



# An *in silico*–*in vitro* Pipeline Identifying an HLA-A\*02:01<sup>+</sup> KRAS G12V<sup>+</sup> Spliced Epitope Candidate for a Broad Tumor-Immune Response in Cancer Patients

Michele Mishto<sup>1,2\*</sup>, Artem Mansurkhodzhaev<sup>3</sup>, Ge Ying<sup>4</sup>, Aruna Bitra<sup>4</sup>, Robert A. Cordfunke<sup>5</sup>, Sarah Henze<sup>3</sup>, Debdas Paul<sup>3</sup>, John Sidney<sup>6</sup>, Henning Urlaub<sup>7,8</sup>, Jacques Neefjes<sup>9</sup>, Alessandro Sette<sup>6,10</sup>, Dirk M. Zajonc<sup>11,12</sup> and Juliane Liepe<sup>3,13\*</sup>

<sup>1</sup> Centre for Inflammation Biology and Cancer Immunology (CIBCI) & Peter Gorer Department of Immunobiology, King's College London, London, United Kingdom, <sup>2</sup> Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin Institute of Health, Institut für Biochemie, Berlin, Germany, <sup>3</sup> Quantitative and Systems Biology, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany, <sup>4</sup> Division of Immune Regulation, La Jolla Institute for Allergy and Immunology, La Jolla, CA, United States, <sup>5</sup> Department of Immunohematology and Bloodbank, Leiden University Medical Center LUMC, Leiden, Netherlands, <sup>6</sup> Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, La Jolla, CA, United States, <sup>7</sup> Bioanalytical Mass Spectrometry Group, Max-Planck-Institute for Biophysical Chemistry, Goettingen, Germany, <sup>8</sup> Institut für Klinische Chemie, University Medical Center Goettingen Bioanalytics, Goettingen, Germany, <sup>9</sup> Department of Cell and Chemical Biology, Oncode Institute, Leiden University Medical Center LUMC, Leiden, Netherlands, <sup>10</sup> Department of Medicine, University of California, San Diego, San Diego, CA, United States, <sup>11</sup> Division of Immune Regulation, La Jolla Institute for Immunology, La Jolla, CA, United States, <sup>12</sup> Department of Internal Medicine, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium, <sup>13</sup> Department of Life Sciences, Centre for Integrative Systems Biology and Bioinformatics, Imperial College London, London, United Kingdom

## OPEN ACCESS

### Edited by:

John M. Maris,  
University of Pennsylvania,  
United States

### Reviewed by:

Graham Robert Leggatt,  
University of Queensland, Australia  
Jan Joseph Melenhorst,  
University of Pennsylvania,  
United States

### \*Correspondence:

Michele Mishto  
michele.mishto@kcl.ac.uk  
Juliane Liepe  
jliepe@mpibpc.mpg.de

### Specialty section:

This article was submitted to  
Cancer Immunity and Immunotherapy,  
a section of the journal  
Frontiers in Immunology

**Received:** 07 August 2019

**Accepted:** 16 October 2019

**Published:** 15 November 2019

### Citation:

Mishto M, Mansurkhodzhaev A, Ying G, Bitra A, Cordfunke RA, Henze S, Paul D, Sidney J, Urlaub H, Neefjes J, Sette A, Zajonc DM and Liepe J (2019) An *in silico*–*in vitro* Pipeline Identifying an HLA-A\*02:01<sup>+</sup> KRAS G12V<sup>+</sup> Spliced Epitope Candidate for a Broad Tumor-Immune Response in Cancer Patients. *Front. Immunol.* 10:2572. doi: 10.3389/fimmu.2019.02572

Targeting CD8<sup>+</sup> T cells to recurrent tumor-specific mutations can profoundly contribute to cancer treatment. Some of these mutations are potential tumor antigens although they can be displayed by non-spliced epitopes only in a few patients, because of the low affinity of the mutated non-spliced peptides for the predominant HLA class I alleles. Here, we describe a pipeline that uses the large sequence variety of proteasome-generated spliced peptides and identifies spliced epitope candidates, which carry the mutations and bind the predominant HLA-I alleles with high affinity. They could be used in adoptive T cell therapy and other anti-cancer immunotherapies for large cohorts of cancer patients. As a proof of principle, the application of this pipeline led to the identification of a KRAS G12V mutation-carrying spliced epitope candidate, which is produced by proteasomes, transported by TAPs and efficiently presented by the most prevalent HLA class I molecules, HLA-A\*02:01 complexes.

**Keywords:** proteasome, peptide splicing, adoptive T cell therapy targets, antigen presentation, cancer epitopes, KRAS, tumor immunology

## INTRODUCTION

Activating CD8<sup>+</sup> T cells to recurrent tumor-specific mutations is one of a number of cutting-edge strategies to treat cancer. It can be achieved by immunotherapy approaches such as adoptive T cell therapy (ATT), peptide vaccination and dendritic cell (DC) vaccination. Neoepitopes that carry cancer recurrent mutations and efficiently bind common Human Leukocyte Antigen class I (HLA-I) variants are ideal targets for tumor immunology vaccination of large cohorts of patients.

Peptide epitopes are generally produced by proteasomes, which are the final effectors of the ubiquitin-proteasome system (1). Epitope production is the first step of the antigen processing and presentation (APP) pathway, which accounts for the epitope translocating into the endoplasmic reticulum (ER) lumen through mediation by transporters associated with antigen processing (TAPs), binding to the peptide loading complex, trimming by exopeptidases, binding to the HLA-I complex, and transport to the cell surface for recognition by cytotoxic T lymphocytes (CTLs) (2).

There are several proteasome isoforms that can be involved in APP. The most active proteasome isoform is a large (26S) protease consisting of a 20S proteasome core coupled to one or two 19S regulatory complexes. The 19S subunit contains the ubiquitin recognition and removal system as well as an unfolding activity, the 20S form is the actual protease. The 20S proteasome is constituted of four rings; two  $\alpha$  rings at the apexes; and two  $\beta$  rings forming the central chamber. Each ring has seven distinct subunits. Each  $\beta$  ring carries three catalytic (i.e.,  $\beta 1$ ,  $\beta 2$ , and  $\beta 5$ ) subunits, which have distinct preferences for peptide sequence motifs (1). Human cells can express different isoforms of catalytic subunits, which are incorporated in distinct proteasome isoforms. Standard proteasomes (s-proteasomes) contain  $\beta 1$ ,  $\beta 2$ , and  $\beta 5$  subunits. Immunoproteasomes (i-proteasomes) contain  $\beta 1i$ ,  $\beta 2i$ , and  $\beta 5i$  subunits and are constitutively present in immune cells, such as mature DCs, as well as in cells exposed to an inflammatory milieu (3). Tumors express various intermediate-type proteasome isoforms, in which standard- and immuno-subunits are assembled in one proteasome complex (4, 5). 20S proteasome is also functional alone in cells or coupled to other regulatory subunits such as PA28  $\alpha\beta$  (3, 6, 7).

