

TECHNICAL ADVANCES AND RESOURCES

Targeting infection-specific peptides in immunopeptidomics studies for vaccine target discovery

Owen Leddy^{1,2,3} , Yuko Yuki^{4,5} , Mary Carrington^{2,4,5} , Bryan D. Bryson^{1,2*} , and Forest M. White^{1,3,6*} 

Vaccine-elicited T cell responses can contribute to immune protection against emerging infectious disease risks such as antimicrobial-resistant (AMR) microbial pathogens and viruses with pandemic potential, but rapidly identifying appropriate targets for T cell priming vaccines remains challenging. Mass spectrometry (MS) analysis of peptides presented on MHCs can identify potential targets for protective T cell responses in a proteome-wide manner. However, pathogen-derived peptides are outnumbered by self-peptides in the MHC repertoire and may be missed in untargeted MS analyses. Here, we present a novel approach, termed PathMHC, that uses computational analysis of untargeted MS data followed by targeted MS to discover novel pathogen-derived MHC peptides more efficiently than untargeted methods alone. We applied this workflow to identify MHC peptides derived from multiple microbes, including potential vaccine targets presented on MHC-I by human dendritic cells infected with *Mycobacterium tuberculosis* (*Mtb*), finding that all *Mtb* peptides detected in the MHC-I repertoire derived from proteins exported by type VII secretion systems. PathMHC will facilitate antigen discovery campaigns for vaccine development.

Introduction

Antimicrobial-resistant (AMR) bacterial, fungal, and parasitic infections pose a substantial threat to global public health, causing millions of deaths per year (Murray et al., 2022). The impact of AMR infections is expected to further increase by 2050, reaching USD 0.33–1.2 trillion per year in healthcare expenditures (Jonas et al., 2017). Effective vaccines against AMR pathogens could mitigate these impacts and limit further opportunities for resistance to develop by preventing infections that would otherwise need to be treated with antimicrobials. T cell responses are essential and, in some cases, sufficient for immune protection against many common AMR pathogens and other globally significant microbial infections, including *Mycobacterium tuberculosis* (*Mtb*) (Flynn and Chan, 2022), uropathogenic *Escherichia coli* (Rousseau et al., 2023), *Salmonella* sp. (Benoun et al., 2018), *Staphylococcus aureus* (Brown et al., 2015), *Streptococcus pneumoniae* (Wang et al., 2017), *Klebsiella pneumoniae* (Chen et al., 2011; Iwanaga et al., 2021), *Trypanosoma cruzi* (Acosta Rodríguez et al., 2019), and others. In pathogens with multiple serotypes, T cell responses can confer broad-spectrum protection against distinct strains that lack antibody cross-reactivity (Chen et al., 2011; Wang et al., 2017). However, it is

often not known which proteins are optimal targets for T cell-priming vaccines. Methods to identify appropriate targets can therefore contribute to accelerating vaccine development against microbial pathogens.

T cell responses also contribute to immune protection against classes of viruses with pandemic potential, such as coronaviruses (Carter et al., 2023; McMahan et al., 2021; Sette et al., 2023) and influenza viruses (La Gruta and Turner, 2014; Mallajosyula et al., 2024; McElhaney et al., 2006), and often target epitopes that are conserved in variants that evade antibody neutralization (Carter et al., 2023; La Gruta and Turner, 2014; Sette et al., 2023). The US Biomedical Advanced Research and Development Authority have made establishing the capability to develop and deploy a vaccine within 100 days of a novel infectious disease outbreak a key strategic goal (BARDA, 2022). Methods that can rapidly identify potential targets of protective T cell responses could contribute to achieving this goal and enhancing future outbreak preparedness.

Mass spectrometry (MS) analysis of MHC-bound peptides (termed immunopeptidomics) can directly identify which pathogen-derived antigens are presented on MHCs in infected

¹Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA; ²Ragon Institute of MGH, Harvard, and MIT, Cambridge, MA, USA; ³Koch Institute for Integrative Cancer Medicine, Cambridge, MA, USA; ⁴Basic Science Program, Frederick National Laboratory for Cancer Research, National Cancer Institute, Frederick, MD, USA; ⁵Laboratory of Integrative Cancer Immunology, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA; ⁶Center for Precision Cancer Medicine, Massachusetts Institute of Technology, Cambridge, MA, USA.

*B.D. Bryson and F.M. White contributed equally to this paper. Correspondence to Bryan D. Bryson: bryand@mit.edu; Forest M. White: fwhite@mit.edu.

© 2025 Leddy et al. This article is distributed under the terms as described at <https://rupress.org/pages/terms102024/>.

cells and/or professional antigen-presenting cells and could serve as targets for protective T cell responses (Leddy et al., 2021; Mayer and Impens, 2021). Immunopeptidomics can have particular advantages as a vaccine target discovery approach in the context of microbial infections, since microbial pathogens may have thousands of protein-coding genes that can be challenging or costly to comprehensively screen for T cell reactivity. By identifying MHC peptides in an untargeted manner, immunopeptidomics allows proteome-wide discovery of potential antigenic targets without an a priori bias toward any particular subset of the pathogen proteome. Immunopeptidomics can serve as a faster, less costly, and higher-throughput way to identify candidate human T cell antigens than approaches that rely on directly measuring T cell reactivity, which often require recruiting cohorts of human subjects with prior exposure to the pathogen of interest and may involve synthesizing large libraries of peptides (Law and Watts, 2023).

Pathogen-derived peptides can be outnumbered by self-peptides in the MHC repertoire of infected cells by as much as 1,000 to 1 and may be missed in untargeted immunopeptidome analyses performed using data-dependent analysis (DDA) (Leddy et al., 2023; Mayer et al., 2022; Weingarten-Gabbay et al., 2021). In DDA analyses, instrument control software selects precursor ions for fragmentation and MS/MS spectrum acquisition in real time, generally prioritizing more abundant precursor ions first with no prior knowledge of which peptides are pathogen derived. Most instrument time during an analysis may be dedicated to MS/MS acquisition of self-peptides, which are not candidate vaccine targets. In principle, targeted MS methods could selectively and sensitively identify pathogen-derived peptides while ignoring self-peptides, but these methods require prior knowledge of peptides of interest and therefore cannot be used for novel target discovery.

Here, we describe a novel immunopeptidomics approach (termed “PathMHC”) that facilitates discovery of pathogen-derived MHC class I (MHC-I) or class II (MHC-II) peptides by combining computational analysis of untargeted MS data with follow-up targeted MS analyses to identify previously missed pathogen-derived peptides. We first benchmark PathMHC against a comparable approach based on untargeted MS alone, using MHC-II peptide presentation by dendritic cells after phagocytosis of the nonpathogenic mycobacterium *Mycobacterium smegmatis* (*Msmeg*) as a model system. We next demonstrate that the PathMHC computational pipeline generalizes across infectious disease areas by applying it to publicly available immunopeptidomics datasets. Finally, we apply PathMHC to analyze the MHC-I repertoire of human dendritic cells infected with *Mtb*, a globally important pathogen against which T cell-mediated immunity is an essential component of protective immune responses (Flynn and Chan, 2022).

Results

Combining computational enrichment with targeted MS facilitates identification of microbe-derived MHC peptides

The PathMHC workflow (Fig. 1 a) requires MHC peptide samples isolated from cells infected with a pathogen of interest and from

mock-infected control cells. First, each sample is divided in two equal parts, and one half of each sample (infected and mock-infected) is analyzed by untargeted MS (DDA). Second, the untargeted MS data are analyzed using a computational pipeline that finds precursor ions that are specific to infected cells and absent in the mock-infected control, excluding those already identified in the untargeted analysis (see Materials and methods). Finally, this set of putatively infection-specific precursor ions is used as an inclusion list for a targeted MS analysis of the remaining half of each sample using parallel reaction monitoring (PRM) to identify pathogen-derived peptides previously missed in the untargeted analysis.

The PathMHC computational pipeline comprises three steps (Fig. 1 b): First, liquid chromatography (LC)-MS features (i.e., precursor ions) are annotated using the feature detection algorithm Dinosaur (Teleman et al., 2016). Second, the LC-MS features are aligned across the two samples using DeepRTAlign (Liu et al., 2023), and features in the infected sample that are not paired with any feature in the mock-infected control are selected for further analysis. Finally, LC-MS features that have a charge state and retention time consistent with MHC peptides are scored for quality (including signal-to-noise ratio [SNR]) using the autoencoder-based scoring algorithm AutoMS (Ji and Tian, 2022). Peaks that meet scoring thresholds are added to the inclusion list for the follow-up targeted analysis with PRM.

To benchmark the PathMHC approach, we used *Msmeg*, a nonpathogenic mycobacterium. We incubated primary human monocyte-derived dendritic cells (hMDCs) with *Msmeg* at a multiplicity of infection (MOI) of 10 for 2 h to allow phagocytosis of the bacteria, harvested and lysed the cells 24 h after phagocytosis, and isolated MHC-II-associated peptides (Fig. 2 a).

In early tests of prototype PathMHC workflows, injecting 100% of an MHC-II peptide sample in a single DDA MS analysis identified fewer bacterial peptides (24 unique bacterial peptides out of 2,318 total unique identifications [IDs]) than did two consecutive DDA analyses, each using 50% of the sample (34 unique bacterial peptides out of 2,875 total unique IDs, with 19 shared between the two runs) (Fig. S1 a). The inherent variability in the selection of peptides for MS/MS spectrum acquisition during DDA analyses means that the second analysis will capture some bacterial peptides not identified in the first, resulting in more IDs as long as the bacterial peptides are still above the limit of detection when using 50% of the sample. We therefore adopted two consecutive DDA analyses as the baseline method against which we would benchmark PathMHC (Fig. 2 b).

Peptides identified in MHC-II pulldowns from *Msmeg*-infected hMDCs or uninfected controls had a length distribution typical of MHC-II peptides (Fig. 2 c). An average of 65,293 LC-MS features (precursor ion peaks) were identified in untargeted DDA analyses of MHC-II peptides from *Msmeg*-infected hMDCs, 14.4% of which were classified as specific to infected cells and absent in the corresponding mock-infected control by DeepRTAlign and had an appropriate charge state (+2 or +3). After discarding precursor ion peaks that were already identified in the initial untargeted MS analysis or failed to meet AutoMS quality scoring thresholds, an average of 7.57% of

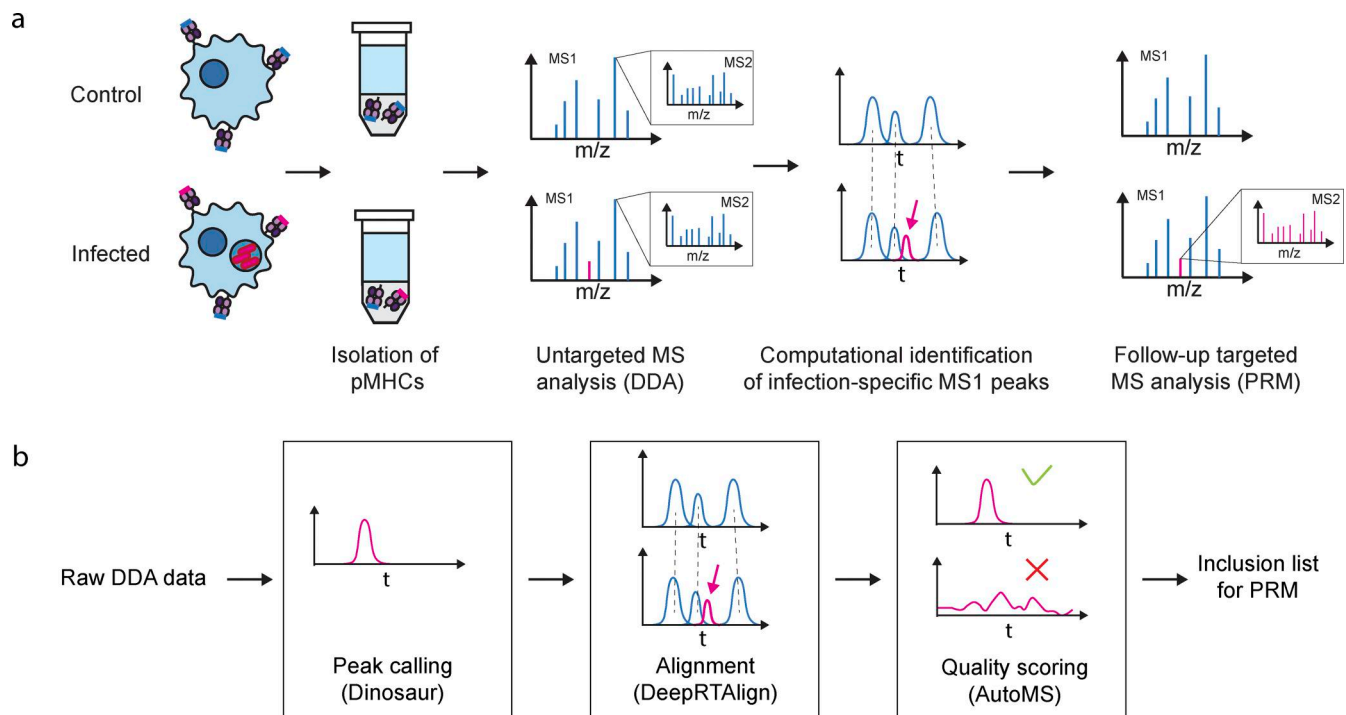


Figure 1. Overview of the PathMHC workflow. (a) A schematic representation of the PathMHC workflow, including MHC peptide isolation from infected cells and mock-infected control cells, untargeted DDA analysis, computational ID of infection-specific precursor ions, and targeted follow-up PRM analyses. **(b)** A schematic representation of the steps of the PathMHC computational pipeline, including peak calling using Dinosaur (Teleman et al., 2016), chromatographic alignment using DeepRTAlign (Liu et al., 2023), and peak quality filtering using AutoMS (Ji and Tian, 2022).

total precursor ions were ultimately added to the inclusion list for targeted analysis by PRM (Fig. 2 d).

