



An open-source computational and data resource to analyze digital maps of immunopeptidomes

Etienne Caron^{1*}, Lucia Espona¹, Daniel J Kowalewski^{2,3}, Heiko Schuster^{2,3}, Nicola Ternette⁴, Adán Alpízar⁵, Ralf B Schittenhelm⁶, Sri H Ramarathinam⁶, Cecilia S Lindestam Arlehamn⁷, Ching Chiek Koh¹, Ludovic C Gillet¹, Armin Rabsteyn^{2,3}, Pedro Navarro⁸, Sangtae Kim⁹, Henry Lam¹⁰, Theo Sturm¹, Miguel Marcilla⁵, Alessandro Sette⁷, David S Campbell¹¹, Eric W Deutsch¹¹, Robert L Moritz¹¹, Anthony W Purcell⁶, Hans-Georg Rammensee^{2,3}, Stefan Stevanovic^{2,3}, Ruedi Aebersold^{1,12*}

¹Department of Biology, Institute of Molecular Systems Biology, ETH Zürich, Zurich, Switzerland; ²Department of Immunology, Interfaculty Institute for Cell Biology, University of Tübingen, Tübingen, Germany; ³DKFZ partner site Tübingen, German Cancer Consortium, Tübingen, Germany; ⁴Target Discovery Institute Mass Spectrometry Laboratory, University of Oxford, Oxford, United Kingdom; ⁵Proteomics Unit, Spanish National Biotechnology Centre, Madrid, Spain; ⁶Department of Biochemistry and Molecular Biology, Monash University, Clayton, Australia; ⁷La Jolla Institute for Allergy and Immunology, La Jolla, United States; ⁸Institute for Immunology, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany; ⁹Pacific Northwest National Laboratory, Richland, United States; ¹⁰Division of Biomedical Engineering and Department of Chemical and Biomolecular Engineering, Hong Kong University of Science and Technology, Hong Kong, China; ¹¹Institute for Systems Biology, Seattle, United States; ¹²Faculty of Science, University of Zurich, Zurich, Switzerland

*For correspondence: caron@imsb.biol.ethz.ch (EC); aebersold@imsb.biol.ethz.ch (RA)

Competing interests: The authors declare that no competing interests exist.


Funding: See page 13

Received: 23 March 2015

Accepted: 07 July 2015

Published: 08 July 2015

Reviewing editor: Arup K Chakraborty, Massachusetts Institute of Technology, United States

 Copyright Caron et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Abstract We present a novel mass spectrometry-based high-throughput workflow and an open-source computational and data resource to reproducibly identify and quantify HLA-associated peptides. Collectively, the resources support the generation of HLA allele-specific peptide assay libraries consisting of consensus fragment ion spectra, and the analysis of quantitative digital maps of HLA peptidomes generated from a range of biological sources by SWATH mass spectrometry (MS). This study represents the first community-based effort to develop a robust platform for the reproducible and quantitative measurement of the entire repertoire of peptides presented by HLA molecules, an essential step towards the design of efficient immunotherapies.

DOI: [10.7554/eLife.07661.001](https://doi.org/10.7554/eLife.07661.001)

Introduction

Next-generation immune-based therapies are expected to facilitate the eradication of intractable pathogens, cancer and autoimmune diseases (Koff et al., 2013). T cells play a critical role in such therapies by their ability to detect the presence of disease-specific antigens/peptides presented by major histocompatibility complex (MHC) molecules (human leukocyte antigen [HLA] molecules in humans). Under steady-state or pathological conditions, thousands of HLA class I-associated peptides of 8–12 amino acids in length are displayed on the surface of virtually all nucleated cells for scrutiny by

eLife digest The cells of the immune system protect us by recognizing telltale molecules produced by damaged and diseased cells, or by infection-causing microorganisms (which are also called pathogens). To help with this process, the cells in our bodies display small fragments of proteins (called peptides) on their surface that are then checked by the immune cells. Collectively, these peptides are referred to as the 'immunopeptidome', and deciphering the complexity of the human immunopeptidome is important for both basic research and medical science. Such an achievement would help to guide the development of next-generation vaccines and therapies against autoimmune disorders, infectious diseases and cancers.

In the past, immune peptides were mostly identified using a technique that is commonly called 'shotgun' mass spectrometry. However, this approach doesn't always provide reproducible results. In 2012, researchers reported the development of a new approach—which they called 'SWATH' mass spectrometry—that could yield more reproducible data.

Now, Caron et al.—including many of the researchers involved in the 2012 study—have developed a large collection of standardized tests that use SWATH mass spectrometry to analyze the human immunopeptidome. The workflow and the computational and data resources developed as part of this international effort are the first steps toward highly reproducible and measurable analyses of the immunopeptidome across many samples. Moreover, the large repository of assays generated by the project has been made public and will serve a large community of researchers, which should enable better collaborations.

In the future, SWATH mass spectrometry could be used as a robust technology for the reproducible detection and measurement of pathogen-specific or cancer-specific immune peptides. This could greatly help in the design of personalized immune-based therapies.

DOI: [10.7554/eLife.07661.002](https://doi.org/10.7554/eLife.07661.002)

CD8+ T cells. HLA class II-associated peptides are 10–25 amino acids in length and are normally found on the surface of specialized antigen-presenting cells including macrophages and dendritic cells for presentation to CD4+ T cells. Collectively, HLA class I and class II peptides are referred to as the immunopeptidome, also known as HLA ligandome/peptidome (Caron et al., 2011; Kowalewski et al., 2014). The composition of the immunopeptidome in the human population is complicated by the presence of more than 3000 HLA alleles, resulting in a high diversity of peptide repertoires characterized by the presence of HLA allele-specific binding motifs (Falk et al., 1991). To be successful in designing efficient immunotherapies against autoimmunity, cancer and infectious diseases, it is becoming increasingly important to comprehensively map the complexity of the human immunopeptidome and to gain a more quantitative understanding of its dynamics in various disease states.

Mass spectrometry (MS) has evolved as the method of choice for the exploration of the human immunopeptidome (Hunt et al., 1992; Admon and Bassani-Sternberg, 2011; Granados et al., 2015). The largest HLA peptidomes reported to date using MS contain more than 10,000 class I or class II peptides (Hassan et al., 2013; Bergseng et al., 2014; Bassani-Sternberg et al., 2015). Estimates from various analytical and cell-based techniques also indicate that individual peptides are expressed on average at 50 copies per cell with extremes ranging from 1 to 10,000 copies per cell (Granados et al., 2015). Until recently, the most common strategy for the analysis of immunopeptidomes by MS has focused on the isolation of HLA-bound peptides by immunoaffinity chromatography and the collection of fragment ion spectra of selected peptides through automated MS operated in data-dependent acquisition (DDA) mode. Although DDA is a powerful strategy for exploring the peptidomic content of various cell and tissue types, it is not a reliable platform for solving problems that require the comparison of comprehensive, quantitative, and reproducible data sets across many samples or conditions. In fact, analyses of complex/unfractionated digests of cell lysate using DDA have shown that as many as 84% of peptides may remain unselected for fragmentation even though they are clearly detectable by the mass spectrometer (Michalski et al., 2011). Although the complexity of isolated HLA peptides is hardly comparable with that of cell lysate digests, as many as 20% of the selected HLA peptides can vary between replicate analyses of the same sample (Granados et al., 2014) (Figure 1—figure supplement 1A). A second strategy, referred

to as selected/multiple reaction monitoring (S/MRM), is a targeting MS technique capable of generating highly reproducible, quantitatively accurate and sensitive datasets (Picotti and Aebersold, 2012). S/MRM is, however, limited by its capacity to detect only tens to hundreds of peptides per sample injection and thus is not ideally suited to comprehensively quantify HLA peptidomes. To overcome this limitation, we recently introduced SWATH-MS, a new mass spectrometric technique that combines data-independent acquisition (DIA) with a targeted data extraction strategy (Gillet et al., 2012; Röst et al., 2014). In DIA mode, all peptides in a sample are fragmented and the corresponding fragment ion spectra are acquired, resulting in a digital recording of the peptide sample. DIA is an unbiased MS technique and therefore represents a suitable strategy for efficiently generating consistent, reproducible and quantitatively accurate measurements of peptides across multiple samples (Gillet et al., 2012; Collins et al., 2013; Rosenberger et al., 2014; Röst et al., 2014; Guo et al., 2015; Liu et al., 2015; Schubert et al., 2015a).

To extract quantitative information from digital SWATH-MS data, high-quality assay libraries are required. Such libraries contain retention-time and fragmentation information of the peptides to be targeted. Assay libraries are generated from native and/or synthetic peptides using a SWATH compatible mass spectrometer operated in DDA mode. To date, several generic SWATH assay libraries were generated for the analysis of proteomes in various species. These include *Mycobacterium tuberculosis* (Schubert et al., 2015a), *Saccharomyces cerevisiae* (Selevsek et al., 2015), and *Homo sapiens* (Rosenberger et al., 2014). Assay libraries were successfully employed to measure a limited number of MHC class I peptides by S/MRM in various contexts—that is, viral infection (Croft et al., 2013), autoimmunity (Schittenhelm et al., 2014a) and cancer (Gubin et al., 2014)—but have never been created for robust quantitative and high-throughput measurement of HLA-associated peptides by SWATH-MS.

For the SWATH-MS technology to meet its potential to support rapid advances in the design of next-generation vaccines and immunotherapies, comprehensive HLA peptide assay libraries have to be created and made readily available to basic and translational scientists. Generating such assay libraries could ultimately enable the fast and reproducible quantification of the entire repertoire of HLA peptides across many samples. Towards this end, we developed a workflow to (1) generate a pilot repository of HLA allele-specific peptide spectral and assay libraries, and to (2) analyze SWATH-MS HLA peptidomic data acquired from multiple international laboratories (Figure 1). In this study, libraries were created from natural and/or synthetic HLA class I and II peptides whereas analysis of SWATH-MS HLA peptidomic data focused mainly on naturally presented class I peptides.

