

# Identification of the Cryptic HLA-I Immunopeptidome

Florian Erhard<sup>1</sup>, Lars Dölken<sup>1</sup>, Bastian Schilling<sup>2</sup>, and Andreas Schlosser<sup>3</sup>



## ABSTRACT

The success of cancer immunotherapy relies on the ability of cytotoxic T cells to specifically recognize and eliminate tumor cells based on peptides presented by HLA-I. Although the peptide epitopes that elicit the corresponding immune response often remain unidentified, it is generally assumed that neoantigens, due to tumor-specific mutations, are the most common targets. Here, we used a mass spectrometric approach to show an underappreciated class of epitopes that accounts for up to 15% of HLA-I peptides for certain HLA alleles in various tumors and patients. These peptides are translated from cryptic open reading frames in supposedly noncoding regions in the genome and are mostly unidentifiable with conventional computational analyses of mass spectrometry (MS) data. Our approach, Peptide-PRISM, identi-

fied thousands of such cryptic peptides in tumor immunopeptidomes. About 20% of these HLA-I peptides represented the C-terminus of the corresponding translation product, suggesting frequent proteasome-independent processing. Our data also revealed HLA-I allele-dependent presentation of cryptic peptides, with HLA-A\*03 and HLA-A\*11 presenting the highest percentage of cryptic peptides. Our analyses refute the reported frequent presentation of HLA peptides generated by proteasome-catalyzed peptide splicing. Thus, Peptide-PRISM represents an important step toward comprehensive identification of HLA-I immunopeptidomes and reveals cryptic peptides as an abundant class of epitopes with potential relevance for novel immunotherapeutic approaches.

## Introduction

Immunotherapeutic approaches, such as adoptive T-cell therapy or therapeutic peptide vaccination, are among the most promising approaches to treat cancer (1). Their success demonstrates that cytotoxic T cells are able to specifically recognize and eliminate tumor cells via peptides presented by HLA-I. These epitopes include peptides carrying a tumor-specific mutation (neoantigens), or are derived from germline antigens. However, other sources, such as epitopes generated by proteasome-catalyzed peptide splicing (PCPS) have been proposed (2). HLA-I presented peptides can be analyzed by mass spectrometry (MS) of purified HLA-I complexes. However, the identification of peptides that are not encoded in the proteome is challenging. We demonstrated here that this can lead to erroneous identification of neoantigens, and, in line with previously published concerns (3), to unwarranted claims of frequent presentation of PCPS peptides. The lack of advanced approaches for proteogenomic identification of HLA-I epitopes precludes the comprehensive analysis of the contribution of aberrant transcription and translation to HLA-I immunopeptidomes.

Ribosome profiling (Ribo-seq) has provided compelling evidence for the translation of thousands of open reading frames (ORF) outside of

the annotated proteome (4). These cryptic ORFs are encoded in 5'- and 3'-untranslated regions (UTR), noncoding RNAs, intronic and intergenic regions, and within coding sequences shifted with respect to the canonical reading frame. In spite of overwhelming evidence for their translation, only a limited number of their translation products have been detected by MS so far, even when specifically adopted isolation methods and database search approaches have been applied (5, 6). This indicates that cryptic translation products are short-lived, but are exploited as a source for HLA-I peptides, as has been described for defective ribosomal products (7–10). Alternatively, cryptic translation products might arise from noncanonical translation events, potentially translated from a specialized subpopulation of ribosomes (immunoribosomes; ref. 11). These might channel their translation products directly to the HLA-I antigen-processing machinery (12). Both, the short half-life and direct channeling explain why cryptic translation products are hardly detected in the proteome. Both mechanisms should lead to a higher incidence of cryptic HLA-I peptides. By analyzing HLA-I immunopeptidome data, we previously found that at least 2% of the peptides bound to HLA-I in human fibroblasts originate from cryptic (i.e., unknown) ORFs identified by Ribo-seq (13). This indicates that cryptic peptides are indeed enriched in the HLA-I immunopeptidome. A study has employed RNA sequencing (RNA-seq) in parallel to HLA immunopeptidomics and identified 168 cryptic peptides from an Epstein-Barr virus-transformed B-cell line (14). Both studies were based on the established software developed for proteomics data and on sequence databases derived from the Ribo-seq or RNA-seq data. Here, we present a fundamentally different strategy, termed Peptide-PRISM, to identify cryptic peptides based on mass spectrometric data alone. This enabled us to analyze a large collection of HLA immunopeptidomes without using additional sequencing data, and to identify 6,636 cryptic peptides from various types of tumors from several patients with high confidence. This reveals cryptic HLA peptides to be a substantial part of tumor immunopeptidomes.

<sup>1</sup>Institute for Virology and Immunobiology, Julius-Maximilians-Universität Würzburg, Würzburg, Germany. <sup>2</sup>Department of Dermatology, Venereology, and Allergology, University Hospital Würzburg, Würzburg, Germany. <sup>3</sup>Rudolf Virchow Center for Experimental Biomedicine, Julius-Maximilians-Universität Würzburg, Würzburg, Germany.

**Note:** Supplementary data for this article are available at Cancer Immunology Research Online (<http://cancerimmunolres.aacrjournals.org/>).

**Corresponding Authors:** Florian Erhard, Institute for Virology and Immunobiology, University of Würzburg, Würzburg 97078, Germany. Phone: 49 931 31 86523; E-mail: Florian.Erhard@uni-wuerzburg.de; and Andreas Schlosser, Rudolf Virchow Center for Experimental Biomedicine, Josef-Schneider-Str. 2, 97080 Würzburg, Germany. Phone: 49 931 31 86888; E-mail: Andreas.Schlosser@uni-wuerzburg.de

Cancer Immunol Res 2020;8:1018–26

doi: 10.1158/2326-6066.CIR-19-0886

©2020 American Association for Cancer Research.

## Materials and Methods

### Datasets for analysis

All immunopeptidome datasets analyzed in this study are listed in Supplementary Table S1. We included datasets from different types of

## Identification of the Cryptic HLA-I Immunopeptidome

tumor samples (melanoma, ref. 15; lung cancer, ref. 16; glioblastoma, ref. 17; triple-negative breast cancer, ref. 18; and mantle cell lymphoma, ref. 19). We also included two datasets from publications that report the identification of HLA-I peptides derived from PCPS (20, 21). We only included datasets with fragment ion spectra providing high mass accuracy (i.e., fragment ion spectra acquired with Orbitrap mass analyzer). For all analyses, we exclusively used raw MS data.

### De novo sequencing

*De novo* sequencing was performed with PEAKS X (refs. 22, 23; Bioinformatics Solutions Inc.). Raw data refinement was performed with the following settings: (i) Merge Options: no merge; (ii) Precursor Options: corrected; (iii) Charge Options: no correction; (iv) Filter Options: no filter; (v) Process: true; (vi) Default: true; and (vii) Associate Chimera: yes. *De novo* sequencing was performed with Parent Mass Error Tolerance set to 15 ppm for dataset PXD004894 (melanoma) and with 10 ppm for all other datasets. Fragment Mass Error Tolerance was set to 0.015 Da, and Enzyme was set to none. The following variable modifications were used: Oxidation (M), pyro-Glu from Q (N-term Q), and carbamidomethylation (C). A maximum of three variable posttranslational modifications were allowed per peptide. Up to 10 *de novo* sequencing candidates were reported for each identified fragment ion mass spectrum, with their corresponding average local confidence (ALC) score. Because we applied the chimeric spectra option of PEAKS X, two or more TOP10 candidate lists could be assigned to a single fragment ion spectrum. Two tables (“all *de novo* candidates” and “*de novo* peptides”) were exported from PEAKS for further analysis.