Proteasomes can break proteins and release the peptide fragments or re-ligate them, thereby forming “spliced peptides,” which have sequences that do not recapitulate the linear sequence of the parental protein (3, 8, 9). Spliced peptides may represent a sizeable portion of the peptide pool bound to HLA-I molecules—i.e., the HLA-I immunopeptidome—of tumor and non-tumor human cell lines (10–12). This hypothesis is currently a matter of debate since different analytical approaches obtained discordant results (10–16). According to our previous analysis, however, several antigens displayed by HLA-I immunopeptidomes are represented only by spliced peptides. The antigens represented by spliced peptides seem to be longer, more hydrophobic and more basic than those represented by non-spliced peptides. Antigens that are represented by both spliced and non-spliced peptides show antigenic hot spots, i.e., antigenic areas from which both spliced and non-spliced peptides derive (12). HLA-I-bound spliced peptides are generally less abundant than non-spliced peptides (10–12, 17). Proteasome-generated spliced epitopes can trigger specific CTL responses *ex vivo/in vivo* against tumor- and type 1 diabetes-associated antigens (17, 18) as well as pathogens (19). ATT targeting a spliced epitope successfully treated a melanoma patient (20, 21).

Proteasome-catalyzed peptide splicing (PCPS; see **Figure 1A**) can occur by combining two non-contiguous sequences of the same molecule (*cis* PCPS) or of two distinct molecules (*trans* PCPS). The latter seems to be efficiently catalyzed *in vitro* by

purified proteasomes (22–24) and may constitute a large portion of the HLA-I associated spliced immunopeptidomes (11).

Although the role of spliced peptides in central tolerance still has to be investigated, the theoretically large sequence variability of spliced peptides makes them attractive for anti-cancer immunotherapy (25). Indeed, some of the most recurrent driver mutations in tumors often cannot be efficiently represented by canonical non-spliced peptides bound to the predominant HLA-I variants because of antigen sequence restrictions. On the contrary, they might be represented on the cell surface by tumor-specific spliced epitopes. Therefore, the identification of tumor antigen-specific spliced epitopes might represent a unique opportunity to treat a large cohort of patients.

We here present a pipeline combining *in silico* and *in vitro* approaches. It successfully identifies tumor-specific spliced and non-spliced epitope candidates, which can be further validated as targets for anti-cancer immunotherapies, as illustrated by the HLA-A\*02:01<sup>+</sup> KRAS G12V<sup>+</sup> spliced epitope candidate here described (**Figure 1B**).

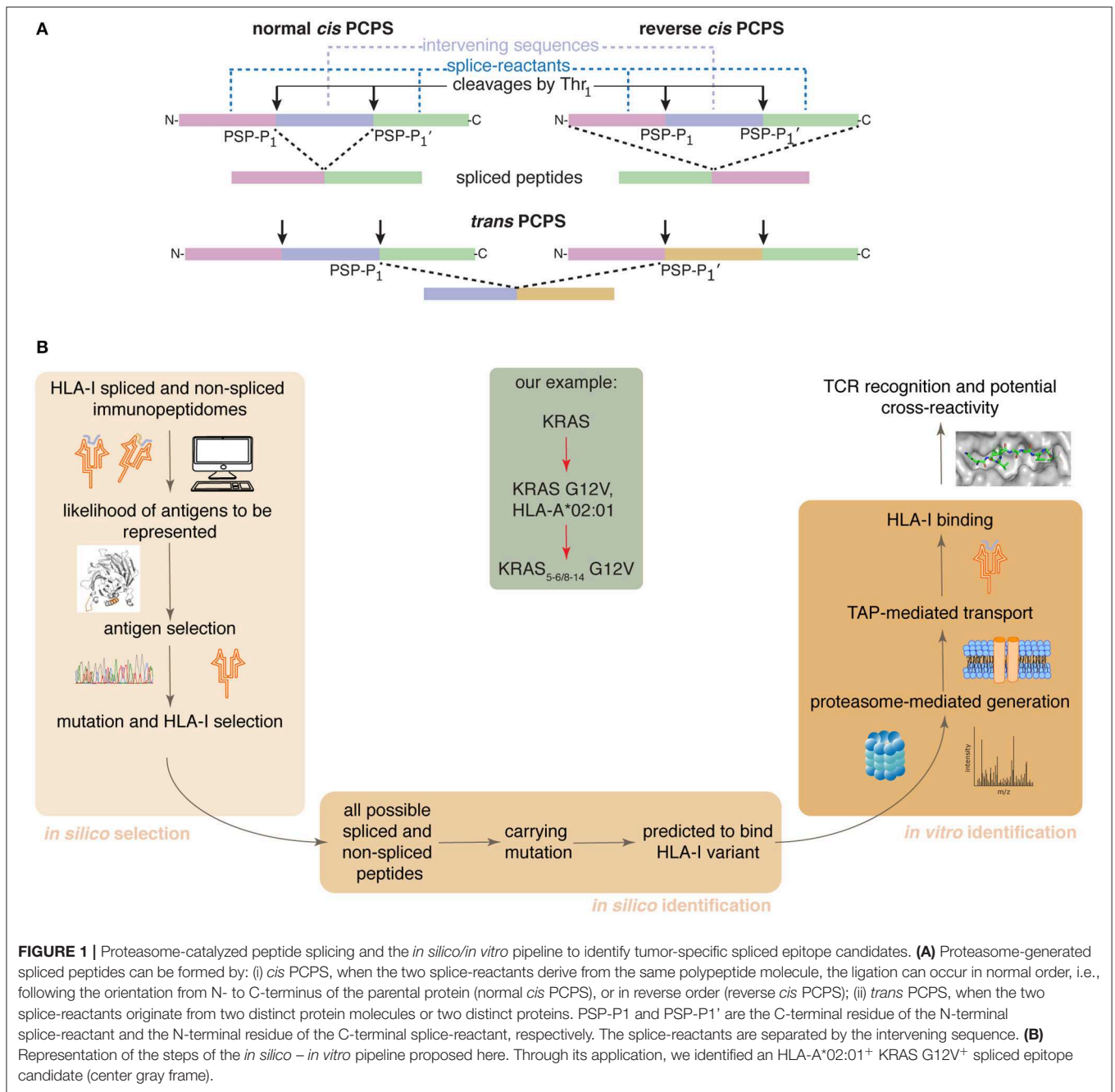
## MATERIALS AND METHODS

### Antigen Selection and Spliced Epitope Candidate Identification

To rank antigens that are over-represented in HLA-I immunopeptidomes by spliced and non-spliced peptides according to their protein characteristics, we generated a simple model based on the following characteristics: length, hydrophobicity index, isoelectric point, and instability index.

