following sections, we will mention spectrum search engines suited to address these issues. Additionally, there are also various postprocessors available for PTM localization such as A-score (Beausoleil et al., 2006), phosphoRS (Taus et al., 2011), and DeepFLR (Zong et al., 2023). Notably, phosphoRS and DeepFLR are specifically designed for localization of phosphorylated residues.

## 5.2 | Overview of the data analysis workflow

Figure 4 presents a general workflow for analyzing mass spectrometry data. In this discussion, we will provide detailed insights into the data analysis aspects, specifically focusing on nonspecific enzymatic digestion and open modifications searching. We will begin by exploring various peptide identification methods and then examine the potential benefits of preprocessing steps, such as spectral clustering and proteogenomics. Finally, we will delve into rescoring as a post-processing technique.

### 5.2.1 | Peptide identification

There are a multitude of strategies available to identify peptides, each having their advantages and disadvantages. They can be classified into three large categories: Sequence database search algorithms, spectral library search algorithms and de novo search algorithms. The choice of identification algorithm can influence the identification of immunopeptides. A benchmarking study by Parker et al. benchmarked multiple search engines, finding varying sensitivities and reproducibility between the different search engines (Parker, Tailor, et al., 2021). Another benchmark study by Shahbazy et al. found an increased coverage using DIA instead of DDA and varying coverage and reproducibility between various spectral library-based DIA tools. They concluded that a consensus approach yielded the highest confidence in identifications (Shahbazy et al., 2023). Different algorithms are discussed more in depth with their advantages regarding open modification searching and identifying PTMs.

*Sequence database search algorithms*

Database searching remains the most popular peptide identification method. Database search algorithms compare the experimental spectrum against a theoretical

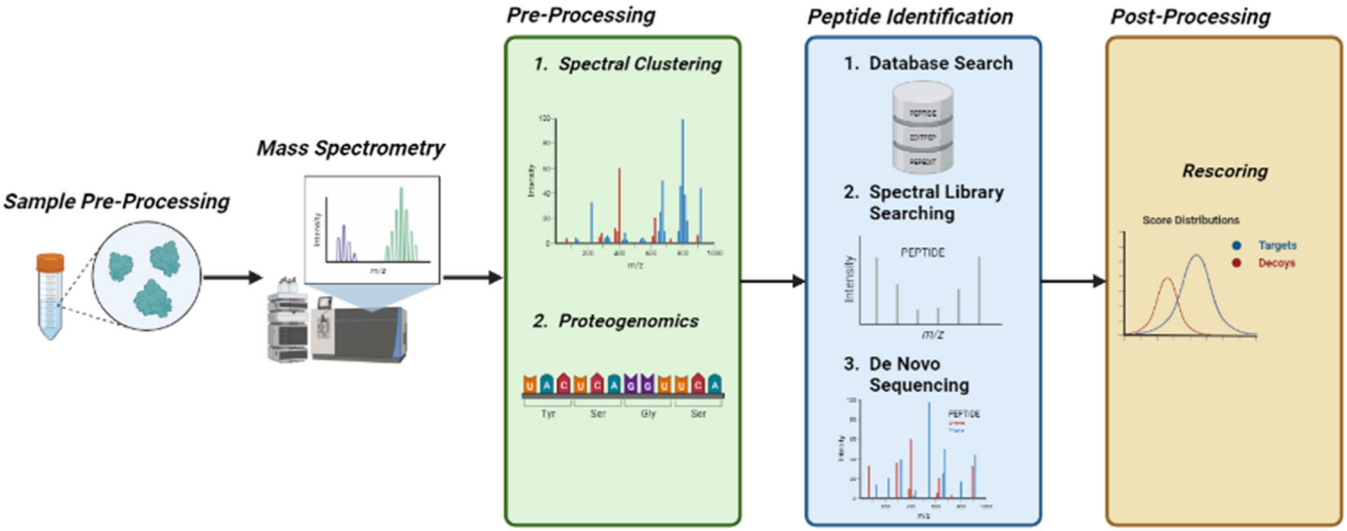


**FIGURE 4** A general description of an immunopeptidomics workflow. The preprocessing, peptide identification, and postprocessing steps show potential methods of analyzing the highly complex nature of immunopeptidomics data. While some of these aspects have not yet been commonly applied, their potential will be discussed with respect to immunopeptidomics and open modification searching. Created with BioRender.com. [Color figure can be viewed at wileyonlinelibrary.com]

fragmentation spectrum generated from candidate peptides in a database. When performing a database search, there is the option for traditional database search engines, who match the masses obtained from the spectrum to the theoretical masses of the database or matching the observed fragment ions to the theoretical fragment ions of the database. Another method is leveraging sequence tag-based approaches, which use short sequences of amino-acids, called tags, to identify a spectrum. There are also hybrid approaches available who combine both traditional and sequence tag-based strategies for peptide identification. The appropriate database can be chosen based on the topic of the research question at hand. Several popular sequence database search algorithms are shown in Table 1.

Regarding user-friendliness and accessibility, neither PROMISE nor PRiSM offers a web-based version or GUI for their software. Even though both algorithms are very well suited for the immunopeptidomics setting concerning open modification searching and nonenzymatic cleavage enzymes, accessibility is essential for most end-users. Both algorithms are easily usable when the user has little coding experience and, therefore, is still noteworthy. PROMISE was built upon the MSFragger database search engine and has shown promising results in identifying modified HLA-1 peptides. Since the developers of MSFragger developed PROMISE, it may still be incorporated in the Fragpipe GUI, making it very suitable for immunopeptidomics research. On the other hand, MASCOT, Spectrum Mill, and Ionbot offer cloud computing possibilities through their web applications, which is advantageous for researchers who lack the necessary computational power for large, exhaustive database searches.

MASCOT, SEQUEST, X!Tandem, and Ionbot do not allow for nonspecific enzymatic digestion of the theoretical database, which will interfere with identifying nontryptic immunopeptides. Regarding the open modification searching, MASCOT, SEQUEST, X!Tandem, MS-GF+, Spectrum Mill, and Andromeda only allow for a limited set of possible modifications, leading to possible missed identifications of unknown modified peptides. Hence, taking both the absence of a specific cleavage enzyme and the large number of modifications in immunopeptides into account, usage of these algorithms will be disadvantageous, leading to a reduced number of identifications. Of course, this still depends on the research question at hand. MetaMorpheus, MSFragger, PROMISE, SAGE, Open-pFind, and PRiSM will be more suitable for the challenges in immunopeptidomics research.

There are still other influential parameters when considering the ideal database search tool. Novel research has shown that TimsTOF data has an edge over Orbitrap data for immunopeptidomics experiments due to increased sensitivity and speed (Hoenisch Gravel et al., 2023). Hence, compatibility with TimsTOF data is advantageous. To our knowledge, Spectrum Mill, Andromeda, and MSFragger are the only algorithms geared towards TimsTOF data. When the research is focused on the localization of PTMs, MSFragger with PTM-Shepherd (Han et al., 2011), MetaMorpheus with

**TABLE 1** A summary of popular database search algorithms.

| Algorithm | Principle | Nonspecific enzymatic digestion | Open modification search possible | References | Accessibility (web application/GUI) |
|---|---|---|---|---|---|
| MASCOT | Probabilistic scoring function | No | No | Perkins et al. (1999) | Web application, GUI |
| SEQUEST | Cross-correlation scoring function | No | No | Eng et al. (1994) | GUI |
| X!Tandem | Cross-correlation scoring function | No | No | Craig and Beavis (2003) | GUI |
| COMET | Cross-correlation scoring function | Yes | No | Eng et al. (2013) | GUI |
| MS-GF+ | Probabilistic scoring function | Yes | No | Kim and Pevzner (2014) | GUI |
| Spectrum Mill | Cross-correlation scoring function | Yes | No | *Spectrum Mill MS Proteomics Software* (n.d.) | Web application |
| MetaMorpheus | Multinotch search | Yes | Large search space possible | Solntsev et al. (2018) | GUI |
| Andromeda | Probabilistic scoring function | Yes | No | Cox et al. (2011) | GUI |
| MSFragger | Cross-correlation scoring function | Yes | Large search space possible | Kong et al. (2017) | GUI |
| PROMISE | Cross-correlation scoring function | Yes | Large search space possible | Kacen et al. (2022) | Neither |
| SAGE | Cross-correlation scoring function | Yes | Large search space possible | Lazear (2023) | GUI |
| Open-pFind | Sequence tag<br>Dot-product | Yes | Large search space possible | Chi et al. (2018) | GUI |
| Ionbot | Decision trees | No | Yes | Sven Degroeve et al. (2022) | Web application |
| PRiSM | Pseudo-ions<br>Mass-shift paradigm | Yes | Yes | Van Houtven et al. (2020) | Neither |

*Note*: Their core principle is mentioned and their capabilities towards non-specific enzymatic digestion and open modification searching is shown. The table contains information on the algorithm and their capabilities with respect to nonenzymatic digestion searches and open modification searching, either yes or no. For the open modification searches, it is also mentioned if the algorithm is uncapable of open modification searches but does allow for large search spaces, containing a large amount of PTMs. The last column shows how accessible the algorithm is for users with respect to web-applications or a GUI. If neither a web-application or GUI is available, the algorithm is available through coding.

G-PTM-D (Li et al., 2017), Andromeda, Mascot, and Spectrum Mill, are all valid options with separate PTM localization scores. Lastly, given the increased number of identifications in immunopeptidomics through leveraging DIA approaches, Fragpipe and MSFragger offer the availability to identify spectra from DIA using DIA-Umpire (Tsou et al., 2015).

As a concluding remark regarding database search engines, the Fragpipe GUI currently offers the most complete framework for identifying immunopeptides and is under continuous development to tune in to the most recent changes within the field. Other algorithms are also very suitable and have already been applied within the field, and therefore, mention worthy.

### *Spectral library search algorithms*

Spectral library search engines identify experimental spectra by comparing these spectra to spectral libraries. These spectral libraries are often constructed from previously identified experimental spectra. Spectral libraries can also be generated using spectral library searching algorithms or stand-alone tools such as $MS^2PIP$ (Degroeve & Martens, 2013) or Prosit (Gessulat et al., 2019). A recent publication by Cox J. discusses the prediction of spectral libraries with machine and deep learning applications (Cox, 2023). The advantage of spectral libraries is that they also contain information about the intensities of fragment ions and contain data from noncanonical peptide fragment ions. Spectral libraries are available from the National Institute of Standards and Technology (NIST), PRIDE repository (Perez-Riverol et al., 2022), MassIVE-KB (Wang et al., 2018), and SysteMHC (Shao et al., 2018). For a long time, spectral library searching was the only way of identifying spectra acquired through DIA. Nowadays, DIA-Umpire allows for the analysis of DIA data with tools specifically designed for DDA data (Tsou et al., 2015). Spectral library searching in proteomics is reviewed by Griss (2016) and Shao and Lam (2017). Table 2 summarizes spectral library search algorithms usable in immunopeptidomics.

Spectral library search algorithms have yet to find their application in the field of immunopeptidomics. They are less popular in comparison to database search algorithms. Their usefulness has been observed in proteomics, especially for DIA experiments, and it may be worthwhile to investigate their usefulness in immunopeptidomics. SpectraST and Scribe provide a simple GUI to perform rapid spectral library searches. While ANN-SoLo does not have a GUI available, it does provide a step-by-step guide to use ANN-SoLo in Python. The flexibility of spectral library search engines to perform open modification searches and nonspecific enzymatic

**TABLE 2** Three popular and well-maintained spectral library search algorithms, allowing for both nonspecific enzymatic digestion and open modification searching.

| Algorithm | Principle | Nonspecific enzymatic digestion | Open modification search possible | References | Accessibility (web application/GUI) |
|---|---|---|---|---|---|
| SpectraST | Probabilistic model and dot-product | Yes | Yes | Cox (2023) | GUI |
| ANN-SoLo | Shifted dot-product | Yes | Yes | Arab et al. (2023) | Neither |
| Scribe | Mean Squared Error | Yes | No | Searle et al. (2023) | GUI |

digestion is suitable for the immunopeptidomics scene. Scribe is limited towards open modification searching concerning the dependency of the presence of the modifications in the spectral library.

Each of the aforementioned algorithms is novel and actively maintained. A proper benchmark study employing spectral library search algorithms with immunopeptidomics data would be required to advise on the usage of these algorithms.

