The Journal of Immunology

Upstream Position of Proline Defines Peptide –HLA Class I Repertoire Formation and CD8 + T Cell Responses

Ayumi Hongo, Takayuki Kanaseki, Serina Tokita, Vitaly Kochin, Sho Miyamoto, Yuiko Hashino, Amy Codd, Noriko Kawai, Munehide Nakatsugawa, Yoshihiko Hirohashi, Noriyuki Sato and Toshihiko Torigoe

J Immunol published online 1 April 2019
http://www.jimmunol.org/content/early/2019/03/29/jimmunol.1900029

Supplementary Material **http://www.jimmunol.org/content/suppl/2019/03/29/jimmunol.1900029.DCSupplemental**

The Journal of Immunology is published twice each month by The American Association of Immunologists, Inc., 1451 Rockville Pike, Suite 650, Rockville, MD 20852

Print ISSN: 0022-1767 Online ISSN: 1550-6606.

Published April 1, 2019, doi:10.4049/jimmunol.1900029

Upstream Position of Proline Defines Peptide–HLA Class I Repertoire Formation and CD8$^+$ T Cell Responses

Ayumi Hongo,* Takayuki Kanaseki,* Serina Tokita,* Vitaly Kochin,† Sho Miyamoto,‡ Yuiko Hashino,* Amy Codd,§ Noriko Kawai,* Munehide Nakatsugawa,* Yoshihiko Hirohashi,* Noriyuki Sato,* and Toshihiko Torigoe*

Cytotoxic CD8$^+$ T lymphocytes (CTLs) recognize peptides displayed by HLA class I molecules on cell surfaces, monitoring pathological conditions such as cancer. Difficulty in predicting HLA class I ligands is attributed to the complexity of the Ag processing pathway across the cytosol and the endoplasmic reticulum. By means of HLA ligandome analysis using mass spectrometry, we collected natural HLA class I ligands on a large scale and analyzed the source-protein sequences flanking the ligands. This comprehensive analysis revealed that the frequency of proline at amino acid positions 1–3 upstream of the ligands was selectively decreased. The depleted proline signature was the strongest among all the upstream and downstream profiles. Experiments using live cells demonstrated that the presence of proline at upstream positions 1–3 attenuated CTL responses against a model epitope. Other experiments, in which N-terminal–flanking Ag precursors were confined in the endoplasmic reticulum, demonstrated an inability to remove upstream prolines regardless of their positions, suggesting a need for synergistic action across cellular compartments for making the proline signature. Our results highlight, to our knowledge, a unique role and position of proline for inhibiting downstream epitope presentation, which provides a rule for defining natural peptide–HLA class I repertoire formation and CTL responses. ***The Journal of Immunology*****, 2019, 202: 000–000.**

Cytotoxic CD8$^+$ T lymphocytes (CTLs) recognize peptides presented by HLA molecules on cell surfaces, monitoring cellular abnormalities such as viral infection or tumor transformation. Ag processing begins in the cytosol, where endogenous proteins are digested by the proteasomes and other peptidases (1–4). Generated peptide precursors of protein fragments are then transported into the endoplasmic reticulum (ER) through the TAP, where the aminopeptidase ERAAP (ERAP1) customizes the peptides and peptide–HLA class I (pHLA I) complexes are formed (5–7). These pHLA I complexes are sorted onto a surface and subsequently serve as targets of CTL recognition. Thus, HLA class I peptide presentation is a selective process comprising multiple steps across distinct cellular compartments.

Accurate prediction of CTL epitopes presented by HLA class I molecules is an urgent issue for therapeutic purposes. Cancer patients' T cells recognize neoantigens that arise from gene mutations (8); however, it is currently challenging to predict those epitopes from hundreds or thousands of gene mutations (9). The difficulty of accurate prediction is mostly due to the complexity of Ag processing. Although attempts to predict enzymatic properties of the proteasomes or TAP transport have been made (10, 11), many of the currently used algorithms focus on the affinity or stability between ligands and HLA class I molecules (12–15). This leaves out the intracellular mechanisms by which peptides are selected from the vast number of candidate protein fragments. Meanwhile, recent advances in proteomics have revealed a landscape of naturally presented HLA class I ligands on a large scale (16–19). The sequences detected by HLA ligandome analysis are the "footprints" of the Ag processing machinery, representing the intracellular selection process.

In this study, we isolated naturally presented HLA ligands from tumor cells and searched the source-protein sequences flanking the ligands for the footprints of Ag processing. To maintain natural presentation signatures, if any, we leveraged affinity Abs against specific HLA class I alleles and avoided the conventional deconvolution process of assigning the mixture of peptides isolated from multiple HLA alleles into their cognate alleles using algorithms (20). Our results uncovered an upstream signature that was shared across different HLA class I types. The frequency of proline located at amino acid positions 1–3 upstream decreased and was an unfavorable signature for HLA presentation, suggesting that the presence of a preceding proline inhibited HLA class I presentation of downstream epitopes. This finding was validated by model experiments using live cells that demonstrated the sequential action of processing machinery across cellular compartments in making the proline signature. Our results demonstrate how proline at upstream positions influences Ag processing and surface pHLA I repertoire formation. Integration of this rule could reduce the false positive rate and thereby improve the accuracy of epitope prediction.

*Department of Pathology, Sapporo Medical University, Sapporo, Hokkaido 060-8556, Japan; †Department of Immunology, Nagoya University, Nagoya 466-8550, Japan; ‡Department of Oral Surgery, Sapporo Medical University, Sapporo, Hokkaido 060-8556, Japan; and §Division of Cancer and Genetics, School of Medicine, Cardiff University, Cardiff CF14 4XN, Wales, United Kingdom

ORCIDs: 0000-0001-6747-6036 (A.H.); 0000-0002-9463-5917 (T.T.).

Received for publication January 8, 2019. Accepted for publication March 8, 2019.

This work was supported by Japan Agency for Medical Research and Development (AMED) Grant JP18cm0106334 (to T.K.), the Takeda Science Foundation (to T.K.), the Kobayashi Foundation for Cancer Research (to T.K.), AMED Grant JP18cm0106309 (to T.T.), and a Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (to T.T.).

Address correspondence and reprint requests to Dr. Takayuki Kanaseki, Department of Pathology, Sapporo Medical University, S1W17, Sapporo, Hokkaido 060-8556, Japan. E-mail address: kanaseki@sapmed.ac.jp

The online version of this article contains supplemental material.

Abbreviations used in this article: ATCC, American Type Culture Collection; CTL, cytotoxic CD8$^+$ T lymphocyte; ER, endoplasmic reticulum; FC, fold change; FDR, false-discovery rate; MS, mass spectrometry; MS/MS, tandem MS; m/z, mass-to-charge ratio; pHLA I, peptide–HLA class I; WT, wild-type.