The PathMHC workflow (one untargeted DDA analysis and one targeted PRM analysis) on average identified ~91% more bacterial peptides than did two consecutive untargeted DDA analyses (Fig. 2 e and Data S1). PathMHC PRM analyses detected, on average, ~270% more new bacterial peptides (i.e., peptides that were not detected in the initial DDA analysis) than did a second DDA analysis of the same sample. On average, 1.01% of peptides identified in DDA analyses were *Msmeg* derived, whereas 5.45% of peptides identified in PRM analyses using inclusion lists generated by the PathMHC computational pipeline were *Msmeg* derived, showing that the PathMHC pipeline successfully enriched for bacterial peptides ($P < 0.05$; Fig. 2 f). These results show that PathMHC can increase the number of bacterial MHC peptides identified from a given sample compared with DDA alone.

To maximize the ease of adoption of PathMHC, we also asked whether an analogous computational workflow implemented in a commercial proteomics analysis program could produce comparable results to our pipeline based on open-source tools. We implemented a workflow (PathMHC-Proteome Discoverer [PD]) in Thermo Scientific's proteomics analysis program PD that can also perform alignment of precursor ion peaks across samples and select candidate peaks that are specific to the MHC repertoire of infected cells (see Materials and methods). A workflow based on DDA alone identified 20.5 bacterial peptides on average, while PathMHC, using our computational pipeline based on open-source tools, identified 40 bacterial peptides on

average, and PathMHC-PD identified 34 bacterial peptides on average (Fig. S1 b). Whereas on average 1.02% of peptides identified in DDA runs were bacterial peptides, 5.97% of peptides identified in PathMHC PRM runs using our open-source computational workflow, and 8.20% of peptides identified in PathMHC-PD PRM runs were bacterial peptides (Fig. S1 c). These results show that either our open-source implementation of the PathMHC computational pipeline or an analogous workflow implemented in a commercial proteomics analysis program can provide comparable performance in enriching and identifying pathogen-derived MHC peptides.

By design, the overlap in bacterial peptides identified between our initial DDA analysis and follow-up PRM analysis was lower than that between consecutive DDA analyses (Fig. S1 d), since precursor ions already identified by DDA were excluded from our PRM inclusion lists. Overlapping peptides in this context could reflect targeting of precursor ions with different charge states derived from the same peptide and/or peptides incidentally captured in the mass-to-charge ratio (m/z) and retention time tolerance window of another precursor ion on the inclusion list. Peptide-spectrum matches (PSMs) for bacterial peptides that were uniquely identified by PRM and were previously missed by DDA on average had lower precursor ion intensities than PSMs for bacterial peptides identified by DDA (Fig. S1 e; $P < 10^{-6}$, Wilcoxon rank-sum test). Additionally, in DDA runs, instrument cycle time was fully utilized for MS/MS scans throughout most of the gradient (i.e., the maximum possible number of MS/MS scans were acquired in each cycle), indicating that not all peptides could be isolated and sequenced (Fig. S1 f).

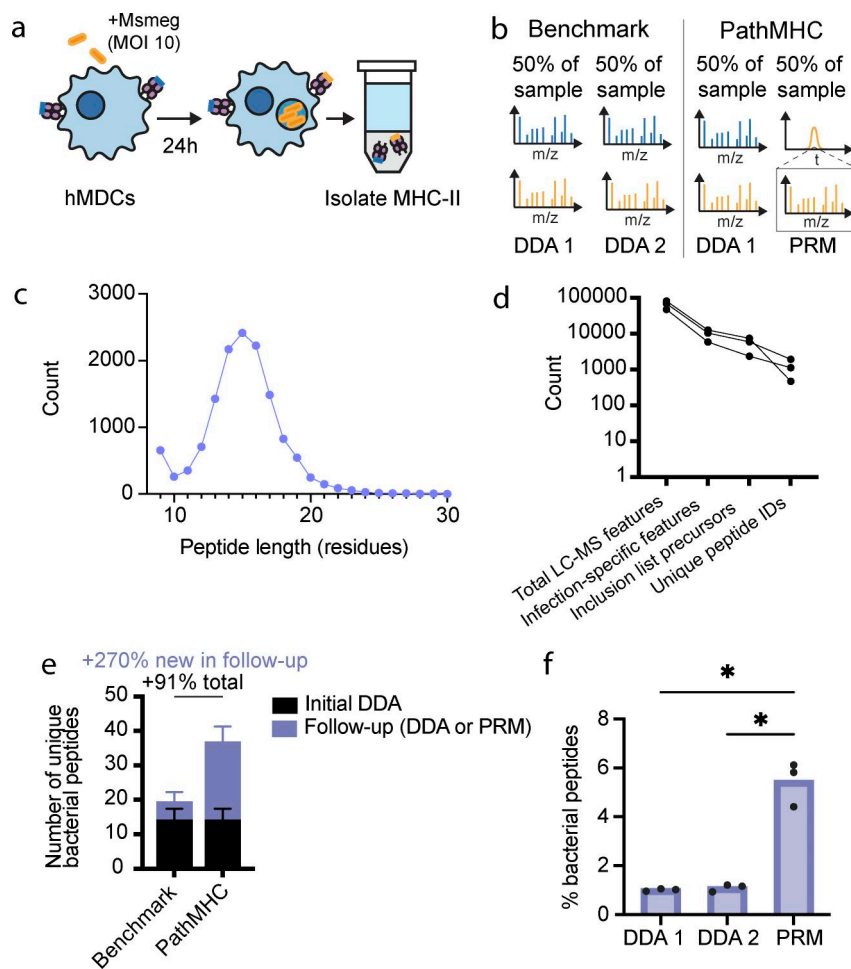


Figure 2. PathMHC identifies more bacterial peptides presented by professional antigen-presenting cells than does DDA alone in a model system. (a) Schematic representation of the model system used to benchmark PathMHC. hMDCs were allowed to phagocytose *Msmeg* at an MOI of 10 for 2, and 24 h later MHC-II peptides were isolated for MS analysis. (b) Outline of the benchmark DDA only workflow and the corresponding PathMHC workflow. (c) Length distribution of peptide IDs in MHC-II analyses. (d) Number of LC-MS (precursor ion) features identified by Dinosaur, precursors with charge state +2 or +3 identified as specific to infected cells by DeepRTalign ("infection-specific features"), precursors selected for inclusion list, and final number of peptide IDs obtained by PRM, for $n = 3$ PathMHC analyses of hMDCs that phagocytosed *Msmeg*. (e) Number of unique bacterial peptides identified in the initial DDA analysis (black bar) and the number of new unique bacterial peptides identified in the follow-up DDA or PRM analysis (lavender bar) for the benchmark or PathMHC workflow, respectively. Bars represent the mean $\pm$ SD of $n = 3$ replicates using cells from different healthy donors. (f) The percentage of peptides derived from *Msmeg* in each DDA or PRM analysis, showing enrichment of bacterial peptides in PathMHC PRM follow-up analyses for $n = 3$ replicates using cells from different healthy donors (* $P < 0.05$, one-way paired ANOVA).

Taken together, lower precursor ion intensity and limited instrument cycle time may explain why some bacterial peptides were missed in DDA runs but detected by PRM.

The PathMHC computational pipeline generalizes across pathogens

We next wanted to test whether our PathMHC computational pipeline generalizes across different infectious disease areas. We therefore used the PathMHC computational pipeline to identify infection-specific precursor ions in published untargeted MS analyses of the MHC-I repertoire of the A549 human epithelial cell line infected with SARS-CoV-2 (Weingarten-Gabbay et al., 2021). In this case, we could not know what the identities of all the selected precursors would have turned out to be if they had been targeted in a follow-up PRM analysis, so we instead analyzed the subset of precursors selected by the PathMHC pipeline that had already been identified in the untargeted MS analysis. SARS-CoV-2-derived peptides were enriched among the precursors selected by the PathMHC computational pipeline that had known identities, relative to the overall set of peptides identified (Fig. 3 a). Self-peptides selected by the PathMHC pipeline were enriched for proteins associated with interferon responses, which are known to be induced by SARS-CoV-2 infection (Lei et al., 2020) (Fig. 3 b). These results indicate that self-peptides selected by the PathMHC computational pipeline

are enriched for pathogen-derived peptides and peptides derived from proteins associated with the host response to infection.

An analogous computational analysis of an MS dataset of MHC-I peptides isolated from the HCT116 human colon epithelial cell lines infected with *Listeria monocytogenes* also enriched for *Listeria*-derived peptides relative to the overall proportion of *Listeria*-derived peptides identified in the study (Mayer et al., 2022) (Fig. 3 c). (In this case, the study authors did not report IDs separately for each replicate.) Although not statistically significant after multiple hypothesis test correction, the mTORC1 signaling pathway was the most enriched among source proteins of self-peptides selected by the PathMHC computational pipeline (Fig. 3 d). *L. monocytogenes* has previously been shown to induce mTORC1 signaling to facilitate entry into host cells (Bhalla et al., 2017). The fact that the PathMHC computational pipeline enriches for SARS-CoV-2- and *Listeria*-derived MHC peptides in publicly available datasets suggests that PathMHC can help identify pathogen-derived MHC peptides across a range of infectious disease areas.

PathMHC identifies potential TB vaccine targets presented on MHC-I

We next applied PathMHC to analyze the MHC-I repertoire of hMDCs infected with *Mtb* to identify potential TB vaccine

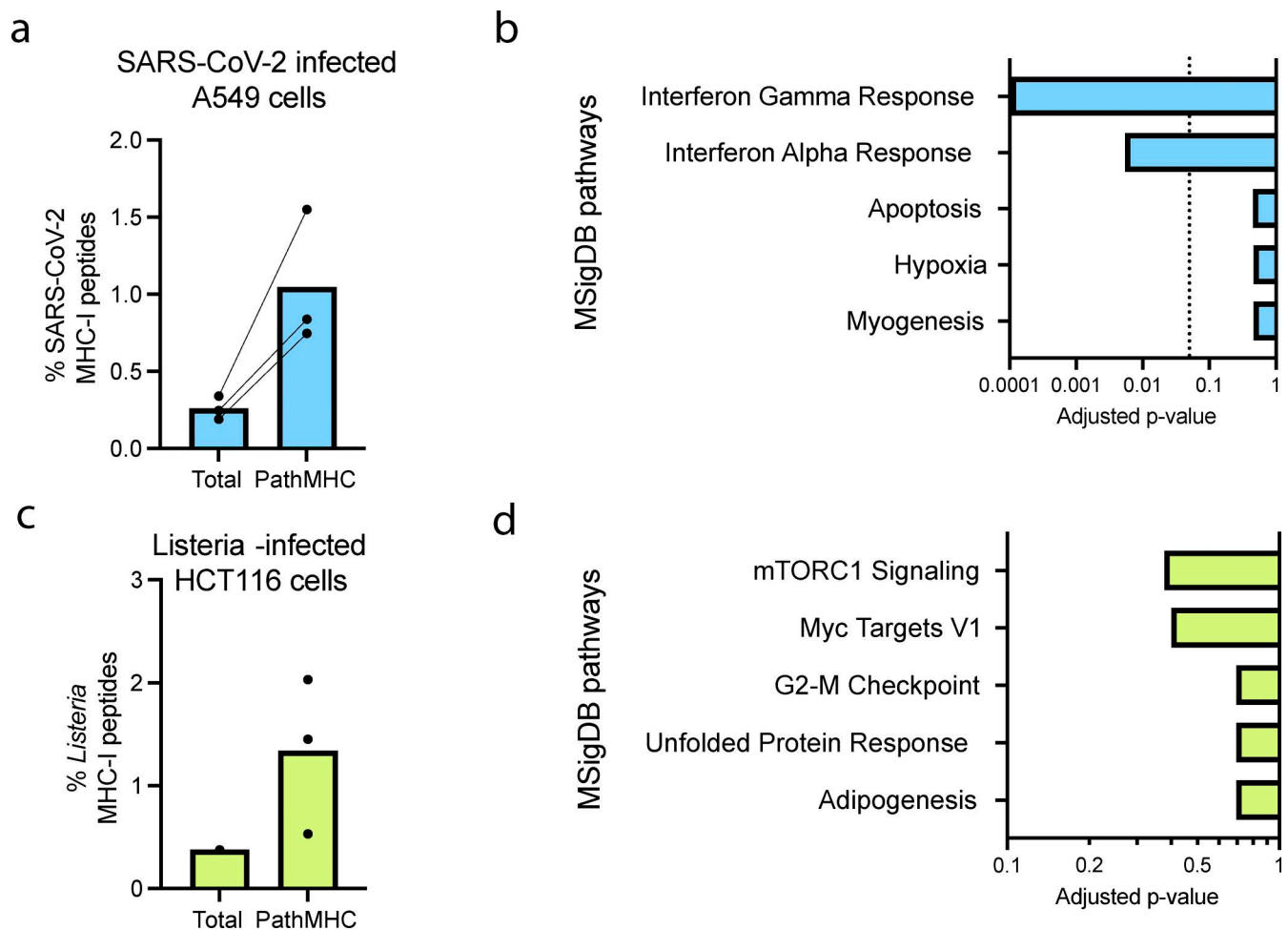


Figure 3. PathMHC enriches for pathogen-derived and infection-associated peptides in publicly available immunopeptidomics datasets. (a and c) Percentage of pathogen-derived peptides in published DDA analyses of MHC-I peptides isolated from (a) SARS-CoV-2-infected A549 cells ($n = 3$ fractions) (Weingarten-Gabbay et al., 2021) and (c) *L. monocytogenes*-infected HCT116 cells ($n = 3$ replicates) (Mayer et al., 2022), compared with the percentage of pathogen-derived peptides among identified precursors selected by the PathMHC computational pipeline (not excluding peptides identified by DDA; see Materials and methods). **(b and d)** Pathway enrichment analysis of precursors selected by the PathMHC computational pipeline that were identified as self-peptides in (b) SARS-CoV-2-infected A549 cells and (d) *L. monocytogenes*-infected HCT116 cells. Enrichment analysis was performed using Enrichr (Chen et al., 2013) against the MSigDB 2020 database (Liberzon et al., 2011), relative to the overall set of source proteins of peptides identified in the study as a background. The dashed line indicates the threshold for statistical significance after multiple hypothesis testing correction (Benjamini-Hochberg).