### *De novo search algorithms*

Given that numerous immunopeptides stem from low-abundance RNA expression, the necessity for transcriptomic-independent data in neoantigen discovery becomes evident. A potential solution to identify non-canonical immunopeptides without expanding the database with proteogenomics is De novo sequencing. De novo sequencing, being a method independent of databases, holds relevance for immunopeptidomics. Unlike database searching, it directly extracts sequence tags from spectra, enabling the identification of non-tryptic and noncanonical peptides. MARS, a recently introduced method for filtering de novo identified immunopeptides, aids in preselecting those likely originating from non-canonical sources (Liao et al., 2024). De novo sequencing does suffer from several pitfalls, leading to a reduced integration of de novo methods for peptide antigen discovery. De novo sequencing algorithms require high-quality spectra to identify peptides correctly. In immunopeptidomics, this is challenging due to the nature of immunopeptides. Because immunopeptides aren't digested by a specific enzyme, the peptide fragmentation by MS/MS is never fully "complete" and frequently lacks a consecutive ion series for the nontryptic peptides. Gaps in the fragment ion series allow alternative sequence interpretation, resulting in identical probability scores for isobaric amino-acid combinations and sequence permutations. Additionally, due to the lower abundance of immunopeptides, the distinction between signal and noise is more challenging. Lastly, de novo sequencing is typically more computationally burdening than database and spectral library search. Note that with the continuous computational advancements, de novo methods have become more viable. This has led to an increasing number of novel algorithms being released over the past few years. Muth et al. published a review of de novo algorithms in proteomics in 2018 (Muth & Renard, 2018). Table 3 shows several algorithms in immunopeptidomics for open modification searching.

The proliferation of numerous de novo algorithms makes selecting a sole optimal algorithm impossible. Each algorithm demonstrates adequate performance in its respective settings. Hence, we can only consider several pivotal arguments when selecting a de novo algorithm to analyse immunopeptidomics data. Pertinent to user-friendliness, PEAKS, Novor and PepNet offer a web application for analysing immunopeptidomics data, solving the computational burden of de novo methods. Notably, pNovo offers a very intuitive GUI. Other algorithms require little coding knowledge and Python to

**TABLE 3** Comparative overview of de novo algorithms for immunopeptidomics.

| Algorithm | Principle | Open modification search possible | References | Accessibility (web application/GUI) |
|---|---|---|---|---|
| PEAKS | Reward/penalty scoring | Yes | Ma et al. (2003) | Web application |
| Novor | Fragmentation and residue scoring | No | Ma (2015) | Web application |
| DeepNovo | Neural network | No | Tran et al. (2017) | Neither |
| pNovo 3 | Support Vector Machine | Yes | Yang et al. (2019) | GUI |
| TagGraph | Hybrid | Yes | Devabhaktuni et al. (2019) | Neither |
| SMSNet | Neural network | No | Karunratanakul et al. (2019) | Neither |
| Casanovo | Neural networks | Yes | Yilmaz et al. (2022) | Neither |
| InstaNovo | Neural network | No | Kevin Eloff et al. (2024) | Neither |
| GraphNovo | Graph neural networks | No | Mao et al. (2023) | Neither |
| PepNet | Neural network | No | Liu et al. (2023) | Web application |
| ContraNovo | Neural Network | No | Jin et al. (2023) | Neither |

implement them. For the methods based on a neural network, decent hardware with sufficient memory is needed to run the models, as the developers have mentioned memory-related challenges.

Despite the performance adequacy demonstrated on immunopeptidomics data validation, DeepNovo, SMSNet, and GraphNovo are not suited for true open searches with open modification searches, making them unfit for the challenges within immunopeptidomics. Conversely, Novor, InstaNovo, PepNet, and ContraNovo permit nonenzymatic digestion but impose restrictions on the number of allowed or possible modifications within their models. Lastly, PEAKS, pNovo, TagGraph, and Casanovo allow for true open searches, positioning them potentially as optimal models for identifying immunopeptides.

Presently, several de novo algorithms combine their sequencing results with database searches to enhance accuracy and sensitivity, technically making them hybrid methods, such as PEAKS DB (Zhang et al., 2012), Novor, DeepNovo, TagGraph, and SMSNet. Considerations such as the localization of PTMs may also be an influential factor, a service offered by PEAKS through PEAKS PTM (Han et al., 2011), pNovo, and TagGraph. Moreover, the capability to deal with DIA data is pertinent, achievable within PEAKS, DeepNovo, Casanovo, GraphNovo, and PepNet.

While individual assessments of each algorithm may invariably depict it as the superior performer, this review paper acknowledges the subjectivity inherent in such claims. An independent benchmarking study, such as that conducted by Beslic et al. (2023), remains integral for a fair and impartial evaluation. They compared Novor, pNovo 3, DeepNovo, SMSNet, PointNovo, and Casanovo on three datasets. It is still available through different pipelines. Their findings indicated that Casanovo performed the best.

In summation, each de novo algorithm enumerated herein demonstrates commendable performance, as only validated options are referenced. Our first recommendation is PEAKS, as they offer the most complete service. They continuously develop their framework further to address issues and offer cloud computing. However, the caveat lies in the commercial nature of PEAKS. Alternatives such as Casanovo and TagGraph, tailored to the unique requirements of immunopeptidomics, present valid choices, boasting commendable performance.

### 5.2.2 | Processing steps in MS data analysis

While the identification method is crucial in identifying the peptides, other factors may influence the number of peptides identified and control for valid identifications. Several concepts are discussed concerning immunopeptidomics.

#### *Spectral clustering*

The potential of applying spectral clustering to immunopeptidomics data is discussed as a preprocessing step for immunopeptidomics. Spectral clustering algorithms group spectra based on their similar peak patterns and generate consensus spectra for each cluster. It assumes that all spectra within a cluster belong to the same peptide in the case of immunopeptidomics. A wide variety of spectral clustering algorithms are available such as Bonanza (Falkner et al., 2008), MS-Cluster (Frank et al., 2011), PRIDE Cluster (Griss et al., 2013), Spectra-cluster (Griss et al., 2016), MaRaCluster (The & Käll, 2016), msCRUSH (Wang et al., 2019), Falcon (Bittremieux et al., 2021), and GLEAMS (Bittremieux et al., 2022). The algorithms differ due to distinctions in their preprocessing of the spectra, clustering algorithms, and similarity metrics.

While spectral clustering has yet to find an application in the field of immunopeptidomics, primarily due to typically relatively long computation times, the concept has sparked a lot of interest. In the initial article on the Bonanza algorithm, Falkner et al. detected more PTMs, expected and unexpected, and amino-acid substitutions when using spectral clustering. One year later, in 2009, this was confirmed in a standalone study by Menschaert et al. using the Bonanza algorithm specifically on peptidome data (Menschaert et al., 2009). Spectra-cluster, msCRUSH, and GLEAMS had similar findings in their papers. Luo et al. performed a benchmark study in 2022 using MaRaCluster and spectra-cluster on various public data sets. They evaluated multiple consensus spectrum generation methods to find the most optimal methodology to increase the number of peptide identifications. Among the data used in the research, a phosphorylated data set was also used, showing adequate results in clustering phosphorylated peptides (Luo et al., 2022). With a vision towards identifying more of the immunopeptidome, spectral clustering is an exciting tool for identifying more immunopeptides.

#### *Proteogenomics*

A compelling concept within immunopeptidomics that holds promise for improving the detection of immunopeptides is proteogenomics. Despite the advancements made by MS-based immunopeptidomics in identifying these peptides, it is widely acknowledged that a significant number of immunopeptides still elude recognition. As previously mentioned, spectra originating from unknown immunopeptides, mutations, and PTMs

continue to pose identification challenges when relying on databases. The term "dark proteome" is frequently invoked in the literature to describe these proteomics hurdles. Utilizing proteogenomics emerges as a potentially convenient solution to address and overcome these persistent challenges.

Proteogenomics is a concept from proteomics where one enriches the database used for search algorithms using genomics, transcriptomics, or translatomics (riboseq) data. Coding the nucleotide sequences into amino-acid sequences combined with variant calling can add noncanonical sequences and mutations to the database. This generates a personalized database adjusted to the research question or sample. Using an annotated database will allow for the identification of mutations such as single nucleotide polymorphisms and single amino-acid polymorphisms. Recently, these approaches have expanded to the identification of unannotated open reading frames (ORFs), or out-of-frame translated sequences, which present potential for the development for immunotherapies. Within the MaxQuant GUI, there is a dedicated option to incorporate whole exome sequencing data into the database. MaxQuant utilizes this sequencing information for the purpose of mutation calling. Subsequently, the database enriched with this data can be explored using the previously outlined algorithms. Proteogenomics offers the potential to discern personalized immunopeptides, frequently of interest in the context of tumor-specific immunopeptides. An example of a proteogenomics workflow is explained by Scull et al. (2021). They enriched a database using transcriptomics data followed by a search using PEAKS and PEAKS DB (Ma et al., 2003; Xin et al., 2022). The use of proteogenomics resulted in more identifications, including noncanonical HLA peptides. Similar results were obtained using the MaxQuant workflow (Chong et al., 2020; Kalaora et al., 2016), MASCOT combined with PEAKS (Attig et al., 2019; Ternette et al., 2016), and MaxQuant combined with PEAKS (Qi et al., 2021). For a more in-depth discussion on proteogenomics, we refer to the article of Chong et al. (2022). Chong et al. cover different strategies and articles which apply proteogenomics within immunopeptidomics.

It is crucial to highlight that incorporating proteogenomic data into a database for a target-decoy search strategy significantly expands the search space. This expansion, in turn, leads to an elevated false discovery rate (FDR), rendering the approach impractical. The challenges associated with FDR have been extensively deliberated, and for a more in-depth exploration, we recommend referring to the publication cited (Aggarwal et al., 2022).

However, an effective solution has been demonstrated in reducing the search space. By employing a filtering mechanism that prioritizes immunopeptides, researchers can circumvent the complications arising from the enlarged search space. This strategic reduction not only mitigates the issues related to FDR but also enhances the practicality and efficiency of the methodology (Scull et al., 2021).

### *Rescoring*

Rescoring is a widely adopted practice in immunopeptidomics and proteomics overall. The extensive search spaces resulting from nontryptic immunopeptides, PTMs, and the inclusion of noncanonical sequences pose challenges for traditional target-decoy approaches, leading to adverse effects in accurately estimating the FDR. Rescoring tools aim to circumvent this problem by taking the scores of identification algorithms and rescoring them to improve the identification rate. Some algorithms incorporate additional features such as retention time or peak intensity predictions to enhance the identification rate. Each rescoring tool takes a different approach and requires specific output files of search algorithms. For an extensive review on rescoring within immunopeptidomics, we refer the readers to the review paper of Adams et al. (2023). Table 4 shows several popular rescoring algorithms.

Percolator is the most used rescoring algorithm. It has been applied within immunopeptidomics with the MASCOT, MS-GF+, and COMET database search algorithms, although it can take input from various commonly used algorithms. PeptideProphet is another traditional rescoring algorithm available. Like Percolator, it can perform rescoring for a wide range of database search algorithms. Percolator and PeptideProphet are integrated within the Fragpipe GUI to combine with MSFragger. It must be noted that when performing an open search for immunopeptides within Fragpipe, Percolator is used as the default rescoring algorithm. Prosit, MS$^2$Rescore, and MSBooster are novel tools. Prosit, relies on Percolator to rescore and can only directly work with the output from MaxQuant. However, with the release of Oktoberfest, the tool supports the rescoring of results from MSAmanda, Mascot, MSFRagger and Andromeda. Wilhelm et al. recently evaluated the performance of Prosit on MS-based immunopeptidomics. Prosit increased the identification of immunopeptides by sevenfold as compared to standard MaxQuant searches (Wilhelm et al., 2021). MS$^2$Rescore is a rescoring tool combining output from several sources to enhance peptide identifications. It uses peak intensity predictions from MS$^2$PIP (Degroeve & Martens, 2013), retention time

**TABLE 4** An overview of various rescoring algorithms and their principles.

| Algorithm | Principle | References | References in immunopeptidomics |
|---|---|---|---|
| Percolator | Support vector machine | Käll et al. (2007) | Bichmann et al. (2019); Mishto et al. (2022) |
| PeptideProphet | Discriminant function and Bayes Theorem | Keller et al. (2002) | Kacen et al. (2022) |
| Prosit/Oktoberfest | Neural Network | Gessulat et al. (2019); Picciani et al. (2023) | Wilhelm et al. (2021) |
| $MS^2$Rescore | Decision trees | Declercq et al. (2022) | Declercq et al. (2022) |
| MSBooster | Neural Network | Yang et al. (2023) | Yang et al. (2023) |
| AlphaPeptDeep | Neural Network | Zeng et al. (2022) | Zeng et al. (2022) |
| INFERYS | Neural Network | Zolg et al. (2021) | Zolg et al. (2021) |

*Note*: Each of these algorithms have been used within the immunopeptidomics setting, with references included.

prediction from DeepLC (Bouwmeester et al., 2021), and rescoring from Percolator (Käll et al., 2007) to strengthen identifications. The developers specifically evaluated the tool on immunopeptidomics data. MSBooster is a novel tool taking a similar approach to $MS^2$Rescore in the sense that it combines information and predictions from various sources. MSBooster can incorporate information from the LC retention time and ion mobility when using a timsTOF instrument, for example, and peak intensities combined with Percolator rescoring. This information is processed using neural networks to perform a final rescoring. MSBooster is integrated within the Fragpipe GUI to work properly with MSFragger and Percolator. In the release paper of MSBooster, the algorithm was specifically evaluated on immunopeptidomics data. When performing open searches for immunopeptides within Fragpipe, MSBooster is automatically enabled. Lastly, AlphaPeptDeep is a novel Python framework to construct deep learning applications within proteomics easily. It also contains pre-trained models to predict retention times, MS2 intensities, and collisional cross sections for peptides with and without PTMs. AlphaPeptDeep can use all this information to perform rescoring using Percolator and was shown to perform equally well as existing tools on immunopeptides. INFERYS is a tool for intensity-based rescoring of Sequest results. Using this tool, they claim to increase the number of identified peptides by 50% (Zolg et al., 2021).