We first calculated these characteristics for all proteins in the Uniprot Reference human proteome database. Next, we analyzed the previously published HLA-I immunopeptidomes of GR-LCL, HCC1143 and HCT116 cell lines (12). All identified spliced and non-spliced peptides were mapped to their antigen(s) of origin, thereby determining a set of represented antigens. Among those antigens not represented in the MS-detected HLA-I immunopeptidomes, there are likely many antigens that would be represented if we considered larger HLA-I immunopeptidome datasets. For this reason, we compared the characteristics of the represented antigens to the characteristics of all proteins (control set). The aim was to determine a combination of the four selected protein characteristics that has the largest difference between the represented antigen set and the control set.

We define the feature sum ( $f$ ) as:  $f = \sum p_i c_i$ , where  $c$  is the vector of the four selected features (length, hydrophobicity index, isoelectric point and instability index) and  $p$  is the vector of factors ranging from  $-1$  to  $1$ . A factor of  $-1$  would favor presentation, while a factor of  $1$  would disfavor presentation. We use Bayesian inference in a Markov Chain Monte Carlo scheme to determine the factors that result in a distribution of  $f$  for represented antigens ( $F_1$ ), which is most different to the distribution of  $f$  for control proteins ( $F_0$ ). The difference between the densities  $F_{diff} = F_1 - F_0$  indicates which  $f$  is favoring (positive values) or disfavoring (negative values) protein representation. Next, we calculated  $F_{diff}(f)$  for a set of candidate antigens (BRAF, KRAS, HRAS, NRAS, TP53, CDK4, IDH1, TYR) using



the Kolmogorov-Smirnov distance. To do so, we sampled ( $N = 1,000$ ) from the posterior distribution of factors resulting in distributions of  $F_{diff}$  for each candidate antigen. Antigens with the highest  $F_{diff}$  have characteristics that lead to more likely representation of those antigens in HLA-I immunopeptidomes.

### Peptide Synthesis and Proteasome Purification

All peptides were synthesized using Fmoc solid phase chemistry (**Supplementary Table 2**). 20S proteasome was purified from peripheral blood as follows: (i) 10 ml peripheral blood was

homogenized, lysed and centrifuged; (ii) the supernatant was fractionated by ammonium sulfate precipitation (35% and then 75%); (iii) the latter pellet was fractionated by chromatography on DEAE-Sephacel; (iv) the selected fractions were separated by 10–40% sucrose gradient and followed by (v) anion exchange chromatography on Mono Q in an Akta-FPLC system; (vi) the selected fractions (2–4 mL) were further purified by DEAE-Affi-gel-blue chromatography. In each of the (ii–vi) steps, the fractions were monitored by degradation assays of standard short fluorogenic substrate Suc-LLVY-AMC. Proteasome concentration was measured by Bradford staining

and verified by Coomassie staining of an SDS-Page gel, as shown elsewhere (26). The purity of the preparation using this protocol has been previously shown (27).

### **In vitro Digestions and MS Measurements**

Synthetic polypeptides (40  $\mu$ M) were digested by 3  $\mu$ g 20S proteasomes in 100  $\mu$ l TEAD buffer for different time points (0–20 h) at 37°C, as previously described (27). We performed three independent experiments, each of them measured either 3 times (for the 0–4 h kinetics) or 2 times (for 20 h digestions) by mass spectrometry (MS).

The identification of target peptide products was carried out by targeted MS using a mass to charge ratio (m/z) inclusion list. The inclusion list was comprised of all theoretically possible 8–12 mer spliced and non-spliced peptide products derived from KRAS<sub>2–35</sub> G12V synthetic polypeptide substrate, which carried the G12V mutation and were predicted to bind HLA-A\*02:01 complex with IC<sub>50</sub>  $\leq$  100 nM (see below). The same principle was applied for the peptide products derived from the wild type KRAS<sub>2–35</sub> G12 synthetic polypeptide substrate (**Supplementary Table 1**). To this end, 20 h *in vitro* digestions with 20S proteasomes were measured by Fusion Lumos Mass Spectrometer (Thermo Fisher Scientific). Prior to measurement, the samples were diluted with the loading buffer (2% acetonitrile, 0.05% Trifluoroacetic acid) containing human insulin (Sigma-Aldrich) to a final substrate concentration of 25  $\mu$ M and insulin concentration of 2  $\mu$ M. Insulin was used as a coating polymer to prevent binding of peptides to the glass vials used for measurements and to improve reproducibility between technical replicates. Eight  $\mu$ l of those dilutions (corresponding to 200 pmol of substrate initially present in the sample) were injected. Samples were loaded and separated by a nanoflow HPLC (RSLC Ultimate 3000) on an Easy-spray C18 nano column (30 cm length, 75  $\mu$ m internal diameter) coupled on-line to a nano-electrospray ionization Fusion Lumos mass spectrometer (Thermo Fisher Scientific). Peptides were eluted with a linear gradient of 5–55% buffer B (80% ACN, 0.1% formic acid) over 88 min at 50°C at a flow rate of 300 nl/min. The instrument was programmed within Xcalibur 3.1.66.10 to acquire MS data in a Data Dependent Acquisition mode using Top 20 precursor ions. We acquired one full-scan MS spectrum at a resolution of 120,000 with an automatic gain control (AGC) target value of 1,000,000 ions and a scan range of 300–1,600 m/z with maximum injection time set to 50 ms and intensity threshold set to 50,000. The MS/MS fragmentation was conducted using HCD collision energy (35%) with an orbitrap resolution of 30,000 at 1.4 m/z isolation window with Fixed First Mass set to 105 m/z. The AGC target value was set up at 100,000 with a maximum injection time of 128 ms. A dynamic exclusion of 30 s and 1–7 included charged states were defined within this method.

*In vitro* proteasome-mediated digestion kinetics (0–4 h) and the 20 h digestions were measured by LC-MS/MS as follows: Prior to measurement, samples were diluted with the loading buffer and insulin as described above. Eight  $\mu$ l (i.e., 200 pmol substrate) of those dilutions were loaded. Samples were loaded and separated by a nanoflow HPLC (RSLC Ultimate 3000) on an Easy-spray C18 nano column (30 cm length, 75  $\mu$ m

internal diameter; Dr. Maisch) coupled on-line to a nano-electrospray ionization Q Exactive Hybrid-Quadrupol-Orbitrap mass spectrometer (Thermo Fisher Scientific). Peptides were eluted with a linear gradient of 5–55% buffer B (80% ACN, 0.1% formic acid) over 88 min at 50°C at a flow rate of 300 nl/min. The instrument was programmed within Xcalibur 3.1.66.10 to acquire MS data in a Data Dependent Acquisition mode using Top 20 precursor ions. We acquired one full-scan MS spectrum at a resolution of 70,000 with an automatic gain control (AGC) target value of 1,000,000 ions and a scan range of 350–1,600 m/z. The MS/MS fragmentation was conducted using HCD collision energy (30%) with an Orbitrap resolution of 35,000 at 2 m/z isolation window with Fixed First Mass set to 110 m/z. The AGC target value was set up at 100,000 with a maximum injection time of 128 ms. For Data Dependent Scans the minimum AGC target value and the Intensity threshold were set to 2,600–20,000 accordingly. A dynamic exclusion of 25 s and 1–6 included charged states were defined within this method.