Materials and Methods

Cells

The histological origins and HLA genotypes of the human tumor cell lines are as follows: MKN45 (Japanese Collection of Research Bioresources Cell Bank), gastric cancer, HLA-A*24:02, B*67:02, B*52:01, and C*12:02; KHM2B (a gift from Dr. T. Sonoki, Wakayama Medical University), Burkitt lymphoma, HLA-A*24:02, B*52:01, and C*12:02; SW480 (American Type Culture Collection [ATCC]), colon cancer, HLA-A*24:02, A*02:01, B*07:02, B*15:18, and C*07:04; and DU145 (ATCC), prostate cancer, HLA-A*33:03, B*50:01, B*57:01, and C*06:02. The HLA genotype of KHM2B was determined using a commercially available service (LSI Medience), and the others were obtained from the TRON Cell Line Portal (http://celllines.tron-mainz.de) (21). HEK293T cells were purchased from the ATCC, and T2 cells expressing HLA-A24 (T2-A24) were a gift from Dr. K. Kuzushima (Aichi Cancer Center Research Institute). 293T-A24 is a 293T clone stably expressing an HLA-A*24:02 gene construct and was established in our laboratory. A CTL clone specific to IV9/HLA-A24 is described elsewhere (22).

Abs for HLA class I ligand isolation

Hybridomas for anti–HLA-A24 (C7709A2; a gift from Dr. P. G. Coulie, Ludwig Institute for Cancer Research), anti–HLA-A02 (BB7.2; ATCC), and pan–HLA class I (W6/32; ATCC) were cultured in Hybridoma-SFM (Life Technologies) supplemented with 1% penicillin/streptomycin in CELLine bioreactor flasks (Corning). Produced mAbs were condensed and collected through a semipermeable membrane during cell culture.

HLA class I ligand elution

Cell lysates were prepared from ~1.0×10^9 KHM2B, MKN45, SW480, and DU145 cells with lysate buffer containing 0.6% CHAPS (Dojindo Molecular Technologies) and 1× complete protease inhibitor (Roche) in PBS. The pHLA I complexes were captured using affinity chromatography of either C7709A2, BB7.2, or W6/32 Abs coupled to cyanogen bromide–activated Sepharose 4B (GE Healthcare). The HLA ligands were eluted with 0.2% trifluoroacetic acid, passed through a 10-kDa ultra-centrifugal filter (Millipore), condensed by vacuum centrifugation, and desalted using C18 ZipTip (Millipore).

Liquid chromatography–tandem mass spectrometry analysis

This procedure has been previously described (19, 22). Samples were loaded into a nanoflow liquid chromatograph (Easy-nLC 1000 system; Thermo Fisher Scientific) online-coupled to an Orbitrap mass spectrometer equipped with a nanospray ion source (Q-Exactive Plus; Thermo Fisher Scientific). We used a 75 μm × 20 cm capillary column with a particle size of 3 μm (NTCC-360; Nikkyo Technos) and separated the samples with a linear gradient ranging from 3 to 30% buffer B (100% acetonitrile and 0.1% formic acid) at a flow rate of 300 nl/min for 80 min. In mass spectrometry (MS) analysis, survey scan spectra were acquired at a resolution of 70,000 at 200 mass-to-charge ratio (m/z) with a target value of 3e6 ions ranging from 350 to 2000 m/z with charge states between 1+ and 4+. We applied a data-dependent top 10 method that generates high-energy collision dissociation fragments for the 10 most intense precursor ions per survey scan. Tandem MS (MS/MS) resolution was 17,500 at 200 m/z with a target value of 1e5 ions. For database matching of the MS/MS spectra, we used both Sequest HT (Thermo Fisher Scientific) and Mascot version 2.6 (Matrix Science) algorithms embedded in the Proteome Discoverer 2.2 platform (Thermo Fisher Scientific). The peak lists were searched against the Swiss-Prot human databases with a tolerance of precursor ions at 10 ppm and fragment ions at 0.02 Da. We applied a false-discovery rate (FDR) of 0.01 as a stringent threshold and removed known contaminant sequences.

Upstream and downstream profiles of HLA ligands

Up to 15 upstream and downstream amino acid residues adjoining all the isolated 9–11-mer HLA ligands were derived from the Swiss-Prot database (https://www.uniprot.org). Fold changes (FC) were calculated as follows: FC = the number of a given amino acid at a given upstream (or downstream) position / the median number of that amino acid across upstream positions 1–15 (or downstream positions 1–15). Frequency in the heat map is shown as $\log_2$(FC).

DNA constructs

The ASB4 gene in SW480 was amplified using a pair of PCR primers containing a FLAG tag (5′-ATTAGGATCCATGGACGGCACCACTGCCC-3′ and 5′-AGATCTCGAGTTACTTATCGTCGTCATCCTTGTAATCATAAATAATTCC-3′) and inserted into pcDNA3.1 (Invitrogen). Single amino acid substitution of residues preceding an IV9 epitope was carried out using site-directed mutagenesis with the following PCR primer pairs: P.U1, 5′-GCTGCCCCTATATACCCTCCACAGTTC-3′ and 5′-GTATATAGGGGCAGCCCCATGGTTCAA-3′; P.U2, 5′-GGGCTCCTCGAATATACCCTCCACAG-3′ and 5′-ATATTCGAGGAGCCCCATGGTTCAACAG-3′; P.U3, 5′-CATGGGCCTGCCCGAATATACCCTCCA-3′ and 5′-TCGGGCAGGCCCATGGTTCAACAGGAG-3′; P.U4, 5′-AACCATCCTGCTGCCCGAATATACCCT-3′ and 5′-GGCAGCAGGATGGTTCAACAGGAGCTG-3′; P.U5, 5′-TTGAACCCTGGGGCTGCCCGAATATAC-3′ and 5′-AGCCCCAGGGTTCAACAGGAGCTGGTA-3′; P.U6, 5′-CTGTTGCCTCATGGGGCTGCCCGAATA-3′ and 5′-CCCATGAGGCAACAGGAGCTGGTAGCA-3′; N.U1, 5′-GCTGCCAATATATACCCTCCACAGTTC-3′ and 5′-GTATATATTGGCAGCCCCATGGTTCAA-3′; D.U2, 5′-GGGCTGATCGAATATACCCTCCACAG-3′ and 5′-ATATTCGATCAGCCCCATGGTTCAACAG-3′; M.U2, 5′-GGGCTATGCGAATATACCCTCCACAG-3′ and 5′-ATATTCGCATAGCCCCATGGTTCAACAG-3′; V.U3, 5′-CATGGGGTAGCCCGAATATACCCTCCA-3′ and 5′-TCGGGCTACCCCATGGTTCAACAGGAG-3′; and M.U3, 5′-CATGGGATGGCCCGAATATACCCTCCA-3′ and 5′-TCGGGCCATCCCATGGTTCAACAGGAG-3′. All the nucleotide substitutions were verified with DNA sequencing. As for ER series constructs, an ER-targeting signal sequence derived from the adenovirus E3/19K protein (5′-ATGAGGTACATGATCCTGGGCCTGCTGGCCCTGGCCGCCGTGTGCAGCGCCGCCGCCGCCATCGTGATGGAC-3′), followed by a sequence encoding flanking residues (5′-AACCATGGGGCTGCCCGA-3′) and the IV9 epitope coupled with a stop codon (5′-ATATACCCTCCACAGTTCCATAAGGTGTAA-3′) were cloned into pLenti-puro (Addgene). Nucleotide substitutions of the flanking sequences were carried out using site-directed mutagenesis with the following PCR primer pairs: ER.P.U1, 5′-GCTGCCCCTATATACCCTCCACAGTTC-3′ and 5′-GTATATAGGGGCAGCCCCATGGTTGTC-3′; ER.P.U2, 5′-GGGGCTCCTCGAATATACCCTCCACAG-3′ and 5′-TATTCGAGGAGCCCCATGGTTGTCCAT-3′; ER.P.U3, 5′-CATGGGCCTGCCCGAATATACCCTCCA-3′ and 5′-TCGGGCAGGCCCATGGTTGTCCATCAC-3′; ER.P.U4, 5′-AACCATCCTGCTGCCCGAATATACCCT-3′ and 5′-GGCAGCAGGATGGTTGTCCATCACGAT-3′; ER.P.U5, 5′-GACAACCCTGGGGCTGCCCGAATATAC-3′ and 5′-AGCCCCAGGGTTGTCCATCACGATGGC-3′; and ER.P.U6, 5′-ATGGACCCTCATGGGGCTGCCCGAATA-3′ and 5′-CCCATGAGGGTCCATCACGATGGCGGC-3′. All the nucleotide substitutions were verified with DNA sequencing. Using the SignalP prediction algorithm (http://www.cbs.dtu.dk/services/SignalP), we confirmed that none of the inserted 5-aa flanking residues were cleaved by a signal peptidase.