Researchers should be aware of the challenges posed by PTMs for traditional search engines in controlling the FDR. As mentioned before, searching for PTMs leads to an expansion of the search space, resulting in computational and statistical challenges. Conventional search engines struggle with handling the sheer number of possible candidates and differentiating between PTMs with similar mass shifts. Additionally, by expanding the search space, the probability of acquiring random matches between a spectrum and theoretical spectra of (un) modified spectra increases. This poses challenges for statistical frameworks to properly control the FDR. In the article of PROMISE, a subgroup FDR was used. The identifications were split into three groups: unmodified, standard search modification types (N-acetylation and methionine oxidation) and the other modification types, making it possible to control the FDR more stringently (Kacen et al., 2022).

A final point for discussion concerning rescoring is within the specific setting of open modification searching. Traditional methods, leveraging target-decoy approaches, typically require well-defined targets to distinguish between true and false positives properly. A second issue is that target-decoy approaches often have one major assumption: homogeneity of the data. To elaborate on the issue, the target-decoy method consists of targets, real peptides, and decoys, artificial data designed to mimic the characteristics of targets. With homogenous data, it is easier to define these categories. In open searches, there is no predefined target in mind. A wide precursor mass tolerance is set to allow for various potential PTMs. The lack of a predefined target makes designing good decoys difficult. Hence, the homogeneity assumption of target-decoy approaches typically fails here as it is harder to distinguish real targets from false positives, leading to an increased FDR. Rescoring does not solve this issue, as heterogeneity is also problematic for rescoring algorithms. While various methods have been explored, for example, the group-walk method of Freestone et al., much improvement is still possible as future research (Freestone et al., 2022). Additionally, it's important to recognize that the majority of identified

PTMs in open modification searching may not receive sufficient support from existing prediction algorithms. To address this challenge, new approaches like transfer learning (Wilburn et al., 2023) and innovative algorithms capable of predicting properties for previously unexplored PTMs have been introduced (Bouwmeester et al., 2021; Zeng et al., 2022).

## 6 | CONCLUSION

Immunopeptidomics stands at the forefront of rapid development, offering immense potential for therapeutic applications and addressing fundamental biological inquiries. Given their unparalleled potential, the primary emphasis within this dynamic research field is the pursuit of neoantigens. Despite this focus, numerous challenges still warrant careful consideration. A pivotal avenue for further exploration lies in expanding the scope of research to encompass immunopeptides-bearing PTMs. Several PTMs have successfully been identified on immunopeptides, with a subset already advancing into clinical phases as targets in T-cell therapy. Research efforts have demonstrated the remarkable capability of T-cells to recognize PTMs on immunopeptides, sometimes serving as the impetus for recognition. The expansion toward PTMs holds promise for uncovering new targets and forging novel avenues in the development of treatments for cancer and other diseases. However, the identification and localization of PTMs pose formidable experimental and computational challenges, a task already demanding in the realm of proteomics and further compounded in the intricate landscape of immunopeptidomics. While substantial progress has been achieved in the field of immunopeptidomics, the journey is far from over. The evolving landscape presents exciting prospects for future research, and it is with anticipation that we await the unfolding possibilities that lie ahead in this dynamic realm.

## AUTHOR CONTRIBUTIONS

**Daniel Flender**: Writing—original draft; writing—review and editing. **Frédérique Vilenne**: Writing—original draft; writing—review and editing. **Charlotte Adams**: Writing—review and editing. **Kurt Boonen**: Supervision; writing—review and editing. **Dirk Valkenborg**: Supervision; writing—review and editing. **Geert Baggerman**: Supervision; writing—review and editing.

## ACKNOWLEDGMENTS

This research was funded by the Research Foundation – Flanders (FWO) under the "Beyond the Genome: Ethical Aspects of Large Cohort Studies" project (Case number G070722N) and the Flemish Institute for Technological Research (VITO).

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## ORCID

*Daniel Flender* https://orcid.org/0009-0008-5154-3060
*Frédérique Vilenne* https://orcid.org/0000-0002-6066-9035
*Charlotte Adams* http://orcid.org/0000-0001-5652-2324
*Kurt Boonen* http://orcid.org/0000-0002-6252-7065
*Dirk Valkenborg* http://orcid.org/0000-0002-1877-3496
*Geert Baggerman* http://orcid.org/0000-0002-0661-931X

## REFERENCES

Abbott, M., & Ustoyev, Y. (2019). Cancer and the immune system: the history and background of immunotherapy. *Seminars in Oncology Nursing*, *35*(5), 150923. https://doi.org/10.1016/j.soncn.2019.08.002

Abelin, J. G., Trantham, P. D., Penny, S. A., Patterson, A. M., Ward, S. T., Hildebrand, W. H., Cobbold, M., Bai, D. L., Shabanowitz, J., & Hunt, D. F. (2015). Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry. *Nature Protocols*, *10*(9), 1308–1318. https://doi.org/10.1038/nprot.2015.086

Adams, C., Laukens, K., Bittremieux, W., & Boonen, K. (2023). Machine learning-based peptide-spectrum match rescoring opens up the immunopeptidome. *PROTEOMICS*, *24*(8), e2300336. https://doi.org/10.1002/pmic.202300336

Admon, A., & Bassani-Sternberg, M. (2011). The Human Immunopeptidome Project, a suggestion for yet another postgenome next big thing. *Molecular & Cellular Proteomics*, *10*(10), O111.011833. https://doi.org/10.1074/mcp.O111.011833

Aggarwal, S., Raj, A., Kumar, D., Dash, D., & Yadav, A. K. (2022). False discovery rate: the Achilles' heel of proteogenomics. *Briefings in Bioinformatics*, *23*(5), bbac163. https://doi.org/10.1093/bib/bbac163

Alspach, E., Lussier, D. M., Miceli, A. P., Kizhvatov, I., DuPage, M., Luoma, A. M., Meng, W., Lichti, C. F., Esaulova, E., Vomund, A. N., Runci, D., Ward, J. P., Gubin, M. M., Medrano, R. F. V., Arthur, C. D., White, J. M., Sheehan, K. C. F., Chen, A., Wucherpfennig, K. W., … Schreiber, R. D. (2019). MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature*, *574*(7780), 696–701. https://doi.org/10.1038/s41586-019-1671-8

An, Z., Zhai, L., Ying, W., Qian, X., Gong, F., Tan, M., & Fu, Y. (2019). PTMiner: localization and quality control of protein modifications detected in an open search and its application to comprehensive post-translational modification characterization in human proteome*. *Molecular & Cellular Proteomics*, *18*(2), 391–405. https://doi.org/10.1074/mcp.RA118.000812

Andersen, M. H., Bonfill, J. E., Neisig, A., Arsequell, G., Søndergaard, I., Neefjes, J., Zeuthen, J., Elliott, T., & Haurum, J. S. (1999). Phosphorylated peptides can be

transported by TAP molecules, presented by Class I MHC molecules, and recognized by phosphopeptide-specific CTL1. *The Journal of Immunology*, *163*(7), 3812–3818. https://doi.org/10.4049/jimmunol.163.7.3812

Annan, R. S., Huddleston, M. J., Verma, R., Deshaies, R. J., & Carr, S. A. (2000). A multidimensional electrospray MS-Based approach to phosphopeptide mapping. *Analytical Chemistry*, *73*(3), 393–404. https://doi.org/10.1021/ac001130t

Apavaloaei, A., Hardy, M.-P., Thibault, P., & Perreault, C. (2020). The origin and immune recognition of tumor-specific antigens. *Cancers*, *12*(9), 2607. https://doi.org/10.3390/cancers12092607

Apostolopoulos, V., Yuriev, E., Ramsland, P. A., Halton, J., Osinski, C., Li, W., Plebanski, M., Paulsen, H., & McKenzie, I. F. C. (2003). A glycopeptide in complex with MHC class I uses the GalNAc residue as an anchor. *Proceedings of the National Academy of Sciences*, *100*(25), 15029–15034. https://doi.org/10.1073/pnas.2432220100

Arab, I., Fondrie, W. E., Laukens, K., & Bittremieux, W. (2023). Semisupervised machine learning for sensitive open modification spectral library searching. *Journal of Proteome Research*, *22*(2), 585–593. https://doi.org/10.1021/acs.jproteome.2c00616

Ardito, F., Giuliani, M., Perrone, D., Troiano, G., & Lo Muzio, L. (2017). The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *International Journal of Molecular Medicine*, *40*(2), 271–280. https://doi.org/10.3892/ijmm.2017.3036

Arnaud, M., Bobisse, S., Chiffelle, J., & Harari, A. (2021). The promise of personalized TCR-based cellular immunotherapy for cancer patients. *Frontiers in Immunology*, *12*, 701636. https://doi.org/10.3389/fimmu.2021.701636

Attig, J., Young, G. R., Hosie, L., Perkins, D., Encheva-Yokoya, V., Stoye, J. P., Snijders, A. P., Ternette, N., & Kassiotis, G. (2019). LTR retroelement expansion of the human cancer transcriptome and immunopeptidome revealed by de novo transcript assembly. *Genome Research*, *29*(10), 1578–1590. https://doi.org/10.1101/gr.248922.119

Avtonomov, D. M., Kong, A., & Nesvizhskii, A. I. (2019). DeltaMass: automated detection and visualization of mass shifts in proteomic open-search results. *Journal of Proteome Research*, *18*(2), 715–720. https://doi.org/10.1021/acs.jproteome.8b00728

Bäcklund, J., Carlsen, S., Höger, T., Holm, B., Fugger, L., Kihlberg, J., Burkhardt, H., & Holmdahl, R. (2002). Predominant selection of T cells specific for the glycosylated collagen type II epitope (263-270) in humanized transgenic mice and in rheumatoid arthritis. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(15), 9960–9965. https://doi.org/10.1073/pnas.132254199

Badgett, M. J., Boyes, B., & Orlando, R. (2017). The separation and quantitation of peptides with and without oxidation of methionine and deamidation of asparagine using hydrophilic interaction liquid chromatography with mass spectrometry (HILIC-MS). *Journal of the American Society for Mass Spectrometry*, *28*(5), 818–826. https://doi.org/10.1007/s13361-016-1565-z

Barbier, A. J., Jiang, A. Y., Zhang, P., Wooster, R., & Anderson, D. G. (2022). The clinical progress of mRNA vaccines and immunotherapies. *Nature Biotechnology*, *40*(6), 840–854. https://doi.org/10.1038/s41587-022-01294-2

Bassani-Sternberg, M., Barnea, E., Beer, I., Avivi, I., Katz, T., & Admon, A. (2010). Soluble plasma HLA peptidome as a potential source for cancer biomarkers. *Proceedings of the National Academy of Sciences*, *107*(44), 18769–18776. https://doi.org/10.1073/pnas.1008501107

Beausoleil, S. A., Villén, J., Gerber, S. A., Rush, J., & Gygi, S. P. (2006). A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nature Biotechnology*, *24*(10), 1285–1292. https://doi.org/10.1038/nbt1240

Beltrami, A., Rossmann, M., Fiorillo, M. T., Paladini, F., Sorrentino, R., Saenger, W., Kumar, P., Ziegler, A., & Uchanska-Ziegler, B. (2008). Citrullination-dependent differential presentation of a self-peptide by HLA-B27 subtypes. *The Journal of Biological Chemistry*, *283*(40), 27189–27199. https://doi.org/10.1074/jbc.M802818200

Bernhardt, M., Cruz-Garcia, Y., Rech, A., Meierjohann, S., Erhard, F., Schilling, B., & Schlosser, A. (2022). Extending the mass spectrometry-detectable landscape of MHC peptides by use of restricted access material. *Analytical Chemistry*, *94*(41), 14214–14222. https://doi.org/10.1021/acs.analchem.2c02198

Beslic, D., Tscheuschner, G., Renard, B. Y., Weller, M. G., & Muth, T. (2023). Comprehensive evaluation of peptide de novo sequencing tools for monoclonal antibody assembly. *Briefings in Bioinformatics*, *24*(1), bbac542. https://doi.org/10.1093/bib/bbac542

Bichmann, L., Nelde, A., Ghosh, M., Heumos, L., Mohr, C., Peltzer, A., Kuchenbecker, L., Sachsenberg, T., Walz, J. S., Stevanović, S., Rammensee, H.-G., & Kohlbacher, O. (2019). MHCquant: automated and reproducible data analysis for immunopeptidomics. *Journal of Proteome Research*, *18*(11), 3876–3884. https://doi.org/10.1021/acs.jproteome.9b00313

Bittremieux, W., Laukens, K., Noble, W. S., & Dorrestein, P. C. (2021). Large-scale tandem mass spectrum clustering using fast nearest neighbor searching. *Rapid Communications in Mass Spectrometry*, e9153. https://doi.org/10.1002/rcm.9153

Bittremieux, W., May, D. H., Bilmes, J., & Noble, W. S. (2022). A learned embedding for efficient joint analysis of millions of mass spectra. *Nature Methods*, *19*, 6. https://doi.org/10.1038/s41592-022-01496-1

Blees, A., Januliene, D., Hofmann, T., Koller, N., Schmidt, C., Trowitzsch, S., Moeller, A., & Tampé, R. (2017). Structure of the human MHC-I peptide-loading complex. *Nature*, *551*, 7681. https://doi.org/10.1038/nature24627

Blum, J. S., Wearsch, P. A., & Cresswell, P. (2013). Pathways of antigen processing. *Annual Review of Immunology*, *31*(1), 443–473. https://doi.org/10.1146/annurev-immunol-032712-095910

Bouwmeester, R., Gabriels, R., Hulstaert, N., Martens, L., & Degroeve, S. (2021). DeepLC can predict retention times for peptides that carry as-yet unseen modifications. *Nature Methods*, *18*, 11. https://doi.org/10.1038/s41592-021-01301-5

Brodsky, F. M., & Parham, P. (1982). Monomorphic anti-HLA-A,B,C monoclonal antibodies detecting molecular subunits and combinatorial determinants. *Journal of Immunology (Baltimore, Md.: 1950)*, *128*(1), 129–135.