### Peptide-PRISM

To efficiently search ultralarge sequence databases, first, a keyword trie was built from all *de novo* candidate sequences. A trie is a data structure to store keywords and enable parallel searching a text for all keywords using the Aho–Corasick algorithm (24). To account for additional variable modifications (pyro-Glu from N-terminal Glu and deamidation at asparagine when followed by glycine), and for the isobaric Leu and Ile, all possible combinations of the corresponding sequences were inserted into the trie for each *de novo* candidate. Then the Aho–Corasick algorithm was employed, scanning through the 6-frame translation of the genome (reference assembly HG38) and 3-frame translation of the transcriptome (Ensembl 90). Optionally, on-the-fly generated proteasome-spliced peptides (normal and reverse *cis*-spliced and maximal intervening sequence length of 25 amino acids from all annotated proteins, as described ref. 20), all possible on-the-fly generated single amino acid substitutions from annotated proteins (i.e., at each position in each protein each of the 18 substitutions, excluding Leu), and all possible on-the-fly generated frameshift peptides (-3,-2,-1,1,2,3 nucleotide shifts at each position in each annotated protein) were scanned as well. All considered sequences were additionally reversed (called the decoy database) and scanned.

All identified string matches from the 6-frame translated genome and 3-frame translated transcriptome were categorized as follows: (i) Coding sequence (CDS): in-frame with annotated protein; (ii) 5′-UTR: contained in annotated mRNA, consistently with its introns, overlapping with 5′-UTR; (iii) Off-Frame: off-frame contained in the coding sequence, consistently with its introns; (iv) 3′-UTR: all others that are introns, consistently contained in an mRNA; (v) noncoding (nc) RNA: consistently contained in an annotated ncRNA; (vi) Intronic: intersecting any annotated intron; or (vii) Intergenic. For each fragment ion mass spectrum, the category with highest priority (CDS > 5′-UTR > Off-Frame > Frameshift > 3′-UTR > ncRNA > Substitution >

Intronic > Intergenic > PCPS) was identified, and all other hits among the 10 *de novo* candidates were discarded.

Next, if more than one fragment ion mass spectrum for the same peptide was seen in the dataset, only the fragment ion mass spectrum with maximal score among all these was retained. Finally, if the remaining candidate for a spectrum was *a*, the originally best candidate (irrespective of whether it was found in the sequence database or not) was *f*, and the next best distinct remaining candidate in the same category but a distinct sequence was *n*, the spectrum was discarded if  $ALC(a) < ALC(f) - \delta_f$  or  $ALC(a) < ALC(n) + \delta_n$ . Here, we set  $\delta_f = 15$  (the maximal difference to the originally top candidate), and  $\delta_n = 16$  (the minimal difference to the next best candidate sequence from the same category). Isobaric Ile/Leu and additionally considered modifications were not treated as distinct here.

This procedure resulted in a list of unique peptide sequences annotated with its best ALC score, its category, and whether it is from the target or decoy database. In principle, after discarding all categories but CDS, a standard target-decoy approach can be utilized to filter proteome-derived peptides by the classic FDR.

### FDR control in Peptide-PRISM

Peptide-PRISM was built on the following mixture modeling approach. For a given length-specific score distribution with density  $p_l(x)$  modeling true hits (i.e., peptides that actually were in the sample), and a second length-specific score distribution with density  $n_l(x)$  modeling false hits, the overall score distribution of all filtered peptides of a given length *l* and category/stratum *c* is given by the density  $f_{l,c}(x) = w_{l,c} p_l(x) + (1 - w_{l,c}) n_l(x)$ . Here,  $w_{l,c}$  is the total fraction of true hits with length *l* from category *c*. Note that the same two component distributions were used for all categories. For the rationale of this model, see Supplementary Data S1 (2, 3, 20, 25–30).

For each observed peptide length, category, and target/decoy status, Peptide-PRISM first built histograms of the (integer) ALC scores of the filtered peptides. Then, for each peptide length, all decoy histograms were summed up to fit the false-hit score distribution using unimodal penalized B-spline regression (31). The true hit score distribution was fit in the same manner after subtracting the CDS decoy histogram from the CDS target histogram. For the rationale of these approaches, see the Supplementary Data S1. Then, for each peptide length *l* and category *c*,  $w_{l,c}$  was estimated by maximum likelihood (30) based on the respective score histograms of hits in target sequences. The expected number of true and false targets per ALC score was computed by multiplying  $w_{l,c}$  and  $(1 - w_{l,c})$ , respectively, with the total number of identified peptides with length *l* in category *c*. Finally, these expected numbers of true and false targets were then used to compute FDRs per peptide length and category.

### Putative ORF identification

Each remaining peptide was identified to originate from a location on the genome, or potentially several locations if the same sequence occurred multiple times in its category (or any Ile, Leu/variable modification variant). For each location, all potential ORFs were identified. More than one ORF can exist because all paths via exon–exon junctions of annotated transcript isoforms were considered in addition to the path on the genome. Because translation might initiate at non-AUG start codons (13), two kinds of ORFs were considered. For nonprioritized ORFs, along each path in the genome or transcriptome upstream of the peptide location, the closest in-frame start codon candidate was identified (one of AUG, CUG, GUG, ATC, and ACG). For prioritized ORFs, any closest in-frame AUG was identified. If none was found, the closest CUG was identified, followed

Erhard et al.

by GUG, ATC, and ACG. From all remaining start codon candidates, the one closest to the peptide was chosen for both kinds of ORFs. The same procedure was repeated downstream of the peptide location to identify the stop codon (only nonprioritized). Finally, for peptides with more than one location, the location giving rise to the shortest prioritized ORF was chosen.

### Prediction of HLA-I peptide binding

NetMHCpan 4.0 (32) was run on all final remaining peptide sequences with the HLA-I alleles given in the original publication for the corresponding patient or cell line. The allele with the minimal rank reported by NetMHCpan was annotated. As per default, we used a cutoff of 0.5% rank for strong binders and 2% rank for weak binders. For samples with unknown HLA-I genotype, Gibbs clustering was performed with GibbsCluster 2.0 (33), and alleles were manually assigned by comparing motifs from clustering results with known HLA-I peptide motifs (34).

### MS analysis of synthetic peptides

Peptides were ordered from JPT Peptide Technologies. Peptides were dissolved in 2% acetonitrile, and nanoLC-MS/MS analyses were performed on an Orbitrap Fusion (Thermo Fisher Scientific) equipped with a PicoView Ion Source (New Objective) and coupled to an EASY-nLC 1000 (Thermo Fisher Scientific). Peptides were loaded on capillary columns (PicoFrit, 30 cm × 150 μm ID, New Objective) self-packed with ReproSil-Pur 120 C18-AQ, 1.9 μm (Dr. Maisch GmbH), and separated with a 30-minute linear gradient from 3% to 30% acetonitrile and 0.1% formic acid at a flow rate of 500 nL/minutes. Both MS and MS-MS scans were acquired in the Orbitrap analyzer with a resolution of 60,000 for MS scans and 15,000 for MS-MS scans. Higher energy collisional dissociation (HCD) with 35% normalized collision energy was applied. A top speed data-dependent MS-MS method with a fixed cycle time of 3 seconds was used. Dynamic exclusion was applied with a repeat count of 1 and an exclusion duration of 30 seconds. Singly charged precursors were excluded from selection. EASY-IC was used for internal calibration.