Results and discussion

Large-scale DDA-based identification of immunoaffinity purified HLA class I peptides is supported by several software tools (e.g., MaxQuant, Perseus or X-PRESIDENT) and results in thousands of unclassified peptides of various lengths. Since large HLA peptidomic datasets are generated at an increasing pace, additional computational frameworks facilitating the HLA annotation and storage of such datasets need to be developed. Here, we first created a computational workflow to support the identification, classification/annotation, visualization and storage of HLA peptidomic data in an allele-dependent manner. The software tools described in the section below enable (1) systematic annotation of peptides to their respective HLA allele, (2) visualization of HLA peptidomic datasets, and (3) generation of HLA class I allele-specific peptide spectral libraries, which can be converted into high quality assay libraries for the processing of SWATH-data (Figure 2, Figure 2—figure supplement 1, Figure 2—source data 2 and Supplementary file 1).

To test our workflow, the generated data and computational resources, we first assessed the feasibility of generating HLA class I allele-specific peptide spectral libraries from a panel of fourteen PBMC samples (PBMC #1–14) expressing different combinations of HLA class I alleles. HLA class I-bound peptides were isolated from HLA-typed PBMC's by immunoaffinity chromatography and analyzed by DDA on an Orbitrap-XL mass spectrometer (Figure 2 and Figure 2—source data 1). Peptides were identified using multiple open-source database search engines. The search identifications were combined and statistically scored using PeptideProphet and iProphet within the Trans-Proteomic Pipeline (TPP) as shown previously (Figure 1) (Shteynberg et al., 2011, 2013). We next annotated the identified peptides to their respective HLA allele. Previously, HLA binding prediction algorithms such as SYFPETHI, NetMHC and SMM were used for manual or semi-automated

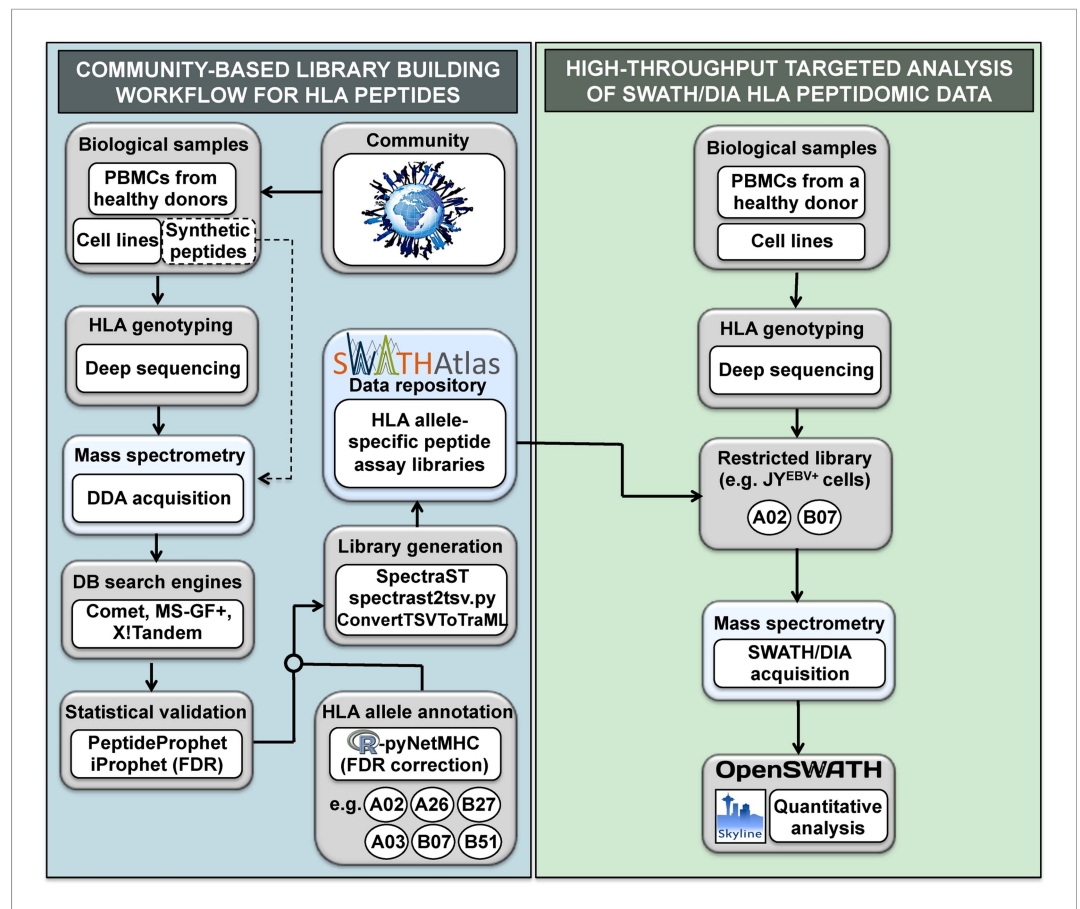


Figure 1. General workflow for building HLA allele-specific peptide assay libraries and for analyzing SWATH-MS HLA peptidomic data. (Left panel) A community-based repository of HLA class I allele-specific peptide spectral and assay libraries was created and stored in the SWATHAtlas database. HLA typed-biological samples and synthetic HLA peptides were used to build the repository. Our workflow integrates (1) data-dependent acquisition (DDA) of HLA peptidomic data, (2) multiple open-source database search engines and statistical validation tools, (3) HLA allele annotation of the identified peptides, and (4) spectral and assay library generation tools. (Right panel) HLA peptidomic data from HLA-typed biological samples were acquired in data-independent acquisition (DIA) mode. The matching HLA class I allele-specific peptide assay libraries were combined and DIA data were analyzed using the OpenSWATH and the Skyline software.

DOI: [10.7554/eLife.07661.003](https://doi.org/10.7554/eLife.07661.003)

The following source data and figure supplements are available for figure 1:

Source data 1. Comparative analysis of DDA and SWATH-MS for the identification of HLA class I peptides.

DOI: [10.7554/eLife.07661.004](https://doi.org/10.7554/eLife.07661.004)

Figure supplement 1. Reproducibility of DDA and SWATH-MS for the identification of HLA class I peptides.

DOI: [10.7554/eLife.07661.005](https://doi.org/10.7554/eLife.07661.005)

Figure supplement 2. Combining results of three open-source database search engines in immunopeptidomics using iProphet.

DOI: [10.7554/eLife.07661.006](https://doi.org/10.7554/eLife.07661.006)

Figure supplement 3. Combining both open-source and commercial database search engines in immunopeptidomics.

DOI: [10.7554/eLife.07661.007](https://doi.org/10.7554/eLife.07661.007)

annotation of HLA peptides (Fortier et al., 2008; Berlin et al., 2014; Granados et al., 2014). Here, we designed a fully automated annotation strategy integrating the stand-alone software package of the HLA binding prediction algorithm NetMHC 3.4 with a set of in-house software tools (Figure 2—figure supplement 1). The in-house software tools enable an automated, consistent and effective annotation of the majority of the identified peptides to their respective HLA allele