Capietto, A.-H., Hoshyar, R., & Delamarre, L. (2022). Sources of cancer neoantigens beyond single-nucleotide variants. *International Journal of Molecular Sciences*, *23*,17. https://doi.org/10.3390/ijms231710131

Caron, E., Aebersold, R., Banaei-Esfahani, A., Chong, C., & Bassani-Sternberg, M. (2017). A case for a human immuno-peptidome project consortium. *Immunity*, *47*(2), 203–208. https://doi.org/10.1016/j.immuni.2017.07.010

Caron, E., Kowalewski, D. J., Chiek Koh, C., Sturm, T., Schuster, H., & Aebersold, R. (2015). Analysis of major histocompatibility complex (MHC) immunopeptidomes using mass spectrometry*. *Molecular & Cellular Proteomics*, *14*(12), 3105–3117. https://doi.org/10.1074/mcp.O115.052431

Caron, E., Espona, L., Kowalewski, D. J., Schuster, H., Ternette, N., Alpízar, A., Schittenhelm, R. B., Ramarathinam, S. H., Lindestam Arlehamn, C. S., Chiek Koh, C., Gillet, L. C., Rabsteyn, A., Navarro, P., Kim, S., Lam, H., Sturm, T., Marcilla, M., Sette, A., Campbell, D. S., … Aebersold, R. (2015). An open-source computational and data resource to analyze digital maps of immunopeptidomes. *eLife*, *4*, e07661. https://doi.org/10.7554/eLife.07661

Chakraborty, C., Majumder, A., Bhattacharya, M., Chatterjee, S., & Lee, S.-S. (2024). The landscape of neoantigens and its clinical applications: from immunobiology to cancer vaccines. *Current Research in Biotechnology*, *7*, 100177. https://doi.org/10.1016/j.crbiot.2024.100177

Chapman, J. D., Goodlett, D. R., & Masselon, C. D. (2014). Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. *Mass Spectrometry Reviews*, *33*(6), 452–470. https://doi.org/10.1002/mas.21400

Chen, R., Fauteux, F., Foote, S., Stupak, J., Tremblay, T.-L., Gurnani, K., Fulton, K. M., Weeratna, R. D., Twine, S. M., & Li, J. (2018). Chemical derivatization strategy for extending the identification of MHC Class I immunopeptides. *Analytical Chemistry*, *90*(19), 11409–11416. https://doi.org/10.1021/acs.analchem.8b02420

Chen, X., & Jensen, P. E. (2008). MHC class II antigen presentation and immunological abnormalities due to deficiency of MHC class II and its associated genes. *Experimental and Molecular Pathology*, *85*(1), 40–44. https://doi.org/10.1016/j.yexmp.2008.03.011

Chi, H., Liu, C., Yang, H., Zeng, W.-F., Wu, L., Zhou, W.-J., Wang, R.-M., Niu, X.-N., Ding, Y.-H., Zhang, Y., Wang, Z.-W., Chen, Z.-L., Sun, R.-X., Liu, T., Tan, G.-M., Dong, M.-Q., Xu, P., Zhang, P.-H., & He, S.-M. (2018). Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nature Biotechnology*, *36*, 11. https://doi.org/10.1038/nbt.4236

Chicz, R. M., Urban, R. G., Gorga, J. C., Vignali, D. A., Lane, W. S., & Strominger, J. L. (1993). Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. *Journal of Experimental Medicine*, *178*(1), 27–47. https://doi.org/10.1084/jem.178.1.27

Chong, C., Coukos, G., & Bassani-Sternberg, M. (2022). Identification of tumor antigens with immunopeptidomics. *Nature Biotechnology*, *40*, 2. https://doi.org/10.1038/s41587-021-01038-8

Chong, C., Müller, M., Pak, H., Harnett, D., Huber, F., Grun, D., Leleu, M., Auger, A., Arnaud, M., Stevenson, B. J., Michaux, J., Bilic, I., Hirsekorn, A., Calviello, L., Simó-Riudalbas, L., Planet, E., Lubiński, J., Bryśkiewicz, M., Wiznerowicz, M., … Bassani-Sternberg, M. (2020). Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nature Communications*, *11*, 1. https://doi.org/10.1038/s41467-020-14968-9

Chu, J., Gao, F., Yan, M., Zhao, S., Yan, Z., Shi, B., & Liu, Y. (2022). Natural killer cells: a promising immunotherapy for cancer. *Journal of Translational Medicine*, *20*(1), 240. https://doi.org/10.1186/s12967-022-03437-0

Cobbold, M., De La Peña, H., Norris, A., Polefrone, J. M., Qian, J., English, A. M., Cummings, K. L., Penny, S., Turner, J. E., Cottine, J., Abelin, J. G., Malaker, S. A., Zarling, A. L., Huang, H.-W., Goodyear, O., Freeman, S. D., Shabanowitz, J., Pratt, G., Craddock, C., … Engelhard, V. H. (2013). MHC Class I–associated phosphopeptides are the targets of memory-like immunity in leukemia. *Science Translational Medicine*, *5*(203), 203ra125. https://doi.org/10.1126/scitranslmed.3006061

Corthay, A., Bäcklund, J., Broddefalk, J., Michaëlsson, E., Goldschmidt, T. J., Kihlberg, J., & Holmdahl, R. (1998). Epitope glycosylation plays a critical role for T cell recognition of type II collagen in collagen-induced arthritis. *European Journal of Immunology*, *28*(8), 2580–2590. https://doi.org/10.1002/(SICI)1521-4141(199808)28:08<2580::AID-IMMU2580>3.0.CO;2-X

Cox, J. (2023). Prediction of peptide mass spectral libraries with machine learning. *Nature Biotechnology*, *41*, 1. https://doi.org/10.1038/s41587-022-01424-w

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., & Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of Proteome Research*, *10*(4), 1794–1805. https://doi.org/10.1021/pr101065j

Craig, R., & Beavis, R. C. (2003). A method for reducing the time required to match protein sequences with tandem mass spectra. *Rapid Communications in Mass Spectrometry*, *17*(20), 2310–2316. https://doi.org/10.1002/rcm.1198

Cresswell, P. (1996). Invariant chain structure and MHC class II function. *Cell*, *84*(4), 505–507. https://doi.org/10.1016/S0092-8674(00)81025-9

Cresswell, P., Ackerman, A. L., Giodini, A., Peaper, D. R., & Wearsch, P. A. (2005). Mechanisms of MHC class I-restricted antigen processing and cross-presentation. *Immunological Reviews*, *207*, 145–157. https://doi.org/10.1111/j.0105-2896.2005.00316.x

Croft, N. P., Smith, S. A., Wong, Y. C., Tan, C. T., Dudek, N. L., Flesch, I. E. A., Lin, L. C. W., Tscharke, D. C., & Purcell, A. W. (2013). Kinetics of antigen expression and epitope presentation during virus infection. *PLOS Pathogens*, *9*(1), e1003129. https://doi.org/10.1371/journal.ppat.1003129

Declercq, A., Bouwmeester, R., Hirschler, A., Carapito, C., Degroeve, S., Martens, L., & Gabriels, R. (2022). MS2Rescore: data-driven rescoring dramatically boosts immunopeptide identification rates. *Molecular & Cellular Proteomics*, *21*(8), 100266. https://doi.org/10.1016/j.mcpro.2022.100266

Degroeve, S., Gabriels, R.,Velghe, K., Bouwmeester, R., Tichshenko, N., & Martens, L. (2022). ionbot: a novel, innovative and sensitive machine learning approach to LC-MS/MS

peptide identification. bioRxiv, 2021.07.02.450686. https://doi.org/10.1101/2021.07.02.450686

Degroeve, S., & Martens, L. (2013). MS2PIP: a tool for MS/MS peak intensity prediction. *Bioinformatics*, *29*(24), 3199–3203. https://doi.org/10.1093/bioinformatics/btt544

Devabhaktuni, A., Lin, S., Zhang, L., Swaminathan, K., Gonzalez, C. G., Olsson, N., Pearlman, S. M., Rawson, K., & Elias, J. E. (2019). TagGraph reveals vast protein modification landscapes from large tandem mass spectrometry datasets. *Nature Biotechnology*, *37*, 4. https://doi.org/10.1038/s41587-019-0067-5

Distler, U., Kuharev, J., & Tenzer, S. (2014). Biomedical applications of ion mobility-enhanced data-independent acquisition-based label-free quantitative proteomics. *Expert Review of Proteomics*, *11*(6), 675–684. https://doi.org/10.1586/14789450.2014.971114

Doll, S., & Burlingame, A. L. (2015). Mass spectrometry-based detection and assignment of protein posttranslational modifications. *ACS Chemical Biology*, *10*(1), 63–71. https://doi.org/10.1021/cb500904b

Dudek, N. L., Croft, N. P., Schittenhelm, R. B., Ramarathinam, S. H., & Purcell, A. W. (2016). A systems approach to understand antigen presentation and the immune response. *Methods in Molecular Biology (Clifton, N.J.)*, *1394*, 189–209. https://doi.org/10.1007/978-1-4939-3341-9_14

Duru, A. D., Sun, R., Allerbring, E. B., Chadderton, J., Kadri, N., Han, X., Peqini, K., Uchtenhagen, H., Madhurantakam, C., Pellegrino, S., Sandalova, T., Nygren, P.-Å., Turner, S. J., & Achour, A. (2020). Tuning antiviral CD8 T-cell response via proline-altered peptide ligand vaccination. *PLOS Pathogens*, *16*(5), e1008244. https://doi.org/10.1371/journal.ppat.1008244

Eng, J. K., McCormack, A. L., & Yates, J. R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*, *5*(11), 976–989. https://doi.org/10.1016/1044-0305(94)80016-2

Eng, J. K., Jahan, T. A., & Hoopmann, M. R. (2013). Comet: an open-source MS/MS sequence database search tool. *PROTEOMICS*, *13*(1), 22–24. https://doi.org/10.1002/pmic.201200439

Engelhard, V. H., Obeng, R. C., Cummings, K. L., Petroni, G. R., Ambakhutwala, A. L., Chianese-Bullock, K. A., Smith, K. T., Lulu, A., Varhegyi, N., Smolkin, M. E., Myers, P., Mahoney, K. E., Shabanowitz, J., Buettner, N., Hall, E. H., Haden, K., Cobbold, M., Hunt, D. F., Weiss, G., … Craig L Slingluff, J. (2020). MHC-restricted phosphopeptide antigens: preclinical validation and first-in-humans clinical trial in participants with high-risk melanoma. *Journal for ImmunoTherapy of Cancer*, *8*(1), e000262. https://doi.org/10.1136/jitc-2019-000262

Falk, K., Rötzschke, O., Deres, K., Metzger, J., Jung, G., & Rammensee, H. G. (1991). Identification of naturally processed viral nonapeptides allows their quantification in infected cells and suggests an allele-specific T cell epitope forecast. *The Journal of Experimental Medicine*, *174*(2), 425–434. https://doi.org/10.1084/jem.174.2.425

Falkner, J. A., Falkner, J. W., Yocum, A. K., & Andrews, P. C. (2008). A spectral clustering approach to MS/MS identification of posttranslational modifications. *Journal of Proteome Research*, *7*(11), 4614–4622. https://doi.org/10.1021/pr800226w

Faridi, P., Purcell, A. W., & Croft, N. P. (2018). In immunopeptidomics we need a sniper instead of a shotgun. *PROTEOMICS*, *18*(12), 1700464. https://doi.org/10.1002/pmic.201700464

Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F., & White, F. M. (2002). Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nature Biotechnology*, *20*, 3. https://doi.org/10.1038/nbt0302-301

Frank, A. M., Monroe, M. E., Shah, A. R., Carver, J. J., Bandeira, N., Moore, R. J., Anderson, G. A., Smith, R. D., & Pevzner, P. A. (2011). Spectral archives: extending spectral libraries to analyze both identified and unidentified spectra. *Nature Methods*, *8*, 7. https://doi.org/10.1038/nmeth.1609

Freestone, J., Short, T., Noble, W. S., & Keich, U. (2022). Group-walk: a rigorous approach to group-wise false discovery rate analysis by target-decoy competition. *Bioinformatics*, *38*(Sup-Supplement_2), ii82–ii88. https://doi.org/10.1093/bioinformatics/btac471

Gallien, S., Bourmaud, A., Kim, S. Y., & Domon, B. (2014). Technical considerations for large-scale parallel reaction monitoring analysis. *Journal of Proteomics*, *100*, 147–159. https://doi.org/10.1016/j.jprot.2013.10.029

Gao, Y., & Wang, Y. (2007). A method to determine the ionization efficiency change of peptides caused by phosphorylation. *Journal of the American Society for Mass Spectrometry*, *18*(11), 1973–1976. https://doi.org/10.1016/j.jasms.2007.08.010

Gatlin, C. L., Eng, J. K., Cross, S. T., Detter, J. C., & Yates, J. R. (2000). Automated identification of amino acid sequence variations in proteins by hplc/microspray tandem mass spectrometry. *Analytical Chemistry*, *72*(4), 757–763. https://doi.org/10.1021/ac991025n

Gessulat, S., Schmidt, T., Zolg, D. P., Samaras, P., Schnatbaum, K., Zerweck, J., Knaute, T., Rechenberger, J., Delanghe, B., Huhmer, A., Reimer, U., Ehrlich, H.-C., Aiche, S., Kuster, B., & Wilhelm, M. (2019). Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nature Methods*, *16*, 6. https://doi.org/10.1038/s41592-019-0426-7

Griss, J. (2016). Spectral library searching in proteomics. *PROTEOMICS*, *16*(5), 729–740. https://doi.org/10.1002/pmic.201500296

Griss, J., Foster, J. M., Hermjakob, H., & Vizcaíno, J. A. (2013). PRIDE cluster: building a consensus of proteomics data. *Nature Methods*, *10*(2), Article 2. https://doi.org/10.1038/nmeth.2343

Griss, J., Perez-Riverol, Y., Lewis, S., Tabb, D. L., Dianes, J. A., del-Toro, N., Rurik, M., Walzer, M., Kohlbacher, O., Hermjakob, H., Wang, R., & Vizcaíno, J. A. (2016). Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets. *Nature Methods*, *13*, 8. https://doi.org/10.1038/nmeth.3902

Gubin, M. M., Zhang, X., Schuster, H., Caron, E., Ward, J. P., Noguchi, T., Ivanova, Y., Hundal, J., Arthur, C. D., Krebber, W.-J., Mulder, G. E., Toebes, M., Vesely, M. D., Lam, S. S. K., Korman, A. J., Allison, J. P., Freeman, G. J., Sharpe, A. H., Pearce, E. L., … Schreiber, R. D. (2014). Checkpoint blockade cancer immunotherapy targets tumour-

specific mutant antigens. *Nature*, *515*, 7528. https://doi.org/10.1038/nature13988

Guilloy, N., Brunet, M. A., Leblanc, S., Jacques, J.-F., Hardy, M.-P., Ehx, G., Lanoix, J., Thibault, P., Perreault, C., & Roucou, X. (2023). OpenCustomDB: integration of unannotated open reading frames and genetic variants to generate more comprehensive customized protein databases. *Journal of Proteome Research*, *22*(5), 1492–1500. https://doi.org/10.1021/acs.jproteome.3c00054

Han, X., He, L., Xin, L., Shan, B., & Ma, B. (2011). PeaksPTM: mass spectrometry-based identification of peptides with unspecified modifications. *Journal of Proteome Research*, *10*(7), 2930–2936. https://doi.org/10.1021/pr200153k

Harriff, M. J., Wolfe, L. M., Swarbrick, G., Null, M., Cansler, M. E., Canfield, E. T., Vogt, T., Toren, K. G., Li, W., Jackson, M., Lewinsohn, D. A., Dobos, K. M., & Lewinsohn, D. M. (2017). HLA-E presents glycopeptides from the Mycobacterium tuberculosis protein MPT32 to human CD8+ T cells. *Scientific Reports*, *7*, 1. https://doi.org/10.1038/s41598-017-04894-0

Hassan, C., Kester, M. G. D., Oudgenoeg, G., de Ru, A. H., Janssen, G. M. C., Drijfhout, J. W., Spaapen, R. M., Jiménez, C. R., Heemskerk, M. H. M., Falkenburg, J. H. F., & van Veelen, P. A. (2014). Accurate quantitation of MHC-bound peptides by application of isotopically labeled peptide MHC complexes. *Journal of Proteomics*, *109*, 240–244. https://doi.org/10.1016/j.jprot.2014.07.009

Haurum, J. S., Høier, I. B., Arsequell, G., Neisig, A., Valencia, G., Zeuthen, J., Neefjes, J., & Elliott, T. (1999). Presentation of cytosolic glycosylated peptides by human Class I major histocompatibility complex molecules in vivo. *Journal of Experimental Medicine*, *190*(1), 145–150. https://doi.org/10.1084/jem.190.1.145

Hensen, S. M. M., & Pruijn, G. J. M. (2014). Methods for the detection of peptidylarginine deiminase (PAD) activity and protein citrullination. *Molecular & Cellular Proteomics*, *13*(2), 388–396. https://doi.org/10.1074/mcp.R113.033746

Hilton, H. G., & Parham, P. (2013). Direct binding to antigen-coated beads refines the specificity and cross-reactivity of four monoclonal antibodies that recognize polymorphic epitopes of HLA class I molecules. *Tissue Antigens*, *81*(4), 212–220. https://doi.org/10.1111/tan.12095

Hoenisch Gravel, N., Nelde, A., Bauer, J., Mühlenbruch, L., Schroeder, S. M., Neidert, M. C., Scheid, J., Lemke, S., Dubbelaar, M. L., Wacker, M., Dengler, A., Klein, R., Mauz, P.-S., Löwenheim, H., Hauri-Hohl, M., Martin, R., Hennenlotter, J., Stenzl, A., Heitmann, J. S., … Walz, J. S. (2023). TOFIMS mass spectrometry-based immunopeptidomics refines tumor antigen identification. *Nature Communications*, *14*, 1. https://doi.org/10.1038/s41467-023-42692-7

Hogan, K. T., Sutton, J. N., Chu, K. U., Busby, J. A. C., Shabanowitz, J., Hunt, D. F., & Slingluff, C. L. (2005). Use of selected reaction monitoring mass spectrometry for the detection of specific MHC class I peptide antigens on A3 supertype family members. *Cancer Immunology, Immunotherapy*, *54*(4), 359–371. https://doi.org/10.1007/s00262-004-0592-y

Hogan, K. T., Eisinger, D. P., Cupp, S. B., Lekstrom, K. J., Deacon, D. D., Shabanowitz, J., Hunt, D. F., Engelhard, V. H., Slingluff, C. L., & Ross, M. M. (1998). The peptide recognized by HLA-A68.2-restricted, squamous cell carcinoma of the lung-specific cytotoxic T lymphocytes is derived from a mutated elongation factor 2 gene. *Cancer Research*, *58*(22), 5144–5150.

Hunt, D. F., Norris, A., English, A. M., Shabanowitz, J., Hilderbrand, W. H., & Hawkins, O. E. (2022). Identification of MHC class I phospho-peptide antigens from breast cancer utilizing SHLA technology and complementary enrichment strategies (United States Patent US11414458B2). https://patents.google.com/patent/US11414458B2/en

Hunt, D. F., Shabanowitz, J., Myers, P. T., Cobbold, M., Büttner, N., Malaker, S. A., & Penny, S. (2019). Target peptides for cancer therapy and diagnostics (United States Patent US20190374627A1). https://patents.google.com/patent/US20190374627A1/en

Hunt, D. F., Shabanowitz, J., Cottine, J., English, A. M., Norris, A., Engelhard, V. H., Cobbold, M., Cummings, K. L., Zarling, A., Obeng, R. C., & Qian, J. (2013). Class i mhc phosphopeptides for cancer immunotherapy and diagnosis (United States Patent US20130259883A1). https://patents.google.com/patent/US20130259883A1/en

Istrail, S., Florea, L., Halldórsson, B. V., Kohlbacher, O., Schwartz, R. S., Yap, V. B., Yewdell, J. W., & Hoffman, S. L. (2004). Comparative immunopeptidomics of humans and their pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(36), 13268–13272. https://doi.org/10.1073/pnas.0404740101

Jin, Z., Xu, S., Zhang, X., Ling, T., Dong, N., Ouyang, W., Gao, Z., Chang, C., & Sun, S. (2023). ContraNovo: a contrastive learning approach to enhance de novo peptide sequencing. arXiv. https://doi.org/10.48550/arXiv.2312.11584

Kaabinejadian, S., Barra, C., Alvarez, B., Yari, H., Hildebrand, W. H., & Nielsen, M. (2022). Accurate MHC Motif deconvolution of immunopeptidomics data reveals a significant contribution of DRB3, 4 and 5 to the total DR immunopeptidome. *Frontiers in Immunology*, *13*, 835454.

Kacen, A., Javitt, A., Kramer, M. P., Morgenstern, D., Tsaban, T., Shmueli, M. D., Teo, G. C., da Veiga Leprevost, F., Barnea, E., Yu, F., Admon, A., Eisenbach, L., Samuels, Y., Schueler-Furman, O., Levin, Y., Nesvizhskii, A. I., & Merbl, Y. (2022). Post-translational modifications reshape the antigenic landscape of the MHC I immunopeptidome in tumors. *Nature Biotechnology*, *41*, 239–251. https://doi.org/10.1038/s41587-022-01464-2

Kalaora, S., Barnea, E., Merhavi-Shoham, E., Qutob, N., Teer, J. K., Shimony, N., Schachter, J., Rosenberg, S. A., Besser, M. J., Admon, A., & Samuels, Y. (2016). Use of HLA peptidomics and whole exome sequencing to identify human immunogenic neo-antigens. *Oncotarget*, *7*(5), 5110–5117. https://doi.org/10.18632/oncotarget.6960

Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., & MacCoss, M. J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature Methods*, *4*, 11. https://doi.org/10.1038/nmeth1113

Karunratanakul, K., Tang, H.-Y., Speicher, D. W., Chuangsuwanich, E., & Sriswasdi, S. (2019). Uncovering thousands of new peptides with sequence-mask-search hybrid de novo peptide sequencing framework. *Molecular & Cellular Proteomics*, *18*(12), 2478–2491. https://doi.org/10.1074/mcp.TIR119.001656

Kastrup, I. B., Andersen, M. H., Elliott, T., & Haurum, J. S. (2001). MHC-restricted T cell responses against posttranslationally modified peptide antigens. In F. J. Dixon (Ed.), *Advances in Immunology* (78, pp. 267–289). Academic Press. https://doi.org/10.1016/S0065-2776(01)78006-6

Keller, A., Nesvizhskii, A. I., Kolker, E., & Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Analytical Chemistry*, *74*(20), 5383–5392. https://doi.org/10.1021/ac025747h

Kevin E., Konstantinos K., Oliver M., Amandla M., Jakob Berg J., Wesley W., Sam P. B. van B., Marcin S., Andreas H. L., Stan J. J. B., Anne L., Erwin M. S., J. Van Goey, U. auf dem Keller, K. Beguir, N. Lopez Carranza, & T. P. Jenkins. (2024). De novo peptide sequencing with InstaNovo: accurate, database-free peptide identification for large-scale proteomics experiments. bioRxiv. https://doi.org/10.1101/2023.08.30.555055

Kim, M.-S., Zhong, J., & Pandey, A. (2016). Common errors in mass spectrometry-based analysis of post-translational modifications. *PROTEOMICS*, *16*(5), 700–714. https://doi.org/10.1002/pmic.201500355

Kim, S., & Pevzner, P. A. (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. *Nature Communications*, *5*, 1. https://doi.org/10.1038/ncomms6277

Kloetzel, P.-M. (2001). Antigen processing by the proteasome. *Nature Reviews Molecular Cell Biology*, *2*, 3. https://doi.org/10.1038/35056572

Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., & Nesvizhskii, A. I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. *Nature Methods*, *14*, 5. https://doi.org/10.1038/nmeth.4256

Kuznetsov, A., Voronina, A., Govorun, V., & Arapidi, G. (2020). Critical review of existing MHC I immunopeptidome isolation methods. *Molecules (Basel, Switzerland)*, *25*(22), 5409. https://doi.org/10.3390/molecules25225409