### Software/code availability

The data and scripts to reproduce all figures are available at Zenodo (DOI: 10.5281/zenodo.3775934); they can be accessed by visiting <https://doi.org/10.5281/zenodo.3775934>. Peptide-PRISM is available free of charge for academic use at <http://software.erhard-lab.de>.

## Results

### Sensitive and reliable identification of HLA-I peptides by Peptide-PRISM

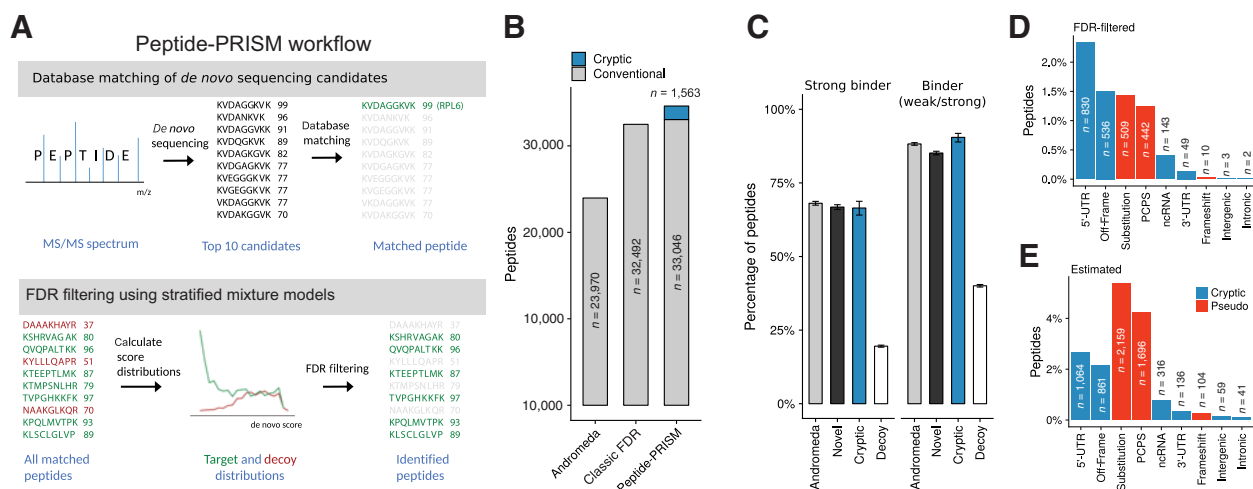
The standard analysis workflow for MS-based proteomics relies on sequence database search coupled with FDR estimation using the target-decoy approach (26). Commonly applied search engines, such as Mascot, Andromeda, or Comet try to identify the best peptide-spectrum match (PSM) by matching experimental and theoretical fragment ion spectra of sequence candidates. This approach works well for tryptic peptides, the most frequently analyzed sample type in proteomics, but shows only moderate sensitivity for HLA-I peptides. On the one hand, this is because of the limited length of HLA-I peptides, which limits the maximum number of matching fragment ions and the maximum matching score. On the other hand, the moderate sensitivity is caused by the missing cleavage specificity (nontryptic), which greatly increases the search space. Because *de novo* peptide sequencing performs well for short peptides, we hypothesized that the length restriction of HLA-I peptides of predominantly

8 to 11 amino acids could enable reliable *de novo* peptide sequencing based on state-of-the-art high mass accuracy tandem mass spectra. To test this, we developed a computational pipeline by combining *de novo* peptide sequencing, highly efficient string search, mixture modeling, and stratification of the search space for the analysis of HLA-I immunopeptidomes and termed it Peptide-PRISM (Fig. 1A). Because the quality of fragment ion spectra is often not high enough to derive a single definite peptide sequence, we generated up to 10 sequence candidates by *de novo* sequencing per spectrum. In a second step, all candidates were matched against a database. For the majority of spectra, not more than one of the sequence candidate was found in the database, and the correct peptide sequence could be identified by database matching of *de novo* sequencing candidates. In cases with multiple matching database peptides, we used a biologically motivated heuristic to select a single candidate: we stratified the database into biologically meaningful categories and prioritized these categories to select the most parsimonious candidate. For instance, if a matching sequence was found in the proteome and as part of a 5'-UTR, we selected the proteome hit as the single candidate. This is equivalent to first searching the proteome, then searching the 5'-UTR with all so far unmapped sequences, etc. (see Materials and Methods for the specifics of the prioritization). To test the performance of our new approach, we applied Peptide-PRISM for the analysis of a tumor immunopeptidome sample from a patient with melanoma (MM15; ref. 15). Database matching of *de novo* sequencing candidates resulted in approximately 35% more high-confidence, conventional peptides derived from proteins (1% FDR based on classic target-decoy approach) compared with the previously employed database searching approach (Fig. 1B; ref. 15). Of note, the identified peptides exhibited the same frequency of HLA-I binders predicted by NetMHCpan 4.0 (32) as the previously identified peptides (Fig. 1C). This provides evidence that the improved sensitivity was not accompanied by a loss of specificity.

In contrast to conventional database search tools, *de novo* peptide sequencing, in principle, allows the systematic identification of cryptic HLA peptides. To achieve this, Peptide-PRISM utilizes a highly efficient string search algorithm to search for millions of peptides in sequence databases in the order of gigabases in less than an hour. For FDR filtering, we designed a statistical method that extends the central idea from Peptide Prophet (29). Peptide-PRISM utilizes mixture modeling to deconvolute the overall *de novo* score distribution into components of false and true identifications. Here, we combined this approach with the stratification of the sequence search space introduced above, which was essential for maintaining statistical power. Our Peptide-PRISM approach identified 1,563 cryptic peptides at 1% FDR, corresponding to more than 4.5% of all identified peptides for MM15 (Fig. 1B), and the predicted binding affinities of the cryptic peptides to HLA-I were indistinguishable from those of the conventional peptides (Fig. 1C). The median intensities of the identified cryptic peptides obtained from peak areas of extracted ion chromatograms did not show any differences compared with conventional peptides (Supplementary Fig. S1). Twenty-four of twenty-four identified cryptic peptides were successfully validated by reference spectra of synthetic peptides (Supplementary Fig. S2). In summary, these results confirmed the stringent FDR control of our approach.