Western blot

Cells were lysed with a buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM EDTA, 1% NP40, and a protease inhibitor mixture (Roche) for 30 min at 4°C. Samples were separated by 12% SDS-PAGE, transferred to PVDF membranes, and blocked with 5% milk in PBS for 1 h at room temperature. Abs used for blotting were as follows: mouse anti-Flag (1:2000), mouse anti–β-actin (1:5000), and HRP-conjugated goat anti-mouse Ig (1:5000). Signals were detected using ECL Western Blotting Detection Reagents (GE Healthcare) and an Odyssey Fc imaging system (LI-COR Biosciences).

IFN-γ ELISPOT assay

IFN-γ production by IV9-specific CTL clones was assessed using the ELISPOT assay (human IFN-γ ELISPOT set; BD Biosciences) according to the manufacturer's instructions. A total of 1–5 × 10^4 IV9-CTL clones were incubated overnight with the equivalent number of target cells that were either 293T-A24 cells transiently expressing one of a series of ASB4 Ag constructs, or T2-A24 cells stably expressing one of a series of ER-IV9 Ag constructs. The wells were incubated with biotinylated anti-human IFN-γ for 2 h at room temperature followed by an ELISPOT streptavidin–HRP Ab for 1 h. Spots were visualized using the ELISPOT AEC Substrate Set (BD Biosciences). For blocking the activity of the proteasomes, target cells were preincubated with 0.02 μM carfilzomib (ApexBio Technology) or 0.02 μM bortezomib (Selleck) for 5 h. For blocking TAP transportation, the ICP47 gene construct (a gift from Dr. N. Hirano, University of Toronto) was transiently expressed in 293T-A24 cells along with ASB4 Ag constructs (23).

RT-PCR

Total RNA was isolated from T2-A24 cells expressing a panel of Ag constructs using an RNeasy Mini Kit (QIAGEN) according to the manufacturer's instructions. The synthesis of cDNA was performed on 2 μg of total RNA by reverse transcription using the RevertAid RT Reverse Transcription Kit (Thermo Fisher Scientific). RT-PCR mixtures were

initially incubated at 94°C for 2 min, followed by 35 cycles of denaturation at 94°C for 15 s, annealing at 60°C for 30 s, and extension at 72°C for 15 s. Primer pairs were as follows: ER-IV9, 5′-ATGAGGTACATGATCCTGGG-3′ and 5′-CACCTTATGGAACTGTGGAG-3′ (product size 117 bp); and G3PDH, 5′-ACCACAGTCCATGCCATCAC-3′ and 5′-TCCACCACCCTGTTGCTGTA-3′ (product size 452 bp).

Results

HLA class I ligandome analysis demonstrates the profile of naturally presented peptides

By means of HLA-A24– and -A02–reactive mAbs, as well as a pan–HLA class I Ab, we directly captured peptides that were naturally presented by corresponding HLA class I molecules from a panel of tumor cell lines. The peptides were subsequently analyzed using MS and sequenced at a stringent FDR threshold of 0.01. The data included 591 and 381 HLA-A*24:02–bound unique peptides from KHM2B and MKN45, respectively, 336 HLA-A*02:01–bound unique peptides from SW480, and 1374 HLA class I (A*33:03, B*50:01, B*57:01, and C*06:02)–bound unique peptides from DU145 (Fig. 1A). Both HLA-A*24:02 and -A*02:01 peptides were directly eluted from tumor cells expressing multiple HLA class I alleles using HLA allele–specific mAbs so as not to be influenced by potential biases yielded by a conventional deconvolution process or manipulation such as artificial HLA gene engineering of host cells. To validate the specificity of the pipeline, we assessed both length profiles and amino acid conservation of detected peptide sequences. Among the detected peptides, 9-mer peptides were enriched and the known HLA-binding anchor motifs were conserved at P2 and P9, both of which suggested the successful isolation of HLA-bound peptides minimizing contamination (if any) of peptides from other HLA alleles (Fig. 1B, 1C).

The frequency of upstream proline is decreased in naturally presented HLA class I ligands

Considering that the number of HLA class I molecules is limited, they cannot afford to present every peptide fragment generated inside a cell. In actuality, the number of the peptides simultaneously presented on a cell surface could be 0.01% or less of the peptides generated inside a cell (24). Endogenous Ag processing across the cytosol and the ER precedes pHLA I formation in the ER, digestion of source proteins, and selection of peptides appropriate for HLA class I binding. To investigate footprints of Ag processing in shaping the pHLA I repertoires in our data sets, we reconstituted the source-protein sequences that had originally surrounded the HLA ligands. Up to 15 aa residues upstream (upstream positions 1–15) and downstream (downstream positions 1–15) of the 2450 nonredundant natural HLA ligands were recovered from the Swiss-Prot database and classified according to HLA or cell type along with the 9–11-mer HLA ligands (Fig. 2A). First, we compared the amino acid composition across the HLA ligand and upstream and downstream sequences. The standard composition in the human proteome registered in the Swiss-Prot database served as a reference (Fig. 2B). The compositions of the HLA class I ligands varied across HLA types, and each composition was different from the reference, mainly because of the conservation of anchor motifs according to HLA types (Fig. 2C). In contrast, the compositions of upstream and downstream sequences were uniformly consistent with the reference, which indicated a lack of particular residues necessary for presentation (Fig. 2D, 2E). Lack of variations across HLA types may also suggest that Ag processing that removes these flanking extensions is not particular between HLA types.