targets. While CD8⁺ T cell reactivity to many *Mtb* epitopes has been evaluated, prior studies found that vaccination with some immunodominant *Mtb* antigens elicited CD8⁺ T cell responses that failed to recognize *Mtb*-infected cells (Nyendak et al., 2016). This finding suggests a need for further assessment of which antigens are presented on MHC-I by infected target cells to facilitate vaccine design. Immunopeptidomics provides a way to fill this knowledge gap in a proteome-wide manner. Unlike *Msmeg*, *Mtb* is adapted to intracellular survival, and much of the bacterial cell contents are likely to be inaccessible to host antigen-processing machinery in the absence of extensive bacterial death. Furthermore, *Mtb* resides primarily in membrane-bound phagosomes (Kinsella et al., 2021), further limiting access to MHC-I antigen-processing pathways. We hypothesized that PathMHC could potentially facilitate ID of the select subset of *Mtb* peptides that can be presented on MHC-I for CD8⁺ T cell recognition in cells infected with live *Mtb*.

hMDCs were infected with *Mtb* at an MOI of 2.5 for 72 h prior to isolation of MHC-I peptides (Fig. 4a). Identified peptides had a length distribution typical of MHC-I peptides (Fig. 4b), and unsupervised clustering of MHC-I peptides using Gibbs clustering (Andreatta et al., 2017; Andreatta et al., 2013) resulted in clusters with sequence motifs that matched the known peptide-binding preferences of HLA alleles expressed by each donor (Fig. S2). *Mtb*-derived peptides comprised 0.116% of MHC-I peptide IDs in infected hMDCs in DDA analyses and 0.156% on average in PRM analyses (Fig. 4c), suggesting a modest enrichment despite the low overall prevalence of *Mtb* peptides in this system. This proportion of *Mtb*-derived peptides relative to self-peptides is in line with our prior studies of the MHC-I repertoire in *Mtb*-infected macrophages (Leddy et al., 2023) and other immunopeptidomics studies focused on other pathogens (Mayer et al., 2022; Weingarten-Gabbay et al., 2021). We obtained a total of 19 *Mtb*-derived peptide IDs (16 unique epitopes), including

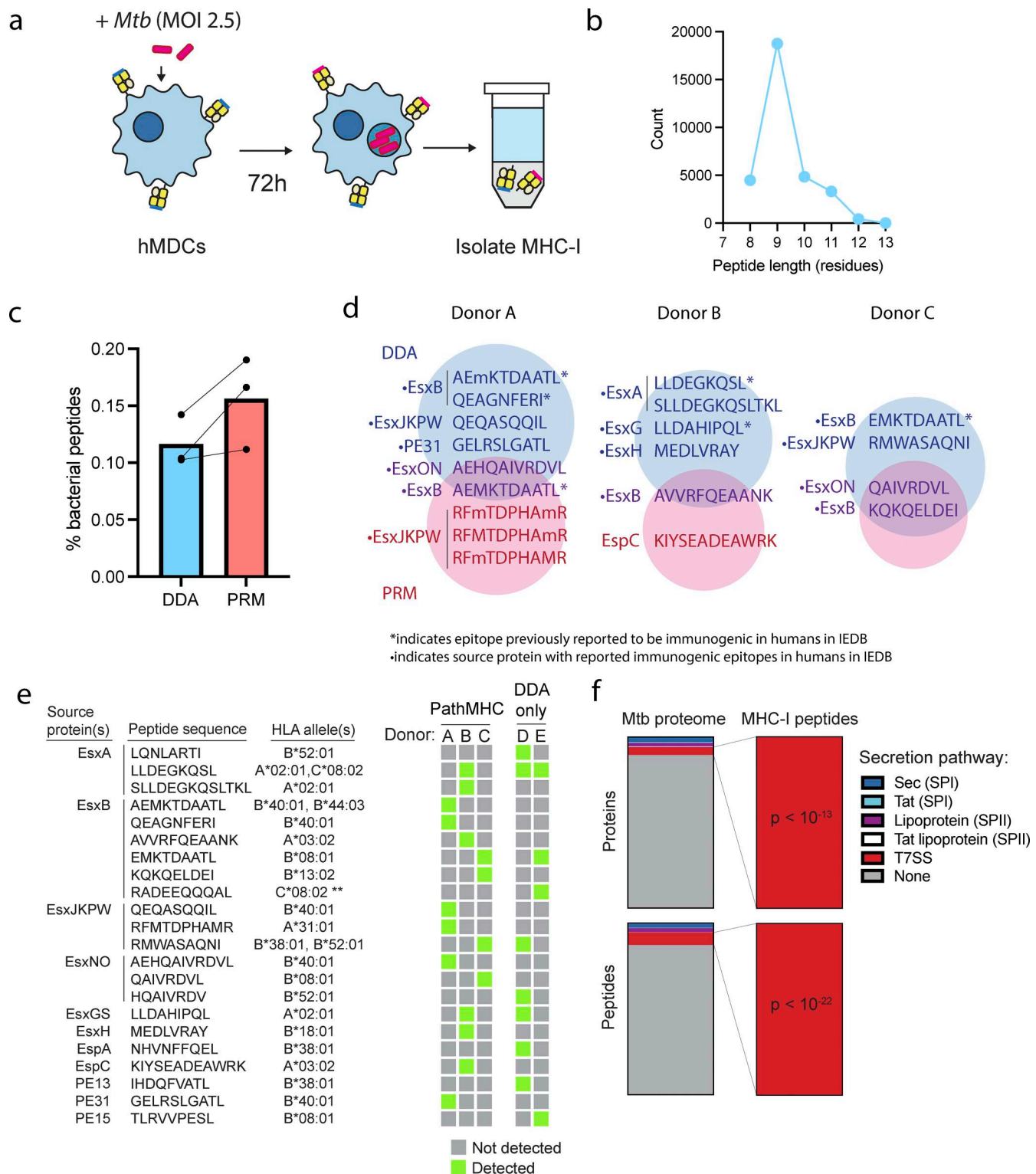


Figure 4. *Mtb*-infected hMDCs present peptides derived from T7SS substrates on MHC-I. (a) Schematic representation of the *Mtb* dendritic cell infection model. hMDCs were allowed to phagocytose *Mtb* at an MOI of 2.5 for 4, and 72 h later MHC-I peptides were isolated for MS analysis. (b) Length distribution of MHC-I peptides IDs. (c) The proportion of *Mtb*-derived peptides in each initial DDA analysis and follow-up PRM analysis for each of $n = 3$ donors. (d) *Mtb* peptide sequences detected in DDA analyses, PRM analyses, or both in the PathMHC workflow (right) and their source proteins (left) for each of three donors. Lowercase m = oxidized methionine. * indicates peptides previously reported to be immunogenic in humans with prior *Mtb* exposure in the IEDB (Vita et al., 2019). (e) Sequences, source proteins, associated HLA alleles, and donors for each validated *Mtb*-derived MHC-I peptide. Green boxes indicate the donor(s) that presented a given peptide. The associated HLA alleles listed are the allele with the top eluted ligand (EL) rank predicted by NetMHCpan 4.1 (Reynisson et al., 2020), for each donor that presented a given peptide. (** indicates that this epitope has previously been reported as an HLA-B*14-restricted CD8⁺ T cell epitope [Lewinsohn et al., 2007]. HLA-B*14:02, expressed by donor E, may in fact be the restricting allele, but the peptide is not predicted to bind HLA-B*14:02 by

NetMHCpan). (f) Enrichment analysis of *Mtb* protein secretion pathways among the detected MHC-I peptides relative to the encoded *Mtb* H37Rv proteome, by peptide (binomial test) and by source protein (hypergeometric test; see Materials and methods). T7SS substrates (red) comprise all detected *Mtb*-derived MHC-I peptides. IEDB, Immune Epitope Database.

4 identified in PRM analyses that had not been previously identified by DDA (2 unique epitopes) (Fig. 4 d and Data S2). As expected, excluding precursor ions previously identified by DDA from PRM analyses resulted in limited overlap in peptide IDs. Peptides identified by PathMHC PRM analyses that were previously missed in DDA analyses derived from EspC and the EsxJ family of proteins—EsxJ, EsxK, EsxP, and EsxW (abbreviated here as EsxJKPW). We next examined whether these *Mtb*-derived peptides identified using our in vitro experiments showed evidence of immunogenicity in individuals with prior evidence of *Mtb* exposure. We queried the Immune Epitope Database (Vita et al., 2019), which catalogs studies of immunogenic peptides identified in human cohorts. Not every *Mtb*-derived peptide has been comprehensively tested for human immunogenicity, so a lack of a reported T cell response to a given peptide in the Immune Epitope Database does not indicate lack of immunogenicity. We found that five of the peptides we detected by PathMHC are immunogenic in humans (LLDEGKQSL—EsxA, LLDAHIPQL—EsxG/EsxS, AEMKTDAATL—EsxB, EMKTDAATL—EsxB, and QEAGNFERI—EsxB) (Lewinsohn et al., 2007; Tully et al., 2005). The overall space of possible 8- to 11-mer *Mtb* peptides that could hypothetically be candidate MHC-I epitopes is large, so matching immunogenic epitopes by coincidence is unlikely. These results demonstrate that the application of immuno-peptidomics to *Mtb*-infected samples in vitro can identify immunogenic peptides with in vivo relevance.

The extremely low proportion of *Mtb* peptides in the MHC-I repertoire relative to self-peptides can result in a high false discovery rate (FDR) for *Mtb*-derived IDs even if overall FDR is well controlled (Leddy et al., 2024). *Mtb* IDs must therefore be carefully validated, as even infrequent misidentification of self-antigens as *Mtb* antigens would greatly increase the FDR for *Mtb* peptides. We manually checked MS/MS spectrum quality and specificity to infected cells alone (see Materials and methods) for the *Mtb* MHC-I peptide IDs we obtained from PathMHC. We further validated peptide IDs that passed these checks using SureQuant (Leddy et al., 2024; Leddy et al., 2023), a targeted MS method that compares biological MHC peptides with stable isotope-labeled (SIL) synthetic standards to demonstrate identical retention times and MS/MS spectra (Data S6). We also similarly validated five other epitopes that had been separately identified in DDA only analysis of MHC-I peptides from hMDCs (Fig. 4 d and Data S6). Each *Mtb*-derived peptide was predicted to bind at least one class I HLA allele expressed by the corresponding donor by the MHC peptide prediction algorithm NetMHCpan (Reynisson et al., 2020) (Fig. S3). These results increase our confidence that the peptides identified are authentic *Mtb* MHC-I epitopes.

The *Mtb*-derived MHC-I peptides presented by hMDCs are exclusively derived from substrates of *Mtb*'s type VII secretion systems (T7SSs), a statistically significant enrichment relative to

the encoded *Mtb* proteome (peptide enrichment: $P < 10^{-22}$, binomial test with Bonferroni correction; source protein enrichment: $P < 10^{-13}$, hypergeometric test with Bonferroni correction; Fig. 4 e). T7SSs are specialized molecular machines that export proteins across the bacterial cell envelope and have roles in nutrient acquisition (Siegrist et al., 2009), outer membrane-selective permeability (Boradia et al., 2022), and virulence (Stanley et al., 2003). This enrichment for T7SS substrates cannot be solely the result of bias in our manual inspection process, as four putative non-T7SS-derived *Mtb* peptides passed manual inspection and were selected for SureQuant validation but did not co-elute with the corresponding synthetic standards, indicating these IDs were incorrect (Table S2). In contrast, all T7SS-derived peptides selected for SureQuant analysis were successfully validated. Based on our secondary screen using SureQuant, the ground truth FDR for peptides that passed our initial filtering and manual inspection process was thus $4/26 = 15.4\%$. We previously observed that *Mtb*-infected human monocyte-derived macrophages also predominantly present T7SS substrates on MHC-I (Leddy et al., 2023). Our results here suggest that the determinants of *Mtb* MHC-I antigen processing and presentation may be similar among multiple classes of monocyte-derived phagocytes (Lee et al., 2020; Wolf et al., 2007).

PathMHC also targets self-peptides that reflect the host cell response to *Mtb* infection. Self-peptides uniquely detected by PathMHC PRM analyses of *Mtb*-infected hMDCs and not in DDA analyses were enriched for proteins associated with interferon responses, relative to self-peptides identified by DDA (Fig. 5 a). The type I interferon response is a dominant component of the myeloid cell response to *Mtb* infection (Akter et al., 2022; Barczak et al., 2017; Wu et al., 2012). Many interferon-associated self-peptides identified by PathMHC (such as peptides derived from MX1, IFIT2, and IRF1) were only detectable in infected cells or highly upregulated during infection (Fig. 5 b), whereas self-MHC-I peptides identified in DDA analyses of mock-infected cells were variously upregulated, downregulated, or unaffected by *Mtb* infection (Fig. S4). These results confirm that PathMHC enriches for MHC-I peptides associated with *Mtb* infection.