To map *de novo* sequencing candidates, we generated a database consisting of the 6-frame translated human genome, 3-frame translated transcriptome from Ensembl (including all coding and noncoding splice variants), PCPS peptide database (normal and reverse *cis*-spliced peptides with a maximal intervening distance of 25), the human proteome with all possible substitutions of a single amino acid, and any peptide that could be generated from ribosomal

## Identification of the Cryptic HLA-I Immunopeptidome

**Figure 1.**

Peptide-PRISM identifies cryptic peptides in a melanoma sample. **A**, Workflow of Peptide-PRISM. **B**, Number of peptides identified for sample MM15. Andromeda represents the originally published numbers; classic FDR and Peptide-PRISM are described in the Materials and Methods. **C**, Frequency of novel and cryptic peptides predicted to be HLA-I binders by NetMHCpan 4.0. Novel peptides are proteome-derived peptides identified by our approach, but not in the original report. Error bars represent 95% binomial confidence intervals. **D**, Distribution of FDR-filtered peptides in the nine cryptic peptide subcategories. **E**, Distribution of the unfiltered total amount of MS-detected peptides estimated by Peptide-PRISM.

frameshifting in the human proteome (for details see Materials and Methods). Peptides in this database can be stratified into nine categories, in addition to conventional peptides from CDSs (Fig. 1D). Each of these strata had a distinct size and likelihood for identifying peptides, which had an impact for estimating FDRs (Supplementary Data S1). Peptide-PRISM allowed for FDR control per stratum as opposed to the global FDR control by the classic target-decoy approaches. As a consequence, the relative frequencies of the peptides above and below the thresholds used for FDR filtering were distinct among strata. Therefore, counting the peptides retained after filtering was insufficient for assessing the relative composition of a sample by peptides from different strata. However, Peptide-PRISM allowed for estimating the number of peptides in each stratum without FDR filtering based on the mixture modeling approach (Fig. 1E).

### PCPS peptides do not represent a large fraction of immunopeptidomes

It has been suggested that proteasomes can catalyze the ligation of two peptides, generating epitopes that do not occur consecutively in the proteome (35, 36). The contribution of such peptides generated by PCPS to HLA-I immunopeptidomes of cell lines or primary samples has raised major controversies. The percentage of PCPS peptides for the same dataset (GR-LCL) was estimated to be as high as 33% (20), 16% (2), or not more than 2% to 6% (3). We identified two major issues with the previous analyses. First, Liepe and colleagues (33% and 16% PCPS peptides; refs. 2, 20) used Mascot to perform sequence database search and the target-decoy approach for controlling the FDR. To reduce the computational burden for the search, only peptides with a precursor mass observed in the experimental data were included in the sequence database as concatameric pseudoprotein sequences. Unfortunately, the randomization of the pseudoproteins used to generate decoys resulted in drastically underestimated FDRs (see Supplementary Data S1). Second, we found that the spectra of 517 (49%) of the reported 1,056 PCPS peptides (2) either matched to conventional peptides ( $n = 377$ ; 36%) or cryptic peptides ( $n = 140$ ; 13%) in our analysis (10% FDR, only binders predicted by NetMHCpan 4.0;

Supplementary Table S2). These included 204 (19%) reported PCPS peptides with a sequence differing only by I/L from conventional nonspliced peptides. In our reanalysis of the GR-LCL dataset with Peptide-PRISM, only 10 spliced peptides, less than 0.1% of the identified HLA-I peptides, survived 1% FDR filtering.

### A variety of peptides undermine FDR control

Even after stringent FDR filtering (1% FDR), Peptide-PRISM identified a number of PCPS peptides for MM15 ( $n = 442$ , 1.3% of all filtered peptides; Fig. 1D), as well as for the other datasets (Supplementary Table S1; Supplementary Fig. S3; 1% FDR). However, the remaining PCPS peptide candidates had a lower fraction of predicted HLA-I binders than that of conventional or cryptic peptides, which was in the range of decoy peptides for MM15 (Fig. 2A) and the other datasets (Supplementary Fig. S4). The *de novo* score (ALC) distribution of PCPS peptides resembled the ALC score distribution of decoys. For instance, approximately 30% of the PCPS peptides and approximately 26% of decoy peptides had an ALC score of  $\geq 50$  ( $A_{50}$  score), indicating that most were false identifications (Fig. 2B). After removing the expected false identifications by subtracting their estimated distribution from the mixture distribution, the remaining ALC scores of PCPS peptides, but not of cryptic peptides, still showed significant differences from conventional peptides (Fig. 2C; peptides with ALC score  $\geq 90$ :  $A_{90}$  PCPS = 13%;  $A_{90}$  cryptic = 38%; and  $A_{90}$  conventional = 40%). The length distribution of PCPS peptides was significantly shifted toward longer peptides (Fig. 2D). All three phenomena indicated that the majority of the remaining PCPS peptides represented false identifications.

Not all fragment ion mass spectra contain full  $y$  ion series. In such cases, two or more sequences explain the spectrum equally well, and the wrong sequence might be identified because of spurious peaks in the spectrum. Indeed, the spliced peptide IEVGNLPSAMR (Fig. 2E) clearly represented such a misidentification due to a missing  $y_{10}$  fragment ion (Fig. 2F), and the parsimonious explanation for the spectrum clearly was the consecutive peptide EIVGNLPSAMR in the same locus. These isobaric ambiguities were not an issue of *de novo*

Erhard et al.

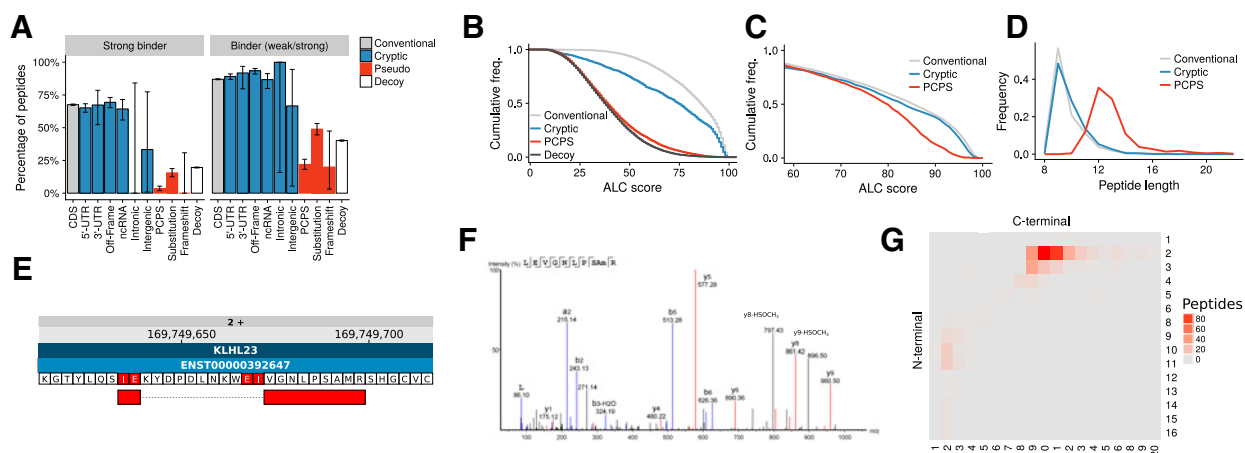


Figure 2.

PCPS peptides represent misidentifications. **A**, Percentages of predicted HLA-I binders among the 10 categories and decoy peptides for MM15. Error bars represent 95% binomial confidence intervals. **B**, Cumulative *de novo* sequencing score (ALC) distributions for 9 amino acid-long peptides for MM15 (higher is better). For cryptic peptides, only the most abundant categories (5'-UTR and Off-Frame) were considered. **C**, Same as **B**, but false-positive peptide identifications were removed by deconvolution. **D**, Distribution of sequence lengths for categories of peptides in MM15. **E**, Example of a PCPS peptide indicated in red. Chromosomal positions refer to human genome assembly Hg38. The gene, coding Ensembl transcript, and the amino acid sequence are indicated. **F**, HCD fragment ion mass spectrum of the peptide in **E**. **G**, Heatmap of the number of PCPS peptides in MM15 with different lengths of the C-terminal and N-terminal parts. freq., frequency.

sequencing, but inherent to MS. The majority of PCPS peptides (>73%) identified at 1% FDR for MM15 resembled this example (Fig. 2G), and the majority of these had a matching conventional peptide in the corresponding locus (Supplementary Fig. S5A and S5B). This was also true for the two datasets used in studies on PCPS (refs. 20, 37; Supplementary Fig. S5C and S5G). After filtering to 1% FDR and removing the most obvious false identifications, less than 0.3% potential PCPS peptides remained for MM15 and even less for the other datasets.