Because no remarkable changes in compositions were found, we further investigated positional biases within upstream or downstream residues (Fig. 3A, 3B). We found that positional variations in most amino acid residues were modest, except for proline. In the upstream profile, proline at upstream positions 1–3 was selectively depleted in a row; its frequency was reduced to one half to one fourth of the other positions (Fig. 3A). This unique signature was not limited to a certain sample or HLA type but was uniformly observed across data sets we have tested (Supplemental Fig. 1). In the cases of HLA-A24 (KHM2B and

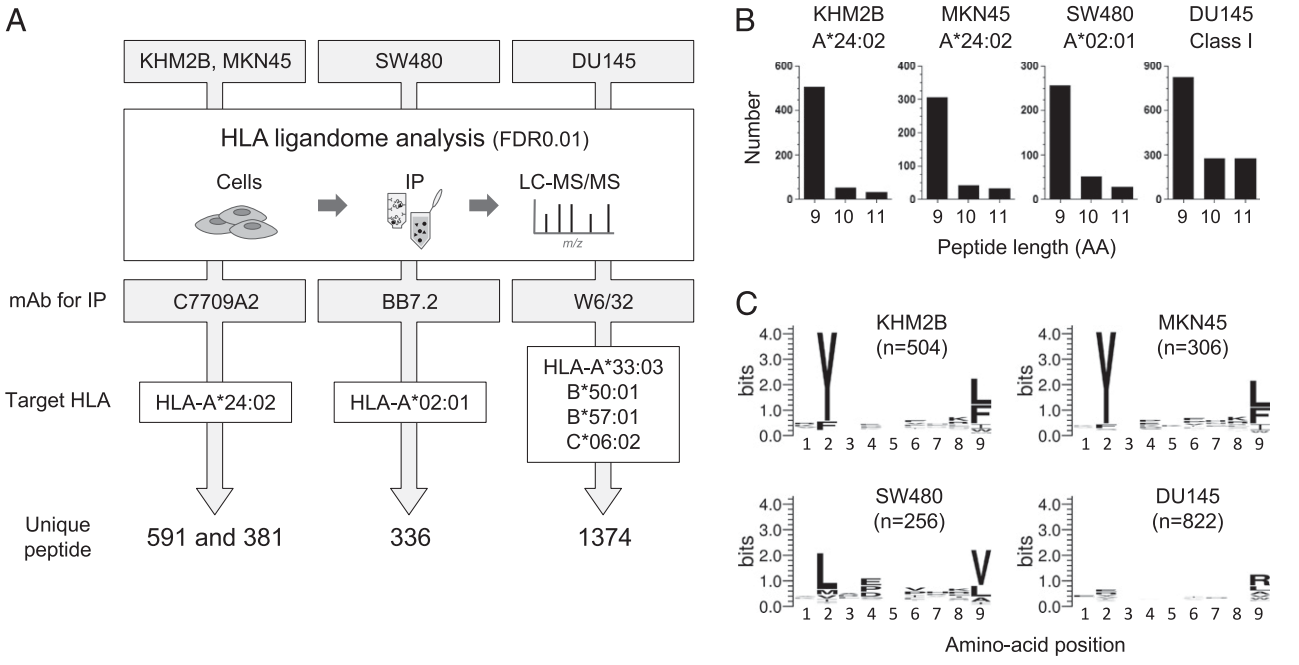

FIGURE 1. HLA ligandome analysis and collection of HLA-A24, -A2, and other HLA class I natural ligands. (**A**) A schematic pipeline of HLA ligandome analysis with cell names, affinity Abs, target HLA alleles, and the resulting numbers of identified nonredundant HLA ligands (FDR of 0.01). (**B**) Length distribution of detected peptides. The *x*-axis, peptide lengths (amino acids); the *y*-axis, numbers of unique peptides. (**C**) Conventional logo sequence analysis of 9-mer ligands (https://weblogo.berkeley.edu/logo.cgi).

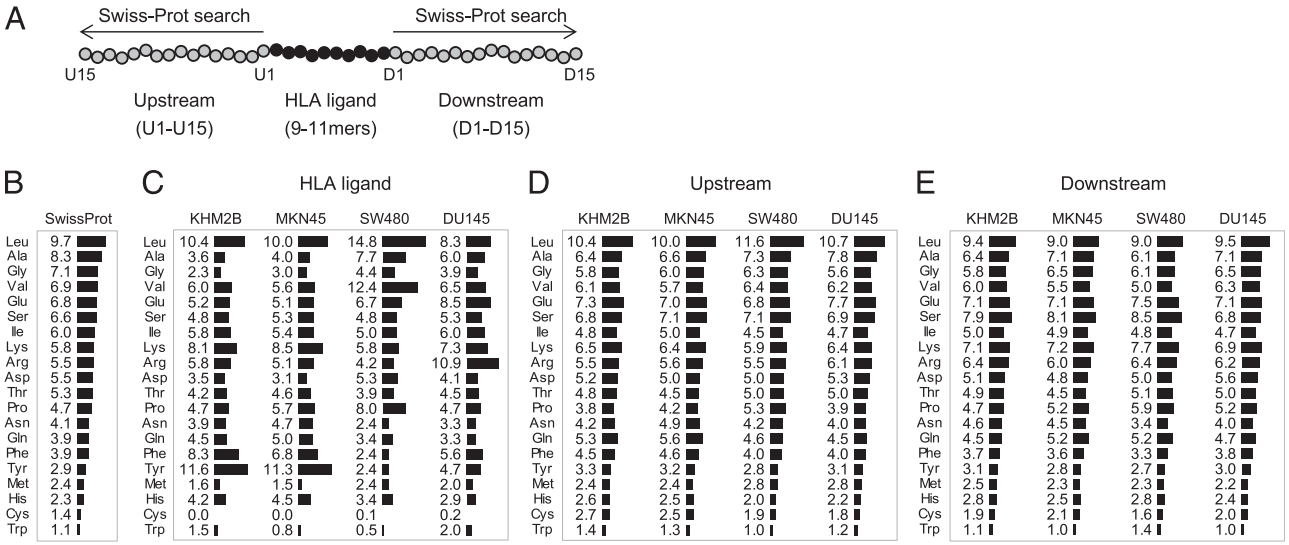

FIGURE 2. Upstream and downstream profiles of HLA class I ligands. (**A**) Up to 15 upstream and downstream residues flanking outside of the 9–11-mer naturally presented HLA class I ligands were recovered from the Swiss-Prot database (https://www.uniprot.org) and reconstituted according to their cell types (HLA types). (**B**) A standard amino acid composition from the UniProtKB/Swiss-Prot data bank (percentage) (https://web.expasy.org/protscale/pscale/A.A.Swiss-Prot.html, release 2013_04). (**C**–**E**) The amino acid composition (percentage) in 9–11-mer HLA ligands (C) and upstream (D) and downstream (E) residues of the HLA ligands.