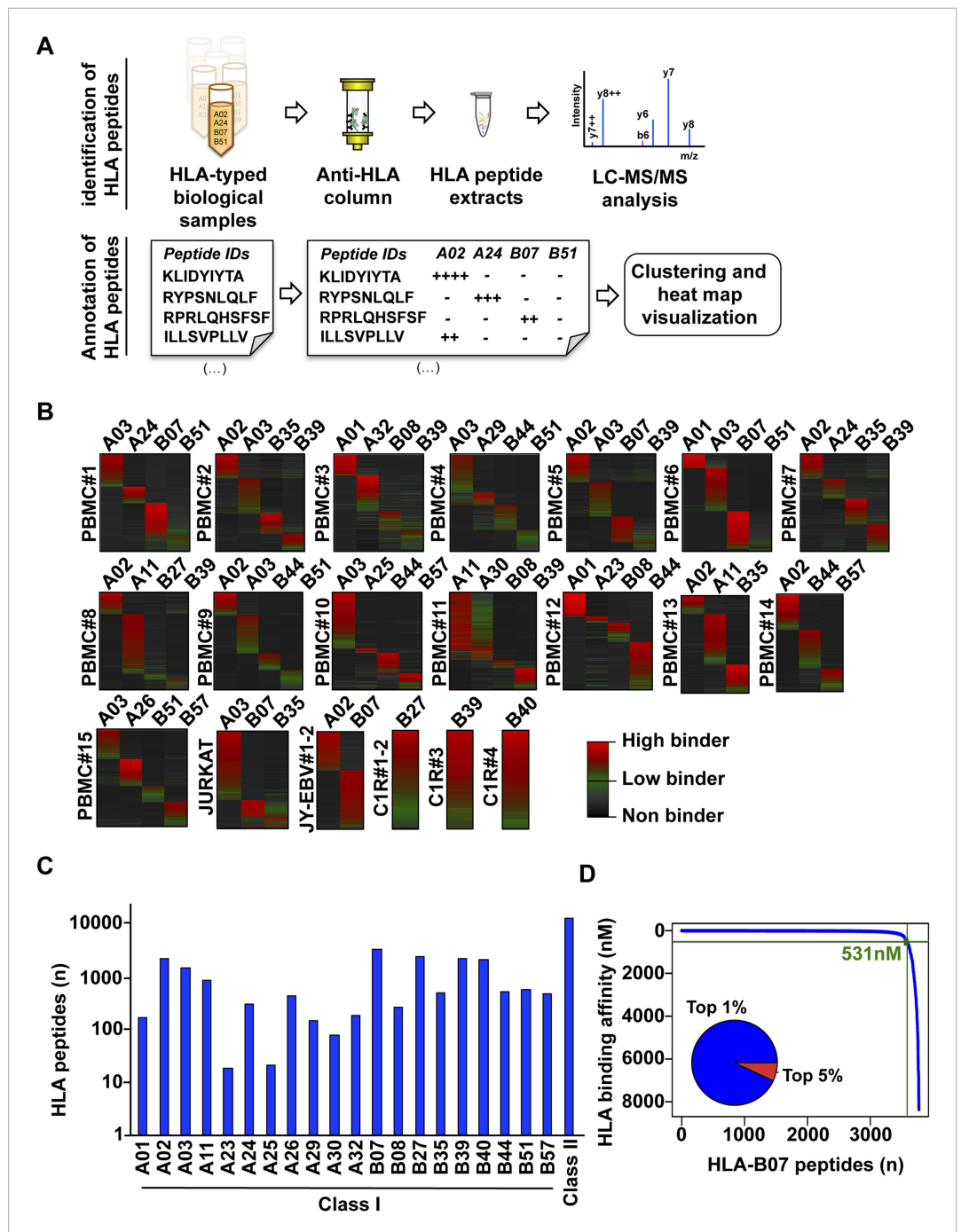


Figure 2. Content and analysis of the pilot repository. **(A)** HLA peptides were isolated by immunoaffinity chromatography and were annotated to their respective HLA alleles following DDA mass spectrometry. **(B)** Heat map visualization of HLA class I peptides identified from 20 HLA-typed biological samples. HLA-A and -B alleles are indicated for each sample. **(C)** 35,812 distinct class I and class II HLA peptides were identified, annotated, and used to build 32 and 11 HLA allele-specific peptide spectral and SWATH assay libraries, respectively. **(D)** The distribution curve shows that 95% of the HLA-B07-annotated peptides were predicted to bind the HLA molecule with an IC50 below 531 nM. Inner pie chart: we assessed the predicted HLA binding affinity of all peptides contained in individual source proteins. The pie chart shows that 92% of naturally presented HLA-B07 peptides were ranked in the top 1% (blue) of predicted peptides (see also *Figure 2—figure supplement 6*).

DOI: [10.7554/eLife.07661.008](https://doi.org/10.7554/eLife.07661.008)

Figure 2. continued on next page

Figure 2. Continued

The following source data and figure supplements are available for figure 2:

Source data 1. Sources of HLA peptides used in this study.

DOI: [10.7554/eLife.07661.009](https://doi.org/10.7554/eLife.07661.009)

Source data 2. Annotation of HLA peptides.

DOI: [10.7554/eLife.07661.010](https://doi.org/10.7554/eLife.07661.010)

Source data 3. List of eluted HLA class I peptides that were identified at 1% and 5% peptide-level FDR.

DOI: [10.7554/eLife.07661.011](https://doi.org/10.7554/eLife.07661.011)

Source data 4. HLA class I allele-specific peptide spectral libraries stored in PeptideAtlas.

DOI: [10.7554/eLife.07661.012](https://doi.org/10.7554/eLife.07661.012)

Source data 5. HLA class I and II allele-specific peptide assay libraries stored in the SWATHAtlas database.

DOI: [10.7554/eLife.07661.013](https://doi.org/10.7554/eLife.07661.013)

Figure supplement 1. Automated NetMHC-based method for annotating and visualizing HLA allele-specific peptides.

DOI: [10.7554/eLife.07661.014](https://doi.org/10.7554/eLife.07661.014)

Figure supplement 2. Identification of HLA class I allele-specific peptides by DDA.

DOI: [10.7554/eLife.07661.015](https://doi.org/10.7554/eLife.07661.015)

Figure supplement 3. Generation of assay libraries from a large collection of synthetic HLA class II peptides.

DOI: [10.7554/eLife.07661.016](https://doi.org/10.7554/eLife.07661.016)

Figure supplement 4. Distribution curves of peptide binding affinities for different HLA-A and -B alleles (1% peptide-level FDR; 0.5% cFDR).

DOI: [10.7554/eLife.07661.017](https://doi.org/10.7554/eLife.07661.017)

Figure supplement 5. Distribution curves of peptide binding affinities for different HLA-A and -B alleles (5% peptide-level FDR; 2.5% cFDR).

DOI: [10.7554/eLife.07661.018](https://doi.org/10.7554/eLife.07661.018)

Figure supplement 6. Binding scores of naturally presented HLA-A and -B peptides contained in individual source proteins.

DOI: [10.7554/eLife.07661.019](https://doi.org/10.7554/eLife.07661.019)

(**Supplementary file 1**). Briefly, each identified peptide was given a predicted HLA binding affinity (IC_{50}) for each of the HLA alleles expressed in the corresponding healthy donor. An HLA annotation score was then computed for each individual peptide by dividing its second best IC_{50} value (i.e., the second best predicted allele) by its best IC_{50} value (i.e., the best predicted allele). The higher this annotation score was, the higher the probability was for the peptide to be correctly annotated to a specific HLA allele. As an example, in PMBC#2, an annotation score of 77 was computed for the KLEEQARAK peptide by dividing 21,400 nM (second best IC_{50} value predicted for HLA-B39) by 278 nM (best IC_{50} value predicted for HLA-A03) (**Figure 2—figure supplement 1A**). Peptides with an HLA annotation score ≥ 3 (selected cutoff value; see ‘Materials and methods’ and **Supplementary file 1**) were systematically annotated to the allele predicted to bind best (e.g., HLA-A03 for the KLEEQARAK peptide). Using this scoring strategy, ~80% of all identified 8–12-mers were annotated to a specific HLA-A or -B allele (**Figure 2—source data 2**). HLA-A and -B alleles were prioritized due to the high reliability of the NetMHC 3.4 predictor for a broad diversity of HLA-A and -B alleles as well as for their high expression levels (*Kim et al., 2014*; *Bassani-Sternberg et al., 2015*; *Trolle et al., 2015*). Peptides with an annotation score below 3 were considered as non-annotated in this study and were discarded for the process of building the HLA allele-specific peptide spectral libraries. Tables including scored peptides were then used to generate heat maps and visualize HLA-A and -B peptidomes of PBMC’s as described (**Figure 2B** and **Supplementary file 1**). Of note, allele-supertype peptides (i.e., peptides predicted to strongly bind more than one allele with an IC_{50} below 500 nM) were curated in the output files but were not visualized on the heat maps in this study. A corrected false discovery rate (cFDR) was estimated for each PBMC sample following removal of all non-annotated contaminant peptides (**Figure 1—figure supplement 2** and **Figure 1—figure supplement 3**), resulting in a total of 4153 (peptide-level FDR 1%; average cFDR 0.5%) or 7921 (peptide-level FDR 5%; average cFDR 2.5%) distinct annotated peptides distributed across eighteen HLA class I alleles (**Figure 2—figure supplement 2A** and **Figure 2—source data 3**). All annotated peptides identified from the 14 PBMC samples were then used in SpectraST (*Lam et al., 2008*) to build

the HLA class I allele-specific peptide spectral libraries ('Materials and methods'). The same procedure was applied to peptides identified from JY^{EBV+} and C1R cells. Notably, endogenous HLA-C04 peptides were recently shown to be significantly expressed on the surface of C1R cells (*Schittenhelm et al., 2014b*) and were therefore considered in this study. In total, 3528 HLA-A peptides, 4208 HLA-B peptides and 205 HLA-C04 peptides were recorded in the spectral libraries, which were then stored in the public PeptideAtlas database (*Figure 2—source data 4*). In summary, we generated a computational workflow to effectively annotate and visualize HLA peptidomic data, which were finally converted and stored into HLA allele-specific peptide spectral libraries consisting of consensus fragment ion spectra. This strategy could be further refined to collect, store and share HLA peptidomic information obtained from various cell lines and from larger cohorts of donors. Importantly, this computational approach can be broadly applied to generate SWATH-compatible assay libraries as described below.