Interchanging the first two amino acids has a strong effect on predicted binding affinities to most HLA alleles. Replacing PCPS peptide sequences by corresponding conventional sequences in the respective locus, as suggested by the above analyses, improved the predicted binding affinities from below the level of decoy peptides to the same level as conventional and cryptic peptides (Supplementary Fig. S5H). This was not only true for cases with two amino acids at the N- or C-terminal part, but also for the remaining peptides. Taken together, this showed that most remaining PCPS peptides undermined FDR control by the target-decoy approach.

We observed the same three phenomena (low predicted HLA-I binding affinity, left-shifted *de novo* score distribution, and longer than expected peptides) for peptides spanning putative ribosomal frameshifting events (see Materials and Methods) and for peptides with single amino acid substitutions. Closer inspection of peptide spectra of the latter category revealed that the majority of identifications could be readily explained by unconsidered modifications, such as deamidation, N-terminal acetylation, and oxidation of cysteine to cysteic acid, or by erroneous identification of the precursor mass (Supplementary Fig. S6). We conclude that frameshift, substitution, and PCPS peptides all predominantly represented pseudoidentifications.

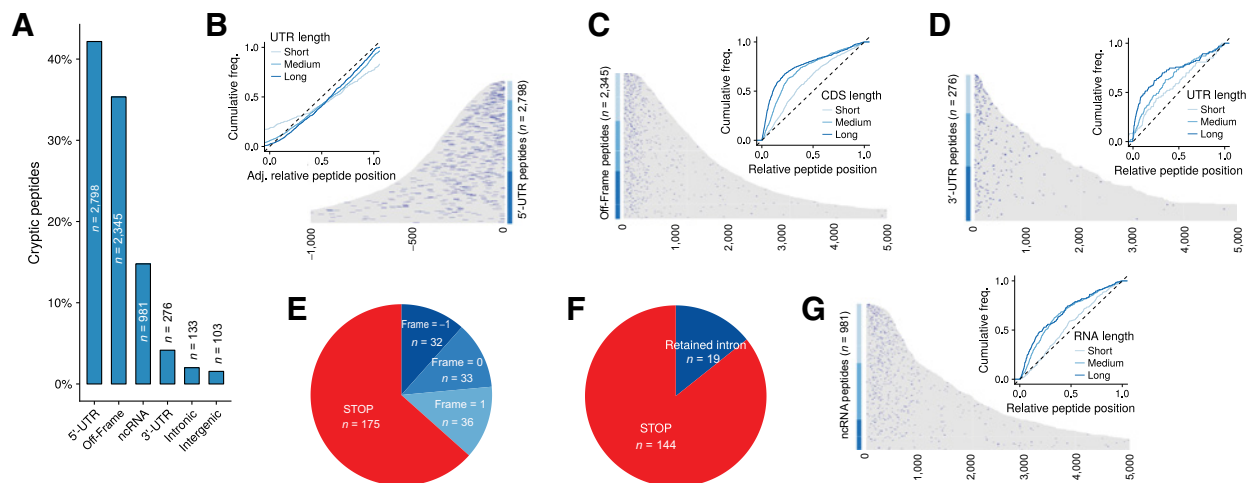
HLA peptides harboring a single amino acid exchange caused by tumor-specific nonsynonymous DNA mutations (neoantigens) are promising candidates for antitumor immunotherapeutic approaches in melanoma and other malignancies (38, 39). We reasoned that neoantigen identification might also suffer from false-positive identifications caused by isobaric ambiguities. Indeed, we found that the

MS-MS spectrum of the previously identified neoantigen candidate ASWVVPIDIK of MM15 corresponded to the cryptic peptide KLWDPLDLK originating from the 5'-UTR of a different gene. Assignment of the spectrum to the cryptic peptide was unequivocally confirmed with synthetic peptides (Supplementary Fig. S7). The cryptic peptide was predicted to be a strong binder to the patient's HLA-A\*03:01 allele. Both peptides had the exact same mass, and the C-terminal part of their sequence was identical (except for I/L), explaining why the incorrect neoantigen sequence gained a relatively high matching score from the Andromeda search engine. This example demonstrates that isobaric ambiguities can easily lead to false-positive neoantigen identification if the correct (cryptic) peptide sequence is excluded from the search.

### Cryptic peptides are a substantial part of tumor immunopeptidomes

After excluding PCPS peptides, peptides with single amino acid substitutions, and frameshift peptides, we concentrated on the remaining six categories of cryptic peptides. In a collection of HLA-I tumor immunopeptidome datasets (Supplementary Table S1), we identified more than 6,500 cryptic HLA-I peptides after stringent filtering ( $n = 6,636$ , FDR 10%, strong binders as predicted by NetMHCpan 4.0; Fig. 3A; Supplementary Table S3). The largest category with 2,798 peptides consisted of cryptic peptides located in the 5'-UTR of coding transcripts. The first 50 nucleotides (nt) of the UTR were mostly devoid of peptides, and peptides were uniformly distributed across the rest of the UTR (Fig. 3B). This is consistent with these peptides being translated from upstream ORFs (uORF) scattered throughout 5'-UTRs. The second largest category consisted of peptides located inside, but out-of-frame, of ORFs encoding for proteins ("Off-Frame"). For those, the majority was located within the first 200 nt of the protein ORF. This indicated that they are encoded either by uORFs ending within the protein ORF or by internal ORFs due to leaky scanning through the AUG of the protein ORF (Fig. 3C). A total of 276 cryptic peptides were derived from 3'-UTRs (Fig. 3D), of which 101 had no additional upstream in-frame stop codon and thus

## Identification of the Cryptic HLA-I Immunopeptidome

**Figure 3.**

Locations of cryptic peptides. **A**, Distribution of categories for identified cryptic peptides (FDR 10%, strong binders as predicted by NetMHCpan 4.0) from all analyzed datasets. **B**, Heatmap of the positions of cryptic peptides in 5'-UTRs. Each row represents a UTR sorted by length (gray area). Blue bars represent locations of peptides. Inlay, cumulative distribution of the relative peptide positions discarding the first 50 nt ["adjusted" (Adj.)] for 5'-UTR-derived peptides grouped by UTR length (indicated next to the heatmap). **C**, Same as **B** but for Off-Frame peptides. Locations within the corresponding protein ORF are shown. **D**, Same as **B** but for 3'-UTR peptides. **E**, Classification of 3'-UTR peptides. Blue, no additional in-frame stop codon upstream of the peptide. The frame indicates the shift relative to the protein ORF frame. Red, at least one additional in-frame stop codon upstream of the peptide. **F**, Classification of intronic peptides. Blue, no in-frame stop codon upstream of the peptide and no frameshift relative to the upstream coding exon. Red, either the upstream exon is noncoding, the peptide is out of frame, or there is at least one in-frame stop codon upstream of it. **G**, Heatmap and cumulative position distribution for peptides from noncoding RNAs. freq., frequency.

might originate from stop codon read-through (Fig. 3E; Frame = 0) or frameshifting (Frame = 1, -1). From the 133 cryptic peptides located within introns, 19 had no upstream in-frame stop codon and were likely generated by translation into a retained intron (Fig. 3F; Supplementary Fig. S8 for an example), consistent with a report identifying 17 HLA-I peptides from retained introns in cell lines (40). Almost 1,000 peptides were encoded by annotated ncRNAs. Most of the peptides were located within the first 700 nt (Fig. 3G) of the ncRNA, indicating that they were indeed translated. Finally, we identified 103 HLA-I peptides from intergenic regions.