### **Spliced and Non-spliced Peptide Identification and Quantification As Well as Computation of SCS-P1 and PSP-P1**

Peptides were identified using the Mascot version 2.6.1 (Matrix Science) search engine. Mass spectra were searched against a customized database that includes all theoretically possible spliced and non-spliced peptides (28). M oxidation, N-terminal acetylation and NQ deamidation were set as variable modification. For the peptide identification in the Orbitrap Q Exactive measurements, we set as mass tolerances for MS and MS/MS 6 ppm and 20 ppm, respectively. For the peptide identification in the Fusion Lumos measurements, we set as mass tolerances for MS and MS/MS 5 ppm and 0.03 Da, respectively.

Peptide hits were filtered using an ion score cut-off of 20, a *q*-value cut off of 0.05 and a delta score between two spliced peptide hits or between a top scoring spliced peptide and a lower scoring non-spliced peptide of 30% (12). Mascot Distiller's label-free quantification toolbox was used to automatically extract MS ion peak areas of all identified peptides for all five time points (0–4 h) and all three technical replicates simultaneously. Biological replicates were processed separately. The resulting peptide kinetics were filtered for peptide synthesis artifacts and non-reproducible peptide kinetics between technical replicates. Furthermore, peptides that showed unrealistic generation kinetic behavior (such as alternating MS ion peak areas between consecutive time points) were removed. In the final analysis, only peptides that were detected and quantified in two biological replicates were considered.

KRAS<sub>5–6/8–14</sub> G12V and KRAS<sub>5–14</sub> G12V generation kinetics were manually quantified by extraction of an ion chromatogram (XIC) corresponding to the peptides monoisotopic peaks, using instrument precursor tolerance and retention time information (from the identified peptides in the 20 h digestions) via Mascot Distiller, followed by determination of the area under the peak at each time point in the kinetics series.



Absolute peptide quantification was carried out through the application of the method QPuB on detected MS ion peak areas for each peptide product, as described elsewhere (see Data availability section). In the specific case of the two epitope candidates KRAS<sub>5–6/8–14</sub> G12V and KRAS<sub>5–14</sub> G12V, we computed their amount using a titration curve of the cognate synthetic peptides since their amount was too low to be estimated with high confidence using QPuB. Synthetic peptide concentration for titration ranged from 0 to 10 pmol injected. Each titration sample was measured twice and right after measuring *in vitro* digestion samples.

SCS-P1 (site specific cleavage strength after amino acid residue P1) and PSP-P1 (frequency of peptide splicing catalyzed using the C-terminus of the N-terminal splice-reactant as splicing site) were calculated based on the absolute amount of each product (resulting from QPuB) identified in the proteasome-catalyzed digestions (23). Briefly, for each time point and each amino acid in the substrate, the sum over all product (non-spliced and spliced peptides) amount that have the corresponding substrate amino acid at their C-terminus has been computed and normalized, so that they add up to 100%, resulting in SCS-P1. For each time point and each amino acid in the substrate, the sum over all spliced peptide amount that have the corresponding substrate amino acid at their C-terminus of the N-terminal splice-reactant was computed and normalized, so that they add up to 100%, resulting in PSP-P1.

## TAP Assay

The transport efficiency of target peptides (**Supplementary Table 2**) into the ER lumen mediated by TAPs was carried out as previously described (29) although some modifications were introduced. These include the use of a fluorescent peptide tracer and the use of microsomes rather than Streptolysin O permeabilized cells. Briefly, peptides were dissolved in DMSO and different concentrations were distributed in a final volume of 10  $\mu$ l DMSO. At the same time, a mixture of 10 mM ATP, 100 mM Tris-HCl pH7.5 and 5 mM MgCl<sub>2</sub> and fluorescent tracer peptide was prepared. 60  $\mu$ l of this mixture was added to the 10  $\mu$ l competing peptide mixture to a final volume of 70  $\mu$ l. This was prewarmed to 37°C and 30  $\mu$ l of pre-warmed microsomes were added. Microsomes were derived from LCL721 cells, as previously described (29).

The mixture was incubated for 20 min at 37°C followed by cell lysis with lysis mixture (0.5% TX100, 5 mM MgCl<sub>2</sub> in 100 mM Tris-HCl pH7.5) at 4°C. After at least 30 min incubation at 4°C, DNA was pelleted and the supernatant transferred to a new vial including ConA-beads. After at least 30 min incubation on ice, cells were washed four times with lysis mixture and the last time with 100 mM Tris-HCl pH7.5 before transfer to 96 wells plates (Corning) followed by fluorescence measurements in a plate reader. To detect background signals, a sample without competing peptide and ATP was included and fully performed as described above. This signal was subtracted from the detected signal. The curves were normalized to the highest value set at 100% and EC<sub>50</sub> values were calculated.

## HLA-I–Peptide Binding Affinity Prediction and Measurement

The binding affinity between theoretical spliced and non-spliced peptides and HLA-A\*02:01 was predicted using the NetMHCpan 3.0 algorithm (30). We restricted the prediction to 8–12 mer peptides and imposed an IC<sub>50</sub> cut-off of 100 nM. The binding affinity between the synthetic peptides and HLA-A\*02:01 complexes was measured using purified HLA-I molecules, as described elsewhere (10).

## HLA-I–Peptide Crystal Structure and Analysis

The ectodomains of HLA-A\*02:01 (residues 21–274) and human  $\beta_2$ -microglobulin (h $\beta_2$ m) (residues 1–99) were expressed in *Escherichia coli* BL21 DE3 cells as inclusion bodies after 4 h induction with 1 mM isopropyl 1-thio-d-galactopyranoside (IPTG) at A<sub>600</sub> of 0.6. Cells were harvested by centrifugation (5,000  $\times$  g for 20 min), resuspended in lysis buffer (100 mM Tris-HCl, pH7.0, 5 mM EDTA, 5 mM DTT, 0.5 mM PMSF), and broken through a microfluidizer (Microfluidics). Inclusion bodies were collected from cell lysate (50,000  $\times$  g for 30 min at 4°C), washed 3 times in 100 mM Tris-HCl, pH7.0, 5 mM EDTA, 5 mM DTT, 2 M urea, 2% (w/v) Triton X-100 plus 1 time in 100 mM Tris-HCl, pH7.0, 5 mM EDTA, 2 mM DTT, and finally dissolved in 50 mM Tris-HCl, pH7.0, 5 mM EDTA, 2 mM DTT, 6 M guanidine HCl) for the following refolding.