Libraries of consensus fragment ion spectra can be converted into high quality assays for high-throughput targeted analysis of SWATH-MS data, an emerging approach for reproducible, consistent and accurate quantitative measurements of peptides (*Gillet et al., 2012; Collins et al., 2013; Rosenberger et al., 2014; Röst et al., 2014; Guo et al., 2015; Liu et al., 2015; Selevsek et al., 2015; Schubert et al., 2015a*). Here, we aimed at initiating a worldwide community-based effort to generate pilot HLA allele-specific peptide assay libraries that could be further used for the analysis of SWATH-MS HLA peptidomic data. Naturally presented and/or synthetic HLA class I and class II peptides were provided from six independent laboratories and were analyzed using four distinct TripleTOF 5600 MS instruments operated in DDA acquisition mode in four different international institutions. Naturally presented HLA class I peptides from JY^{EBV+} (HLA-A02 and -B07), PBMC (HLA-A03, -A26, -B51 and -B57), and Jurkat (HLA-A03, -B07 and -B35) cells were isolated by immunoaffinity chromatography (*Figure 2—source data 1*). Natural class I peptides from three C1R cell lines—stably expressing HLA-C04 as well as HLA-B27, -B39 or -B40 molecules—were also isolated using the same procedure. Synthetic EBV-derived peptides known to bind HLA-A02 or -B07 were also used to build the libraries (*Figure 2—source data 2*). All laboratories used the spiked-in landmark iRT peptides for retention time normalization (*Escher et al., 2012*). The DDA data generated by the different groups were shared and pipelined through the computational workflow described above, resulting in the identification of 7668 (peptide-level FDR 1%; average cFDR 0.5%) or 11,275 (peptide-level FDR 5%; average cFDR 2.5%) distinct HLA class I peptides distributed across eleven different HLA class I alleles (*Figure 2—figure supplement 2B* and *Figure 2—source data 3*). To properly assess the efficiency of generating HLA peptide assay libraries from synthetic peptides, a large collection of 20,176 synthetic HLA class II peptides was analyzed by DDA using different mass spectrometers and fragmentation methods (*Figure 2—figure supplement 3* and *Figure 2—source data 2*). Our results show that a total of 15,875 peptides (~79%) were identified (*Figure 2—source data 2*). A large collection of synthetic HLA class I peptides was not available but could be used in the future to extend the contents of the present class I libraries derived from native peptides. All identified peptides were used to build the HLA allele-specific peptide assay libraries ('Materials and methods'). To date, the pilot libraries contain a total of 223,735 transitions for 26,857 unique peptides and were stored by class and allele in the SWATHAtlas database (*Figure 2—source data 5* and <http://www.swathatlas.org>). By using the automated HLA peptide annotation method described above, we observed that similar binding affinities were predicted for HLA class I peptides identified at peptide-level FDR 1% and peptide-level FDR 5% (*Figure 2—figure supplement 4* and *Figure 2—figure supplement 5*), suggesting that a large fraction of true positives were excluded at peptide-level FDR 1%. Our data also show that 95% of the annotated class I peptides in this study were predicted to bind their respective HLA molecules with an IC₅₀ ranging from 72 nM (for HLA-A01) to 5682 nM (for HLA-B51) at peptide-level FDR 1% (*Figure 2—figure supplement 4*). Similar results were obtained at peptide-level FDR 5% (*Figure 2—figure supplement 5*). This result supports a recent study indicating that HLA class I alleles are associated with peptide-binding repertoires of different affinity (*Paul et al., 2013*). Altogether, we demonstrated the feasibility of collecting DDA data from multiple international laboratories to generate standardized HLA allele-specific peptide assay libraries. We anticipate this global effort as a first step towards the development of a standardized Pan-human HLA peptide assay library, which could be used to rapidly and reproducibly quantify the entire repertoire of peptides presented by HLA molecules using SWATH-MS.

SWATH-MS is emerging as a robust next-generation proteomics technique for efficiently generating reproducible, consistent and quantitatively accurate measurements of peptides across multiple samples (Gillet *et al.*, 2012; Collins *et al.*, 2013; Rosenberger *et al.*, 2014; Röst *et al.*, 2014; Guo *et al.*, 2015; Liu *et al.*, 2015; Selevsek *et al.*, 2015; Schubert *et al.*, 2015a). To promote the worldwide development of SWATH-based MS platforms towards robust quantitative measurements of HLA peptidomes, we assessed whether the HLA allele-specific assay libraries described above could be used to extract quantitative information from digital SWATH maps acquired by different laboratories. Importantly, four independent laboratories generated their own digital SWATH maps using TripleTOF 5600 MS operated in DIA acquisition mode. Naturally presented HLA class I peptides were isolated from the cell types mentioned above (i.e., JY^{EBV+}, Jurkat, PBMC and C1R). Precursors in the range of 400–1200 Th were divided into 32 SWATH windows of 25 Da (Gillet *et al.*, 2012). All ionized peptide precursors in this mass range were fragmented, generating comprehensive and quantitative digital fragment ion maps. The HLA peptidome of JY^{EBV+} cells was analyzed using the OpenSWATH (Röst *et al.*, 2014) software tool and a combined assay library containing 22,206 transitions for 1507 HLA-A02 and 2194 HLA-B07 peptides—the two dominant HLA alleles expressed on these cells. At an estimated peptide-level FDR of 1% (m-score < 0.01), a total of 3150 unique HLA class I peptides were identified from the digital SWATH map (Figure 3A,B,C, Figure 3—figure supplement 1A,B, Figure 3—figure supplement 7 and Figure 3—source data 1). Notably, assays generated from the synthetic EBV-related class I peptides enabled the identification of one EBV-derived HLA-A02 peptide (Figure 3C), thereby demonstrating that building high-quality assay libraries from synthetic class I peptides of pathogen origin could be useful for the identification of non-self HLA-bound peptides by SWATH-MS. To analyze self-HLA peptides isolated from PBMC (HLA-A03, -A26, -B51 and -B57), Jurkat (HLA-A03, B07 and -B35), C1R-B27 (HLA-B27) and C1R-B40 (HLA-B40) cells, the matching HLA class I allele-specific peptide assay libraries were combined accordingly using SpectraST and then processed in the OpenSWATH software. High-throughput targeted analysis from these four additional peptidomic datasets indicated that ~81% of HLA class I peptides present in an assay library could be extracted from a quantitative digital SWATH map in a cell type-independent manner (peptide-level FDR 1%) (Figure 3—figure supplement 1C, Figure 3—figure supplements 2–6 and Figure 3—source data 1). We next optimized the SWATH acquisition conditions according to the size distribution of HLA class I peptides. Most class I peptide precursors (~98%) fall within the range of 400–700 Th and were divided in 30 SWATH windows of 10 Da width each. Using SWATH data generated from JY^{EBV+} cells, we found that narrowing the size of the windows by 2.5-fold resulted in a ~13% fold-increase in the identification of class I peptides (Figure 3—figure supplement 1A). The R^2 value for SWATH-MS quantification was 0.979 from two technical replicates (Figure 3D). In accordance with previous studies, we also observed that the dynamic range of peptides quantified in different cell types using SWATH-MS, based on their signal intensity, was about 3–4 orders of magnitude (Figure 3E) (Hassan *et al.*, 2013; Bassani-Sternberg *et al.*, 2015). Altogether, we demonstrate the feasibility of an international effort to build standardized HLA allele-specific peptide assay libraries, which were used to extract quantitative information from digital SWATH maps acquired in different sites. We therefore provide a proof of concept that acquisition of SWATH-MS HLA peptidomic data may enable robust analysis of the human immunopeptidome on a global scale.

To further establish the robustness of SWATH-MS for the measurement of HLA-associated peptides, we tested whether the JY^{EBV+} HLA peptidome could be reproducibly detected across multiple MS injections. For this purpose, we prepared a sample of class I peptides by immunoaffinity purification from JY^{EBV+} cells and we acquired three datasets in SWATH mode. The datasets were analyzed using OpenSWATH and a combined HLA-A02 and -B07 peptide assay library as described above. At an estimated peptide-level FDR of 1%, a total of 2933 unique HLA class I peptides were identified by SWATH-MS and 2832 peptides (97%) were found in all the SWATH analyses (Figure 1—figure supplement 1B, Figure 1—source data 1). We then conducted a comparative analysis by acquiring three additional datasets in DDA mode from the same sample of class I peptides using the same chromatographic conditions. In total, 3153 HLA-A and -B peptides were identified at 1% peptide-level FDR and 1261 peptides (40%) were found in all the DDA analyses (Figure 1—figure supplement 1A, Figure 1—source data 1). Thus, the SWATH method clearly outperformed the DDA approach for the reproducible identification of JY^{EBV+} HLA class I peptides across several technical replicates. Overall, our results indicate that SWATH-MS has the capability of detecting large numbers of HLA peptides across multiple injections at a high degree of reproducibility. By providing