Lanoix, J., Durette, C., Courcelles, M., Cossette, É., Comtois-Marotte, S., Hardy, M.-P., Côté, C., Perreault, C., & Thibault, P. (2018). Comparison of the MHC I immunopeptidome repertoire of B-cell lymphoblasts using two isolation methods. *Proteomics*, *18*(12), e1700251. https://doi.org/10.1002/pmic.201700251

Larsen, M. R., Trelle, M. B., Thingholm, T. E., & Jensen, O. N. (2006). Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *BioTechniques*, *40*(6), 790–798. https://doi.org/10.2144/000112201

Lazear, M. R. (2023). Sage: an open-source tool for fast proteomics searching and quantification at scale. *Journal of Proteome Research*, *22*(11), 3652–3659. https://doi.org/10.1021/acs.jproteome.3c00486

Leddy, O. K., White, F. M., & Bryson, B. D. (2021). Leveraging immunopeptidomics to study and combat infectious disease. *mSystems*, *6*(4), e0031021. https://doi.org/10.1128/mSystems.00310-21

Li, Q., Shortreed, M. R., Wenger, C. D., Frey, B. L., Schaffer, L. V., Scalf, M., & Smith, L. M. (2017). Global post-translational modification discovery. *Journal of Proteome Research*, *16*(4), 1383–1390. https://doi.org/10.1021/acs.jproteome.6b00034

Li, X., Pak, H. S., Huber, F., Michaux, J., Taillandier-Coindard, M., Altimiras, E. R., & Bassani-Sternberg, M. (2023). A microfluidics-enabled automated workflow of sample preparation for MS-based immunopeptidomics. *Cell Reports Methods*, *3*(6), 100479. https://doi.org/10.1016/j.crmeth.2023.100479

Liao, H., Barra, C., Zhou, Z., Peng, X., Woodhouse, I., Tailor, A., Parker, R., Carré, A., Borrow, P., Hogan, M. J., Paes, W., Eisenlohr, L. C., Mallone, R., Nielsen, M., & Ternette, N. (2024). MARS an improved de novo peptide candidate selection method for non-canonical antigen target discovery in cancer. *Nature Communications*, *15*(1), 661. https://doi.org/10.1038/s41467-023-44460-z

Lim Kam Sian, T. C. C., Goncalves, G., Steele, J. R., Shamekhi, T., Bramberger, L., Jin, D., Shahbazy, M., Purcell, A. W., Ramarathinam, S., Stoychev, S., & Faridi, P. (2023). SAPrIm, a semi-automated protocol for mid-throughput immunopeptidomics. *Frontiers in Immunology*, *14*, 1107576. https://www.frontiersin.org/articles/10.3389/fimmu.2023.1107576

Linnemann, C., van Buuren, M. M., Bies, L., Verdegaal, E. M. E., Schotte, R., Calis, J. J. A., Behjati, S., Velds, A., Hilkmann, H., Atmioui, D. el, Visser, M., Stratton, M. R., Haanen, J. B. A. G., Spits, H., van der Burg, S. H., & Schumacher, T. N. M. (2015). High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4+ T cells in human melanoma. *Nature Medicine*, *21*, 1. https://doi.org/10.1038/nm.3773

Liu, H., Ponniah, G., Neill, A., Patel, R., & Andrien, B. (2013). Accurate determination of protein methionine oxidation by stable isotope labeling and LC-MS analysis. *Analytical Chemistry*, *85*(24), 11705–11709. https://doi.org/10.1021/ac403072w

Liu, K., Ye, Y., Li, S., & Tang, H. (2023). Accurate de novo peptide sequencing using fully convolutional neural networks. *Nature Communications*, *14*(1), 7974. https://doi.org/10.1038/s41467-023-43010-x

Lou, R., & Shui, W. (2024). Acquisition and analysis of DIA-based proteomic data: a comprehensive survey in 2023. *Molecular & Cellular Proteomics*, *23*(2), 100712. https://doi.org/10.1016/j.mcpro.2024.100712

Luo, X., Bittremieux, W., Griss, J., Deutsch, E. W., Sachsenberg, T., Levitsky, L. I., Ivanov, M. V., Bubis, J. A., Gabriels, R., Webel, H., Sanchez, A., Bai, M., Käll, L., & Perez-Riverol, Y. (2022). A comprehensive evaluation of consensus spectrum generation methods in proteomics. *Journal of Proteome Research*, *21*(6), 1566–1574. https://doi.org/10.1021/acs.jproteome.2c00069

Ma, B. (2015). Novor: real-time peptide de novo sequencing software. *Journal of the American Society for Mass Spectrometry*, *26*(11), 1885–1894. https://doi.org/10.1007/s13361-015-1204-0

Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., & Lajoie, G. (2003). PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, *17*(20), 2337–2342. https://doi.org/10.1002/rcm.1196

Maes, E., Oeyen, E., Boonen, K., Schildermans, K., Mertens, I., Pauwels, P., Valkenborg, D., & Baggerman, G. (2019). The challenges of peptidomics in complementing proteomics in a clinical context. *Mass Spectrometry Reviews*, *38*(3), 253–264. https://doi.org/10.1002/mas.21581

Mahoney, K. E., Shabanowitz, J., & Hunt, D. F. (2021). MHC phosphopeptides: promising targets for immunotherapy of cancer and other chronic diseases. *Molecular & Cellular Proteomics*, *20*, 100112. https://doi.org/10.1016/j.mcpro.2021.100112

Malaker, S. A., Ferracane, M. J., Depontieu, F. R., Zarling, A. L., Shabanowitz, J., Bai, D. L., Topalian, S. L., Engelhard, V. H., & Hunt, D. F. (2017). Identification and characterization of complex glycosylated peptides presented by the MHC Class II processing pathway in melanoma. *Journal of Proteome Research*, *16*(1), 228–237. https://doi.org/10.1021/acs.jproteome.6b00496

Malaker, S. A., Penny, S. A., Steadman, L. G., Myers, P. T., Loke, J. C., Raghavan, M., Bai, D. L., Shabanowitz, J., Hunt, D. F., & Cobbold, M. (2017). Identification of glycopeptides as posttranslationally modified neoantigens in leukemia. *Cancer Immunology Research*, *5*(5), 376–384. https://doi.org/10.1158/2326-6066.CIR-16-0280

Mann, M., & Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. *Nature Biotechnology*, *21*, 3. https://doi.org/10.1038/nbt0303-255

Mao, Z., Zhang, R., Xin, L., & Li, M. (2023). Mitigating the missing-fragmentation problem in de novo peptide sequencing with a two-stage graph-based deep learning model. *Nature Machine Intelligence*, *5*(11), 1250–1260. https://doi.org/10.1038/s42256-023-00738-x

Marcu, A., Bichmann, L., Kuchenbecker, L., Kowalewski, D. J., Freudenmann, L. K., Backert, L., Mühlenbruch, L., Szolek, A., Lübke, M., Wagner, P., Engler, T., Matovina, S., Wang, J., Hauri-Hohl, M., Martin, R., Kapolou, K., Walz, J. S., Velz, J., Moch, H., … Neidert, M. C. (2021). HLA ligand Atlas: a benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy. *Journal for ImmunoTherapy of Cancer*, *9*(4), e002071. https://doi.org/10.1136/jitc-2020-002071

Marino, F., Bern, M., Mommen, G. P. M., Leney, A. C., van Gaans-van den Brink, J. A. M., Bonvin, A. M. J. J., Becker, C., van Els, C. A. C. M., & Heck, A. J. R. (2015). Extended O-GlcNAc on HLA Class-I-bound peptides. *Journal of the American Chemical Society*, *137*(34), 10922–10925. https://doi.org/10.1021/jacs.5b06586

Mazhar, S., Taylor, S. E., Sangodkar, J., & Narla, G. (2019). Targeting PP2A in cancer: combination therapies. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1866*(1), 51–63. https://doi.org/10.1016/j.bbamcr.2018.08.020

Mei, S., Ayala, R., Ramarathinam, S. H., Illing, P. T., Faridi, P., Song, J., Purcell, A. W., & Croft, N. P. (2020). Immunopeptidomic analysis reveals that deamidated HLA-bound peptides arise predominantly from deglycosylated precursors. *Molecular & Cellular Proteomics*, *19*(7), 1236–1247. https://doi.org/10.1074/mcp.RA119.001846

Meier, F., Brunner, A.-D., Koch, S., Koch, H., Lubeck, M., Krause, M., Goedecke, N., Decker, J., Kosinski, T., Park, M. A., Bache, N., Hoerning, O., Cox, J., Räther, O., & Mann, M. (2018). Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer*. *Molecular & Cellular Proteomics*, *17*(12), 2534–2545. https://doi.org/10.1074/mcp.TIR118.000900

Menschaert, G., Vandekerckhove, T. T. M., Baggerman, G., Schoofs, L., Luyten, W., & Criekinge, W. V. (2010). Peptidomics coming of age: a review of contributions from a bioinformatics angle. *Journal of Proteome Research*, *9*(5), 2051–2061. https://doi.org/10.1021/pr900929m

Menschaert, G., Vandekerckhove, T. T. M., Landuyt, B., Hayakawa, E., Schoofs, L., Luyten, W., & Van Criekinge, W. (2009). Spectral clustering in peptidomics studies helps to unravel modification profile of biologically active peptides and enhances peptide identification rate. *PROTEOMICS*, *9*(18), 4381–4388. https://doi.org/10.1002/pmic.200900248

Mishto, M., Horokhovskyi, Y., Cormican, J. A., Yang, X., Lynham, S., Urlaub, H., & Liepe, J. (2022). Database search engines and target database features impinge upon the identification of post-translationally cis-spliced peptides in HLA class I immunopeptidomes. *PROTEOMICS*, *22*(10), 2100226. https://doi.org/10.1002/pmic.202100226

Mohammed, F., Cobbold, M., Zarling, A. L., Salim, M., Barrett-Wilt, G. A., Shabanowitz, J., Hunt, D. F., Engelhard, V. H., & Willcox, B. E. (2008). Phosphorylation-dependent interaction between antigenic peptides and MHC class I: a molecular basis for the presentation of transformed self. *Nature Immunology*, *9*(11), 1236–1243. https://doi.org/10.1038/ni.1660

Mohammed, F., Stones, D. H., Zarling, A. L., Willcox, C. R., Shabanowitz, J., Cummings, K. L., Hunt, D. F., Cobbold, M., Engelhard, V. H., & Willcox, B. E. (2017). The antigenic identity of human class I MHC phosphopeptides is critically dependent upon phosphorylation status. *Oncotarget*, *8*(33), 54160–54172. https://doi.org/10.18632/oncotarget.16952

Mommen, G. P. M., Frese, C. K., Meiring, H. D., van Gaans-van den Brink, J., de Jong, A. P. J. M., van Els, C. A. C. M., & Heck, A. J. R. (2014). Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proceedings of the National Academy of Sciences*, *111*(12), 4507–4512. https://doi.org/10.1073/pnas.1321458111

Mommen, G. P. M., Marino, F., Meiring, H. D., Poelen, M. C. M., van Gaans-van den Brink, J. A. M., Mohammed, S., Heck, A. J. R., & van Els, C. A. C. M. (2016). Sampling from the proteome to the human leukocyte antigen-DR (HLA-DR) ligandome proceeds via high specificity. *Molecular & Cellular Proteomics*, *15*(4), 1412–1423. https://doi.org/10.1074/mcp.M115.055780

Morand, K., Talbo, G., & Mann, M. (1993). Oxidation of peptides during electrospray ionization. *Rapid Communications in Mass Spectrometry*, *7*(8), 738–743. https://doi.org/10.1002/rcm.1290070811

Mukherjee, S., Jankevics, A., Busch, F., Lubeck, M., Zou, Y., Kruppa, G., Heck, A. J. R., Scheltema, R. A., & Reiding, K. R. (2023). Oxonium ion–guided optimization of ion mobility–assisted glycoproteomics on the timsTOF Pro. *Molecular & Cellular Proteomics*, *22*(2), 100486. https://doi.org/10.1016/j.mcpro.2022.100486

Muth, T., & Renard, B. Y. (2018). Evaluating de novo sequencing in proteomics: already an accurate alternative to database-driven peptide identification? *Briefings in Bioinformatics*, *19*(5), 954–970. https://doi.org/10.1093/bib/bbx033

Nielsen, B. V., & Abaye, D. A. (2013). Influence of electrolytes and a supercharging reagent on charge state distribution and

response of neuropeptide ions generated during positive electrospray ionisation mass spectrometry. *European Journal of Mass Spectrometry*, *19*(5), 335–344. https://doi.org/10.1255/ejms.1246

Nilsson, J. B., Kaabinejadian, S., Yari, H., Peters, B., Barra, C., Gragert, L., Hildebrand, W., & Nielsen, M. (2023). Machine learning reveals limited contribution of trans-only encoded variants to the HLA-DQ immunopeptidome. *Communications Biology*, *6*, 442. https://doi.org/10.1038/s42003-023-04749-7