For all peptides, we identified the first downstream in-frame stop codon and the closest upstream canonical (AUG) or noncanonical (CUG, GUG, ATC, and ACG) start codon (either in an unbiased manner or by prioritizing AUG > CUG > GUG > ATC > ACG; see Materials and Methods). For all cryptic categories, we observed the same expected distribution of start codon frequencies (Fig. 4A), providing strong evidence that all categories originated from *bona fide* ORFs and that translation initiation at noncanonical start codons could produce HLA-I-presented peptides.

The canonical pathway for HLA-I peptide presentation includes protein degradation by the proteasome into 9 to 20 amino acid-long peptides, transport into the endoplasmic reticulum (ER) by the transporter associated with antigen processing (TAP), N-terminal trimming by ERAP1/2 to 9 to 10 amino acids, and binding to HLA-I. For most cryptic peptides ( $n = 3,634$ ; 54%) independent of their category, the predicted start codon (with prioritization) was within 10 amino acids from the N-terminus of the peptide (Fig. 4B), and for 1,240 (>18%) cases, the peptide represented the C-terminus of the translation product (Fig. 4C). Thus, at least 10% ( $n = 710$ ) of cryptic peptides could directly enter the ER via TAP independently of processing by the proteasome or any cytoplasmic peptidases. This is in contrast to conventional peptides, which had uniform distance distributions for translation initiation sites (TIS) and stop codons (Fig. 4B

and C) with the exception that about twice as many peptides as expected from the uniform distribution were located at the C-terminus of the protein. We speculate that this reflects the number of cleavage events necessary to produce the peptide.

### The percentage of cryptic peptides varies among different alleles

The amino acid frequency distributions of conventional and cryptic peptide showed clear differences for both anchor residues (B and F pocket), and most prominently indicated more frequent basic residues in the F pocket for cryptic peptides (Supplementary Fig. S9A). This was also true after controlling for different background distributions of amino acid frequencies in and outside of the proteome (Fig. 4D). The HLA-I locus is highly polymorphic, and each HLA-I allele has a distinct sequence preference for ligands. Intriguingly, the percentage of cryptic peptides was highly HLA-I allele-specific and varied greatly (<1% to >15%) between different HLA-I alleles (Fig. 4E). For instance, consistently across experiments (Supplementary Table S4), approximately 10% of the ligands bound to alleles from the common A03 HLA-I supertype were cryptic peptides. The major determinant for ligand specificity of A03 is a basic residue at the C-terminus. We hypothesized that this might be due to a specific processing mechanism of cryptic peptides, such as a protease with specificity for basic amino acids. However, the percentage of processing-independent peptides (peptides ending directly at a stop codon) was constant for all alleles including A03 (Supplementary Fig. S9B), and the percentage of basic residues at the C-terminus of processed and processing-independent cryptic peptides was the same (Supplementary Fig. S9C). Thus, the higher frequency of cryptic peptides from A03 is not due to a distinct processing mechanism. In summary, we found that the percentage for cryptic peptide presentation is allele specific. The reasons for the preference of certain alleles to cryptic peptides, however, remain elusive.

Erhard et al.

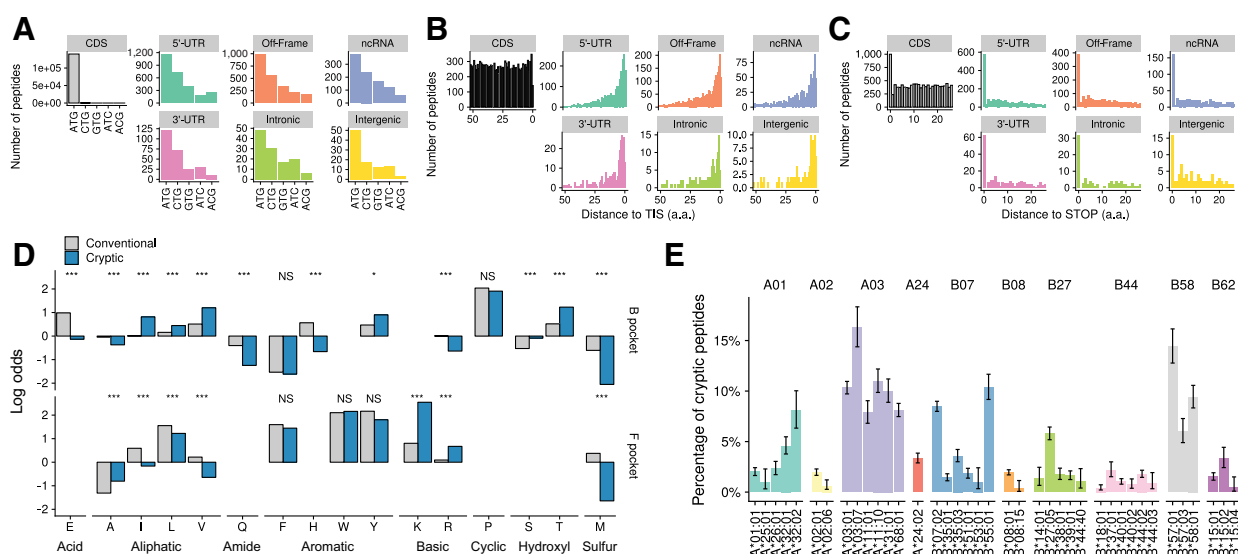


Figure 4.

Characteristics of ORFs encoding for cryptic peptides. **A**, Distribution of annotated start codons for conventional (CDS) peptides and putative start codons for cryptic peptides. In peptide cryptic ORF, the closest upstream canonical or noncanonical candidate was counted (unbiased start codon selection). **B**, Distribution of distance to the putative TIS. For conventional peptides, the annotated TIS was used. For cryptic peptides, the closest in-frame ATG was preferred over the closest CTG, which in turn was preferred over the subsequent candidate codons (prioritized start codon selection; for details see Materials and Methods). **C**, Distribution of distances to the next in-frame stop codon. **D**,  $\text{Log}_2$  odds of all amino acids with frequencies  $>1\%$  in B and F pockets are shown. For the odds, the raw frequency of amino acids at the second (B pocket) or last (F pocket) position in conventional or cryptic peptides was divided by the raw frequency of the respective amino acid four codon triplets upstream of the peptide location. Benjamini-Hochberg corrected  $P$  values according to Fisher exact test are indicated (NS, not significant; \*,  $P < 0.05$ ; \*\*\*,  $P < 0.001$ ). **E**, Percentages of cryptic peptides grouped by HLA supertypes for different HLA-I alleles in the merged peptide data. Error bars represent the 95% binomial confidence interval.