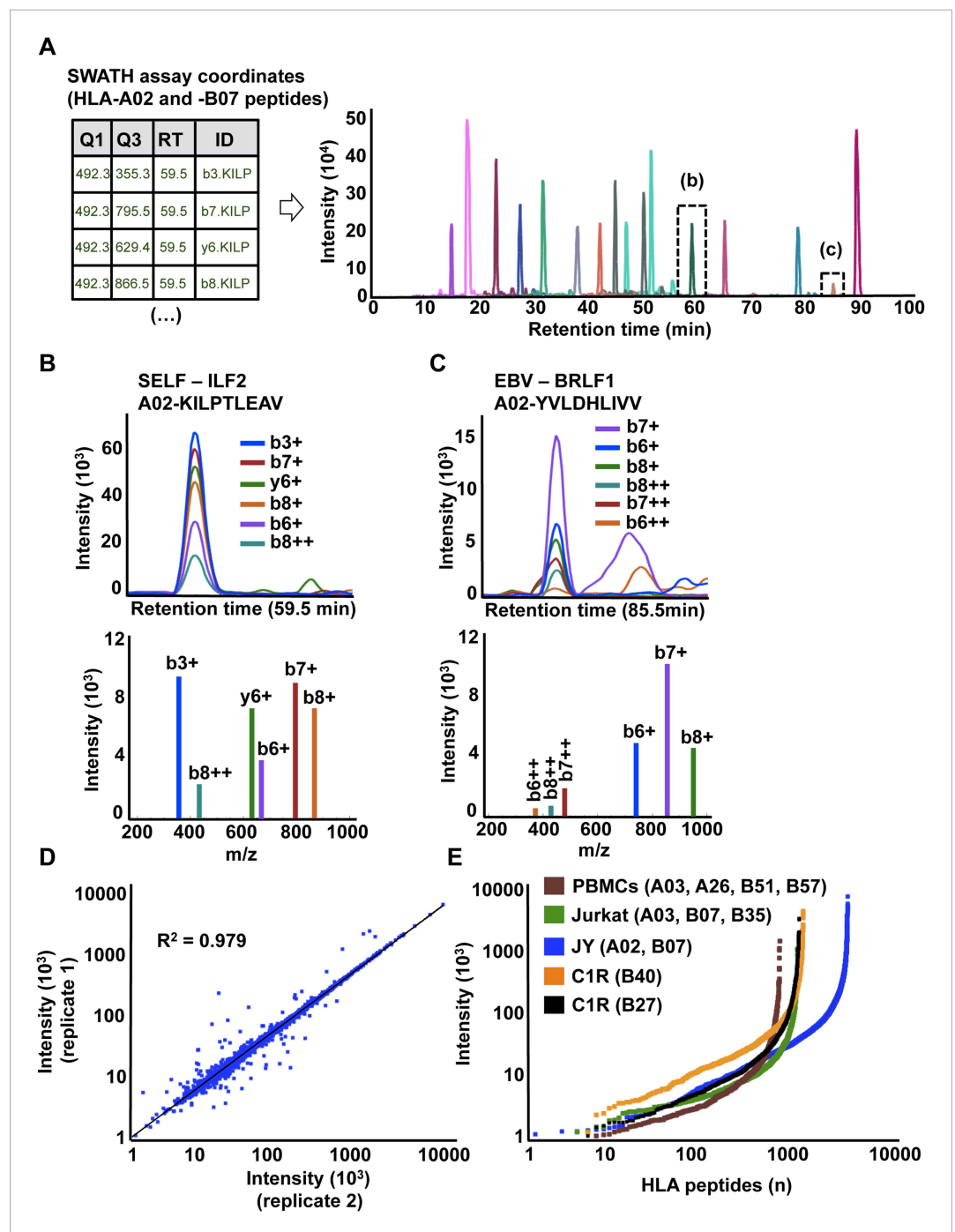


Figure 3. High-throughput targeted analysis of HLA peptidomic data by SWATH-MS. **(A)** SWATH-MS coordinates of two HLA class I allele-specific assay libraries (HLA-A02 and -B07) were combined to extract SWATH data generated from the HLA peptidome of JY^{EBV+} cells. Sixteen summed transition groups are shown here for simplicity. **(B, C)** Visualization of two extracted SWATH transition groups corresponding to the self-HLA-A02 peptide, KILPTLEAV and the non-self HLA-A02 EBV peptide, YVLDHLIVV. **(D)** Reproducibility of intensity measurements for technical replicates. **(E)** Dynamic range of transition group intensities following targeted analysis of SWATH-MS HLA peptidomic data generated from various cell types expressing different combinations of HLA alleles. SWATH/DIA data were acquired in four independent international laboratories.

DOI: [10.7554/eLife.07661.020](https://doi.org/10.7554/eLife.07661.020)

The following source data and figure supplements are available for figure 3:

Source data 1. OpenSWATH analysis.

DOI: [10.7554/eLife.07661.021](https://doi.org/10.7554/eLife.07661.021)

Figure 3. continued on next page

Figure 3. Continued

Figure supplement 1. OpenSWATH analysis of HLA peptidomic data.

DOI: [10.7554/eLife.07661.022](https://doi.org/10.7554/eLife.07661.022)

Figure supplement 2. OpenSWATH analysis and PyProphet statistics of HLA peptidomic data acquired at ETH Zurich, Switzerland.

DOI: [10.7554/eLife.07661.023](https://doi.org/10.7554/eLife.07661.023)

Figure supplement 3. OpenSWATH analysis and PyProphet statistics of HLA peptidomic data acquired at ETH Zurich, Switzerland.

DOI: [10.7554/eLife.07661.024](https://doi.org/10.7554/eLife.07661.024)

Figure supplement 4. OpenSWATH analysis and PyProphet statistics of HLA peptidomic data acquired at University of Oxford, UK.

DOI: [10.7554/eLife.07661.025](https://doi.org/10.7554/eLife.07661.025)

Figure supplement 5. OpenSWATH analysis and PyProphet statistics of HLA peptidomic data acquired at Monash University, Australia.

DOI: [10.7554/eLife.07661.026](https://doi.org/10.7554/eLife.07661.026)

Figure supplement 6. OpenSWATH analysis and PyProphet statistics of HLA peptidomic data acquired at Centro Nacional de Biotecnología, Madrid, Spain.

DOI: [10.7554/eLife.07661.027](https://doi.org/10.7554/eLife.07661.027)

Figure supplement 7. Visualization and analysis of SWATH-MS HLA peptidomic data in Skyline.

DOI: [10.7554/eLife.07661.028](https://doi.org/10.7554/eLife.07661.028)

a community resource for the continuous expansion of the library contents and by improving the performance of the OpenSWATH software, it can be expected that additional HLA peptides—including cryptic and mutant peptides—will be reproducibly identified and quantified from the same digital SWATH maps in the future.

The life sciences community greatly benefits from robust technologies such as microarrays and RNA-seq. Similarly, robust generation and analysis of quantitative digital maps of HLA peptidomes is expected to have important implications in basic and translational research as these will allow research groups to accurately investigate the dynamics of immunopeptidomes in various immune-related diseases such as autoimmunity, infectious diseases and cancers. For instance, reproducible digital mapping of tumor-specific mutant HLA peptides during cancer progression will facilitate stratification of patients who might best benefit from innovative immunotherapeutic interventions (*Gubin et al., 2014; Snyder et al., 2014; Schumacher et al., 2015*). The workflow and the computational and data resources presented in this community-based study is a first step towards highly reproducible and quantitative MS-based measurements of HLA peptidomes across many samples and could therefore be greatly beneficial in the design of personalized immune-based therapies. Moreover, the storage of HLA peptide spectral and assay libraries by class and allele in the SWATHAtlas database provides an initial framework to collect, organize and share HLA peptidomic data, thereby supporting the recently proposed Human Immunopeptidome and Vaccines Projects (*Admon and Bassani-Sternberg, 2011; Koff et al., 2014*).

Materials and methods

Blood samples, cell lines and synthetic peptides

PBMCs from healthy donors were isolated by density gradient centrifugation. Informed consent was obtained in accordance with the Declaration of Helsinki protocol. HLA typing was carried out by the Department of Hematology and Oncology, Tübingen, Germany. PBMCs were stored at -80°C until further use. JY^{EBV+}, Jurkat and C1R cells were cultured in RPMI supplemented with 10% fetal bovine serum, 50 IU/ml penicillin, and 50ug/ml streptomycin (Invitrogen, Life Technologies Europe BV, Zug, Switzerland). C1R cells were stably transfected with -B2705, -B3901 and -B4002 constructs, as described previously (*Marcilla et al., 2014; Schittenhelm et al., 2014a*). The EBV peptide was synthesized by Thermo Fischer Scientific (Ulm, Germany). The collection of 20,176 MTB peptides was synthesized by Mimotopes (Victoria, Australia) as described (*Lindestam Arlehamn et al., 2013*).

Isolation of HLA peptides

HLA class I peptide complexes were isolated by standard immunoaffinity purification as described previously using the pan-HLA class I-specific mAb W6/32 (Hunt et al., 1992; Croft et al., 2013; Kowalewski and Stevanovic, 2013; Marcilla et al., 2014).

RT normalization peptides

For the RT normalization and analysis, the peptides from the iRT Kit (Biognosys AG, Schlieren, Switzerland) were added to samples (see Figure 2—source data 1) prior to MS injection according to vendor instructions (Escher et al., 2012).

DDA mass spectrometry

AB SCIEX TripleTOF 5600⁺

Both naturally presented and synthetic HLA peptides were analyzed using a TripleTOF system (see Figure 2—source data 1) as described before (Gillet et al., 2012; Röst et al., 2014). Samples were analyzed on an Eksigent nanoLC (AS-2/1Dplus or AS-2/2Dplus) system coupled with a SWATH-MS-enabled AB SCIEX TripleTOF 5600⁺ System. The HPLC solvent system consisted of buffer A (2% acetonitrile and 0.1% formic acid in water) and buffer B (2% water with 0.1% formic acid in acetonitrile). The samples were separated in a 75 μm -diameter PicoTip emitter (New Objective, Woburn, MA) packed with 20 cm of Magic 3 μm , 200 Å C18 AQ material (Bischoff Chromatography, Leonberg, Germany). The loaded material was eluted from the column at a flow rate of 300 nl/min with the following gradient: linear 2–35% B over 120 min, linear 35–90% B for 1 min, isocratic 90% B for 4 min, linear 90–2% B for 1 min and isocratic 2% solvent B for 9 min. The mass spectrometer was operated in DDA top20 mode, with 500 and 150 ms acquisition time for the MS1 and MS2 scans respectively, and 20 s dynamic exclusion. Rolling collision energy with a collision energy spread of 15 eV was used for fragmentation.

Thermo scientific orbitrap ELITE

Mtb synthetic peptides were analyzed on an Eksigent LC system coupled to an LTQ-Orbitrap ELITE mass spectrometer. Peptides were separated on a custom C18 reversed phase column (150 mm i.d. \times 100 mm, Jupiter Proteo 4 mm, Phenomenex) using a flow rate of 600 nl min⁻¹ and a linear gradient of 3–60% aqueous ACN (0.2% formic acid) in 120 min. Full mass spectra were acquired with the Orbitrap analyser operated at a resolving power of 30,000 (at m/z 400). Mass calibration used an internal lock mass (protonated (Si(CH₃)₂O)₆; m/z 445.120029) and mass accuracy of peptide measurements was within 5 p.p.m. MS/MS spectra were acquired in CID and HCD mode with a normalized collision energy of 35%. Up to ten precursor ions were accumulated to a target value of 50,000 with a maximum injection time of 300 ms and fragment ions were transferred to the Orbitrap analyser operating at a resolution of 15,000 at m/z 400.