Ouspenskaia, T., Law, T., Clauser, K. R., Klaeger, S., Sarkizova, S., Aguet, F., Li, B., Christian, E., Knisbacher, B. A., Le, P. M., Hartigan, C. R., Keshishian, H., Apffel, A., Oliveira, G., Zhang, W., Chen, S., Chow, Y. T., Ji, Z., Jungreis, I., … Regev, A. (2022). Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nature Biotechnology*, *40*(2), 209–217. https://doi.org/10.1038/s41587-021-01021-3

Pace, A. L., Wong, R. L., Zhang, Y. T., Kao, Y.-H., & Wang, Y. J. (2013). Asparagine deamidation dependence on buffer type, pH, and temperature. *Journal of Pharmaceutical Sciences*, *102*(6), 1712–1723. https://doi.org/10.1002/jps.23529

Pak, H., Michaux, J., Huber, F., Chong, C., Stevenson, B. J., Müller, M., Coukos, G., & Bassani-Sternberg, M. (2021). Sensitive immunopeptidomics by leveraging available large-scale multi-HLA spectral libraries, data-independent acquisition, and MS/MS prediction. *Molecular & Cellular Proteomics*, *20*, 100080. https://doi.org/10.1016/j.mcpro.2021.100080

Parker, R., Tailor, A., Peng, X., Nicastri, A., Zerweck, J., Reimer, U., Wenschuh, H., Schnatbaum, K., & Ternette, N. (2021). The choice of search engine affects sequencing depth and HLA Class I allele-specific peptide repertoires. *Molecular & Cellular Proteomics*, *20*, 100124. https://doi.org/10.1016/j.mcpro.2021.100124

Parker, R., Partridge, T., Wormald, C., Kawahara, R., Stalls, V., Aggelakopoulou, M., Parker, J., Powell Doherty, R., Ariosa Morejon, Y., Lee, E., Saunders, K., Haynes, B. F., Acharya, P., Thaysen-Andersen, M., Borrow, P., & Ternette, N. (2021). Mapping the SARS-CoV-2 spike glycoprotein-derived peptidome presented by HLA class II on dendritic cells. *Cell Reports*, *35*(8), 109179. https://doi.org/10.1016/j.celrep.2021.109179

Peixoto, A., Relvas-Santos, M., Azevedo, R., Santos, L. L., & Ferreira, J. A. (2019). Protein glycosylation and tumor microenvironment alterations driving cancer hallmarks. *Frontiers in Oncology*, *9*, 380. https://www.frontiersin.org/articles/10.3389/fonc.2019.00380

Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D. J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., Walzer, M., Wang, S., Brazma, A., & Vizcaíno, J. A. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Research*, *50*(D1), D543–D552. https://doi.org/10.1093/nar/gkab1038

Perkins, D. N., Pappin, D. J., Creasy, D. M., & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, *20*(18), 3551–3567. https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2

Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S., & Coon, J. J. (2012). Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics*. *Molecular & Cellular Proteomics*, *11*(11), 1475–1488. https://doi.org/10.1074/mcp.O112.020131

Pfammatter, S., Bonneil, E., Lanoix, J., Vincent, K., Hardy, M.-P., Courcelles, M., Perreault, C., & Thibault, P. (2020). Extending the comprehensiveness of immunopeptidome analyses using isobaric peptide labeling. *Analytical Chemistry*, *92*(13), 9194–9204. https://doi.org/10.1021/acs.analchem.0c01545

Picciani, M., Gabriel, W., Giurcoiu, V.-G., Shouman, O., Hamood, F., Lautenbacher, L., Jensen, C. B., Müller, J., Kalhor, M., Soleymaniniya, A., Kuster, B., The, M., & Wilhelm, M. (2023). Oktoberfest: open-source spectral library generation and rescoring pipeline based on Prosit. *PROTEOMICS*, *24*(8), 2300112. https://doi.org/10.1002/pmic.202300112

Picotti, P., & Aebersold, R. (2012). Selected reaction monitoring–based proteomics: workflows, potential, pitfalls and future directions. *Nature Methods*, *9*, 6. https://doi.org/10.1038/nmeth.2015

Piovesana, S., Montone, C. M., Cavaliere, C., Crescenzi, C., La Barbera, G., Laganà, A., & Capriotti, A. L. (2019). Sensitive untargeted identification of short hydrophilic peptides by high-performance liquid chromatography on porous graphitic carbon coupled to high resolution mass spectrometry. *Journal of Chromatography A*, *1590*, 73–79. https://doi.org/10.1016/j.chroma.2018.12.066

Pollock, S. B., Rose, C. M., Darwish, M., Bouziat, R., Delamarre, L., Blanchette, C., & Lill, J. R. (2021). Sensitive and quantitative detection of MHC-I displayed neoepitopes using a semi-automated workflow and TOMAHAQ mass spectrometry. *Molecular & Cellular Proteomics*, *20*, 100108. https://doi.org/10.1016/j.mcpro.2021.100108

Prus, G., Hoegl, A., Weinert, B. T., & Choudhary, C. (2019). Analysis and interpretation of protein post-translational modification site stoichiometry. *Trends in Biochemical Sciences*, *44*(11), 943–960. https://doi.org/10.1016/j.tibs.2019.06.003

Purcell, A. W., Ramarathinam, S. H., & Ternette, N. (2019). Mass spectrometry–based identification of MHC-bound peptides for immunopeptidomics. *Nature Protocols*, *14*(6), Article 6. https://doi.org/10.1038/s41596-019-0133-y

Qi, Y. A., Maity, T. K., Cultraro, C. M., Misra, V., Zhang, X., Ade, C., Gao, S., Milewski, D., Nguyen, K. D., Ebrahimabadi, M. H., Hanada, K., Khan, J., Sahinalp, C., Yang, J. C., & Guha, U. (2021). Proteogenomic analysis unveils the HLA Class I-presented immunopeptidome in melanoma and EGFR-mutant lung adenocarcinoma. *Molecular & Cellular Proteomics*, *20*, 100136. https://doi.org/10.1016/j.mcpro.2021.100136

Racle, J., Michaux, J., Rockinger, G. A., Arnaud, M., Bobisse, S., Chong, C., Guillaume, P., Coukos, G., Harari, A., Jandus, C., Bassani-Sternberg, M., & Gfeller, D. (2019). Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes. *Nature Biotechnology*, *37*(11), 1283–1286. https://doi.org/10.1038/s41587-019-0289-6

Rammensee, H. G., Falk, K., & Rötzschke, O. (1993). Peptides naturally presented by MHC class I molecules. *Annual Review of Immunology*, *11*, 213–244. https://doi.org/10.1146/annurev.iy.11.040193.001241

Reily, C., Stewart, T. J., Renfrow, M. B., & Novak, J. (2019). Glycosylation in health and disease. *Nature Reviews Nephrology*, *15*, 6. https://doi.org/10.1038/s41581-019-0129-4

Roche, P. A., & Furuta, K. (2015). The ins and outs of MHC class II-mediated antigen processing and presentation. *Nature Reviews Immunology*, *15*(4), 203–216. https://doi.org/10.1038/nri3818

Rose, C. M., Erickson, B. K., Schweppe, D. K., Viner, R., Choi, J., Rogers, J., Bomgarden, R., Gygi, S. P., & Kirkpatrick, D. S. (2019). TomahaqCompanion: a tool for the creation and analysis of isobaric label based multiplexed targeted assays. *Journal of Proteome Research*, *18*(2), 594–605. https://doi.org/10.1021/acs.jproteome.8b00767

Rötzschke, O., Falk, K., Deres, K., Schild, H., Norda, M., Metzger, J., Jung, G., & Rammensee, H.-G. (1990). Isolation and analysis of naturally processed viral peptides as recognized by cytotoxic T cells. *Nature*, *348*, 6298. https://doi.org/10.1038/348252a0

Sajic, T., Liu, Y., & Aebersold, R. (2015). Using data-independent, high-resolution mass spectrometry in protein biomarker research: perspectives and clinical applications. *PROTEOMICS Clinical Applications*, *9*(3–4), 307–321. https://doi.org/10.1002/prca.201400117

Sandalova, T., Sala, B. M., & Achour, A. (2022). Structural aspects of chemical modifications in the MHC-restricted immunopeptidome; implications for immune recognition. *Frontiers in Chemistry*, *10*, 861609. https://www.frontiersin.org/articles/10.3389/fchem.2022.861609

Scull, K. E., Pandey, K., Ramarathinam, S. H., & Purcell, A. W. (2021). Immunopeptidogenomics: harnessing RNA-Seq to illuminate the dark immunopeptidome. *Molecular & Cellular Proteomics*, *20*, 100143. https://doi.org/10.1016/j.mcpro.2021.100143

Searle, B. C., Shannon, A. E., & Wilburn, D. B. (2023). Scribe: next generation library searching for DDA experiments. *Journal of Proteome Research*, *22*(2), 482–490. https://doi.org/10.1021/acs.jproteome.2c00672

Seo, J., & Lee, K.-J. (2004). Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. *Journal of Biochemistry and Molecular Biology*, *37*(1), 35–44. https://doi.org/10.5483/bmbrep.2004.37.1.035

Shahbazy, M., Ramarathinam, S. H., Illing, P. T., Jappe, E. C., Faridi, P., Croft, N. P., & Purcell, A. W. (2023). Benchmarking bioinformatics pipelines in data-independent acquisition mass spectrometry for immunopeptidomics. *Molecular & Cellular Proteomics*, *22*(4), 100515. https://doi.org/10.1016/j.mcpro.2023.100515

Shao, W., & Lam, H. (2017). Tandem mass spectral libraries of peptides and their roles in proteomics research. *Mass Spectrometry Reviews*, *36*(5), 634–648. https://doi.org/10.1002/mas.21512

Shao, W., Pedrioli, P. G. A., Wolski, W., Scurtescu, C., Schmid, E., Vizcaíno, J. A., Courcelles, M., Schuster, H., Kowalewski, D., Marino, F., Arlehamn, C. S. L., Vaughan, K., Peters, B., Sette, A., Ottenhoff, T. H. M., Meijgaarden, K. E., Nieuwenhuizen, N., Kaufmann, S. H. E., Schlapbach, R., … Caron, E. (2018). The SysteMHC Atlas project. *Nucleic Acids Research*, *46*(D1), D1237–D1247. https://doi.org/10.1093/nar/gkx664

Shapiro, I. E., & Bassani-Sternberg, M. (2023). The impact of immunopeptidomics: from basic research to clinical implementation. *Seminars in Immunology*, *66*, 101727. https://doi.org/10.1016/j.smim.2023.101727

Sidney, J., Vela, J. L., Friedrich, D., Kolla, R., von Herrath, M., Wesley, J. D., & Sette, A. (2018). Low HLA binding of diabetes-associated CD8+ T-cell epitopes is increased by post-translational modifications. *BMC Immunology*, *19*(1), 12. https://doi.org/10.1186/s12865-018-0250-3

Smith, C. C., Selitsky, S. R., Chai, S., Armistead, P. M., Vincent, B. G., & Serody, J. S. (2019). Alternative tumour-specific antigens. *Nature Reviews Cancer*, *19*(8), Article 8. https://doi.org/10.1038/s41568-019-0162-4

Smith, L. E., & Rogowska-Wrzesinska, A. (2020). The challenge of detecting modifications on proteins. *Essays in Biochemistry*, *64*(1), 135–153. https://doi.org/10.1042/EBC20190055

Solntsev, S. K., Shortreed, M. R., Frey, B. L., & Smith, L. M. (2018). Enhanced global post-translational modification discovery with MetaMorpheus. *Journal of Proteome Research*, *17*(5), 1844–1851. https://doi.org/10.1021/acs.jproteome.7b00873

Stern, L. J., & Santambrogio, L. (2016). The melting pot of the MHC II peptidome. *Current Opinion in Immunology*, *40*, 70–77. https://doi.org/10.1016/j.coi.2016.03.004

van Stipdonk, M. J. B., Badia-Martinez, D., Sluijter, M., Offringa, R., van Hall, T., & Achour, A. (2009). Design of agonistic altered peptides for the robust induction of CTL directed towards H-2Db in complex with the melanoma-associated epitope gp100. *Cancer Research*, *69*(19), 7784–7792. https://doi.org/10.1158/0008-5472.CAN-09-1724

Sturm, T., Sautter, B., Wörner, T. P., Stevanović, S., Rammensee, H.-G., Planz, O., Heck, A. J. R., & Aebersold, R. (2021). Mild acid elution and MHC immunoaffinity chromatography reveal similar albeit not identical profiles of the HLA Class I immunopeptidome. *Journal of Proteome Research*, *20*(1), 289–304. https://doi.org/10.1021/acs.jproteome.0c00386

Subramanian, A. (2002). Immunoaffinity chromatography. *Molecular Biotechnology*, *20*(1), 41–47. https://doi.org/10.1385/MB:20:1:041

Sugawara, S., Abo, T., & Kumagai, K. (1987). A simple method to eliminate the antigenicity of surface class I MHC molecules from the membrane of viable cells by acid treatment at pH 3. *Journal of Immunological Methods*, *100*(1–2), 83–90. https://doi.org/10.1016/0022-1759(87)90175-X