MKN45), HLA-A2 (SW480), and a mixture of class I (DU145) ligands, the depletion of proline was concentrated at upstream positions 1–3 but not found at upstream position 4 or more upstream positions (Fig. 3C). As for downstream profiles, we found that arginine, histidine, and serine were slightly enriched at downstream position 1. These signatures were consistent with a recent report in which single HLA class I–expressing B721.221 cell lines were used that suggested their conservation across different cell types and HLA alleles (Fig. 3B) (25).

Presence of proline at upstream positions 1–3 inhibits CTL responses against the following epitope

Depletion of proline at upstream positions 1–3 yielded its positional bias in upstream flanking sequences, suggesting that the

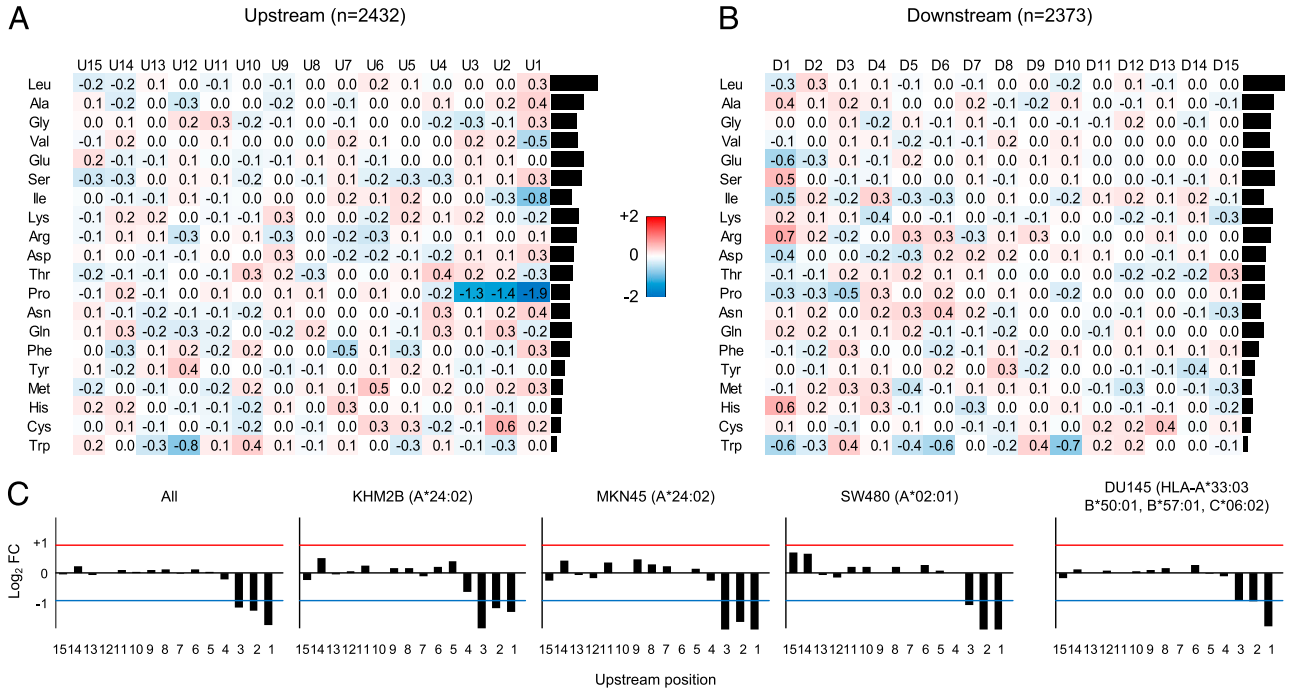

FIGURE 3. Positional biases of amino acids in upstream and downstream profiles. (**A** and **B**) Distribution of amino acids across upstream (A) and downstream (B) positions. The heat maps indicate the frequency of a given residue at a given position relative to the other position, shown as $\log_2$(FC). FC was calculated as follows: the number of a given amino acid at a given upstream (or downstream) position / the median number of that amino acid across upstream positions 1–15 (or downstream positions 1–15). Enriched and depleted residues are indicated as red and blue, respectively. Shown is the summary of KHM2B, MKN45, SW480, and DU145. Appended bar graphs indicate amino acid composition (percentage) in each data set. (**C**) Proline distribution in upstream sequences (upstream positions 1–15) across samples that carried different sets of peptides from different HLA alleles.

presence of proline had deleterious effects on Ag processing of the following peptides. Because this hypothesis has not been tested, to our knowledge, we prepared 293T-A24 cells expressing a panel of ASB4 Ag constructs (Fig. 4A). ASB4 is a known cancer-specific Ag containing an IV9 epitope that is naturally presented by HLA-A24 of colon cancer cells (22). The original ASB4 protein harbors proline at upstream position 15, but not at upstream positions 1–14, of the IV9 epitope. In each construct, the HLA-A24 ligand (IV9) followed the wild-type (WT) or a panel of substituted sequences in which proline or other indicated residues were substituted across upstream positions 1–6. The uniform protein expression of each Ag construct was verified using a tag as a surrogate marker (Fig. 4B). Any substitutions made no detectable changes in total HLA-A24 expression levels on cell surfaces (Supplemental Fig. 2A). The IV9-specific CTL clone readily responded to the WT construct and produced IFN-γ; however, the presence of proline at upstream positions 1–3 selectively attenuated the CTL responses down to 21–40% of WT (Fig. 4C). The CTL responses were not disturbed by the presence of proline at upstream positions 4–6. Likewise, the CTL responded well to the epitopes following randomly substituted amino acids with charged (aspartic acid), polar uncharged (asparagine), or hydrophobic (methionine and valine) side chains.