Thermo scientific orbitrap XL

Naturally presented HLA class I peptides from several PBMC samples (see Figure 2—source data 1) were also analyzed by reversed-phase liquid chromatography (nano-UHPLC, UltiMate 3000 RSLCnano; Thermo Fisher, Waltham, MA, USA) coupled with an LTQ Orbitrap XL hybrid mass spectrometer. Samples were analyzed in five technical replicates. Sample volumes of 5 μl (sample shares of 20%) were injected onto a 75 μm \times 2 cm trapping column (Acclaim PepMap RSLC; Thermo Fisher) at 4 $\mu\text{l}/\text{min}$ for 5.75 min. Peptide separation was subsequently performed at 50°C and a flow rate of 175 nl/min on a 50 μm \times 50 cm separation column (Acclaim PepMap RSLC; Thermo Fisher) applying a gradient ranging from 2.4 to 32.0% of acetonitrile over the course of 140 min. Eluting peptides were ionized by nanospray ionization and analyzed in the mass spectrometer implementing a top five CID method generating fragment spectra for the five most abundant precursor ions in the survey scans. Resolution was set to 60,000. For HLA class I ligands, the mass range was limited to 400–650 m/z with charge states 2 and 3 permitted for fragmentation.

Database search engines and statistical validation

All raw instrument data were centroided and processed as described previously (Collins et al., 2013; Rosenberger et al., 2014). The datasets were searched individually using X!tandem (Craig et al., 2004), MS-GF+ (Kim and Pevzner, 2014) and Comet (Eng et al., 2012) against the full non-redundant, canonical human genome as annotated by the UniProtKB/Swiss-Prot (2014_02) with

20,270 ORFs and appended iRT peptide and decoy sequence. Oxidation (M) was the only variable modification. Parent mass error was set to ± 5 p.p.m., fragment mass error was set to ± 0.5 Da. The search identifications were then combined and statistically scored using PeptideProphet (Keller *et al.*, 2002) and iProphet (Shteynberg *et al.*, 2011) within the TPP (4.7.0) (Keller *et al.*, 2005). All peptides with an iProbability/iProphet score above 0.7 were exported in Excel. Assumed charges were also exported, as this information is needed in SpectraST. Length considered was 8–12 residues for class I HLA peptides. FDR was manually estimated based on the target-decoy approach (Elias and Gygi, 2007). Peptides (1% and 5% peptide-level FDR) were then exported to a .txt file for annotation to their respective HLA allele.

HLA allele annotation

Annotation of the identified peptides (1% and 5% peptide-level FDR) to their respective HLA allele was performed automatically by integrating the stand-alone software package of NetMHC 3.4 (Lundegaard *et al.*, 2008) with our in-house software tools (Supplementary file 1 and Source code 1). An HLA annotation score was computed by the software tools for individual peptides (Figure 2—figure supplement 1). A predefined cutoff score of 3 was then used to annotate each peptide to their respective HLA allele. A cutoff value of 3 was selected because >90% of the identified peptides with an annotation score above 3 have a predicted IC₅₀ below 1000 nM. FDR was corrected from the list of annotated HLA peptides based on the target-decoy approach (Elias and Gygi, 2007). The software tools were used to process and visualize the peptidomic datasets. The final lists of HLA-allele specific peptides were exported into a .txt file and used in SpectraST for library generation.

Generation of HLA allele-specific peptide spectral and assay libraries

This section was adapted from Schubert *et al.* (2015b). The parameters below were used for Spectrast (Lam *et al.*, 2008). Exact meaning of each parameter can be found in the following link: <http://tools.proteomecenter.org/wiki/index.php?title=Software:SpectraST>. Spectrast was used in library generation mode with CID-QTOF settings (-cCID-QTOF) for the Triple-TOF 5600+ or CID (default) settings for the Orbitrap-XL and Orbitrap-ELITE. Retention times were normalized against the iRT Kit peptide sequences (-c_IRTiRT.txt -c_IRR). Only HLA-allele specific peptide ions were included for library generation (-cT):

```
spectrast -cNSpecLib_celltype_allele_fdr_iRT -cCID-QTOF -cTReference_celltype_allele_fdr.txt -cP0.7 -c_IRTiRT.txt -c_IRR iprophet.pep.xml
```

A consensus library was then generated:

```
spectrast -cNSpecLib_cons_celltype_allele_fdr_iRT -cCID-QTOF -cAC SpecLib_celltype_allele_fdr_iRT.splib
```

HLA-allele specific consensus libraries were merged:

```
spectrast -cNSpecLib_cons_celltype_alleles_fdr_iRT -cJU -cAC SpecLib_celltype_allele1_fdr_iRT.splib SpecLib_celltype_allele2_fdr_iRT.splib SpecLib_celltype_allele3_fdr_iRT.splib SpecLib_celltype_allele4_fdr_iRT.splib
```

The script spectrast2tsv.py (msproteomicstools 0.2.2; <https://pypi.python.org/pypi/msproteomicstools>) was then used to generate the HLA-allele specific peptide assay library with the following recommended settings:

```
spectrast2tsv.py -l 350,2000 -s b,y -x 1,2 -o 6 -n 6 -p 0.05 -d -e -w swaths.txt -k openswath -a SpecLib_cons_celltype_alleles_fdr_iRT_openswath.csv SpecLib_cons_celltype_alleles_fdr_iRT.sptxt
```

The _openswath.csv file was then converted into a .tsv file and opened in Excel. Reference coordinates for the 11 iRT peptides were confirmed and any remaining decoy sequences were removed. The file was then saved in .txt format and then converted back in .csv format. The OpenSWATH tool ConvertTSVToTraML converted the TSV/CSV file to TraML:

```
ConvertTSVToTraML -in SpecLib_cons_celltype_alleles_fdr_iRT_openswath.csv -out SpecLib_cons_celltype_alleles_fdr_iRT.TraML
```

Decoys were appended to the TraML assay library with the OpenSWATH tool OpenSwathDecoyGenerator as described before (Rosenberger *et al.*, 2014; Röst *et al.*, 2014; Schubert *et al.*, 2015b) in reverse mode with a similarity threshold of 0.05 Da and an identity threshold of 1:

```
OpenSwathDecoyGenerator -in SpecLib_cons_celltype_alleles_fdr_iRT.TraML -out SpecLib_cons_celltype_alleles_fdr_iRT_decoy.TraML -method shuffle -append -exclude_similar
```

The library was then uploaded into the iPortal workflow for SWATH data analysis (see below).

DIA mass spectrometry (SWATH-MS)

For SWATH-MS data acquisition, the same mass spectrometer and LC-MS/MS setup was operated essentially as described before (*Collins et al., 2013; Rosenberger et al., 2014*) using 32 windows of 25 Da effective isolation width (with an additional 1 Da overlap on the left side of the window) and with a dwell time of 100 ms to cover the mass range of 400–1200 m/z in 3.3 s. Before each cycle, an MS1 scan was acquired, and then the MS2 scan cycle started (400–425 m/z precursor isolation window for the first scan, 424–450 m/z for the second... 1,174–1200 m/z for the last scan). The collision energy for each window was set using the collision energy of a 2+ ion centered in the middle of the window with a spread of 15 eV. Four independent international laboratories acquired their own SWATH maps using the settings described above: (1) Antony Purcell, Monash University; (2) Nicola Ternette, University of Oxford; (3) Miguel Marcilla, Spanish National Biotechnology Center; (4) Ruedi Aebersold, ETH-Zurich.

SWATH-MS data analysis

The iPortal workflow was used for data analyses (*Kunszt et al., 2014*). The OpenSWATH analysis workflow (OpenSWATHWorkflow) (<http://www.openswath.org>) was implemented in the iPortal workflow. The parameters were selected analogously to the ones described before (*Röst et al., 2014*): min_rsq: 0.95, min_coverage: 0.6, min_upper_edge_dist: 1, mz_extraction_window: 0.05, rt_extraction_window: 600, extra_rt_extraction_window: 100. pyprophet (<https://pypi.python.org/pypi/pyprophet>) was run on the OpenSwathWorkflow output adjusted to contain the previously described scores (xx_swath_prelim_score, bseries_score, elution_model_fit_score, intensity_score, isotope_correlation_score, isotope_overlap_score, library_corr, library_rmsd, log_sn_score, massdev_score, massdev_score_weighted, norm_rt_score, xcorr_coelution, xcorr_coelution_weighted, xcorr_shape, xcorr_shape_weighted, yseries_score) (*Röst et al., 2014*). Assay libraries were loaded into Skyline and SWATH traces were analyzed as described previously (*Schubert et al., 2015b*). Advanced protocols for analysis of SWATH/DIA data can be downloaded from the website: <http://skyline.maccosslab.org>.

Acknowledgements

We thank Ben Collins, Yansheng Liu, Tatjana Sajic and Olga Schubert for instrument maintenance and for technical support. We thank Emanuel Schmid, Lorenz Blum, Hannes Röst, George Rosenberger and Ulrich Omasits for assistance with the computational analysis. We thank Valeria de Azcoitia for commenting this manuscript as well as all members of the Aebersold laboratory for discussions.