Refolding was performed in a 100 ml system. Briefly, 1.2 mg of h $\beta_2$ m was loaded dropwise into refolding buffer (0.1 M Tris-HCl, pH8.0, 2 mM EDTA, 400 mM l-arginine, 5 mM oxidized glutathione, 5 mM reduced glutathione) and stirred for 1 h at 4°C. Then, 6 mg of HLA-A\*02:01 mixed with 1.2 mg of individual peptide (**Supplementary Table 2**) was added dropwise into the refolding system and stirred at 4°C for up to 72 h. The refolding system was concentrated to 0.5 mL for size exclusion chromatography using a Superdex S200 Increase 10/300 GL column in 20 mM Tris-HCl pH7.5, 150 mM NaCl. Fractions containing refolded HLA-A\*02:01- $\beta_2$ m-peptide complexes were pooled and concentrated to 5–10 mg/ml for subsequent crystallization.

Thick plate-like or 3-dimensional crystals of HLA-A\*02:01- $\beta_2$ m-peptide complexes were obtained by setting drop vapor diffusion at 1:1–1.5 ratio with 30% PEG 4000, 0.1 M Tris-HCl, pH8.5, 0.2 M lithium sulfate at room temperature after 3 days. The crystals were flash frozen in crystallization solution plus glycerol (25% v/v) using liquid nitrogen.

Diffraction data for HLA-A\*02:01- $\beta_2$ m-peptides were collected remotely at beam line 9.2 at the Stanford Synchrotron Radiation Light source and processed to 1.40–1.55 Å resolution, using HKL2000. Phases were obtained by molecular replacement with Phaser MR in ccp4 using the protein coordinates from a former HLA-A\*02:01- $\beta_2$ m-peptide structure (Protein Data Bank code 5ENW) (31). The model was built with COOT (32) and refined with REFMAC5 (33). Data collection and refinement statistics are shown in **Supplementary Table 3**.

## Statistical Analysis

If not stated otherwise, all statistical tests have been done in R and differences in distributions have been tested using the Kolmogorov-Smirnov test.

## Dataset and Software Availability

A summary of the RAW files of the LC-MS/MS measurements of the *in vitro* digestions accessible via repository is reported in the following Mendeley dataset: <http://doi.org/10.17632/63rj3xczmb.1>.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (34) partner repository with the dataset identifier PXD015580.

The HLA-I immunopeptidome elution MS files used in the first step of the pipeline are available at the PRIDE repository with the dataset identifier PXD000394 (files: 20120321\_EXQ1\_MiBa\_SA\_HCC1143\_1.raw, 20120321\_EXQ1\_MiBa\_SA\_HCC1143\_2.raw, 20120322\_EXQ1\_MiBa\_SA\_HCC1143\_1\_A.raw, 20120515\_EXQ3\_MiBa\_SA\_HCT116\_mHLA-1.raw, 20120515\_EXQ3\_MiBa\_SA\_HCT116\_mHLA-2.raw, 20120617\_EXQ0\_MiBa\_SA\_HCT116\_1\_mHLA\_2hr.raw, 20120617\_EXQ0\_MiBa\_SA\_HCT116\_2\_mHLA\_2hr.raw) and at the Datadryad.org archive (doi: 10.5061/dryad.r984n) and were generated by Bassani et al. (35) and Mommen et al. (36).

QPuB software is available at GitHub (<https://github.com/QuantSysBio/QPuB>).

## RESULTS

### Prioritization of KRAS as Antigen Over-represented on Cell Surface

Antigens represented by spliced peptides in HLA-I immunopeptidomes tend to be preferentially long, hydrophobic and basic, thereby suggesting that the chemical and physical characteristics of antigens can impinge upon spliced peptide generation and presentation (12, 24). To select suitable antigens from which a spliced epitope candidate might be derived, we first investigated which combination of protein features may result in more likely potential over-representation in HLA-I immunopeptidomes by spliced and non-spliced peptides.

Accordingly, we used a previously published HLA-I spliced and non-spliced immunopeptidome database (12), which includes 13,666 unique non-spliced and 1,318 unique spliced peptides, as well as 7,328 represented antigens. With this dataset, we developed a simple model based on protein length, hydrophobicity, isoelectric point and instability index to determine the possible over-representation of a given antigen by spliced and non-spliced peptides in HLA-I immunopeptidomes. These characteristics were previously described to influence the probability of observing peptides of a protein being presented in HLA-I immunopeptidomes determined by antigen gene expression level and antigen abundance as key determinants for efficient presentation (12, 37). However, we here opted to focus on protein intrinsic characteristics that are conserved independently of cell types and cell status, to obtain a model for antigen selection that can be generalized.