Ag processing begins in the cytosol, where epitope precursors are generated and subsequently transported into the ER through the TAP. To clarify the cellular compartment responsible for the proline signature, we took advantage of TAP and proteasome inhibitors. Coexpression of a TAP inhibitor, HSV-derived ICP47, successfully decreased total HLA class I expression levels of the host cells without changing the level of Ag expression (Supplemental Fig. 2B, 2C) (23). As we expected, coexpression of ICP47 influenced the CTL response, reducing IFN-γ production down to a nearly undetectable level (Fig. 4D). Moreover, treatment with two types of proteasome inhibitors also reduced the CTL response (Fig. 4E). These results suggested that proteasomal digestion and TAP transportation were necessary for Ag processing in this setting. Therefore, we found that the presence of proline at upstream positions 1–3, but not other amino acids, selectively inhibited CTL responses when an Ag protein was expressed in the cytosol and endogenously processed.

Inability of Ag processing in the ER to remove upstream prolines

C-terminal ends of HLA ligands are determined in the cytosol prior to transportation into the ER, whereas N-terminal–flanking extensions of potential ligands are trimmed by the ER-resident aminopeptidase ERAAP (ERAP1) (26–29). Therefore, the ER is the final compartment for shaping the pHLA I repertoire on cell surfaces, and processing defects in the ER lead directly to alterations in T cell recognition (30). To investigate the link between the Ag processing in the ER and generation of the proline signature, we modified a panel of Ag constructs to confine their Ag expression within the ER. The series of constructs had an ER-targeting signal sequence followed by N-terminal–flanking residues containing a proline at upstream positions 1–6, and the IV9 epitope was coupled with a stop codon at the C-terminal end (Fig. 5A). Each construct was introduced into TAP-deficient T2-A24, and we confirmed that the gene expression levels of all the mutant Ags (ER.P.U1–ER.P.U6) were uniform and comparable to the WT (ER.WT) (Fig. 5B). The IV9-specific CTL clone recognized T2-A24–expressing ER.WT, suggesting the successful removal of flanking residues in this setting; however, the IFN-γ production against any of the mutant Ags (ER.U1–U6) was significantly decreased (Fig. 5C). This result indicated a deleterious effect of upstream prolines on Ag processing in the ER, and in sharp contrast to the cytosolic Ag series in Fig. 4, prolines at any of the six upstream positions equally impaired the CTL responses. Carfilzomib or bortezomib treatments did not influence the response against ER.WT, suggesting that proteasomal digestion was not involved in this setting (Fig. 5D). We did not observe a decrease in surface HLA-A24 expression levels of T2-A24–expressing ER.U1–U6, presumably because the Ab used for flow

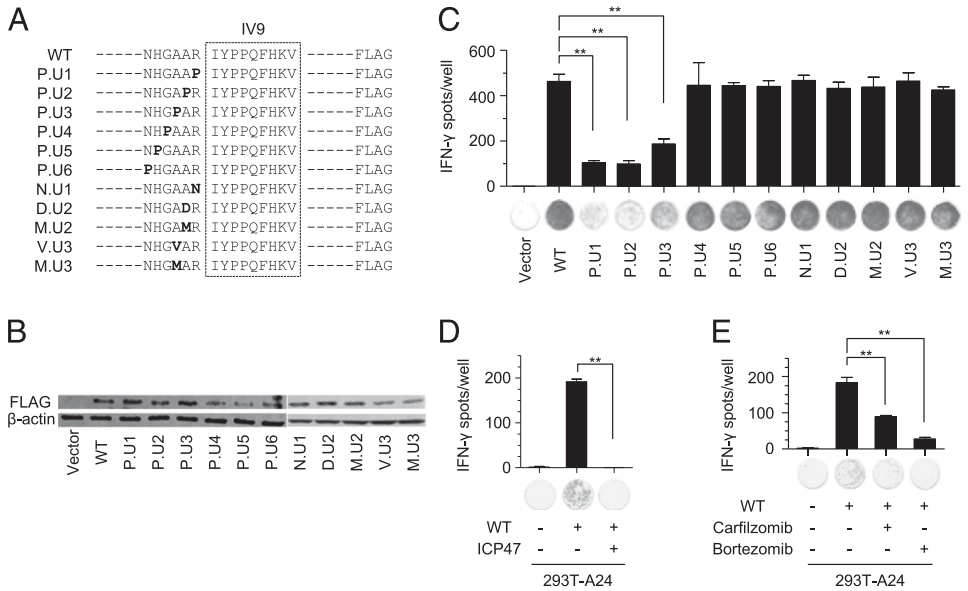

FIGURE 4. Presence of proline at upstream positions 1–3 selectively inhibits CTL responses against the following epitope. (**A**) A panel of ASB4 Ag constructs. The WT construct contains an HLA-A24–restricted 9-mer CTL epitope (IV9) as well as a FLAG sequence. In a series of mutant ASB4 constructs, bold letters indicate the substitution at upstream positions. (**B**) Western blot of 293T-A24 expressing indicated ASB4 constructs. Shown is a representative blot of three independent experiments. (**C**) IFN-γ production of an IV9-specific CTL clone against 293T-A24 expressing indicated ASB4 constructs. (**D** and **E**) IFN-γ production by the CTL clone against 293T-A24 expressing WT in the presence or absence of ICP47 expression (D) or in the presence or absence of 20 nM carfilzomib or bortezomib (E). IFN-γ spots in the ELISPOT assay are shown as mean ± SEM (n = 3), and p values were calculated using a two-tailed t test. $^{**}p < 0.01$.

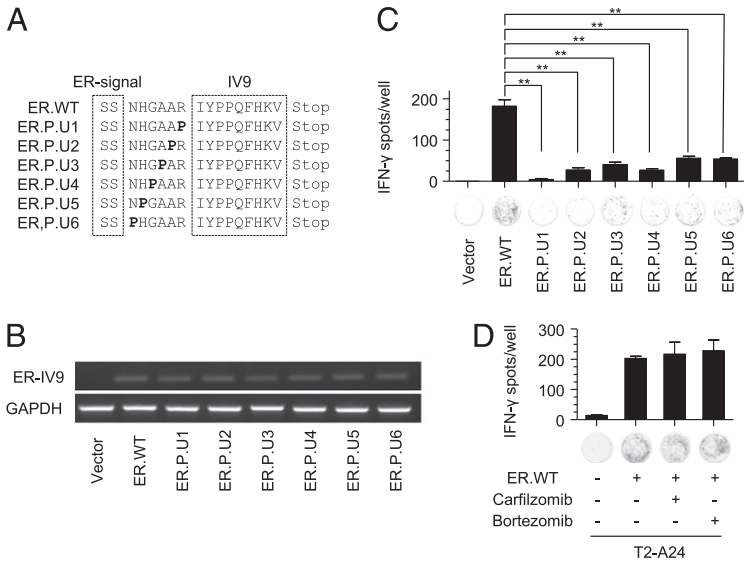

FIGURE 5. Ag processing in the ER fails to remove upstream prolines. (**A**) A panel of ER-IV9 constructs. Each construct contains an ER signal sequence followed by N-terminal–flanking residues as well as IV9 coupled with a stop codon. (**B**) RT-PCR of T2-A24 cells stably expressing indicated ER-IV9 constructs. Shown is a representative image of three independent experiments. (**C**) IFN-γ production of the IV9-specific CTL clone against 293T-A24 cells expressing indicated ER-IV9 constructs. (**D**) IFN-γ production of the CTL clone against 293T-A24 cells expressing ER.WT in the presence or absence of 20 nM carfilzomib or bortezomib. IFN-γ spots from the ELISPOT assay are shown as mean ± SEM ($n = 3$), and p values were calculated using a two-tailed t test. $^{**}p < 0.01$.

cytometry was not specific to the IV9-presenting HLA-A24 but widely reacted with any HLA-A24 molecules on cell surfaces (Supplemental Fig. 3). Taken together, these results demonstrate impaired ER processing for removing upstream prolines, which is consistent with ERAAP's enzymatic preference (26). In contrast to processing beginning in the cytosol, proline at any of up to six upstream positions inhibited HLA class I presentation of the following CTL epitope in the ER.