## Discussion

The success of cancer immunotherapies, such as immune checkpoint blockade (41) or adoptive T-cell transfer (42), demonstrates that cytotoxic T cells can efficiently recognize and eliminate tumor cells *in vivo*. However, the tumor-specific antigens that are recognized by the corresponding cytotoxic T cells mediating the antitumor effect are mostly unknown. Neoantigens came into focus as a potentially relevant class of tumor-specific targets (38, 39). Despite tremendous efforts, identified neoantigens are scarce. In contrast, cryptic HLA peptides that were first discovered more than 30 years ago (43) have largely been neglected as potential tumor-specific targets. One crucial reason for this is the lack of computational methods and tools for their efficient identification. Laumont and colleagues identified a number of tumor-specific cryptic HLA peptides in murine cancer cell lines, as well as in human tumor tissue. Their individualized proteogenomic approach was based on customized databases derived from RNA-seq of tumor cells and medullary thymic epithelial cells (16). In contrast, our *de novo* sequencing-based approach enables more efficient and more sensitive identification of cryptic peptides without the need for DNA or RNA-sequencing of the corresponding tumor sample. This enabled us to uncover that the frequency of these epitopes was highly HLA-I allele-specific and varied considerably between different HLA-I allomorphs. A large fraction of cryptic peptides apparently takes a shortcut in antigen processing. Their C-terminus is defined by translation termination at the stop codon directly downstream of the peptide and not by the proteasome. It is unclear whether the proteasome is required to cleave upstream of the peptide. However, a large number of cryptic ORFs are translated into poly-

peptides shorter than 25 amino acids. Thus, in contrast to all known peptides from large protein-coding ORFs, they might be presented in a proteasome-independent manner. A subset of the cryptic peptides presumably might derive from aberrant, tumor-specific expression, such as intron retention (40), frameshift mutations (44), alternative splicing (45), or translation events such as translation from unconventional 5' start sites (46) that might generate cryptic neoantigens. Further evidence for the existence of tumor-specific translation events comes from a study, in which comparative Ribo-seq analyses of hepatocellular carcinoma versus normal tissue were performed (47). It has been shown that knockdown of the 40S ribosomal protein S28 increases translation from noncanonical ORFs and initiation from non-AUG start codons (48). This indicates that variations in the translation machinery of tumor cells, for example, by mutation of ribosomal proteins or translation initiation factors, can lead to tumor-specific (cryptic) translation events. Hence, similar to classic neoantigens, all these types of peptides may result from tumor-specific mutations (e.g., by frameshift mutations, deletions/insertions, or by mutations generating new transcription or translation start sites). However, in contrast to classic neoantigens, they do not resemble self-peptides that only differ by a single amino acid and are, thus, more likely to induce tumor-specific immune responses (49). Thus, cryptic HLA peptides that can now be efficiently and reliably identified with Peptide-PRISM provide a rich source of potential targets for cancer immunotherapy that can be tested for tumor specificity and immunogenicity.

Peptides originating from PCPS were claimed to contribute up to  $>30\%$  of HLA-I ligandomes (20). However, the contribution of PCPS to immunopeptidomes has raised considerable controversy in the field.

## Identification of the Cryptic HLA-I Immunopeptidome

Here, we unequivocally showed that >99% of the reported PCPS peptides are false identifications and resulted from various kinds of methodologic errors. Our work refutes reported evidence that PCPS occurs frequently *in vivo* (2, 20).

## Disclosure of Potential Conflicts of Interest

F. Erhard reports grants from Deutsche Forschungsgemeinschaft during the conduct of the study and a patent for EP 20 170 185.1 pending to the European Patent Office. B. Schilling reports grants from Interdisciplinary Center for Clinical Research (IZKF) Würzburg during the conduct of the study and a patent for EP 20 170 185.1 pending to the European Patent Office. A. Schlosser reports a patent for EP 20 170 185.1 pending to the European Patent Office. No potential conflicts of interest were disclosed by the other author.

## Authors' Contributions

F. Erhard: Conceptualization, data curation, software, formal analysis, funding acquisition, visualization, methodology, writing—original draft. L. Dölken: Conceptualization, funding acquisition, writing—review and editing. B. Schilling:

Conceptualization, funding acquisition, writing—review and editing. A. Schlosser: Conceptualization, data curation, formal analysis, funding acquisition, investigation, visualization, methodology, writing—original draft.

## Acknowledgments

The authors thank Wolfgang Kastenmüller, Georg Gasteiger, and Elmar Wolf for critical comments on this article. This work was supported by a grant from the Interdisciplinary Center for Clinical Research (IZKF) Würzburg (to B. Schilling and A. Schlosser) and a grant from the Deutsche Forschungsgemeinschaft (FOR 2830, DO 1275/7-1; to F. Erhard and L. Dölken).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 11, 2019; revised March 11, 2020; accepted June 10, 2020; published first June 19, 2020.