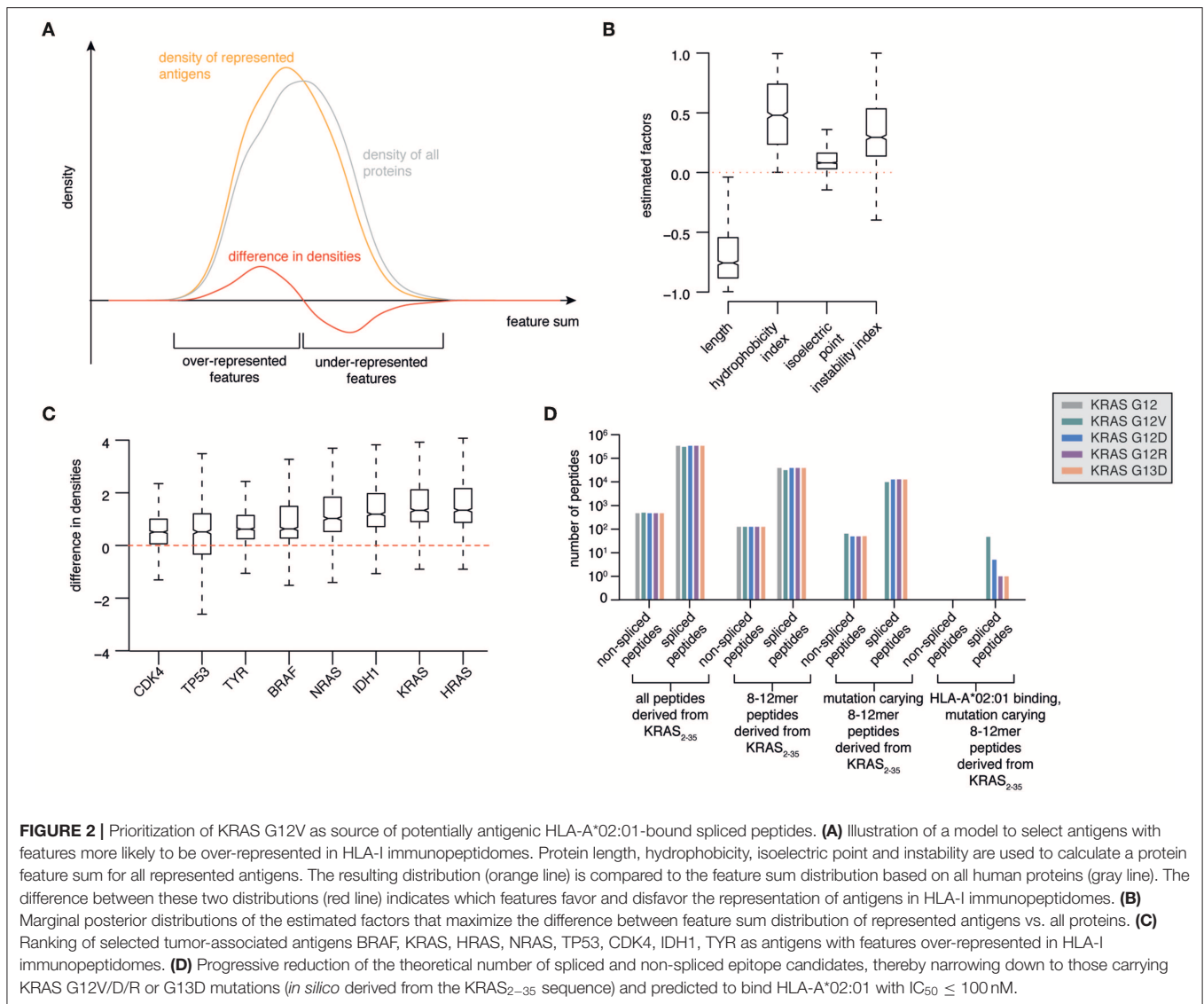
Combining these four selected protein characteristics yielded a distribution of known represented antigens, which can be compared to the feature distribution of all proteins (Figure 2A). Proteins with feature values that show a higher density for represented antigens compared to all proteins are more likely to be favored for antigen presentation than those proteins with feature values that show a lower density for represented antigens compared to all proteins (Figure 2A). We therefore aimed to find a combination of features that maximizes the difference between the two distributions. We defined a model calculating a feature sum  $\sum p_i c_i$ , where  $p_i$  are factors ranging from  $-1$  (favoring representation) to  $1$  (disfavoring representation) and  $c_i$  are the protein characteristic values.

Using Bayesian inference, we estimated the factors that provide the largest distance between the resulting feature sum distributions for represented antigens compared to all proteins. We found that protein length favors representation. On the contrary, very hydrophobic or instable proteins are disfavored during representation. The isoelectric point appeared to have minor influence (Figure 2B).

As proof-of-principle, we focused our analysis on a series of major tumor antigens - BRAF, KRAS, HRAS, NRAS, TP53, CDK4, IDH1, TYR - which all carry recurrent oncogenic mutations. We calculated the feature sums for those eight antigens and determined corresponding density differences for each of those feature sums, which allowed us to rank the candidate antigens (Figure 2C). Among them, HRAS and KRAS are the two antigens that are most likely over-represented in HLA-I immunopeptidomes as compared to the whole proteome (Figure 2C).

### Prioritization of KRAS G12V neoantigen as Source of Potentially Antigenic HLA-A\*02:01-Bound Tumor-specific Spliced Peptides

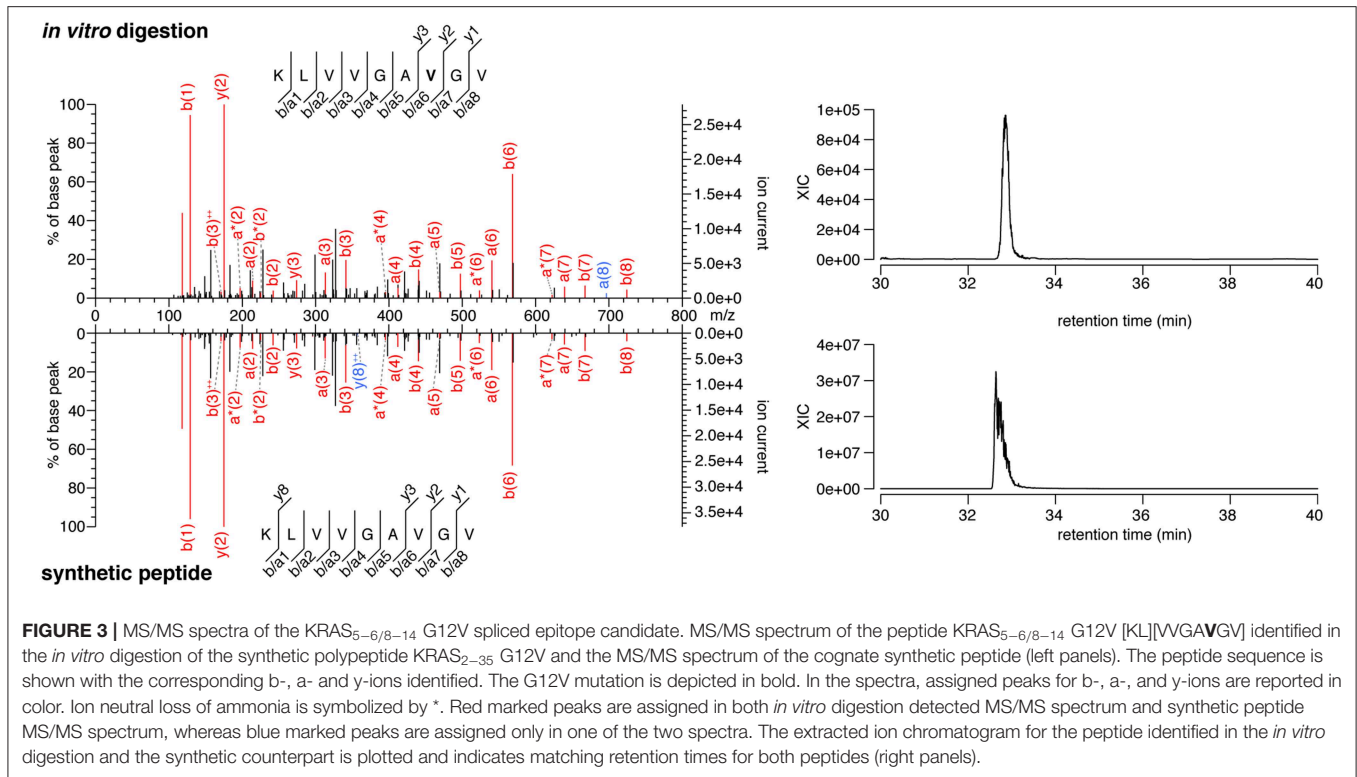
HRAS and KRAS are two GTPases that function as molecular switches in regulatory pathways responsible for proliferation and survival. In particular, KRAS is frequently mutated in cancers with an average of 22% cancers carrying a KRAS mutation, a frequency that rises to 33–61% in colorectal cancer and pancreatic adenocarcinoma (38). The mutations often occur in the KRAS G12 and G13 residues, which impairs the KRAS GTPase activity and renders the mutants persistently in the GTP-bound active form, thereby promoting tumorigenesis and tumor malignancy (39). KRAS G12/13 is a driver in tumors and in combination with its high frequency in cancers makes KRAS an ideal target for immunotherapies. For example, ATT using multiple T cell Receptors (TCRs), which recognize HLA-A\*02:01<sup>+</sup> spliced epitopes carrying KRAS mutations, could treat around 30% of pancreatic adenocarcinoma patients and a large portion of colorectal carcinoma patients. The efficacy of ATT has been demonstrated in a xenograft mouse model (40) and a metastatic colorectal cancer patient (41) by targeting non-spliced epitopes carrying KRAS G12D mutations and presented by HLA-A\*11:01 or -C\*08:02 molecules, respectively.