Suri, A., Lovitch, S. B., & Unanue, E. R. (2006). The wide diversity and complexity of peptides bound to class II MHC molecules. *Current Opinion in Immunology*, *18*(1), 70–77. https://doi.org/10.1016/j.coi.2005.11.002

Tan, C. T., Croft, N. P., Dudek, N. L., Williamson, N. A., & Purcell, A. W. (2011). Direct quantitation of MHC-bound peptide epitopes by selected reaction monitoring. *PROTEOMICS*, *11*(11), 2336–2340. https://doi.org/10.1002/pmic.201000531

Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C., & Mechtler, K. (2011). Universal and confident phosphorylation site localization using phosphoRS. *Journal of Proteome Research*, *10*(12), 5354–5362. https://doi.org/10.1021/pr200611n

Ternette, N., Yang, H., Partridge, T., Llano, A., Cedeño, S., Fischer, R., Charles, P. D., Dudek, N. L., Mothe, B.,

Crespo, M., Fischer, W. M., Korber, B. T. M., Nielsen, M., Borrow, P., Purcell, A. W., Brander, C., Dorrell, L., Kessler, B. M., & Hanke, T. (2016). Defining the HLA class I-associated viral antigen repertoire from HIV-1-infected human cells. *European Journal of Immunology*, *46*(1), 60–69. https://doi.org/10.1002/eji.201545890

The, M., & Käll, L. (2016). MaRaCluster: a fragment rarity metric for clustering fragment spectra in shotgun proteomics. *Journal of Proteome Research*, *15*(3), 713–720. https://doi.org/10.1021/acs.jproteome.5b00749

Ting, Y. T., Petersen, J., Ramarathinam, S. H., Scally, S. W., Loh, K. L., Thomas, R., Suri, A., Baker, D. G., Purcell, A. W., Reid, H. H., & Rossjohn, J. (2018). The interplay between citrullination and HLA-DRB1 polymorphism in shaping peptide binding hierarchies in rheumatoid arthritis. *Journal of Biological Chemistry*, *293*(9), 3236–3251. https://doi.org/10.1074/jbc.RA117.001013

Topalian, S. L., Depontieu, F. A., Hunt, D. F., Shabanowitz, J., Qian, J., Engelhard, V. H., & Zarling, A. L. (2016). Phosphopeptides as melanoma vaccines (United States Patent US9279011B2). https://patents.google.com/patent/US9279011B2/en

Tran, N. H., Zhang, X., Xin, L., Shan, B., & Li, M. (2017). De novo peptide sequencing by deep learning. *Proceedings of the National Academy of Sciences*, *114*(31), 8247–8252. https://doi.org/10.1073/pnas.1705691114

Tretter, C., Krätzig, N. de A., Pecoraro, M., Lange, S., Seifert, P., von Frankenberg, C., Untch, J., Zuleger, G., Wilhelm, M., Zolg, D. P., Dreyer, F. S., Bräunlein, E., Engleitner, T., Uhrig, S., Boxberg, M., Steiger, K., Slotta-Huspenina, J., Ochsenreither, S., von Bubnoff, N., … Krackhardt, A. M. (2023). Proteogenomic analysis reveals RNA as a source for tumor-agnostic neoantigen identification. *Nature Communications*, *14*(1), 4632. https://doi.org/10.1038/s41467-023-39570-7

Tsou, C.-C., Avtonomov, D., Larsen, B., Tucholska, M., Choi, H., Gingras, A.-C., & Nesvizhskii, A. I. (2015). DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nature Methods*, *12*, 3. https://doi.org/10.1038/nmeth.3255

Unanue, E. R., Turk, V., & Neefjes, J. (2016). Variations in MHC Class II antigen processing and presentation in health and disease. *Annual Review of Immunology*, *34*(1), 265–297. https://doi.org/10.1146/annurev-immunol-041015-055420

Van Houtven, J., Boonen, K., Baggerman, G., Askenazi, M., Laukens, K., Hooyberghs, J., & Valkenborg, D. (2020). PRiSM: a prototype for exhaustive, restriction-free database searching for mass spectrometry-based identification. *Rapid Communications in Mass Spectrometry*, e8962. https://doi.org/10.1002/rcm.8962

Van Wanseele, Y., Bongaerts, J., Segers, K., Viaene, J., De Bundel, D., Vander Heyden, Y., Smolders, I., & Van Eeckhaut, A. (2019). Assessing mixtures of supercharging agents to increase the abundance of a specific charge state of Neuromedin U. *Talanta*, *198*, 206–214. https://doi.org/10.1016/j.talanta.2019.01.098

Vizcaíno, J. A., Kubiniok, P., Kovalchik, K. A., Ma, Q., Duquette, J. D., Mongrain, I., Deutsch, E. W., Peters, B., Sette, A., Sirois, I., & Caron, E. (2020). The human immunopeptidome project: a roadmap to predict and treat immune diseases. *Molecular & Cellular Proteomics*, *19*(1), 31–49. https://doi.org/10.1074/mcp.R119.001743

Vlad, A. M., Muller, S., Cudic, M., Paulsen, H., Otvos, Jr., L., Hanisch, F.-G., & Finn, O. J. (2002). Complex carbohydrates are not removed during processing of glycoproteins by dendritic cells: processing of tumor antigen MUC1 glycopeptides for presentation to major histocompatibility complex Class II–restricted T cells. *Journal of Experimental Medicine*, *196*(11), 1435–1446. https://doi.org/10.1084/jem.20020493

Wang, L., Li, S., & Tang, H. (2019). msCRUSH: fast tandem mass spectral clustering using locality sensitive hashing. *Journal of Proteome Research*, *18*(1), 147–158. https://doi.org/10.1021/acs.jproteome.8b00448

Wang, M., Wang, J., Carver, J., Pullman, B. S., Cha, S. W., & Bandeira, N. (2018). Assembling the community-scale discoverable human proteome. *Cell Systems*, *7*(4), 412–421.e5. https://doi.org/10.1016/j.cels.2018.08.004

Wang, W., Yuan, T., Ma, L., Zhu, Y., Bao, J., Zhao, X., Zhao, Y., Zong, Y., Zhang, Y., Yang, S., Qiu, X., Shen, S., Wu, R., Wu, T., Wang, H., Gao, D., Wang, P., & Chen, L. (2022). Hepatobiliary tumor organoids reveal HLA class I neoantigen landscape and antitumoral activity of neoantigen peptide enhanced with immune checkpoint inhibitors. *Advanced Science*, *9*(22), 2105810. https://doi.org/10.1002/advs.202105810

Wieczorek, M., Abualrous, E. T., Sticht, J., Álvaro-Benito, M., Stolzenberg, S., Noé, F., & Freund, C. (2017). Major histocompatibility complex (MHC) Class I and MHC Class II proteins: conformational plasticity in antigen presentation. *Frontiers in Immunology*, *8*, 292. https://www.frontiersin.org/articles/10.3389/fimmu.2017.00292

Wilburn, D. B., Shannon, A. E., Spicer, V., Richards, A. L., Yeung, D., Swaney, D. L., Krokhin, O. V., & Searle, B. C.. (2023). Deep learning from harmonized peptide libraries enables retention time prediction of diverse posttranslational modifications. bioRxiv. https://doi.org/10.1101/2023.05.30.542978

Wilhelm, M., Zolg, D. P., Graber, M., Gessulat, S., Schmidt, T., Schnatbaum, K., Schwencke-Westphal, C., Seifert, P., de Andrade Krätzig, N., Zerweck, J., Knaute, T., Bräunlein, E., Samaras, P., Lautenbacher, L., Klaeger, S., Wenschuh, H., Rad, R., Delanghe, B., Huhmer, A., … Kuster, B. (2021). Deep learning boosts sensitivity of mass spectrometry-based immunopeptidomics. *Nature Communications*, *12*, 1. https://doi.org/10.1038/s41467-021-23713-9

Xin, L., Qiao, R., Chen, X., Tran, H., Pan, S., Rabinoviz, S., Bian, H., He, X., Morse, B., Shan, B., & Li, M. (2022). A streamlined platform for analyzing tera-scale DDA and DIA mass spectrometry data enables highly sensitive immunopeptidomics. *Nature Communications*, *13*, 3108. https://doi.org/10.1038/s41467-022-30867-7

Yamamoto, T. N., Kishton, R. J., & Restifo, N. P. (2019). Developing neoantigen-targeted T cell–based treatments for solid tumors. *Nature Medicine*, *25*(10), 1488–1499. https://doi.org/10.1038/s41591-019-0596-y

Yang, H., Chi, H., Zeng, W.-F., Zhou, W.-J., & He, S.-M. (2019). pNovo 3: precise de novo peptide sequencing using a learning-to-rank framework. *Bioinformatics*, *35*(14), i183–i190. https://doi.org/10.1093/bioinformatics/btz366

Yang, K. L., Yu, F., Teo, G. C., Li, K., Demichev, V., Ralser, M., & Nesvizhskii, A. I. (2023). MSBooster: improving peptide identification rates using deep learning-based features. *Nature Communications*, *14*, 1. https://doi.org/10.1038/s41467-023-40129-9

Yang, Y., & Fang, Q. (2024). Prediction of glycopeptide fragment mass spectra by deep learning. *Nature Communications*, *15*(1), 2448. https://doi.org/10.1038/s41467-024-46771-1

Yilmaz, M., Fondrie, W., Bittremieux, W., Oh, S., & Noble, W. S. (2022). De novo mass spectrometry peptide sequencing with a transformer model. Proceedings of the 39th International Conference on Machine Learning, 25514–25522. https://proceedings.mlr.press/v162/yilmaz22a.html

Yu, L.-R., & Veenstra, T. D. (2021). Characterization of phosphorylated proteins using mass spectrometry. *Current Protein & Peptide Science*, *22*(2), 148–157. https://doi.org/10.2174/1389203721999201123200439

Zarling, A. L., Ficarro, S. B., White, F. M., Shabanowitz, J., Hunt, D. F., & Engelhard, V. H. (2000). Phosphorylated peptides are naturally processed and presented by major histocompatibility complex class i molecules in vivo. *Journal of Experimental Medicine*, *192*(12), 1755–1762. https://doi.org/10.1084/jem.192.12.1755

Zarling, A. L., Polefrone, J. M., Evans, A. M., Mikesh, L. M., Shabanowitz, J., Lewis, S. T., Engelhard, V. H., & Hunt, D. F. (2006). Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(40), 14889–14894. https://doi.org/10.1073/pnas.0604045103

Zavala-Cerna, M. G., Martínez-García, E. A., Torres-Bugarín, O., Rubio-Jurado, B., Riebeling, C., & Nava, A. (2014). The clinical significance of posttranslational modification of autoantigens. *Clinical Reviews in Allergy & Immunology*, *47*(1), 73–90. https://doi.org/10.1007/s12016-014-8424-0

Zeneyedpour, L., Sten-van 't Hoff, J., & Luider, T. (2020). Using phosphoproteomics and next-generation sequencing to discover novel therapeutic targets in patient antibodies. *Expert Review of Proteomics*, *17*(9), 675–684. https://doi.org/10.1080/14789450.2020.1845147

Zeng, W.-F., Zhou, X.-X., Willems, S., Ammar, C., Wahle, M., Bludau, I., Voytik, E., Strauss, M. T., & Mann, M. (2022). AlphaPeptDeep: a modular deep learning framework to predict peptide properties for proteomics. *Nature Communications*, *13*, 1. https://doi.org/10.1038/s41467-022-34904-3

Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G. A., & Ma, B. (2012). PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification *. *Molecular & Cellular Proteomics*, *11*(4), M111.010587. https://doi.org/10.1074/mcp.M111.010587

Zhou, C., Zhu, C., & Liu, Q. (2019). Toward in silico identification of tumor neoantigens in immunotherapy. *Trends in Molecular Medicine*, *25*(11), 980–992. https://doi.org/10.1016/j.molmed.2019.08.001

Zolg, D. P., Gessulat, S., Paschke, C., Graber, M., Rathke-Kuhnert, M., Seefried, F., Fitzemeier, K., Berg, F., Lopez-Ferrer, D., Horn, D., Henrich, C., Huhmer, A., Delanghe, B., & Frejno, M. (2021). INFERYS rescoring: boosting peptide identifications and scoring confidence of database search results. *Rapid Communications in Mass Spectrometry*, e9128. https://doi.org/10.1002/rcm.9128

Zong, Y., Wang, Y., Yang, Y., Zhao, D., Wang, X., Shen, C., & Qiao, L. (2023). DeepFLR facilitates false localization rate control in phosphoproteomics. *Nature Communications*, *14*(1), 2269. https://doi.org/10.1038/s41467-023-38035-1

**How to cite this article:** Flender D, Vilenne F, Adams C, Boonen K, Valkenborg D, Baggerman G. Exploring the dynamic landscape of immunopeptidomics: Unravelling posttranslational modifications and navigating bioinformatics terrain. *Mass Spectrom Rev*. 2025;*44*:599–629. https://doi.org/10.1002/mas.21905