Discussion

PathMHC represents an approach to immuno-peptidomics that is specifically tailored to addressing the challenges of vaccine development for infectious disease. PathMHC accelerates vaccine target discovery by computationally enriching for infection-specific precursor ions in discovery immuno-peptidomics analyses, helping to find the “needle in a haystack” of pathogen-derived peptides in the MHC repertoire. In a model system where *Msmeg*-derived peptides make up ~1% of the MHC-II immuno-peptidome, PathMHC enhanced ID of *Msmeg*-derived peptides by 91% relative

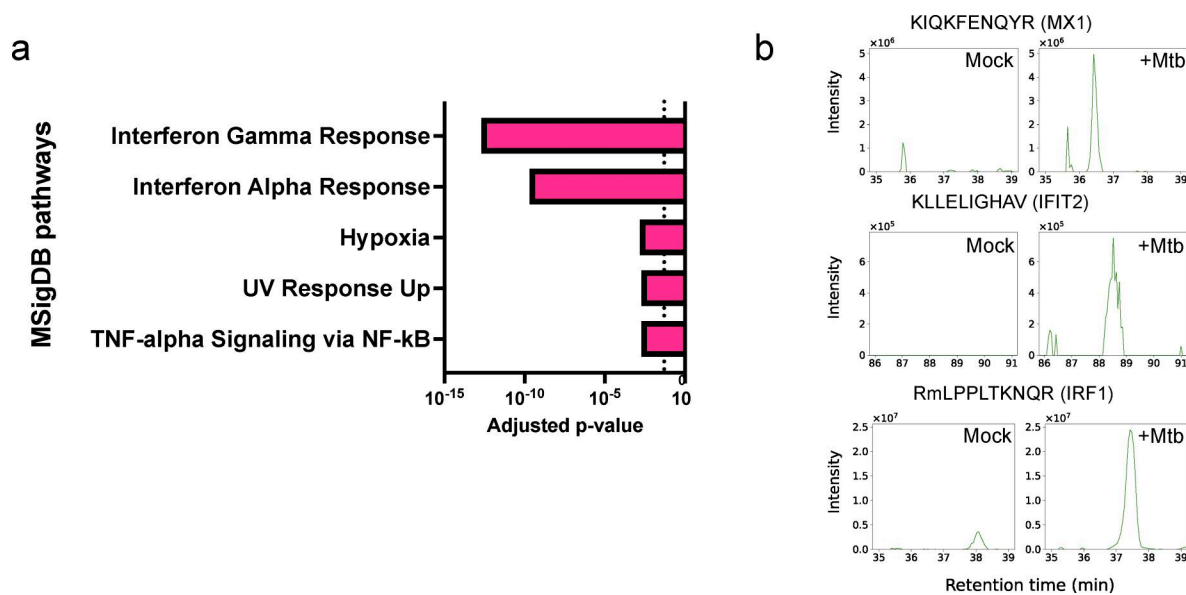


Figure 5. Self-MHC-I peptides uniquely detected by PathMHC PRM analyses reflect the host cell response to *Mtb* infection. (a) Top five enriched pathways among source proteins of self-MHC-I peptides uniquely detected in PathMHC PRM analyses of *Mtb*-infected hMDCs from $n = 3$ donors (i.e., not identified in the initial DDA analysis). P values were computed using Enrichr (Chen et al., 2013; Kuleshov et al., 2016) over the MSigDB Hallmark 2020 pathway database (Liberzon et al., 2011), using the set of source proteins of peptides detected in the corresponding DDA analyses of the *Mtb*-infected cells as a background. The dotted line marks an adjusted P value threshold of 0.05. **(b)** Representative XIC for precursor ions of MHC-I peptides uniquely identified in PathMHC PRM analyses of *Mtb*-infected hMDCs.

to a benchmark method based on DDA alone. The PathMHC computational workflow also generalizes to multiple pathogens, including SARS-CoV-2 infection and *L. monocytogenes* infection. Applying PathMHC to analyze the MHC-I repertoire of *Mtb*-infected hMDCs revealed that substrates of T7SSs were selectively presented on MHC-I. Self-peptides targeted by PathMHC reflected the host cell response to infection, particularly interferon responses.

PathMHC prioritizes pathogen-derived peptides for acquisition, enabling efficient use of instrument cycle time. By narrowing down the set of possible targets that must be scanned, PathMHC may also allow longer ion accumulation times for each peptide in settings where sensitivity is limiting. Further improvements to the computational pipeline will likely increase its ability to enrich for pathogen-derived peptides. DeepRTAlign and AutoMS were not originally trained on LC-MS data from immunopeptidomics studies, and retraining or fine-tuning the machine learning models underlying these tools on more task-specific data or substituting purpose-built tools may increase performance.

For some applications, PathMHC may have advantages over offline fractionation of MHC peptides as a means of increasing depth of coverage in immunopeptidomics studies. Offline fractionation can incur substantial sample losses (Magdeldin et al., 2014), so PathMHC may have advantages when sample input is limited, such as when working with primary cells or tissues. PathMHC also requires less total instrument time, requiring two analyses per condition (4 h) compared with 12 fractions (Leddy et al., 2023) (24 h). Higher throughput could enable broader coverage of HLA alleles or pathogen variants or could enable a faster response to emerging infectious disease risks.

Whereas here we benchmarked PathMHC against DDA as a baseline, library-free data-independent analysis (DIA) is also sometimes used as a method for untargeted discovery of MHC peptides (Shahbazy et al., 2023). Unlike in DDA, all precursor ions are in principle fragmented and analyzed in DIA analyses, obviating the problem of prioritizing pathogen-derived precursors for MS/MS acquisition. However, DIA relies on accurate deconvolution of complex MS/MS spectra comprising fragment ions derived from many precursors. This can be a challenging computational task to perform reliably given the large search space of possible MHC peptides, unless constrained by prior knowledge in the form of a well-defined reference spectral library (Pak et al., 2021). A mismatch between the proportion of pathogen-derived peptides in the search space and the ground-truth proportion of pathogen-derived peptides in the sample can further increase FDRs for pathogen-derived peptides (Prensen et al., 2023). Moreover, the rapid scan speed and limited ion accumulation time in DIA can select against low-abundance peptides, some of which could be pathogen-derived antigens. Augmenting DIA-based discovery approaches with a workflow conceptually analogous to PathMHC, in which candidate pathogen-specific peptides are secondarily screened by PRM, could help confidently identify pathogen-derived peptides.

By applying PathMHC to identify *Mtb*-derived peptides presented on MHC-I, we showed that infected hMDCs preferentially present T7SS substrates on MHC-I and identified several antigens that could potentially serve as vaccine targets. Lewinsohn et al. (2007) previously isolated human CD8⁺ T cell clones from individuals with prior *Mtb* exposure that recognized three MHC-I epitopes derived from EsxB (AEMKTDAATL, EMKTDAATL, and RADEEQQAL) and one epitope conserved in both

EsxG and EsxS (LLDAHIPQL) that we identified in this study. These overlapping results confirm that MHC-I antigens identified by MS can be immunogenic in vivo in humans.

CD8⁺ T cell responses against epitopes derived from non-T7SS antigens such as Ag85B have previously been reported, including epitopes that are restricted by HLA alleles expressed by cells used in this study (Axelsson-Robertson et al., 2015). While we cannot rule out that these peptides went undetected in our study for technical reasons, such as low ionization efficiency, the fact that we were unable to detect these antigens despite using PathMHC to increase coverage of infection-specific peptides suggests that presentation of T7SS substrates is favored in hMDCs for biological reasons, and the low frequency of *Mtb*-derived MHC-I peptides reflects real bottlenecks in access to antigen-processing pathways. It is possible that CD8⁺ T cell responses against other classes of antigens could be primed by antigen-presenting cell types other than monocyte-derived phagocytes or by cells that have taken up dead *Mtb* material rather than live *Mtb*. If this were the case, CD8⁺ T cells targeting T7SS substrates might be more effective at recognizing the monocyte-derived cells that make up a large proportion of infected cells in TB granulomas (Lee et al., 2020).

Mtb infects a range of phagocytic human immune cells, including both macrophages and dendritic cells (Lee et al., 2020; Wolf et al., 2007). Hypothetically, these cell types might express and use different antigen-processing pathways and therefore present different sets of *Mtb* antigens on MHC-I. Heterogeneous antigen presentation would pose an obstacle to efficient clearance of the infection by CD8⁺ T cells, as T cells capable of recognizing one cell type might be unable to recognize the other. However, we previously showed that T7SS substrates predominate among *Mtb*-derived MHC-I antigens presented by human monocyte-derived macrophages (Leddy et al., 2023), and our results here show that the same is true of hMDCs. Some MHC-I peptides (LLDEGKQSL from EsxA and QEQASQQIL from EsxJKPW) and source proteins (EsxB and EspC) were identified in both studies. T cells primed to recognize T7SS substrates may be able to recognize both *Mtb*-infected monocyte-derived dendritic cells and macrophages, making T7SS substrates a promising class of potential vaccine targets.

The workflow underlying PathMHC can in principle be extended to immunopeptidomics in noninfectious disease areas and to other types of mixed host-pathogen proteomic samples. For example, a workflow analogous to PathMHC could potentially be used to identify MHC peptides uniquely presented following treatment of tumor cells with a therapy to nominate potential targets for combination immunotherapy (Stopfer et al., 2022; Stopfer et al., 2020) or could be used to specifically identify pathogen-derived interaction partners of a host protein of interest in immunoprecipitation (IP)-MS or AP-MS studies to better understand host-pathogen interactions (Nicod et al., 2017).

In addition to pathogen-derived peptides, PathMHC enriches for infection-associated MHC peptides derived from self-proteins upregulated during the host cell response to infection. While in general these infection-associated peptide-MHC complexes could in principle be targets for the design of therapies

and diagnostics, targeting the interferon-associated MHC peptides detected by PathMHC in *Mtb*-infected cells could result in low specificity and off-target effects since these peptides could be presented by any cell responding to interferon. Targeting these peptides using vaccines would likely be difficult, as dedicated mechanisms in the thymus ensure deletion of T cells specific for infection-induced self-MHC peptides (You et al., 2024). Using uninfected control samples treated with stimuli that mimic the host response to infection (such as recombinant type I interferon in the case of *Mtb*) could help ensure that a higher proportion of differentially presented antigens targeted by PathMHC are truly pathogen-specific peptides rather than infection-induced self-peptides. In other types of host-pathogen proteomic samples besides MHC peptides, infection-associated host proteins may be of greater biological interest and could reveal novel information about the host response to infection.

Our results show that computational enrichment of infection-specific precursor ions enables immunopeptidomics studies to take advantage of the specificity and sensitivity of targeted MS to more efficiently detect pathogen-derived peptides without prior knowledge of the identities of target epitopes. Selectively targeting infection-associated peptides in discovery immunopeptidomics studies can potentially make target discovery for T cell-priming vaccines more efficient across a range of infectious disease areas.

Materials and methods

Mtb culture

Mtb H37Rv was grown in Difco Middlebrook 7H9 media supplemented with 10% of a supplement containing oleic acid, albumin, dextrose, catalase (OADC), 0.2% glycerol, and 0.05% Tween-80 to mid-log phase.

Msmeg culture

Msmeg mc²155 was grown in Difco Middlebrook 7H9 media supplemented with 10% OADC, 0.2% glycerol, and 0.05% Tween-80 overnight (12–14 h).

Human cell isolation, differentiation, and culture

Buffy coats and leukapheresis samples were procured from the Mass General Brigham Blood Donor Center or STEMCELL Technologies. All samples are collected by StemCell or Mass General staff who are not members of the Bryson Laboratory or Ragon Institute. All specimens are considered not identifiable (truly anonymous). Samples are acquired and provided to research groups with no identifying information. PBMCs were isolated by density-based centrifugation using Ficoll (GE Healthcare). CD14⁺ monocytes were isolated from PBMCs using a CD14 positive-selection kit (Stemcell). Isolated monocytes were differentiated in R10 media (RPMI 1640 without phenol red [Gibco] supplemented with 10% heat-inactivated FBS [Gibco], 1% HEPES [Corning], and 1% L-glutamine [Sigma-Aldrich]) supplemented with 25 ng/ml GM-CSF (572902; BioLegend) and 25 ng/ml IL-4 (574006; BioLegend). Monocytes were cultured on tissue culture-treated T75 flasks (VWR) for 6 days. Media was replaced with fresh cytokine-containing media every 3 days.

HLA genotyping

Genomic DNA was extracted from 5×10^6 PBMCs using a Qiagen DNeasy kit. HLA typing was performed using a targeted next-generation sequencing method. Briefly, locus-specific primers were used to amplify a total of 26 polymorphic exons of HLA-A and B (exons 1–4), C (exons 1–5), E (exon 3), DPA1 (exon 2), DPB1 (exons 2–4), DQA1 (exons 1–3), DQB1 (exons 2 and 3), DRB1 (exons 2 and 3), and DRB3/4/5 (exon 2) genes with Fluidigm Access Array system (Fluidigm Corporation). The 26 Fluidigm PCR amplicons were harvested from Fluidigm Access Array IFC and pooled. Quality and quantity were checked using a Caliper LabChip GX Touch HT Nucleic Acid Analyzer (PerkinElmer). The PCR product library was quantitated and subjected to sequencing on an Illumina MiSeq sequencer (Illumina). HLA alleles and genotypes were called using the Omixon HLA Explore (version 2.0.0) software (Omixon Biocomputing Ltd.). HLA genotyping data for donor C was instead provided by StemCell.

Msmeg infection

The *Msmeg* culture was pelleted by centrifugation, washed once with PBS, and resuspended in R10 media. 25 million hMDCs were infected at a MOI of 10 for 2 h (or mock infected with media not containing *Msmeg*) and then washed with PBS to remove extracellular *Msmeg*. Infected hMDCs were cultured in R10 media for 24 h before harvesting.