All these features define KRAS as an attractive tumor antigen to be further investigated using our pipeline. We investigated *in silico* the sequence surrounding residues 12 and 13 of KRAS wild type, G12V/D/R and G13D antigens. All spliced and non-spliced peptides that could be theoretically generated were computed. From this list, we removed all peptide candidates not carrying the target mutations, as well as those candidates shorter than 8 residues or longer than 12 residues, which is the length range most often observed in HLA-I immunopeptidomes (Figure 2D). Since the HLA-A\*02:01 allele is the predominant allele in Caucasian populations, we predicted the binding affinity of this HLA-I variant to the remaining peptides using the NetMHCpan 3.0 algorithm (30). Finally, we filtered out all peptides that were predicted to bind with IC<sub>50</sub> > 100 nM. None of the non-spliced epitope candidates passed this step, whilst 54 spliced epitope candidates had the required features. Among them, 47 can theoretically carry the KRAS G12V mutation (Figure 2D).

## Identification of KRAS G12V<sup>+</sup> Spliced Epitope Candidates Generated by Proteasomes

The majority of the HLA-I-restricted epitopes are produced by proteasomes. Their production can be verified through *in vitro* digestion of synthetic polypeptides by 20S proteasomes, as measured by MS. Because of the high frequency of putative HLA-A\*02:01<sup>+</sup> spliced epitope candidates carrying the KRAS G12V mutation (KRAS G12V<sup>+</sup>), we focused on this mutation and digested the synthetic KRAS<sub>2-35</sub> wild type and G12V polypeptides with 20S standard proteasomes for 20 h. The digestions were measured by targeted MS, which used a m/z inclusion list of target spliced epitope candidates identified in the previous pipeline step (Supplementary Table 1), and confirmed that one spliced epitope candidate, KRAS<sub>5-6/8-14</sub> G12V [KL][VVGAVGV], is generated by proteasomes under these conditions (Figure 3). This spliced peptide could be



generated by the removal of one of the three V residues in the KRAS sequence, i.e., it could be reported as KRAS<sub>5-6/8-14</sub> G12V, KRAS<sub>5-7/9-14</sub> G12V, or KRAS<sub>5-8/10-14</sub> G12V. We will refer to it as KRAS<sub>5-6/8-14</sub> G12V for the sake of simplicity.

Notably, this spliced peptide is not present in the reaction at  $t = 0$  and the 20 h reaction containing the synthetic polypeptide substrate in absence of proteasomes (data not shown). The cognate spliced peptide KRAS<sub>5-6/8-14</sub> G12 [KL][VVGAGGV] is not produced by 20S proteasomes whilst processing the synthetic wild type KRAS<sub>2-35</sub> polypeptide.

In the KRAS<sub>2-35</sub> G12V polypeptide digestion, we also identified the non-spliced epitope candidate KRAS<sub>5-14</sub> G12V [KLVVVGAVGV] (**Supplementary Figure 1**). The spontaneous response of peripheral blood mononuclear cells (PBMCs) of pancreatic adenocarcinoma patients against this latter epitope candidate was previously described (42). Although this peptide was filtered out in the early steps of our pipeline because it has a predicted HLA-A\*02:01 binding affinity  $IC_{50} > 100$  nM, we compared this epitope candidate to the KRAS<sub>5-6/8-14</sub> G12V spliced epitope candidate in the next validation steps.

### Spliced Peptide and KRAS<sub>5-6/8-14</sub> G12V Spliced Epitope Candidate Production Kinetics by Proteasomes

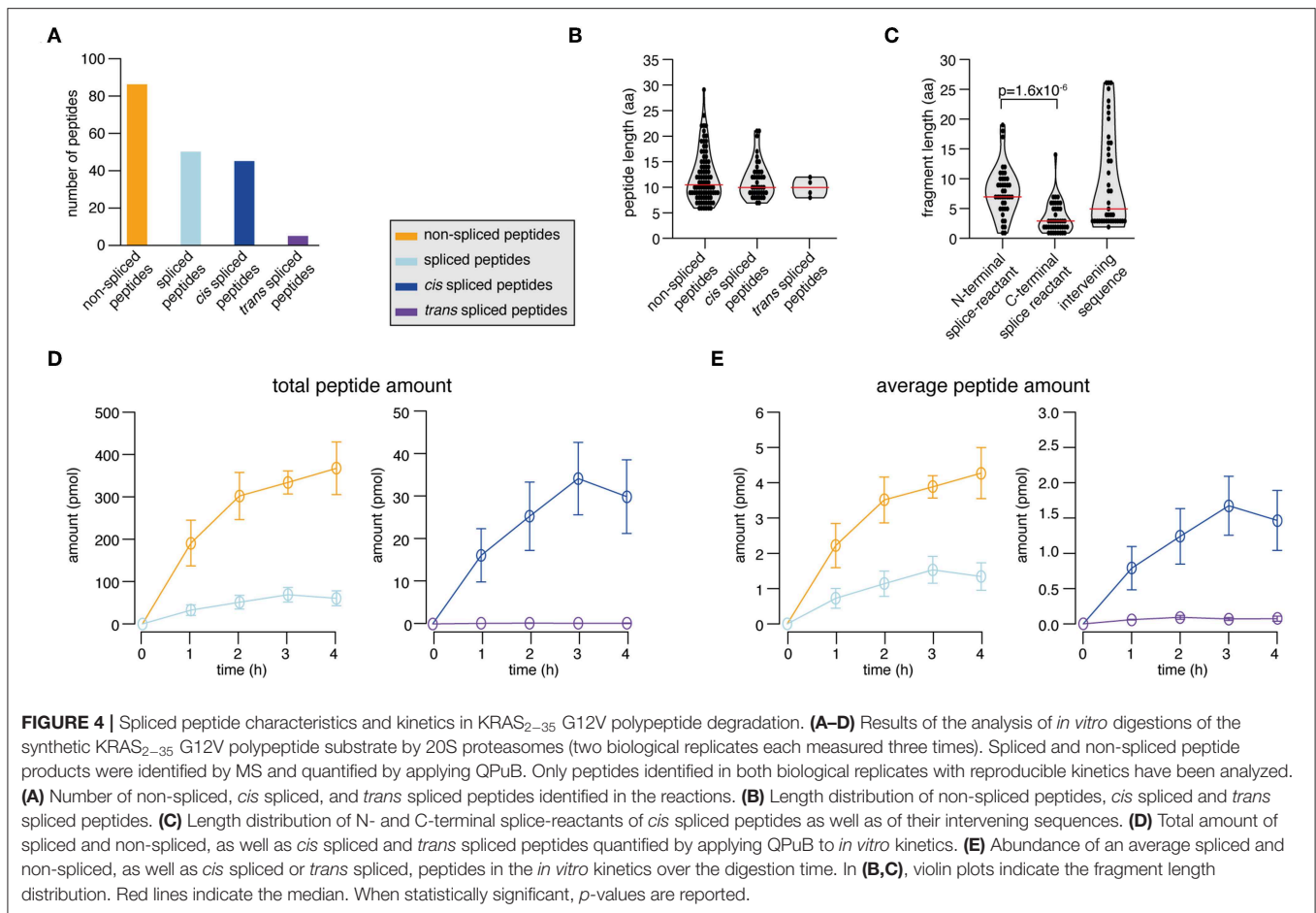
To be a robust epitope candidate, a peptide should be produced *in vitro* by proteasomes in a detectable amount and with consistent kinetics. Correspondence between *in vitro* experiments carried out with purified 20S proteasomes and