Discussion

A previous report using data sets from the SYFPEITHI database suggested that conservation of motifs represents preferences of proteasomal cleavage and TAP transportation in N-terminal–flanking sequences of HLA class I ligands (31). In the current study, we used peptide elution and large-scale liquid chromatography–MS/MS sequencing to generate a collection of naturally presented HLA class I ligands and reconstituted flanking sequences that had been removed by Ag processing. These data revealed a unique signature of pHLA I repertoire on cell surfaces showing that proline at positions upstream positions 1–3 was selectively depleted. Based on the experiments using live cells expressing a model cancer Ag, we demonstrated that 1) the presence of proline at upstream positions 1–6 equally inhibited Ag processing in the ER, but 2) only proline at upstream positions 1–3 predominantly inhibited the processing when the Ag constructs were expressed in the cytosol. These results suggest an inability of the ER to remove upstream prolines and that removal of proline occurs in the cytosol prior to transportation into the ER. Substrate residues preferred for tryptic digestion, such as lysine or arginine, were not enriched at upstream position 1, supporting that many of the 9–11-mer HLA class I ligands are not directly generated by proteasomal digestion (Fig. 3A, Supplemental Fig. 1A). Thus, we reached a conclusion that was different from that of a recent report, in which the authors proposed a direct influence of proteasomal digestion in making upstream profiles (25). Synergistic action of processing in the cytosol generating N-terminal–flanking precursors, and subsequent processing in the ER shapes final pHLA I repertoires displayed on cell surfaces, leaving the proline signature at upstream positions 1–3 in natural pHLA I repertoires.

It was of interest that natural pHLA I repertoires harbored such a signature of a single amino acid at a particular upstream position. Proline is a unique amino acid owing to its circular side chain that confers resistance to trimming by the ER-resident amino peptidase ERAAP (ERAP1) (26, 32). Because ERAAP, along with ERAP2, is the aminopeptidase responsible for peptide trimming in the ER, we consider the proline signature in pHLA I repertoires to mirror its enzymatic preference. Because ERAAP stops trimming at an amino acid preceding proline, leaving X-P-X_n peptides intact, the initial notion proposed that the inability to remove proline leads to an accumulation of X-P-X_n peptides displayed by particular MHC class I alleles (33). In this study, our findings further demonstrate that the enzymatic preference also leads to depletion of some proline-less HLA ligands on display, influencing natural pHLA I repertoires. As far as we have surveyed, this proline signature was observed in multiple HLA class I alleles, including the prevalent HLA-A*24:02 and -A*02:01 alleles, indicating that this at least was not an event specific to a particular allele. Meanwhile, enrichment at upstream position 1 was barely detectable in other residues, including leucine and methionine, both of which are indeed residues ERAAP prefers (34). Although this signature was modest in our data sets, it may imply the presence of positive substrate selection according to ERAAP preference.

The biological significance of the proline signature remains unclear. It has been reported that a mutation from alanine to proline at upstream position 1 of an HIV-Gag protein–derived HLA-B57 epitope impairs CTL recognition, and the mutant strain survives immune pressure (35). Although there is little evidence, it is also plausible that cancer cells make use of the proline signature for immune escape. The proline signature possibly helps avoid HLA class I presentation of endogenous self-proteins derived from proline-rich connective tissue proteins, such as collagens, under physiological conditions; however, pathogens or cancers may take advantage of such a motif unfavorable for HLA presentation for their immune escape. Therefore, this hypothesis should be further investigated.

Nevertheless, our findings in the current study could prove beneficial for cancer immunotherapy and neoantigen/cancer-specific Ag prediction. Although neoantigens that arise from gene mutations are promising CTL targets in tumor immunotherapy, responsible gene mutations are, more often than not, patient-specific; therefore, hundreds or thousands of somatic tumor mutations must be surveyed for every case (36, 37). Considering that only a very limited number of mutations are presented by HLA and elicit T cell responses (38), accurate algorithms accompanied by fewer false-positive hits are in great demand. Integration of the proline rule may decrease false-positive hits and improve current prediction tools.

Disclosures

The authors have no financial conflicts of interest.