Mtb infection

The *Mtb* culture was pelleted by centrifugation, washed once with PBS, resuspended in R10 media, and centrifuged at low speed (500 rpm for 5 min) to pellet clumps, leaving a uniform suspension of bacteria in the supernatant. 50 million hMDCs were infected at MOI 2.5 for 4 h and then washed with PBS to remove extracellular *Mtb*. Infected hMDCs were cultured in R10 media for 72 h before harvesting.

MHC IP

hMDCs were harvested by collecting the culture media (containing any cells in suspension), washing adherent cells with PBS, incubating remaining adherent cells with PBS supplemented with 4 mM EDTA for 15 min at 37°C, gently scraping with a cell scraper, collecting the detached cells, and washing the flask with PBS and collecting the wash. The harvested cells were then pelleted, washed with PBS, and lysed in 1 ml of MHC lysis buffer (20 mM Tris, 150 mM sodium chloride, pH 8.0, supplemented with 1% CHAPS, 1 × HALT protease and phosphatase inhibitor cocktail [Pierce], and 0.2 mM phenylmethylsulfonyl fluoride [Sigma-Aldrich]).

BSL3 protocol

Lysate from *Mtb*-infected (or mock-infected) cells was sonicated using a Q500 ultrasonic water bath sonicator (Qsonica) in five 30-s pulses at an amplitude of 60%, cleared by centrifugation at $16,000 \times g$ for 5 min, and sterile filtered twice using 0.2- μ m filter cartridges (Pall NanoSep).

BSL2 protocol

Lysate from *Msmeg*-infected (or mock-infected) cells was sonicated using a VCX-130 probe ultrasonic processor (Sonics) in

three 10-s pulses at an amplitude of 30% and cleared by centrifugation at $16,000 \times g$ for 5 min.

Lysates were then added to protein A sepharose beads pre-conjugated with 0.5 mg of pan-MHC-I antibody (clone W6/32) or 0.1 mg of pan-MHC-II antibody (clone Tü39), prepared as previously described (Stopfer et al., 2020), and incubated rotating at 4°C overnight (12–14 h). Beads were then washed, and peptide-MHC complexes were eluted as previously described (Stopfer et al., 2020).

MHC-I peptide isolation

C18 SpinTips (Protea) were washed with 0.1% trifluoroacetic acid, activated with 90% acetonitrile supplemented with 0.1% formic acid, and washed with 0.1% formic acid. Eluate from MHC-I IPs was applied to the column by centrifugation. The column was washed with 0.1% formic acid, and peptides were eluted by applying elution solvent (28% acetonitrile with 0.1% formic acid) by centrifugation twice. Eluates were snap-frozen in liquid nitrogen and lyophilized.

MHC-II peptide isolation

MHC-II-associated peptides were purified using 10-kDa molecular weight cutoff filters (Cytiva NanoSep) as previously described (Stopfer et al., 2020), snap-frozen in liquid nitrogen, and lyophilized.

MS analyses

For all MS analyses, samples were analyzed using an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific) coupled with an UltiMate 3000 RSLC Nano LC system (Dionex), Nanospray Flex ion source (Thermo Fisher Scientific), and column oven heater (Sonation). The MHC peptide sample was loaded onto a fused silica capillary chromatography column with an integrated electrospray tip (~1- μ m orifice) prepared and packed in-house with 10 cm of 1.9- μ m C18 beads (ReproSil-Pur). MHC peptide samples were resuspended in 0.1% formic acid and loaded at a flow rate of 300 nl/min. Peptides were eluted using a flow rate of 100 nl/min.

Standard MS parameters were as follows: spray voltage, 2.0 kV; no sheath or auxiliary gas flow; ion transfer tube temperature, 275°C. MHC-I peptides were eluted using a gradient of 6–25% buffer B (70% acetonitrile and 0.1% formic acid) over 75 min, 25–45% over 5 min, 45–100% over 5 min, hold for 1 min, and 100–3% over 2 min. MHC-II peptides were eluted using a gradient of 6–12% buffer B (70% acetonitrile and 0.1% formic acid) over 10 min, 12–20% over 45 min, 20–25% over 10 min, 25–45% over 5 min, 45–97% over 2 min, hold for 1 min, and 97–3% over 2 min. Full scan mass spectra (120,000 resolution, 350–1,200 m/z for MHC-I, 350–1,800 m/z for MHC-II) were detected in the orbitrap analyzer after accumulation of 3×10^6 ions (normalized automatic gain control [AGC] target of 300%) or 25 ms. For every full scan, MS/MS scans were collected during a 3-s cycle time. Ions were isolated (0.4 m/z isolation width) using the standard AGC target and automatic determination of maximum injection time, fragmented by HCD with 30% collision energy (CE), and scanned at a resolution of 120,000. Charge states <2 and >4 were excluded. In DDA analyses, precursors were excluded

from selection for 30 s if fragmented $n = 2$ times within a 20-s window.

For PathMHC PRM analyses, inclusion list masses were matched by m/z and charge state, with a mass tolerance of 5 ppm. Acquisition was scheduled for a window of 3 min centered on the observed retention time of the target precursor ion in the DDA analysis in which it was originally detected. Precursors were excluded from selection for 4 s after being fragmented one time.

For analyses of MHC-II peptides from hMDCs infected with *Msmeg*, 50% of a given sample was injected for each DDA analysis and 50% for a follow-up PRM or DDA analysis. For analyses of MHC-I peptides from hMDCs infected with *Mtb*, 40% of the sample was used for each analysis, while 20% was reserved for subsequent validation using SureQuant.

DDA only MHC-I analyses of MHC-I peptides from *Mtb*-infected hMDCs were performed without prior offline fractionation (donor E) or with offline fractionation by reversed-phase HPLC as previously described (Leddy et al., 2023) (donor D).

PathMHC computational pipeline

The PathMHC pipeline first identifies MS1 peaks (features) using Dinosaur (version 1.2.0), then aligns these features between a sample from infected cells and a mock-infected control using DeepRTAlign (version 1.1.3). MS1 features detected in the infected sample are excluded if they are mapped to a feature in the mock-infected control by DeepRTAlign or if they match a peptide already identified by DDA from a provided list of PSMs to within a mass tolerance of 10 ppm and retention time tolerance defined by the peak boundaries set by Dinosaur. The remaining MS1 features (i.e., those not previously identified by DDA and with no corresponding peak in the mock-infected control) are then scored using AutoMS, using an extracted ion chromatogram (XIC) length of 40 s and an XIC mass tolerance of 10 ppm for MHC-I peptides and 40 ppm for MHC-II peptides. A final inclusion list is generated consisting of precursors meeting the following criteria: For MHC-I, charge state 2–4, retention time <90 min (from the end of sample loading), and AutoMS SNR >1 or AutoMS score >0.2. For MHC-II, charge state 2–3, retention time <90 min, and AutoMS SNR >1.5 or AutoMS score >0.3.

PD PathMHC computational pipeline

To implement an analog of our PathMHC computational pipeline in PD, we created a workflow in which LC-MS features are detected using Minora in the processing step and aligned between the infected and mock-infected samples using the Feature Mapper (Perform RT Alignment = True, Max RT Shift = 10 min, Min S/N Threshold = 2) and Precursor Ions Quantifier nodes in the consensus step. Consensus features are then exported with the following filters applied: charge state = 2 or 3, intensity in mock-infected control sample = 0, and average apex retention time is between 25 and 115 min (should be set depending on your LC gradient and duration of sample loading). The resulting table of precursor ions is reformatted for use as an inclusion list for PRM (see <https://github.com/oleddy/pathMS>).

MS data search and manual inspection

All mass spectra were analyzed with PD (version 3.0) and searched using Sequest with rescoring using INFERYS (Zolg et al., 2021) and Percolator (Brosch et al., 2009) against a custom database comprising the UniProt human proteome (UP000005640) together with either the UniProt *Mtb* H37Rv proteome (UP000001584) or the *Msmeg* mc²155 proteome (UP000006158) (Datas S4 and S5). Each MS run was searched individually. No enzyme was used, and variable modifications included oxidized methionine for all analyses. PSMs from proteomeMHC-I analyses were filtered with the following criteria: search engine rank = 1, length between 8 and 13 amino acids, XCorr ≥ 2.0 , spectral angle ≥ 0.6 , and percolator q-value <0.05. PSMs from MHC-II analyses were filtered with the following criteria: search engine rank = 1, length between 8 and 30 amino acids, XCorr ≥ 2.0 , spectral angle ≥ 0.7 , and percolator q-value <0.05. Bacterial peptides (*Msmeg* or *Mtb*) were checked against the human proteome using the UniProt peptide search tool (<https://www.uniprot.org/peptide-search>), with leucine and isoleucine treated as equivalent to ensure that no human sequences were identified that might differ only by an isobaric substitution.

IDs of putative *Mtb*-derived peptides were rejected if any PSMs for the same peptide were found in the unfiltered DDA MS data for the corresponding mock-infected control. For each putative *Mtb* peptide identified, MS/MS spectra and XIC were manually inspected, and the ID was only accepted for further validation if it met the following criteria: (1) MS/MS spectra contained enough fragment ion peaks separated from one another by the mass of a single amino acid that the majority of amino acids in the peptide sequence could be unambiguously assigned; (2) neutral losses were consistent with the chemical properties of the peptide; (3) manual inspection of MS/MS spectra did not reveal an alternate peptide sequence explaining a greater number of MS/MS peaks (i.e., the MS/MS spectrum did not contain unassigned peaks separated from assigned peaks by an amino acid mass such that including unassigned peaks in either the y-ion or b-ion series would imply a different peptide sequence and result in a greater number of assigned MS/MS spectrum peaks); (4) XIC showed a peak in MS intensity at the m/z of the peptide precursor ion at the retention time at which it was identified that did not appear in the corresponding mock-infected control. Peptides that met these criteria were further validated using SureQuant (see below).

Synthetic standard survey MS analyses

DDA MS analysis of the SIL peptide mixture was performed as described above (see MS analyses) with the following modifications: Peptides were eluted using a gradient of 6–35% buffer B over 30 min, 35–45% over 2 min, 45–100% over 3 min, and 100–2% over 1 min. No dynamic exclusion was used.

A second set of survey analyses was performed on the mixture of SIL peptides with background matrix using the full SureQuant acquisition method (see below). SIL peptides were spiked into a mixture of MHC-I peptides purified as described above from THP-1 cells differentiated into macrophages via 24 h of treatment with 150 nM PMA, which provided a representative background matrix. Because SIL amino acids are not 100% pure, SIL peptide concentrations were adjusted, and survey analyses

were repeated until the SIL peptide could be reliably detected while minimizing background signal detected at the mass of the biological peptide. In the final SIL standard mixture, 100 fmol of each peptide was used per sample for all epitopes except two peptides for which 10 fmol was used per sample (KIYSEA-DEAWRK and RADEEQQAL).

Validation of candidate *Mtb* peptides by SureQuant

MS analysis

SIL synthetic peptide standards were ordered from Biosynth as a crude peptide library. Standard MS parameters and MS1 scan parameters were as described above, except that a scan range of 380–1,200 *m/z* was used. The custom SureQuant method was built based on the template provided in the Thermo Orbitrap Exploris Series 2.0 method editor. After the optimal charge state and most intense product ions were determined via a survey analysis of the synthetic SIL peptide standards alone (see above), a method branch was created for each *m/z* offset between the SIL peptide and biological peptide as previously described (Stopfer et al., 2021). The *m/z* tolerance for detection of SIL standards was set at 10 ppm. A resolution of 15,000 and automatically determined ion accumulation time was used for MS/MS scans of the SIL standards, and a resolution of 240,000 with a maximum accumulation time of 1 s was used for MS/MS scans of the biological peptide. An MS/MS scan of the light (biological) peptide was triggered upon detection of three or more of the top 6 product ions (mass tolerance 20 ppm) in a scan of the corresponding SIL standard. Results were analyzed using Skyline Daily Build 24.1.1.202.

Analysis of publicly available immunopeptidomics datasets

Raw LC-MS/MS data underlying MHC-I immunopeptidomics datasets from *Listeria*-infected cells and SARS-CoV-2-infected cells generated by Mayer et al. (2022) and Weingarten-Gabbay et al. (2021), respectively, were obtained from their respective submissions to the ProteomeXchange PRIDE database. In the case of Mayer et al.'s dataset, for each biological replicate raw, MS data from *Listeria*-infected and uninfected cells was analyzed using the PathMHC computational pipeline as described above, with the exception that precursor ions with existing PSMs were not excluded. Putatively infection-specific precursor ion peaks selected by the pipeline were matched with reported PSMs with an *m/z* tolerance of 10 ppm. The proportion of putative *Listeria*-derived MHC-I peptides among identified precursors selected by the pipeline was compared with the total proportion of *Listeria*-derived peptides in the dataset. One of four provided replicates was excluded because one of the files provided could not be analyzed by Dinosaur, possibly due to an incompatible file format or corruption of the file. The data from Weingarten-Gabbay et al. (2021) were analyzed similarly, but in this dataset each replicate represented paired peptide fractions from offline fractionation rather than biological replicates. For this dataset, the proportion of SARS-CoV-2-derived peptides enriched by the PathMHC computational pipeline was compared with the overall proportion of SARS-CoV-2 peptides for each replicate, as PSMs were reported separately for each fraction. In addition to a 10 ppm *m/z* tolerance, a 1.5-min retention time tolerance was used, as retention times for each PSM were also reported.