Additional information

Funding

Funder	Grant reference	Author
National Health and Medical Research Council (NHMRC)	1022509 and 1085017	Anthony W Purcell
National Institutes of Health (NIH)	HHSN272201200010C and HHSN272200900044C	Cecilia S Lindestam Arlehamn, Alessandro Sette
European Research Council (ERC)	ERC-2008-AdG_20080422	Ruedi Aebersold
Schweizerische Nationalfonds zur Förderung der Wissenschaftlichen Forschung	3100A0-688 107679	Ruedi Aebersold
European Commission (EC)	SysteMtb, 241587	Ruedi Aebersold
German Cancer Consortium (DKTK)		Daniel J Kowalewski, Heiko Schuster, Hans-Georg Rammensee, Stefan Stevanovic

Funder	Grant reference	Author
Bundesministerium für Bildung und Forschung	e:Bio Express2Present, 0316179C	Pedro Navarro
Forschungszentrum Immuntherapie (FZI)	of the Johannes Gutenberg University Mainz	Pedro Navarro
Ministerio de Economía y Competitividad	Carlos III Health Institute (ISCIII) (ProteoRed-PRB2, PT13/0001)	Miguel Marcilla
European Commission (EC)	Marie Curie Intra-European Fellowship	Etienne Caron
Schweizerische Nationalfonds zur Förderung der Wissenschaftlichen Forschung	Postdoc Mobility Fellowship	Ralf B Schittenhelm
National Institute of General Medical Sciences (NIGMS)	R01GM087221 and 2P50 GM076547/Center for Systems Biology	David S Campbell, Eric W Deutsch, Robert L Moritz

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

EC, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; LE, CCK, LCG, PN, SK, HL, Analysis and interpretation of data, Drafting or revising the article; DJK, HS, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; NT, AA, RBS, SHR, MM, AWP, Acquisition of data, Drafting or revising the article; CSLA, AR, Acquisition of data, Drafting or revising the article, Contributed unpublished essential data or reagents; TS, AS, DSC, Analysis and interpretation of data, Drafting or revising the article, Contributed unpublished essential data or reagents; EWD, RLM, Conception and design, Drafting or revising the article, Contributed unpublished essential data or reagents; H-GR, SS, RA, Conception and design, Drafting or revising the article

Author ORCIDs

Etienne Caron,  <http://orcid.org/0000-0003-2770-6970>

Ethics

Human subjects: Informed consent was obtained in accordance with the Declaration of Helsinki protocol. The study was performed according to the guidelines of the local ethics committee (University of Tübingen, Germany).

Additional files

Supplementary files

- Supplementary file 1. Description of the Python and the R scripts for the automated annotation and visualization of HLA peptidomic data.

DOI: [10.7554/eLife.07661.029](https://doi.org/10.7554/eLife.07661.029)

- Source code 1. Python and R scripts.

DOI: [10.7554/eLife.07661.030](https://doi.org/10.7554/eLife.07661.030)

Major datasets

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset ID and/or URL	Database, license, and accessibility information
Caron <i>et al</i> ,	2015	Mass spectrometry discovery peptidomics data (centroided mzXML and identified peptides in pepXML report) used to generate the HLA-allele specific peptide spectral and assay libraries	http://proteomecentral.proteomexchange.org	Publicly available at the ProteomeXchange (Accession no: PXD001872).

Author(s)	Year	Dataset title	Dataset ID and/or URL	Database, license, and accessibility information
Caron <i>et al</i> ,	2015	HLA allele-specific peptide spectral libraries (SpectraST format) and assay libraries (CSV, TraML) available for different SWATH-MS data analysis tools	www.swathatlas.org	Publicly available at the SWATHAtlas. Additional allele-specific spectral libraries (without RT normalization) are available at the PeptideAtlas (www.peptideatlas.org) with the dataset identifier PASS00666.
Caron <i>et al</i> ,	2015	Mass spectrometry SWATH-MS data (instrument raw/wiff files and identified peptides in OpenSWATH report)	http://proteomecentral.proteomexchange.org	Publicly available at the ProteomeXchange (Accession no: PXD001904).

References

- Admon A, Bassani-Sternberg M.** 2011. The Human Immunopeptidome Project, a suggestion for yet another postgenome next big thing. *Molecular & Cellular Proteomics* **10**:O111.011833. doi: [10.1074/mcp.O111.011833](https://doi.org/10.1074/mcp.O111.011833).
- Bassani-Sternberg M, Pletscher-Frankild S, Jensen LJ, Mann M.** 2015. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Molecular & Cellular Proteomics* **14**:658–673. doi: [10.1074/mcp.M114.042812](https://doi.org/10.1074/mcp.M114.042812).
- Bergsgen E, Dørum S, Arntzen MØ, Nielsen M, Nygård S, Buus S, de Souza GA, Sollid LM.** 2014. Different binding motifs of the celiac disease-associated HLA molecules DQ2.5, DQ2.2, and DQ7.5 revealed by relative quantitative proteomics of endogenous peptide repertoires. *Immunogenetics* **67**:73–84. doi: [10.1007/s00251-014-0819-9](https://doi.org/10.1007/s00251-014-0819-9).
- Berlin C, Kowalewski DJ, Schuster H, Mirza N, Walz S, Handel M, Schmid-Horch B, Salih HR, Kanz L, Rammensee HG, Cacace SS, Sticker JS.** 2014. Mapping the HLA ligandome landscape of acute myeloid leukemia: a targeted approach toward peptide-based immunotherapy. *Leukemia* **29**:647–659. doi: [10.1038/leu.2014.233](https://doi.org/10.1038/leu.2014.233).
- Caron E, Vincent K, Fortier M-H, Laverdure J-P, Bramoullé A, Hardy M-P, Voisin G, Roux PP, Lemieux S, Thibault P, Perreault C.** 2011. The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation. *Molecular Systems Biology* **7**:533. doi: [10.1038/msb.2011.68](https://doi.org/10.1038/msb.2011.68).
- Collins BC, Gillet LC, Rosenberger G, Röst HL, Vichalkovski A, Gstaiger M, Aebersold R.** 2013. Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nature Methods* **10**:1246–1253. doi: [10.1038/nmeth.2703](https://doi.org/10.1038/nmeth.2703).
- Craig R, Cortens JP, Beavis RC.** 2004. Open source system for analyzing, validating, and storing protein identification data. *Journal of Proteome Research* **3**:1234–1242. doi: [10.1021/pr049882h](https://doi.org/10.1021/pr049882h).
- Croft NP, Smith SA, Wong YC, Tan CT, Dudek NL, Flesch IE, Lin LC, Tschärke DC, Purcell AW.** 2013. Kinetics of antigen expression and epitope presentation during virus infection. *PLoS Pathogens* **9**:e1003129. doi: [10.1371/journal.ppat.1003129](https://doi.org/10.1371/journal.ppat.1003129).
- Elias JE, Gygi SP.** 2007. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods* **4**:207–214. doi: [10.1038/nmeth1019](https://doi.org/10.1038/nmeth1019).
- Eng JK, Jahan TA, Hoopmann MR.** 2012. Comet: an open-source MS/MS sequence database search tool. *Proteomics* **13**:22–24. doi: [10.1002/pmic.201200439](https://doi.org/10.1002/pmic.201200439).
- Escher C, Reiter L, MacLean B, Ossola R, Herzog F, Chilton J, MacCoss MJ, Rinner O.** 2012. Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **12**:1111–1121. doi: [10.1002/pmic.201100463](https://doi.org/10.1002/pmic.201100463).
- Falk K, Röttschke O, Stevanovic S, Jung G, Rammensee HG.** 1991. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **351**:290–296. doi: [10.1038/351290a0](https://doi.org/10.1038/351290a0).
- Fortier M-H, Caron E, Hardy M-P, Voisin G, Lemieux S, Perreault C, Thibault P.** 2008. The MHC class I peptide repertoire is molded by the transcriptome. *The Journal of Experimental Medicine* **205**:595–610. doi: [10.1084/jem.20071985](https://doi.org/10.1084/jem.20071985).
- Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R.** 2012. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & Cellular Proteomics* **11**:O111.016717. doi: [10.1074/mcp.O111.016717](https://doi.org/10.1074/mcp.O111.016717).
- Granados DP, Laumont CM, Thibault P, Perreault C.** 2015. The nature of self for T cells—a systems-level perspective. *Current Opinion in Immunology* **34**:1–8. doi: [10.1016/j.coi.2014.10.012](https://doi.org/10.1016/j.coi.2014.10.012).
- Granados DP, Sriranganadane D, Daouda T, Zieger A, Laumont CM, Caron-Lizotte O, Boucher G, Hardy MP, Gendron P, Côté C, Lemieux SEB, Thibault P, Perreault C.** 2014. Impact of genomic polymorphisms on the repertoire of human MHC class I-associated peptides. *Nature Communications* **5**:3600. doi: [10.1038/ncomms4600](https://doi.org/10.1038/ncomms4600).
- Gubin MM, Zhang X, Schuster H, Caron E, Ward JP, Noguchi T, Ivanova Y, Hundal J, Arthur CD, Krebber WJ, Mulder GE, Toebe M, Vesely MD, Lam SS, Korman AJ, Allison JP, Freeman GJ, Sharpe AH, Pearce EL, Schumacher TN, Aebersold R, Rammensee HG, Melief CJ, Mardis ER, Gillanders WE, Artyomov MN, Schreiber RD.** 2014. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* **515**:577–581. doi: [10.1038/nature13988](https://doi.org/10.1038/nature13988).