References

1. Shastri, N., S. Schwab, and T. Serwold. 2002. Producing nature's gene-chips: the generation of peptides for display by MHC class I molecules. *Annu. Rev. Immunol.* 20: 463–493.
2. Neefjes, J., M. L. Jongsma, P. Paul, and O. Bakke. 2011. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* 11: 823–836.
3. Blum, J. S., P. A. Wearsch, and P. Cresswell. 2013. Pathways of antigen processing. *Annu. Rev. Immunol.* 31: 443–473.
4. Rock, K. L., E. Reits, and J. Neefjes. 2016. Present yourself! By MHC class I and MHC class II molecules. *Trends Immunol.* 37: 724–737.
5. Elliott, T., and A. Williams. 2005. The optimization of peptide cargo bound to MHC class I molecules by the peptide-loading complex. *Immunol. Rev.* 207: 89–99.
6. Wearsch, P. A., and P. Cresswell. 2008. The quality control of MHC class I peptide loading. *Curr. Opin. Cell Biol.* 20: 624–631.
7. Hammer, G. E., T. Kanaseki, and N. Shastri. 2007. The final touches make perfect the peptide-MHC class I repertoire. *Immunity* 26: 397–406.
8. Schumacher, T. N., and R. D. Schreiber. 2015. Neoantigens in cancer immunotherapy. *Science* 348: 69–74.
9. 2017. The problem with neoantigen prediction. *Nat. Biotechnol.* 35: 97.
10. Nielsen, M., C. Lundegaard, O. Lund, and C. Keşmir. 2005. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* 57: 33–41.
11. Larsen, M. V., C. Lundegaard, K. Lamberth, S. Buus, O. Lund, and M. Nielsen. 2007. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* 8: 424.
12. Lundegaard, C., K. Lamberth, M. Harndahl, S. Buus, O. Lund, and M. Nielsen. 2008. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Res.* 36 (Web Server issue): W509–W512.
13. Rasmussen, M., E. Fenoy, M. Harndahl, A. B. Kristensen, I. K. Nielsen, M. Nielsen, and S. Buus. 2016. *Pan*-specific prediction of peptide-MHC class I complex stability, a correlate of T cell immunogenicity. *J. Immunol.* 197: 1517–1524.
14. Jurtz, V., S. Paul, M. Andreatta, P. Marcatili, B. Peters, and M. Nielsen. 2017. NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.* 199: 3360–3368.
15. Gfeller, D., and M. Bassani-Sternberg. 2018. Predicting antigen presentation-what could we learn from a million peptides? *Front. Immunol.* 9: 1716.
16. Mommen, G. P., C. K. Frese, H. D. Meiring, J. van Gaans-van den Brink, A. P. de Jong, C. A. van Els, and A. J. Heck. 2014. Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proc. Natl. Acad. Sci. USA* 111: 4507–4512.
17. Walz, S., J. S. Stickel, D. J. Kowalewski, H. Schuster, K. Weisel, L. Backert, S. Kahn, A. Nelde, T. Stroh, M. Handel, et al. 2015. The antigenic landscape of multiple myeloma: mass spectrometry (re)defines targets for T-cell-based immunotherapy. *Blood* 126: 1203–1213.
18. Bassani-Sternberg, M., S. Pletscher-Frankild, L. J. Jensen, and M. Mann. 2015. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol. Cell. Proteomics* 14: 658–673.
19. Kochin, V., T. Kanaseki, S. Tokita, S. Miyamoto, Y. Shionoya, Y. Kikuchi, D. Morooka, Y. Hirohashi, T. Tsukahara, K. Watanabe, et al. 2017. HLA-A24 ligandome analysis of colon and lung cancer cells identifies a novel cancer-testis antigen and a neoantigen that elicits specific and strong CTL responses. *OncoImmunology* 6: e1293214.
20. Bassani-Sternberg, M., and D. Gfeller. 2016. Unsupervised HLA peptidome deconvolution improves ligand prediction accuracy and predicts cooperative effects in peptide-HLA interactions. *J. Immunol.* 197: 2492–2499.
21. Scholtalbers, J., S. Boegel, T. Bukur, M. Byl, S. Goerges, P. Sorn, M. Loewer, U. Sahin, and J. C. Castle. 2015. TCLP: an online cancer cell line catalogue integrating HLA type, predicted neo-epitopes, virus and gene expression. *Genome Med.* 7: 118.
22. Miyamoto, S., V. Kochin, T. Kanaseki, A. Hongo, S. Tokita, Y. Kikuchi, A. Takaya, Y. Hirohashi, T. Tsukahara, T. Terui, et al. 2018. The antigen ASB4 on cancer stem cells serves as a target for CTL immunotherapy of colorectal cancer. *Cancer Immunol. Res.* 6: 358–369.
23. Yamashita, Y., M. Anczurowski, M. Nakatsugawa, M. Tanaka, Y. Kagoya, A. Sinha, K. Chamoto, T. Ochi, T. Guo, K. Saso, et al. 2017. HLA-DP^{84Gly} constitutively presents endogenous peptides generated by the class I antigen processing pathway. *Nat. Commun.* 8: 15244.
24. Yewdell, J. W., E. Reits, and J. Neefjes. 2003. Making sense of mass destruction: quantitating MHC class I antigen presentation. *Nat. Rev. Immunol.* 3: 952–961.
25. Abelin, J. G., D. B. Keskin, S. Sarkizova, C. R. Hartigan, W. Zhang, J. Sidney, J. Stevens, W. Lane, G. L. Zhang, T. M. Eisenhaure, et al. 2017. Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity* 46: 315–326.
26. Serwold, T., F. Gonzalez, J. Kim, R. Jacob, and N. Shastri. 2002. ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature* 419: 480–483.
27. Saric, T., S. C. Chang, A. Hattori, I. A. York, S. Markant, K. L. Rock, M. Tsujimoto, and A. L. Goldberg. 2002. An IFN-gamma-induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I-presented peptides. *Nat. Immunol.* 3: 1169–1176.
28. Saveanu, L., O. Carroll, V. Lindo, M. Del Val, D. Lopez, Y. Lepelletier, F. Greer, L. Schomburg, D. Fruci, G. Niedermann, and P. M. van Endert. 2005. Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat. Immunol.* 6: 689–697.
29. Kanaseki, T., N. Blanchard, G. E. Hammer, F. Gonzalez, and N. Shastri. 2006. ERAAP synergizes with MHC class I molecules to make the final cut in the antigenic peptide precursors in the endoplasmic reticulum. *Immunity* 25: 795–806.
30. Shastri, N., N. Nagarajan, K. C. Lind, and T. Kanaseki. 2014. Monitoring peptide processing for MHC class I molecules in the endoplasmic reticulum. *Curr. Opin. Immunol.* 26: 123–127.
31. Schatz, M. M., B. Peters, N. Akkad, N. Ullrich, A. N. Martinez, O. Carroll, S. Bulik, H. G. Rammensee, P. van Endert, H. G. Holzhütter, et al. 2008. Characterizing the N-terminal processing motif of MHC class I ligands. *J. Immunol.* 180: 3210–3217.
32. Nguyen, T. T., S. C. Chang, I. Evnouchidou, I. A. York, C. Zikos, K. L. Rock, A. L. Goldberg, E. Stratikos, and L. J. Stern. 2011. Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAP1. *Nat. Struct. Mol. Biol.* 18: 604–613
33. Serwold, T., S. Gaw, and N. Shastri. 2001. ER aminopeptidases generate a unique pool of peptides for MHC class I molecules. *Nat. Immunol.* 2: 644–651.
34. Hearn, A., I. A. York, and K. L. Rock. 2009. The specificity of trimming of MHC class I-presented peptides in the endoplasmic reticulum. *J. Immunol.* 183: 5526–5536.
35. Draenert, R., S. Le Gall, K. J. Pfafferott, A. J. Leslie, P. Chetty, C. Brander, E. C. Holmes, S. C. Chang, M. E. Feeney, M. M. Addo, et al. 2004. Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. *J. Exp. Med.* 199: 905–915.
36. Liu, X. S., and E. R. Mardis. 2017. Applications of immunogenomics to cancer. *Cell* 168: 600–612.
37. Sahin, U., and Ö. Türeci. 2018. Personalized vaccines for cancer immunotherapy. *Science* 359: 1355–1360.
38. Tran, E., M. Ahmadzadeh, Y. C. Lu, A. Gros, S. Turcotte, P. F. Robbins, J. J. Gartner, Z. Zheng, Y. F. Li, S. Ray, et al. 2015. Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science* 350: 1387–1390.