Gene set enrichment analysis

Self-MHC-I peptides uniquely identified in PathMHC PRM analyses of *Mtb*-infected hMDCs that had not been identified by DDA were mapped to their source proteins. This set of proteins was analyzed using Enrichr (Chen et al., 2013; Kuleshov et al., 2016), using the set of source proteins of self-MHC-I peptides identified by DDA as a background. The MSigDB Hallmark 2020 pathway database was used for pathway enrichment analysis (Liberzon et al., 2011). Pathway enrichment analyses of self-peptides based on computational analyses of publicly available immunopeptidomics datasets using the PathMHC computational pipeline were performed analogously. The precursors selected by PathMHC that had associated PSMs were mapped to their source proteins, and these proteins were analyzed using Enrichr as described above, using the overall set of proteins identified in the analysis as a background.

Protein secretion pathway enrichment analysis

To determine whether *Mtb*-derived MHC-I peptides or their source proteins were enriched for proteins with specific secretion signals, we classified each protein in the reference *Mtb* H37Rv proteome (UP000001584) using SignalP 6.0 (Teufel et al., 2022), and added an additional class of known T7SS substrates (all Esx family proteins; all PE/PPE proteins; and EspA, EspB, EspC, EspE, EspF, EspI, EspJ, and EspK). Under a null model in which all possible *Mtb* peptides are equally likely to be presented, the probability of drawing a peptide from a given class of proteins is approximately the sum of the lengths of the amino acid sequences of proteins in that class divided by the total number of amino acid residues in the *Mtb* proteome. Using this approximation, we used a binomial test to assess whether the number of *Mtb*-derived peptides from each class of proteins detected in the MHC-I was greater or less than would be predicted under the null model. We used a hypergeometric test to assess whether the number of source proteins from each class was greater or less than expected, relative to a null model in which each protein in the *Mtb* proteome is equally likely to contribute to the MHC-I repertoire.

Online supplemental material

Fig. S1 shows the benchmarking of our open-source PathMHC computational pipeline and an analogous pipeline on MHC-II peptides from *Msmeg*-infected hMDCs, as well as an analysis of instrument cycle time usage, sensitivity, and overlap in peptide IDs between PRM and DDA runs. Fig. S2 shows a comparison between motifs identified by Gibbs clustering of MHC-I peptides isolated from *Mtb*-infected hMDCs and the published motifs for the corresponding HLA alleles. Fig. S3 shows the NetMHCpan-binding predictions for *Mtb* MHC-I peptides, for HLA alleles expressed by each donor. Fig. S4 shows the XIC of self-MHC peptides identified in DDA analyses, comparing infected and uninfected cells. Table S1 lists the class I HLA genotype of each donor for experiments involving *Mtb*-infected cells. Table S2 lists the putative *Mtb* peptide IDs that passed manual inspection but failed SureQuant validation. Data S1 provides a table of all PSMs for all MS analyses of MHC-II peptides from *Msmeg*-infected hMDCs. Data S2 provides a table of all PSMs for all

MS analyses of MHC-I peptides from *Mtb*-infected hMDCs. Data S3 provides a sample guide indicating which raw MS files (submitted to the PRIDE database—see Data availability) correspond to each experiment. Data S4 provides the proteomic database against which MS data from *Msmeg*-infected cells were searched, and Data S5 provides the proteomic database against which MS data from *Mtb*-infected cells were searched. Data S6 provides the SureQuant data (mirror plots and chromatograms comparing biological peptides with synthetic SIL standards) validating *Mtb* MHC-I peptide IDs.

Data availability

Raw MS data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2022) with the dataset identifier PXD055726. A sample guide detailing which raw data files correspond to each analysis presented in the paper is provided in Data S3, and the proteomic databases used for MS data searches are provided in Datas S4 and S5. Our computational pipeline for PathMHC and all other code used in this study is available at <https://github.com/oleddy/pathMS>.

Acknowledgments

The authors thank all the staff members of the Ragon Institute, the Koch Institute, and MIT for the essential work they do to make our research possible. The Tü39 hybridoma used for in-house antibody production was generously provided by Hans-Georg Rammensee and Claudia Falkenburger [University of Tübingen, Tübingen, Germany]. We thank Lauren Stopfer, Cameron Flower, Tigist Tamir, Elizabeth Choe, and Ryuhjin Ahn for helpful conversations, training, and technical guidance. We modified code written by Cameron Flower to plot MS/MS spectra. We thank the developers of Dinosaur, DeepRTalign, and AutoMS for making their software free and open-source, as well as the developers of other Python and R libraries used in this study. This work was performed in part in the Ragon Institute BSL3 core facility, which is supported by the National Institutes of Health (NIH)-funded Harvard University Center for AIDS Research (P30 AI060354). We thank Yong Xie, Julie Boucau, and Eliane Shwairi for managing the facility.

This work is supported by funding from the MIT Center for Precision Cancer Research at the Koch Institute and NIH grants U01 CA238720, U54 CA283114, 1R35GM142900, and R01AI022553. O. Leddy is supported by US National Institute of Environmental Health Sciences grant T32-ES007020.

Author contributions: O. Leddy: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, and writing—original draft, review, and editing. Y. Yuki: investigation. M. Carrington: investigation and writing—review and editing. B.D. Bryson: conceptualization, funding acquisition, project administration, resources, supervision, and writing—original draft, review, and editing. F.M. White: conceptualization, funding acquisition, project administration, supervision, and writing—review and editing.

Disclosures: O. Leddy reported a patent to polyepitope TB vaccine, US patent application number 63/693,081 pending. B.D.

Bryson reported a patent to *Mycobacterium tuberculosis* vaccines and methods of use thereof pending. No other disclosures were reported.

Submitted: 26 February 2025

Revised: 22 May 2025

Accepted: 30 June 2025

References

- Acosta Rodríguez, E.V., C.L. Araujo Furlan, F. Fiocca Vernengo, C.L. Montes, and A. Gruppi. 2019. Understanding CD8⁺ T cell immunity to trypanosoma cruzi and how to improve it. *Trends Parasitol.* 35:899–917. <https://doi.org/10.1016/j.pt.2019.08.006>
- Akter, S., K.S. Chauhan, M.D. Dunlap, J.A. Choreño-Parra, L. Lu, E. Esaulova, J. Zúñiga, M.N. Artyomov, D. Kaushal, and S.A. Khader. 2022. *Mycobacterium tuberculosis* infection drives a type I IFN signature in lung lymphocytes. *Cell Rep.* 39:110983. <https://doi.org/10.1016/j.celrep.2022.110983>
- Andreatta, M., B. Alvarez, and M. Nielsen. 2017. GibbsCluster: Unsupervised clustering and alignment of peptide sequences. *Nucleic Acids Res.* 45: W458–W463. <https://doi.org/10.1093/nar/gkx248>
- Andreatta, M., O. Lund, and M. Nielsen. 2013. Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach. *Bioinformatics.* 29:8–14. <https://doi.org/10.1093/bioinformatics/bts621>
- Axelsson-Robertson, R., J.H. Ju, H.-Y. Kim, A. Zumla, and M. Maeurer. 2015. *Mycobacterium tuberculosis*-specific and MHC class I-restricted CD8⁺ T-cells exhibit a stem cell precursor-like phenotype in patients with active pulmonary tuberculosis. *Int. J. Infect. Dis.* 32:13–22. <https://doi.org/10.1016/j.ijid.2014.12.017>
- Barczak, A.K., R. Avraham, S. Singh, S.S. Luo, W.R. Zhang, M.-A. Bray, A.E. Hinman, M. Thompson, R.M. Nietupski, A. Golas, et al. 2017. Systematic, multiparametric analysis of *Mycobacterium tuberculosis* intracellular infection offers insight into coordinated virulence. *PLoS Pathog.* 13:e1006363. <https://doi.org/10.1371/journal.ppat.1006363>
- BARDA. 2022. BARDA Strategic Plan 2022–2026. <https://faolex.fao.org/docs/pdf/us218400.pdf> (accessed July 16, 2025).
- Benoun, J.M., N.G. Peres, N. Wang, O.H. Pham, V.L. Rudisill, Z.N. Fogassy, P.G. Whitney, D. Fernandez-Ruiz, T. Gebhardt, Q.-M. Pham, et al. 2018. Optimal protection against *Salmonella* infection requires noncirculating memory. *Proc. Natl. Acad. Sci. USA.* 115:10416–10421. <https://doi.org/10.1073/pnas.1808339115>
- Bhalla, M., D. Law, G.C. Dowd, and K. Ireton. 2017. Host serine/threonine kinases mTOR and protein kinase C- α promote InlB-mediated entry of *Listeria monocytogenes*. *Infect. Immun.* 85:e00087-17. <https://doi.org/10.1128/IAI.00087-17>
- Boradia, V., A. Frando, and C. Grundner. 2022. The *Mycobacterium tuberculosis* PE15/PPE20 complex transports calcium across the outer membrane. *PLoS Biology.* 20:e3001906. <https://doi.org/10.1371/journal.pbio.3001906>
- Brosch, M., L. Yu, T. Hubbard, and J. Choudhary. 2009. Accurate and sensitive peptide identification with mascot percolator. *J. Proteome Res.* 8: 3176–3181. <https://doi.org/10.1021/pr800982s>
- Brown, A.F., A.G. Murphy, S.J. Lalor, J.M. Leech, K.M. O’Keeffe, M. Mac Aogáin, D.P. O’Halloran, K.A. Lacey, M. Tavakol, C.H. Hearnden, et al. 2015. Memory Th1 cells are protective in invasive *Staphylococcus aureus* infection. *PLoS Pathog.* 11:e1005226. <https://doi.org/10.1371/journal.ppat.1005226>
- Carter, B., P. Huang, G. Liu, Y. Liang, P.J.C. Lin, B.-H. Peng, L.G.A. McKay, A. Dimitrakakis, J. Hsu, V. Tat, et al. 2023. A pan-variant mRNA-LNP T cell vaccine protects HLA transgenic mice from mortality after infection with SARS-CoV-2 Beta. *Front. Immunol.* 14:1135815. <https://doi.org/10.3389/fimmu.2023.1135815>
- Chen, E.Y., C.M. Tan, Y. Kou, Q. Duan, Z. Wang, G.V. Meirelles, N.R. Clark, and A. Ma’ayan. 2013. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics.* 14:128. <https://doi.org/10.1186/1471-2105-14-128>
- Chen, K., J.P. McAleer, Y. Lin, D.L. Paterson, M. Zheng, J.F. Alcorn, C.T. Weaver, and J.K. Kolls. 2011. Th17 cells mediate clade-specific, serotype-independent mucosal immunity. *Immunity.* 35:997–1009. <https://doi.org/10.1016/j.immuni.2011.10.018>
- Flynn, J.L., and J. Chan. 2022. Immune cell interactions in tuberculosis. *Cell.* 185:4682–4702. <https://doi.org/10.1016/j.cell.2022.10.025>