- Guo T**, Kouvonen P, Koh CC, Gillet LC, Wolski WE, Röst HL, Rosenberger G, Collins BC, Lorenz BC, Gillissen S, Joerger M, Jochum W, Aebersold R. 2015. Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nature Medicine* **21**:407–413. doi: [10.1038/nm.3807](https://doi.org/10.1038/nm.3807).
- Hassan C**, Kester MG, de Ru AH, Hombrink P, Drijfhout JW, Nijveen H, Leunissen JA, Heemskerk MH, Falkenburg JH, van Veelen PA. 2013. The human leukocyte antigen-presented ligandome of B lymphocytes. *Molecular & Cellular Proteomics* **12**:1829–1843. doi: [10.1074/mcp.M112.024810](https://doi.org/10.1074/mcp.M112.024810).
- Hunt DF**, Henderson RA, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, Cox AL, Appella E, Engelhard VH. 1992. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* **255**:1261–1263. doi: [10.1126/science.1546328](https://doi.org/10.1126/science.1546328).
- Keller A**, Nesvizhskii AI, 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**:5383–5392. doi: [10.1021/ac025747h](https://doi.org/10.1021/ac025747h).
- Keller A**, Eng J, Zhang N, Li XJ, Aebersold R. 2005. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Molecular Systems Biology* **1**:E1–E8. doi: [10.1038/msb4100024](https://doi.org/10.1038/msb4100024).
- Kim S**, Pevzner PA. 2014. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nature Communications* **5**:5277. doi: [10.1038/ncomms6277](https://doi.org/10.1038/ncomms6277).
- Kim Y**, Sidney J, Buus S, Sette A, Nielsen M, Peters B. 2014. Dataset size and composition impact the reliability of performance benchmarks for peptide-MHC binding predictions. *BMC Bioinformatics* **15**:241. doi: [10.1186/1471-2105-15-241](https://doi.org/10.1186/1471-2105-15-241).
- Koff WC**, Burton DR, Johnson PR, Walker BD, King CR, Nabel GJ, Ahmed R, Bhan MK, Plotkin SA. 2013. Accelerating next-generation vaccine development for global disease prevention. *Science* **340**:1232910. doi: [10.1126/science.1232910](https://doi.org/10.1126/science.1232910).
- Koff WC**, Gust ID, Plotkin SA. 2014. Toward a human vaccines project. *Nature Immunology* **15**:589–592. doi: [10.1038/ni.2871](https://doi.org/10.1038/ni.2871).
- Kowalewski DJ**, Stevanovic S. 2013. Biochemical large-scale identification of MHC class I ligands. *Methods in Molecular Biology* **960**:145–157. doi: [10.1007/978-1-62703-218-6_12](https://doi.org/10.1007/978-1-62703-218-6_12).
- Kowalewski DJ**, Schuster H, Backert L, Berlin C, Kahn S, Kanz L, Salih HR, Rammensee HG, Stevanovic S, Stickel JS. 2014. HLA ligandome analysis identifies the underlying specificities of spontaneous antileukemia immune responses in chronic lymphocytic leukemia (CLL). *Proceedings of the National Academy of Sciences of USA* **112**:E166–E175. doi: [10.1073/pnas.1416389112](https://doi.org/10.1073/pnas.1416389112).
- Kunz P**, Blum L, Hullár B, Schmid E, Srebniak A, Wolski W, Rinn B, Elmer FJ, Ramakrishnan C, Quandt A, Malmström L. 2014. iPortal: the swiss grid proteomics portal. *Concurrency Computation* **27**:433–445. doi: [10.1002/cpe.3294](https://doi.org/10.1002/cpe.3294).
- Lam H**, Deutsch EW, Eddes JS, Eng JK, Stein SE, Aebersold R. 2008. Building consensus spectral libraries for peptide identification in proteomics. *Nature Methods* **5**:873–875. doi: [10.1038/nmeth.1254](https://doi.org/10.1038/nmeth.1254).
- Lindestam Arlehamn CS**, Gerasimova A, Mele F, Henderson R, Swann J, Greenbaum JA, Kim Y, Sidney J, James EA, Taplitz R, McKinney DM, Kwok WW, Grey H, Sallusto F, Peters B, Sette A. 2013. Memory T cells in latent *Mycobacterium tuberculosis* infection are directed against three antigenic islands and largely contained in a CXCR3+CCR6+ Th1 subset. *PLOS Pathogens* **9**:e1003130. doi: [10.1371/journal.ppat.1003130](https://doi.org/10.1371/journal.ppat.1003130).
- Liu Y**, Buil A, Collins BC, Gillet LC, Blum LC, Cheng LY, Vitek O, Mouritsen J, Lachance G, Spector TD, Dermitzakis ET, Aebersold R. 2015. Quantitative variability of 342 plasma proteins in a human twin population. *Molecular Systems Biology* **11**:786. doi: [10.15252/msb.20145728](https://doi.org/10.15252/msb.20145728).
- Lundegaard C**, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. 2008. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acids Research* **36**:W509–W512. doi: [10.1093/nar/gkn202](https://doi.org/10.1093/nar/gkn202).
- Marcilla M**, Alpizar A, Lombardía M, Ramos-Fernandez A, Ramos M, Albar JP. 2014. Increased diversity of the HLA-B40 ligandome by the presentation of peptides phosphorylated at their main anchor residue. *Molecular & Cellular Proteomics* **13**:462–474. doi: [10.1074/mcp.M113.034314](https://doi.org/10.1074/mcp.M113.034314).
- Michalski A**, Cox J, Mann M. 2011. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC–MS/MS. *Journal of Proteome Research* **10**:1785–1793. doi: [10.1021/pr101060v](https://doi.org/10.1021/pr101060v).
- Paul S**, Weiskopf D, Angelo MA, Sidney J, Peters B, Sette A. 2013. HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity. *The Journal of Immunology* **191**:5831–5839. doi: [10.4049/jimmunol.1302101](https://doi.org/10.4049/jimmunol.1302101).
- Picotti P**, Aebersold R. 2012. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature Methods* **9**:555–566. doi: [10.1038/nmeth.2015](https://doi.org/10.1038/nmeth.2015).
- Rosenberger G**, Koh CC, Guo T, Röst HL, Kouvonen P, Collins BC, Heusel M, Liu Y, Caron E, Vichalkovskii A, Faini M, Schubert OT, Faridi P, Ebhardt HA, Matondo M, Lam H, Bader SL, Campbell DS, Deutsch EW, Moritz RL, Tate S, Aebersold R. 2014. A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Scientific Data* **1**:140031. doi: [10.1038/sdata.2014.31](https://doi.org/10.1038/sdata.2014.31).
- Röst HL**, Rosenberger G, Navarro P, Gillet LC, Miladinović SM, Schubert OT, Wolski W, C Collins Ben, Malmström J, Malmström L, Aebersold R. 2014. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nature Biotechnology* **32**:219–223. doi: [10.1038/nbt.2841](https://doi.org/10.1038/nbt.2841).
- Schittenhelm RB**, Lim Kam Sian TCC, Wilmann PG, Dudek NL, Purcell AW. 2014a. Revisiting the arthritogenic peptide theory: quantitative not qualitative changes in the peptide repertoire of HLA-B27 allotypes. *Arthritis & Rheumatology* **67**:702–713. doi: [10.1002/art.38963](https://doi.org/10.1002/art.38963).
- Schittenhelm RB**, Dudek NL, Croft NP, Ramarathinam SH, Purcell AW. 2014b. A comprehensive analysis of constitutive naturally processed and presented HLA-C*04:01 (Cw4)-specific peptides. *Tissue Antigens* **83**:174–179. doi: [10.1111/tan.12282](https://doi.org/10.1111/tan.12282).

- Schubert OT**, Ludwig C, Kogadeeva M, Zimmermann M, Rosenberger G, Gengenbacher M, Gillet LC, Collins BC, Röst HL, Kaufmann SHE, Sauer U, Aebersold R. 2015a. Absolute proteome composition and dynamics during dormancy and resuscitation of *Mycobacterium tuberculosis*. *Cell Host & Microbe* **18**:1–13. doi: [10.1016/j.chom.2015.06.001](https://doi.org/10.1016/j.chom.2015.06.001).
- Schubert OT**, Gillet LC, Collins BC, Navarro P, Rosenberger G, Wolski WE, Lam H, Amodei D, Mallick P, MacLean B, Aebersold R. 2015b. Building high-quality assay libraries for targeted analysis of SWATH MS data. *Nature Protocols* **10**:426–441. doi: [10.1038/nprot.2015-015](https://doi.org/10.1038/nprot.2015-015).
- Schumacher TN**, Keşmir C, van Buuren MM. 2015. Biomarkers in cancer immunotherapy. *Cancer Cell* **27**:12–14. doi: [10.1016/j.ccell.2014.12.004](https://doi.org/10.1016/j.ccell.2014.12.004).
- Selevsek N**, Chang CY, Gillet LC, Navarro P, Bernhardt OM, Reiter L, Cheng LY, Vitek O, Aebersold R. 2015. Reproducible and consistent quantification of the *Saccharomyces cerevisiae* proteome by SWATH-mass spectrometry. *Molecular & Cellular Proteomics* **14**:739–749. doi: [10.1074/mcp.M113.035550](https://doi.org/10.1074/mcp.M113.035550).
- Shteynberg D**, Deutsch EW, Lam H, Eng JK, Sun Z, Tasman N, Mendoza L, Moritz RL, Aebersold R, Nesvizhskii AI. 2011. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Molecular & Cellular Proteomics* **10**:M111.007690. doi: [10.1074/mcp.M111.007690](https://doi.org/10.1074/mcp.M111.007690).
- Shteynberg D**, Nesvizhskii AI, Moritz RL, Deutsch EW. 2013. Combining results of multiple search engines in proteomics. *Molecular & Cellular Proteomics* **12**:2383–2393. doi: [10.1074/mcp.R113.027797](https://doi.org/10.1074/mcp.R113.027797).
- Snyder A**, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, Walsh LA, Postow MA, Wong P, Ho TS, Hollmann TJ, Bruggeman C, Kannan K, Li Y, Elipenahli C, Liu C, Harbison CT, Wang L, Ribas A, Wolchok JD, Chan TA. 2014. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *The New England Journal of Medicine* **371**:2189–2199. doi: [10.1056/NEJMoa1406498](https://doi.org/10.1056/NEJMoa1406498).
- Trolle T**, Metushi IG, Greenbaum JA, Kim Y, Sidney J, Lund O, Sette A, Peter B, Nielsen M. 2015. Automated benchmarking of peptide-MHC class I binding predictions. *Bioinformatics* **31**:2174–2181. doi: [10.1093/bioinformatics/btv123](https://doi.org/10.1093/bioinformatics/btv123).
- Vizcaino JA**, Cote RG, Csordas A, Dianas JA, Fabregat A, Foster JM, Griss J, Alpi E, Birim M, Contell J, O’Kelly G, Schoenegger A, Ovelheiro D, Perez-Riverol Y, Reisinger F, Rios D, Wang R, Hermjakob H. 2012. The proteomics identifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Research* **41**:D1063–D1069. doi: [10.1093/nar/gks1262](https://doi.org/10.1093/nar/gks1262).