## References

- Coulie PG, Van den Eynde BJ, van der Bruggen P, Boon T. Tumour antigens recognized by T lymphocytes: at the core of cancer immunotherapy. *Nat Rev Cancer* 2014;14:135–46.
- Liepe J, Sidney J, Lorenz FKM, Sette A, Mishto M. Mapping the MHC class I-spliced immunopeptidome of cancer cells. *Cancer Immunol Res* 2019;7:62–76.
- Mylonas R, Beer I, Iseli C, Chong C, Pak HS, Gfeller D, et al. Estimating the contribution of proteasomal spliced peptides to the HLA-I ligandome. *Mol Cell Proteomics* 2018;17:2347–57.
- Ingolia NT. Ribosome footprint profiling of translation throughout the genome. *Cell* 2016;165:22–33.
- Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, et al. Peptidomic discovery of short open reading frame–encoded peptides in human cells. *Nat Chem Biol* 2013;9:59–64.
- Orr MW, Mao Y, Storz G, Qian SB. Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Res* 2020;48:1029–42.
- Yewdell JW. DRiPs solidify: progress in understanding endogenous MHC class I antigen processing. *Trends Immunol* 2011;32:548–58.
- Anton LC, Yewdell JW. Translating DRiPs: MHC class I immunosurveillance of pathogens and tumors. *J Leukoc Biol* 2014;95:551–62.
- Starck SR, Shastri N. Nowhere to hide: unconventional translation yields cryptic peptides for immune surveillance. *Immunol Rev* 2016;272:8–16.
- Neeffes J, Reits EAJ, Vos JC, Gromme M. The major substrates for TAP *in vivo* are derived from newly synthesized proteins. *Nature* 2000;404:774–8.
- Wei J, Yewdell JW. Immunoribosomes: where's there's fire, there's fire. *Mol Immunol* 2019;113:38–42.
- Yewdell JW, Dersh D, Fähraeus R. Peptide channeling: the key to MHC class I immunosurveillance? *Trends Cell Biol* 2019;29:929–39.
- Erhard F, Halenius A, Zimmermann C, L'Hernault A, Kowalewski DJ, Weekes MP, et al. Improved Ribo-seq enables identification of cryptic translation events. *Nat Methods* 2018;15:363–6.
- Laumont CM, Daouda T, Laverdure JP, Bonneil É, Caron-Lizotte O, Hardy MP, et al. Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames. *Nat Commun* 2016;7:10238.
- Bassani-Sternberg M, Bräunlein E, Klar R, Engleitner T, Sinitcyn P, Audehm S, et al. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun* 2016;7:13404.
- Laumont CM, Vincent K, Hesnard L, Audemard É, Bonneil É, Laverdure JP, et al. Noncoding regions are the main source of targetable tumor-specific antigens. *Sci Transl Med* 2018;10:eaau5516.
- Shraibman B, Barnea E, Kadosh DM, Haimovich Y, Slobodin G, Rosner I, et al. Identification of tumor antigens among the HLA peptidomes of glioblastoma tumors and plasma. *Mol Cell Proteomics* 2019;18:1255–68.
- Ternette N, Olde Nordkamp MJM, Müller J, Anderson AP, Nicastrì A, Hill AVS, et al. Immunopeptidomic profiling of HLA-A2-positive triple negative breast cancer identifies potential immunotherapy target antigens. *Proteomics* 2018;18:1700465.
- Khodadoust MS, Olsson N, Wagar LE, Haabeth OAW, Chen B, Swaminathan K, et al. Antigen presentation profiling reveals recognition of lymphoma immunoglobulin neoantigens. *Nature* 2017;543:723–7.
- Liepe J, Marino F, Sidney J, Jeko A, Bunting DE, Sette A, et al. A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science* 2016;354:354–8.
- Faridi P, Li C, Ramarathinam SH, Vivian JP, Illing PT, Mifsud NA, et al. A subset of HLA-I peptides are not genomically templated: evidence for cis- and trans-spliced peptide ligands. *Sci Immunol* 2018;3:4–7.
- Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, et al. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol Cell Proteomics* 2012;11:M111.010587.
- Tran NH, Qiao R, Xin L, Chen X, Liu C, Zhang X, et al. Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. *Nat Methods* 2019;16:63–6.
- Aho A V, Corasick MJ. Efficient string matching: an aid to bibliographic search. *Commun ACM* 1975;18:333–40.
- Gupta N, Bandeira N, Keich U, Pevzner PA. Target-decoy approach and false discovery rate: when things may go wrong. *J Am Soc Mass Spectrom* 2011;22:1111–20.
- Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 2007;4:207–14.
- Frank AM, Savitski MM, Nielsen ML, Zubarev RA, Pevzner PA. De novo peptide sequencing and identification with precision mass spectrometry. *J Proteome Res* 2007;6:114–23.
- Nesvizhskii AI. Proteogenomics: concepts, applications and computational strategies. *Nat Methods* 2014;11:1114–25.
- Ma K, Vitek O, Nesvizhskii AI. A statistical model-building perspective to identification of MS/MS spectra with PeptideProphet. *BMC Bioinformatics* 2012;13:S1.
- Jürges C, Dölken L, Erhard F. Dissecting newly transcribed and old RNA using GRAND-SLAM. *Bioinformatics* 2018;34:i218–26.
- Köllmann C, Bornkamp B, Ickstadt K. Unimodal regression using Bernstein-Schoenberg splines and penalties. *Biometrics* 2014;70:783–93.
- Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol* 2017;199:3360–8.
- Andreatta M, Lund O, Nielsen M. Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach. *Bioinformatics* 2013;29:8–14.
- Rapin N, Hoof I, Lund O, Nielsen M. MHC motif viewer. *Immunogenetics* 2008;60:759–65.
- Hanada K, Yewdell JW, Yang JC. Immune recognition of a human renal cancer antigen through post-translational protein splicing. *Nature* 2004;427:252–6.
- Vigneron N. An antigenic peptide produced by peptide splicing in the proteasome. *Science* 2004;304:587–90.
- Faridi P, Li C, Ramarathinam SH, Vivian JP, Illing PT, Mifsud NA, et al. A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands. *Sci Immunol* 2018;3:ear3947.



**Erhard et al.**

38. Sahin U, Derhovannessian E, Miller M, Kloke BP, Simon P, Löwer M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* 2017;547:222–6.
39. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* 2017;547:217–21.
40. Smart AC, Margolis CA, Pimentel H, He MX, Miao D, Adeegbe D, et al. Intron retention is a source of neoepitopes in cancer. *Nat Biotechnol* 2018;36:1056–8.
41. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. *Science* 2018;359:1350–5.
42. Rosenberg SA, Restifo NP. Adoptive cell transfer as personalized immunotherapy for human cancer. *Science* 2015;348:62–8.
43. Boon T, Van Pel A. T cell-recognized antigenic peptides derived from the cellular genome are not protein degradation products but can be generated directly by transcription and translation of short subgenic regions. A hypothesis. *Immunogenetics* 1989;29:75–9.
44. Townsend A, Öhlén C, Rogers M, Edwards J, Mukherjee S, Bastin J. Source of unique tumour antigens. *Nature* 1994;371:662.
45. Kahles A, Lehmann KV, Toussaint NC, Hüser M, Stark SG, Sachsenberg T, et al. Comprehensive analysis of alternative splicing across tumors from 8,705 patients. *Cancer Cell* 2018;34:211–224.
46. Sendoel A, Dunn JG, Rodriguez EH, Naik S, Gomez NC, Hurwitz B, et al. Translation from unconventional 5' start sites drives tumour initiation. *Nature* 2017;541:494–9.
47. Zou Q, Xiao Z, Huang R, Wang X, Wang X, Zhao H, et al. Survey of the translation shifts in hepatocellular carcinoma with ribosome profiling. *Theranostics* 2019;9:4141–55.
48. Wei J, Kishton RJ, Angel M, Conn CS, Dalla-Venezia N, Marcel V, et al. Ribosomal proteins regulate MHC class I peptide generation for immunosurveillance. *Mol Cell* 2019;73:1162–1173.
49. Laumont CM, Perreault C. Exploiting non-canonical translation to identify new targets for T cell-based cancer immunotherapy. *Cell Mol Life Sci* 2018;75:607–21.

# Cancer Immunology Research

## Identification of the Cryptic HLA-I Immunopeptidome

Florian Erhard, Lars Dölken, Bastian Schilling, et al.

*Cancer Immunol Res* 2020;8:1018-1026. Published OnlineFirst June 19, 2020.

<b>Updated version</b>	Access the most recent version of this article at: doi: <a href="https://doi.org/10.1158/2326-6066.CIR-19-0886">10.1158/2326-6066.CIR-19-0886</a>
<b>Supplementary Material</b>	Access the most recent supplemental material at: <a href="http://cancerimmunolres.aacrjournals.org/content/suppl/2020/06/19/2326-6066.CIR-19-0886.DC1">http://cancerimmunolres.aacrjournals.org/content/suppl/2020/06/19/2326-6066.CIR-19-0886.DC1</a>

<b>Cited articles</b>	This article cites 48 articles, 11 of which you can access for free at: <a href="http://cancerimmunolres.aacrjournals.org/content/8/8/1018.full#ref-list-1">http://cancerimmunolres.aacrjournals.org/content/8/8/1018.full#ref-list-1</a>
-----------------------	--

<b>Citing articles</b>	This article has been cited by 6 HighWire-hosted articles. Access the articles at: <a href="http://cancerimmunolres.aacrjournals.org/content/8/8/1018.full#related-urls">http://cancerimmunolres.aacrjournals.org/content/8/8/1018.full#related-urls</a>
------------------------	---

<b>E-mail alerts</b>	<a href="#">Sign up to receive free email-alerts</a> related to this article or journal.
----------------------	--

<b>Reprints and Subscriptions</b>	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at <a href="mailto:pubs@aacr.org">pubs@aacr.org</a> .
-----------------------------------	--

<b>Permissions</b>	To request permission to re-use all or part of this article, use this link <a href="http://cancerimmunolres.aacrjournals.org/content/8/8/1018">http://cancerimmunolres.aacrjournals.org/content/8/8/1018</a> . Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.
--------------------	--