- Iwanaga, N., K. Chen, H. Yang, S. Lu, J.P. Hoffmann, A. Wanek, J.E. McCombs, K. Song, J. Rangel-Moreno, E.B. Norton, and J.K. Kolls. 2021. Vaccine-driven lung TRM cells provide immunity against *Klebsiella* via fibroblast IL-17R signaling. *Sci. Immunol.* 6:eabf1198. <https://doi.org/10.1126/sciimmunol.abf1198>
- Ji, H., and J. Tian. 2022. Deep denoising autoencoder-assisted continuous scoring of peak quality in high-resolution LC-MS data. *Chemom. Intell. Lab. Syst.* 231:104694. <https://doi.org/10.1016/j.chemolab.2022.104694>
- Jonas, O.B., A. Irwin, F.C.J. Berthe, F.G. Le Gall, P.V. Marquez. 2017. Drug-resistant infections: A threat to our economic future. <https://documents.worldbank.org/curated/en/323311493396993758>
- Kinsella, R.L., D.X. Zhu, G.A. Harrison, A.E. Mayer Bridwell, J. Prusa, S.M. Chavez, and C.L. Stallings. 2021. Perspectives and advances in the understanding of tuberculosis. *Annu. Rev. Pathol.* 16:377–408. <https://doi.org/10.1146/annurev-pathol-042120-032916>
- Kuleshov, M.V., M.R. Jones, A.D. Rouillard, N.F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S.L. Jenkins, K.M. Jagodnik, A. Lachmann, et al. 2016. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44:W90–W97. <https://doi.org/10.1093/nar/gkw377>
- La Gruta, N.L., and S.J. Turner. 2014. T cell mediated immunity to influenza: Mechanisms of viral control. *Trends Immunol.* 35:396–402. <https://doi.org/10.1016/j.it.2014.06.004>
- Law, J.C., and T.H. Watts. 2023. Considerations for choosing T cell assays during a pandemic. *J. Immunol.* 211:169–174. <https://doi.org/10.4049/jimmunol.2300129>
- Leddy, O., Y. Cui, R. Ahn, L. Stopfer, E. Choe, D.H. Kim, M. Roerden, S. Spranger, B.D. Bryson, and F.M. White. 2025. Validation and quantification of peptide antigens presented on MHCs using SureQuant. *Nat. Protoc.* 20:1196–1222. <https://doi.org/10.1038/s41596-024-01076-x>
- Leddy, O., F.M. White, and B.D. Bryson. 2023. Immunopeptidomics reveals determinants of *Mycobacterium tuberculosis* antigen presentation on MHC class I. *Elife*. 12:e84070. <https://doi.org/10.7554/eLife.84070>
- Leddy, O.K., F.M. White, and B.D. Bryson. 2021. Leveraging immunopeptidomics to study and combat infectious disease. *mSystems*. 6:e0031021. <https://doi.org/10.1128/msystems.00310-21>
- Lee, J., S. Boyce, J. Powers, C. Baer, C.M. Sasseti, and S.M. Behar. 2020. CD11cHi monocyte-derived macrophages are a major cellular compartment infected by *Mycobacterium tuberculosis*. *PLoS Pathog.* 16:e1008621. <https://doi.org/10.1371/journal.ppat.1008621>
- Lei, X., X. Dong, R. Ma, W. Wang, X. Xiao, Z. Tian, C. Wang, Y. Wang, L. Li, L. Ren, et al. 2020. Activation and evasion of type I interferon responses by SARS-CoV-2. *Nat. Commun.* 11:3810. <https://doi.org/10.1038/s41467-020-17665-9>
- Lewinsohn, D.A., E. Winata, G.M. Swarbrick, K.E. Tanner, M.S. Cook, M.D. Null, M.E. Cansler, A. Sette, J. Sidney, and D.M. Lewinsohn. 2007. Immunodominant tuberculosis CD8 antigens preferentially restricted by HLA-B. *PLoS Pathog.* 3:1240–1249. <https://doi.org/10.1371/journal.ppat.0030127>
- Liberzon, A., A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, and J.P. Mesirov. 2011. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 27:1739–1740. <https://doi.org/10.1093/bioinformatics/btr260>
- Liu, Y., Y. Yang, W. Chen, F. Shen, L. Xie, Y. Zhang, Y. Zhai, F. He, Y. Zhu, and C. Chang. 2023. DeepRTAlign: Toward accurate retention time alignment for large cohort mass spectrometry data analysis. *Nat. Commun.* 14: 8188. <https://doi.org/10.1038/s41467-023-43909-5>
- Magdeldin, S., J.J. Moresco, T. Yamamoto, and J.R. Yates. 2014. Off-line multidimensional liquid chromatography and auto sampling result in sample loss in LC/LC-MS/MS. *J. Proteome Res.* 13:3826–3836. <https://doi.org/10.1021/pr500530e>
- Mallajosyula, V., S. Chakraborty, E. Sola, R.F. Fong, V. Shankar, F. Gao, A.R. Burrell, N. Gupta, L.E. Wagar, P.S. Mischel, et al. 2024. Coupling antigens from multiple subtypes of influenza can broaden antibody and T cell responses. *Science*. 386:1389–1395. <https://doi.org/10.1126/science.ad2396>
- Mayer, R.L., and F. Impens. 2021. Immunopeptidomics for next-generation bacterial vaccine development. *Trends Microbiol.* 29:1034–1045. <https://doi.org/10.1016/j.tim.2021.04.010>
- Mayer, R.L., R. Verbeke, C. Asselman, I. Aernout, A. Gul, D. Eggermont, K. Boucher, F. Thery, T.M. Maia, H. Demol, et al. 2022. Immunopeptidomics-based design of mRNA vaccine formulations against *Listeria monocytogenes*. *Nat. Commun.* 13:6075. <https://doi.org/10.1038/s41467-022-33721-y>
- McElhaney, J.E., D. Xie, W.D. Hager, M.B. Barry, Y. Wang, A. Kleppinger, C. Ewen, K.P. Kane, and R.C. Bleackley. 2006. T cell responses are better correlates of vaccine protection in the elderly. *J. Immunol.* 176: 6333–6339. <https://doi.org/10.4049/jimmunol.176.10.6333>
- McMahan, K., J. Yu, N.B. Mercado, C. Loos, L.H. Tostanoski, A. Chandra-shekar, J. Liu, L. Peter, C. Atyeo, A. Zhu, et al. 2021. Correlates of protection against SARS-CoV-2 in rhesus macaques. *Nature*. 590:630–634. <https://doi.org/10.1038/s41586-020-03041-6>
- Murray, C.J.L., K.S. Ikuta, F. Sharara, L. Swetschinski, G. Robles Aguilar, A. Gray, C. Han, C. Bisignano, P. Rao, E. Wool, et al. 2022. Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. *The Lancet*. 399:629–655. [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0)
- Nicod, C., A. Banaei-Esfahani, and B.C. Collins. 2017. Elucidation of host-pathogen protein-protein interactions to uncover mechanisms of host cell rewiring. *Curr. Opin. Microbiol.* 39:7–15. <https://doi.org/10.1016/j.mib.2017.07.005>
- Nyendak, M., G.M. Swarbrick, A. Duncan, M. Cansler, E.W. Huff, D. Hokey, T. Evans, L. Barker, G. Blatner, J. Sadoff, et al. 2016. Adenovirally-induced polyfunctional T cells do not necessarily recognize the infected target: Lessons from a phase I trial of the AERAS-402 vaccine. *Sci. Rep.* 6:36355. <https://doi.org/10.1038/srep36355>
- Pak, H., J. Michaux, F. Huber, C. Chong, B.J. Stevenson, M. Müller, G. Coukos, and M. Bassani-Sternberg. 2021. Sensitive immunopeptidomics by leveraging available large-scale multi-HLA spectral libraries, data-independent acquisition, and MS/MS prediction. *Mol. Cell. Proteomics* 20:100080. <https://doi.org/10.1016/j.mcpro.2021.100080>
- Perez-Riverol, Y., J. Bai, C. Bandla, D. García-Seisdedos, S. Hewapathirana, S. Kamatchinathan, D.J. Kundu, A. Prakash, A. Frericks-Zipper, M. Eisenacher, et al. 2022. The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* 50: D543–D552. <https://doi.org/10.1093/nar/gkab1038>
- Prensner, J.R., J.G. Abelin, L.W. Kok, K.R. Clauser, J.M. Mudge, J. Ruiz-Orera, M. Bassani-Sternberg, R.L. Moritz, E.W. Deutsch, and S. Van Heesch. 2023. What can ribo-seq, immunopeptidomics, and proteomics tell us about the noncanonical proteome? *Mol. Cell. Proteomics* 22:100631. <https://doi.org/10.1016/j.mcpro.2023.100631>
- Reynisson, B., B. Alvarez, S. Paul, B. Peters, and M. Nielsen. 2020. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* 48:W449–W454. <https://doi.org/10.1093/nar/gkaa379>
- Rousseau, M., L.L. Mariano, T. Canton, and M.A. Ingersoll. 2023. Tissue-resident memory T cells mediate mucosal immunity to recurrent urinary tract infection. *Sci. Immunol.* 8:eabn4332. <https://doi.org/10.1126/sciimmunol.abn4332>
- Sette, A., J. Sidney, and S. Crotty. 2023. T cell responses to SARS-CoV-2. *Annu. Rev. Immunol.* 41:343–373. <https://doi.org/10.1146/annurev-immunol-101721-061120>
- Shahbazy, M., S.H. Ramarathnam, P.T. Illing, E.C. Jappe, P. Faridi, N.P. Croft, and A.W. Purcell. 2023. Benchmarking bioinformatics pipelines in data-independent acquisition mass spectrometry for immunopeptidomics. *Mol. Cell. Proteomics* 22:100515. <https://doi.org/10.1016/j.mcpro.2023.100515>
- Siegrist, M.S., M. Unnikrishnan, M.J. McConnell, M. Borowsky, T.-Y. Cheng, N. Siddiqi, S.M. Fortune, D.B. Moody, and E.J. Rubin. 2009. Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. *Proc. Natl. Acad. Sci. USA*. 106:18792–18797. <https://doi.org/10.1073/pnas.0900589106>
- Stanley, S.A., S. Raghavan, W.W. Hwang, and J.S. Cox. 2003. Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proc. Natl. Acad. Sci. USA*. 100:13001–13006. <https://doi.org/10.1073/pnas.2235593100>
- Stopfer, L.E., A.S. Gajadhar, B. Patel, S. Gallien, D.T. Frederick, G.M. Boland, R.J. Sullivan, and F.M. White. 2021. Absolute quantification of tumor antigens using embedded MHC-I isotopologue calibrants. *Proc. Natl. Acad. Sci. USA*. 118:e2111173118. <https://doi.org/10.1073/pnas.2111173118>
- Stopfer, L.E., J.M. Mesfin, B.A. Joughin, D.A. Lauffenburger, and F.M. White. 2020. Multiplexed relative and absolute quantitative immunopeptidomics reveals MHC I repertoire alterations induced by CDK4/6 inhibition. *Nat. Commun.* 11:2760. <https://doi.org/10.1038/s41467-020-16588-9>
- Stopfer, L.E., N.J. Rettko, O. Leddy, J.M. Mesfin, E. Brown, S. Winski, B. Bryson, J.A. Wells, and F.M. White. 2022. MEK inhibition enhances presentation of targetable MHC-I tumor antigens in mutant melanomas. *Proc. Natl. Acad. Sci. USA*. 119:e2208900119. <https://doi.org/10.1073/pnas.2208900119>
- Teleman, J., A. Chawade, M. Sandin, F. Levander, and J. Malmström. 2016. Dinosaurs: A refined open-source peptide MS feature detector. *J. Proteome Res.* 15:2143–2151. <https://doi.org/10.1021/acs.jproteome.6b00016>

- Teufel, F., J.J. Almagro Armenteros, A.R. Johansen, M.H. Gíslason, S.I. Pihl, K.D. Tsirigos, O. Winther, S. Brunak, G. von Heijne, and H. Nielsen. 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* 40:1023–1025. <https://doi.org/10.1038/s41587-021-01156-3>
- Tully, G., C. Kortsik, H. Höhn, I. Zehbe, W.E. Hitzler, C. Neukirch, K. Freitag, K. Kayser, and M.J. Maeurer. 2005. Highly focused T cell responses in latent human pulmonary *Mycobacterium tuberculosis* infection. *J. Immunol.* 174:2174–2184. <https://doi.org/10.4049/jimmunol.174.4.2174>
- Vita, R., S. Mahajan, J.A. Overton, S.K. Dhanda, S. Martini, J.R. Cantrell, D.K. Wheeler, A. Sette, and B. Peters. 2019. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res.* 47:D339–D343. <https://doi.org/10.1093/nar/gky1006>
- Wang, Y., B. Jiang, Y. Guo, W. Li, Y. Tian, G.F. Sonnenberg, J.N. Weiser, X. Ni, and H. Shen. 2017. Cross-protective mucosal immunity mediated by memory Th17 cells against *Streptococcus pneumoniae* lung infection. *Mucosal Immunol.* 10:250–259. <https://doi.org/10.1038/mi.2016.41>
- Weingarten-Gabbay, S., S. Klaeger, S. Sarkizova, L.R. Pearlman, D.-Y. Chen, K.M.E. Gallagher, M.R. Bauer, H.B. Taylor, W.A. Dunn, C. Tarr, et al. 2021. Profiling SARS-CoV-2 HLA-I peptidome reveals T cell epitopes from out-of-frame ORFs. *Cell.* 184:3962–3980.e17. <https://doi.org/10.1016/j.cell.2021.05.046>
- Wolf, A.J., B. Linas, G.J. Trevejo-Núñez, E. Kincaid, T. Tamura, K. Takatsu, and J.D. Ernst. 2007. *Mycobacterium tuberculosis* infects dendritic cells with high frequency and impairs their function in vivo. *J. Immunol.* 179: 2509–2519. <https://doi.org/10.4049/jimmunol.179.4.2509>
- Wu, K., D. Dong, H. Fang, F. Levillain, W. Jin, J. Mei, B. Gicquel, Y. Du, K. Wang, Q. Gao, et al. 2012. An interferon-related signature in the transcriptional core response of human macrophages to *Mycobacterium tuberculosis* infection. *PLoS One.* 7:e38367. <https://doi.org/10.1371/journal.pone.0038367>
- You, Y., J. Dunst, K. Ye, P.A. Sandoz, A. Reinhardt, I. Sandrock, N.R. Comet, R.D. Sarkar, E. Yang, E. Duprez, et al. 2024. Direct presentation of inflammation-associated self-antigens by thymic innate-like T cells induces elimination of autoreactive CD8⁺ thymocytes. *Nat. Immunol.* 25: 1367–1382. <https://doi.org/10.1038/s41590-024-01899-6>
- Zolg, D.P., S. Gessulat, C. Paschke, M. Graber, M. Rathke-Kuhnert, F. Seefried, K. Fitzemeier, F. Berg, D. Lopez-Ferrer, D. Horn, et al. 2025. INFERYS rescoring: Boosting peptide identifications and scoring confidence of database search results. *Rapid Commun. Mass Spectrom.* 39:e9128. <https://doi.org/10.1002/rcm.9128>

Supplemental material

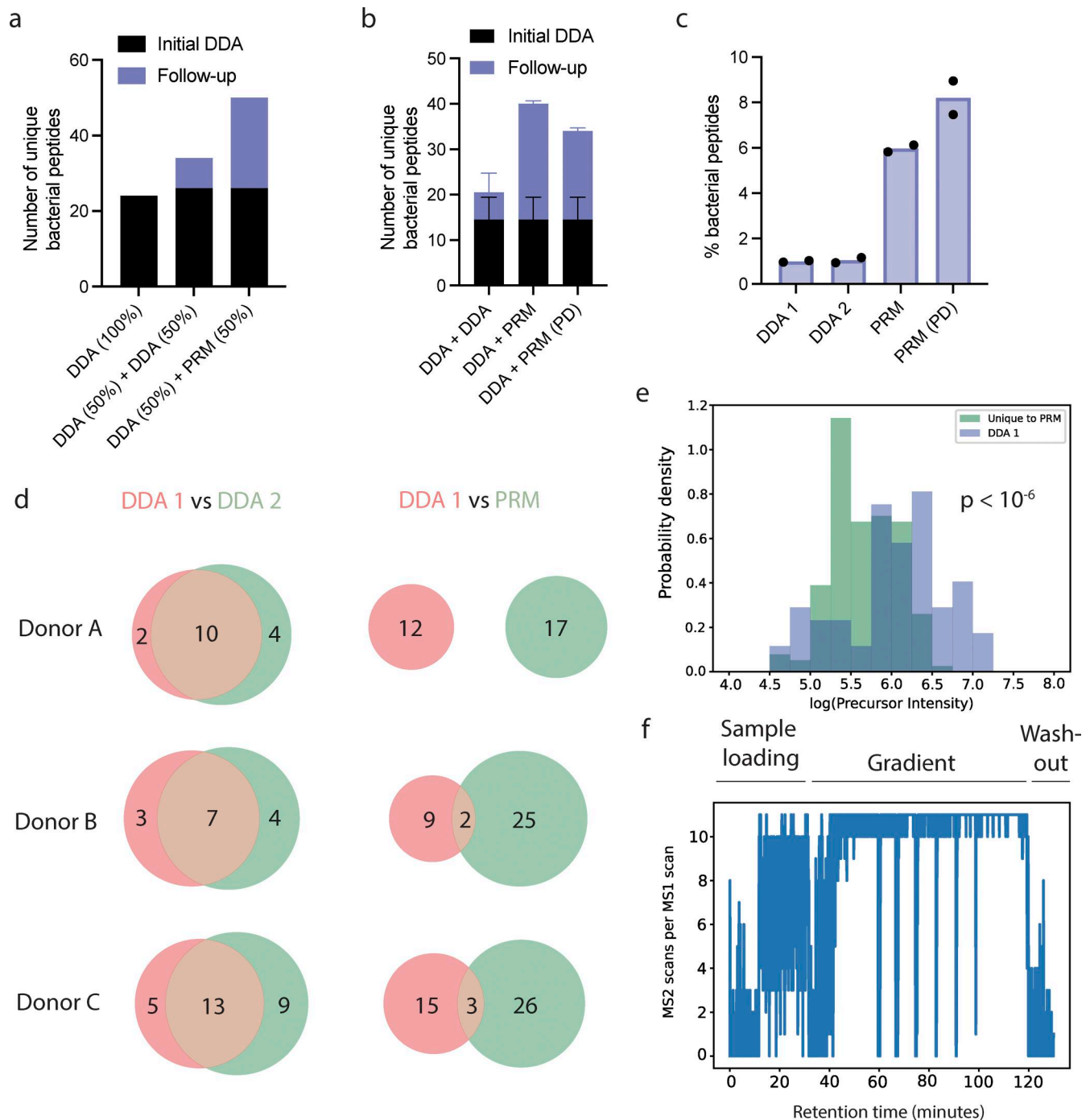


Figure S1. Benchmarking PathMHC against untargeted methods and an alternate computational pipeline. (a) Number of unique bacterial peptides identified in analyses of MHC-II peptides isolated from hMDCs infected with *Msmeg* using three workflows: 100% of the sample injected in a single DDA analysis, 50% of sample injected in each of two DDA analyses, or 50% of the sample injected in a DDA analysis and 50% in a PRM analysis using a prototype version of the PathMHC workflow. **(b and c)** Number of unique bacterial peptides (b) and percentage of bacterial peptides as a proportion of total peptides (c) identified in analyses of MHC-II peptides isolated from hMDCs from $n = 2$ donors infected with *Msmeg*, using a DDA only benchmark protocol, a version of PathMHC using open-source software tools, and a version of PathMHC using a computational pipeline created in PD. **(d)** Venn diagrams of *Msmeg* MHC-II peptide IDs from consecutive DDA analyses (left column) or a DDA analysis and a PathMHC PRM analysis of the same sample (right column). **(e)** Histograms of precursor ion intensities associated with PSMs for *Msmeg* MHC-II peptides identified by DDA analysis or exclusively identified by a follow-up PRM analysis. **(f)** Plot of the number of MS2 (i.e., MS/MS) scans per MS1 scan throughout a DDA analysis of MHC-II peptides isolated from *Msmeg*-infected hMDCs. Representative of DDA analyses in $n = 3$ donors.

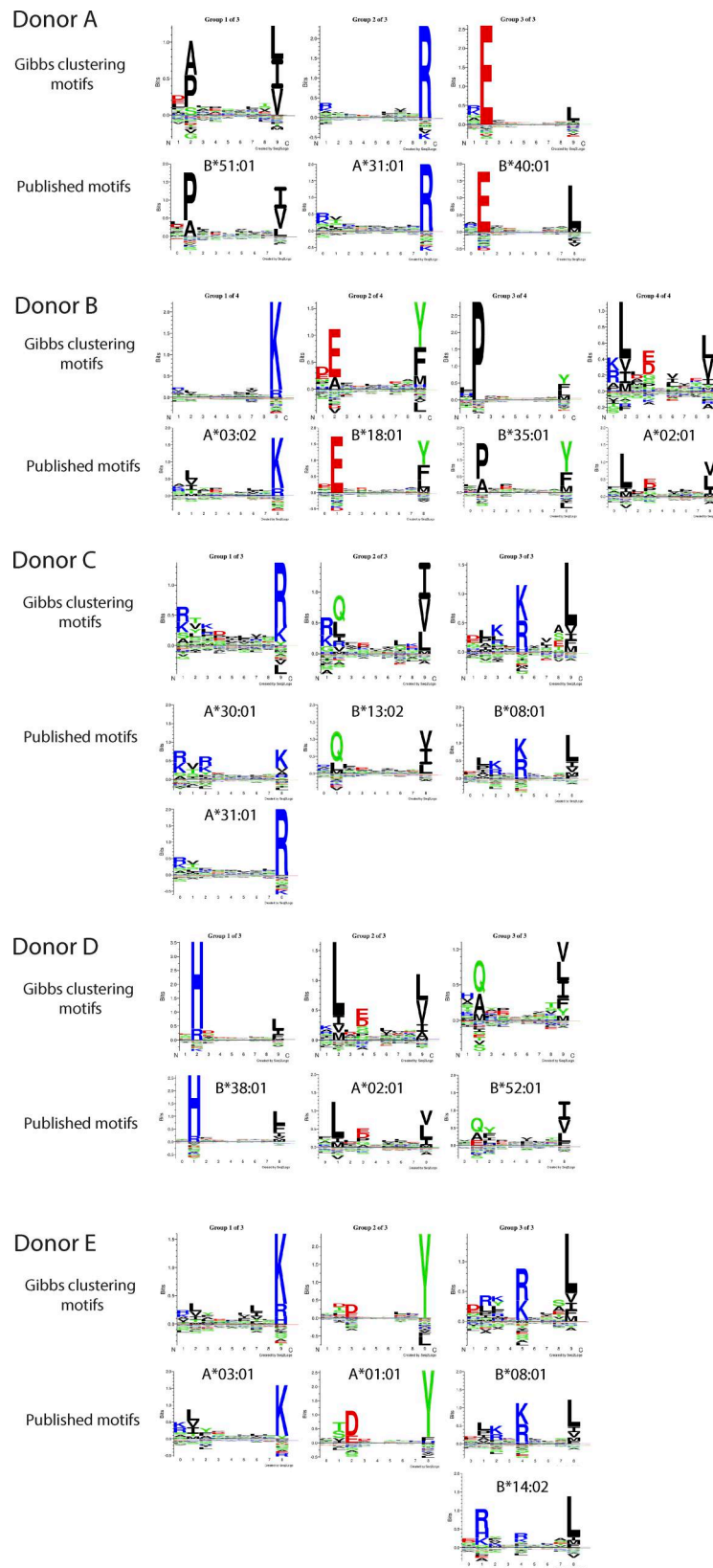


Figure S2. **Gibbs clustering of MHC-I peptides presented by *Mtb*-infected hMDCs recapitulates known binding motifs of class I HLA alleles.** Peptide sequence motif logos associated with each cluster in the optimal clustering solution determined by GibbsCluster 2.0 (Andreatta et al., 2017; Andreatta et al., 2013) (top row for each donor) and the published peptide binding sequence motif logos (obtained from the NetMHCpan 4.1 motif viewer [Reynisson et al., 2020]) for the corresponding HLA alleles expressed by the donor (bottom row for each donor). The optimal number of clusters was defined as the number between 1 and 6 that maximized the Kullback–Leibler distance among the sequence distributions of the clusters (Andreatta et al., 2013).

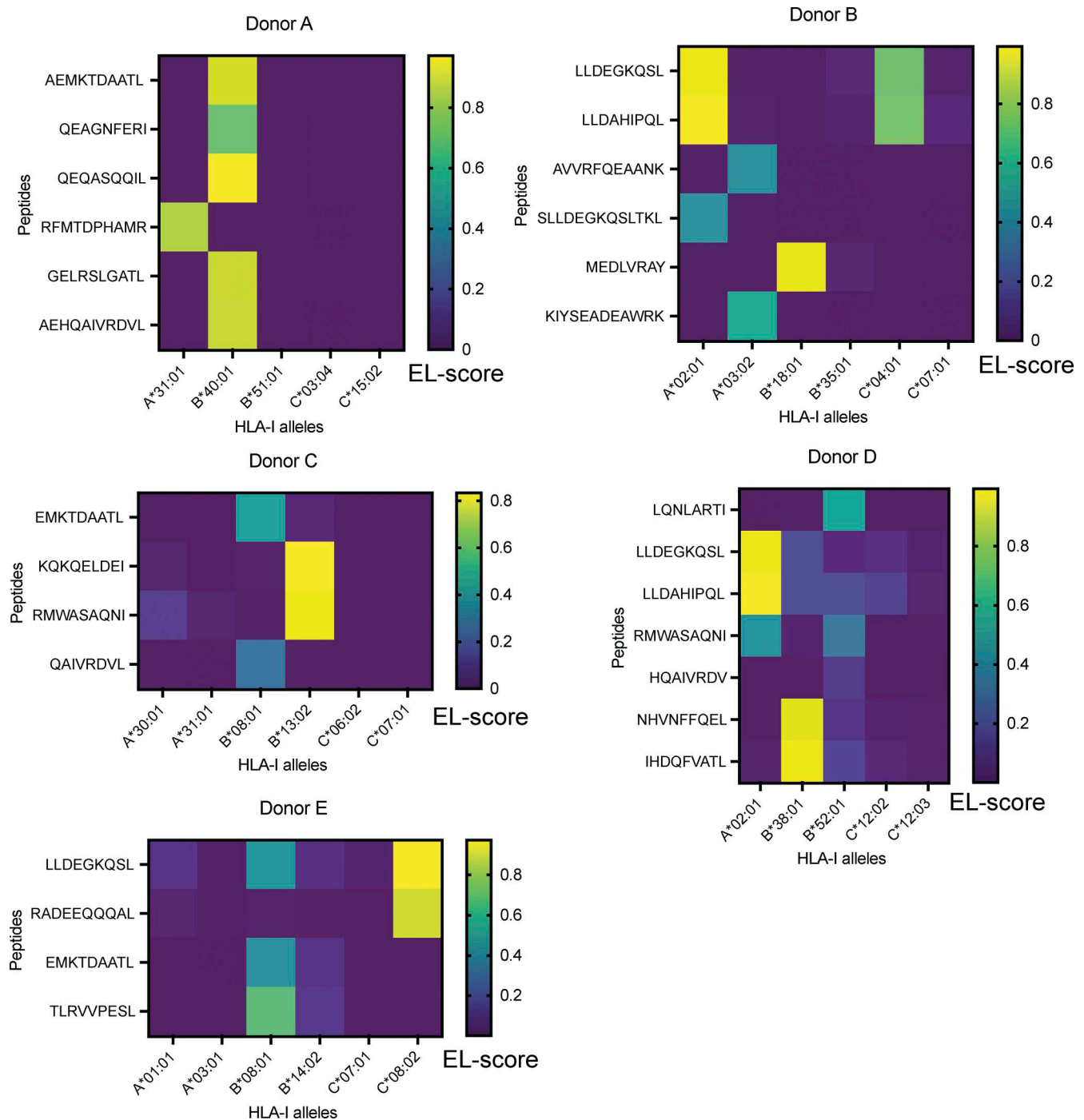


Figure S3. ***Mtb*-derived MHC-I peptides are predicted binders of HLA alleles expressed by each donor.** Heatmaps show the eluted ligand (EL) score assigned to each *Mtb*-derived MHC-I peptide by NetMHCpan4.1 (Reynisson et al., 2020) for each class I HLA allele expressed by a given donor.

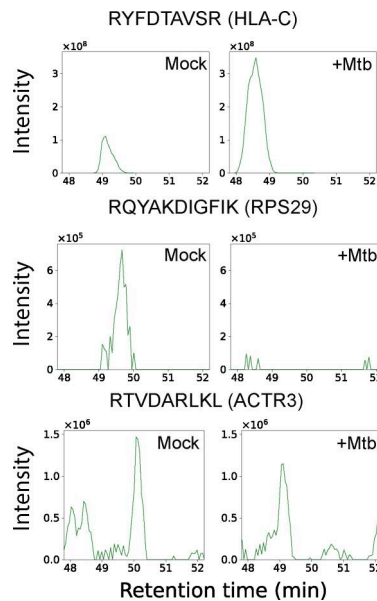


Figure S4. **XIC of self-peptides identified in DDA analyses.** Representative XIC of peptides identified by DDA in mock-infected hMDCs, showing MS1 signal in *Mtb*-infected and mock-infected hMDCs.

Provided online are Table S1, Table S2, Data S1, Data S2, Data S3, Data S4, Data S5, and Data S6. Table S1 shows the HLA types of primary cell donors. Table S2 shows the putative *Mtb* peptide IDs invalidated by SureQuant analyses. Data S1 shows the PSMs for PathMHC analyses of MHC-II peptides from hMDCs following phagocytosis of *Msmeg*. Data S2 shows the PSMs for PathMHC and DDA-only analyses of MHC-I peptides from hMDCs infected with *Mtb*. Data S3 shows the sample guide for raw MS data. Data S4 shows the proteome database for MS data search comprising the human and *Msmeg* proteomes. Data S5 shows the proteome database for MS data search comprising the human and *Mtb* proteomes. Data S6 shows the SureQuant data (mirror plots and chromatograms comparing biological peptides with synthetic SIL standards) validating *Mtb* MHC-I peptide IDs.