*in cellulo* and *in vivo* experiments has been demonstrated in various studies investigating both viral and tumor epitopes (4, 5, 9, 17, 19, 21, 43–50). Therefore, we performed digestion kinetics (0–4 h) of the synthetic KRAS<sub>2-35</sub> wild type and G12V polypeptides with 20S standard proteasomes. The samples were measured by MS to identify all digestion products (via MS/MS). Quantification of peptides was performed using QPUB, a method that uses detected MS ion peak areas to estimate the absolute amount of each spliced and non-spliced peptide products (see Data Availability section), and by comparison with synthetic peptide titration for the two epitope candidates.

In the synthetic KRAS<sub>2-35</sub> G12V polypeptide digestion, we identified and successfully quantified 131 peptide products. 65.6% were non-spliced, 31.3% *cis* spliced and 3.1% *trans* spliced peptides (**Figure 4A**). The length distribution of the non-spliced, *cis* spliced and *trans* spliced peptides did not significantly differ and its median was 10 amino acid residues (**Figure 4B**). N- and C-terminal splice-reactants had a median length of 7 and 3 amino acid residues, respectively (**Figure 4C**). The intervening sequences of *cis* spliced peptides had a median length of 5 amino acid residues (**Figure 4C**). From the quantitative point of view, *cis* and *trans* spliced peptides represent proximately 17.0 and 0.1% of the peptide abundance, respectively (**Figure 4D**). On average, a *trans* spliced peptide is less abundant than a *cis* spliced peptide, which is less abundant than an average non-spliced peptide (**Figure 4E**).

Through the application of QPUB to the synthetic KRAS<sub>2-35</sub> wild type and G12V polypeptide digestions, we could also compute how frequently proteasomes cleaved the substrate after





each of its individual residues (substrate cleavage strength, i.e., SCS-P1) or used each residue for the PCPS reaction (proteasome-generated spliced peptide P1 positions, i.e., PSP-P1). From this analysis we confirmed our previous hypothesis (23), whereby proteasomes splice at sites at which the substrates are less frequently cleaved at (and *vice versa*), as emerged by comparing SCS-P1 and PSP-P1 (Figures 5A,B).

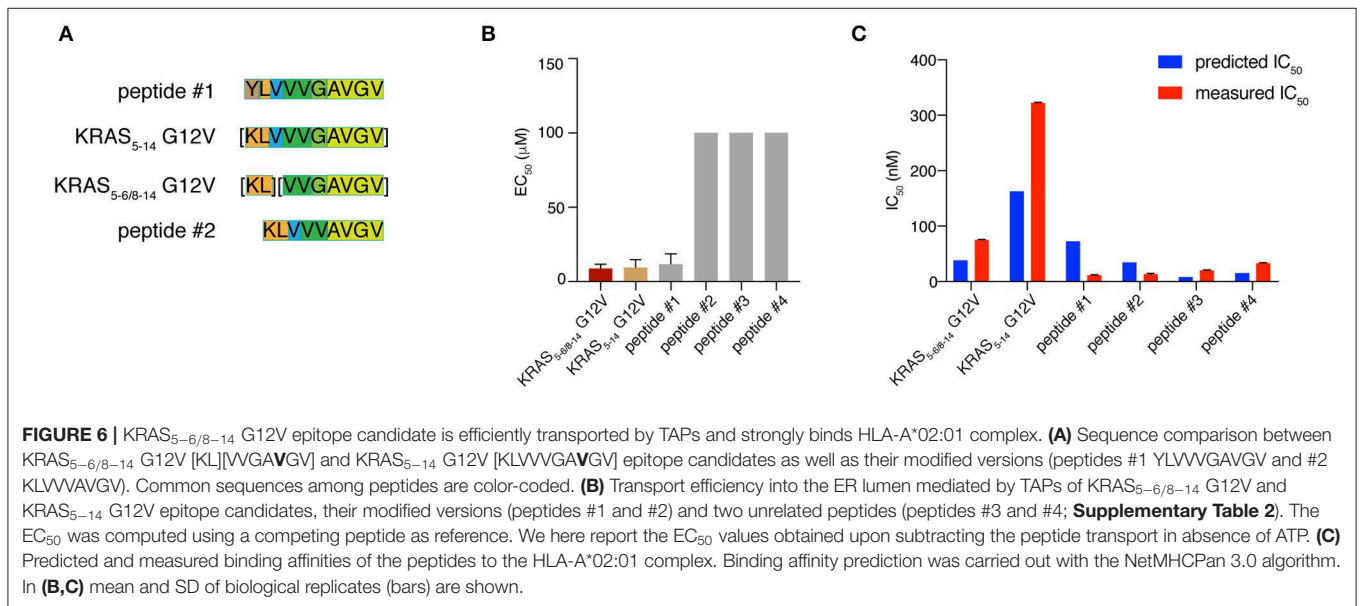
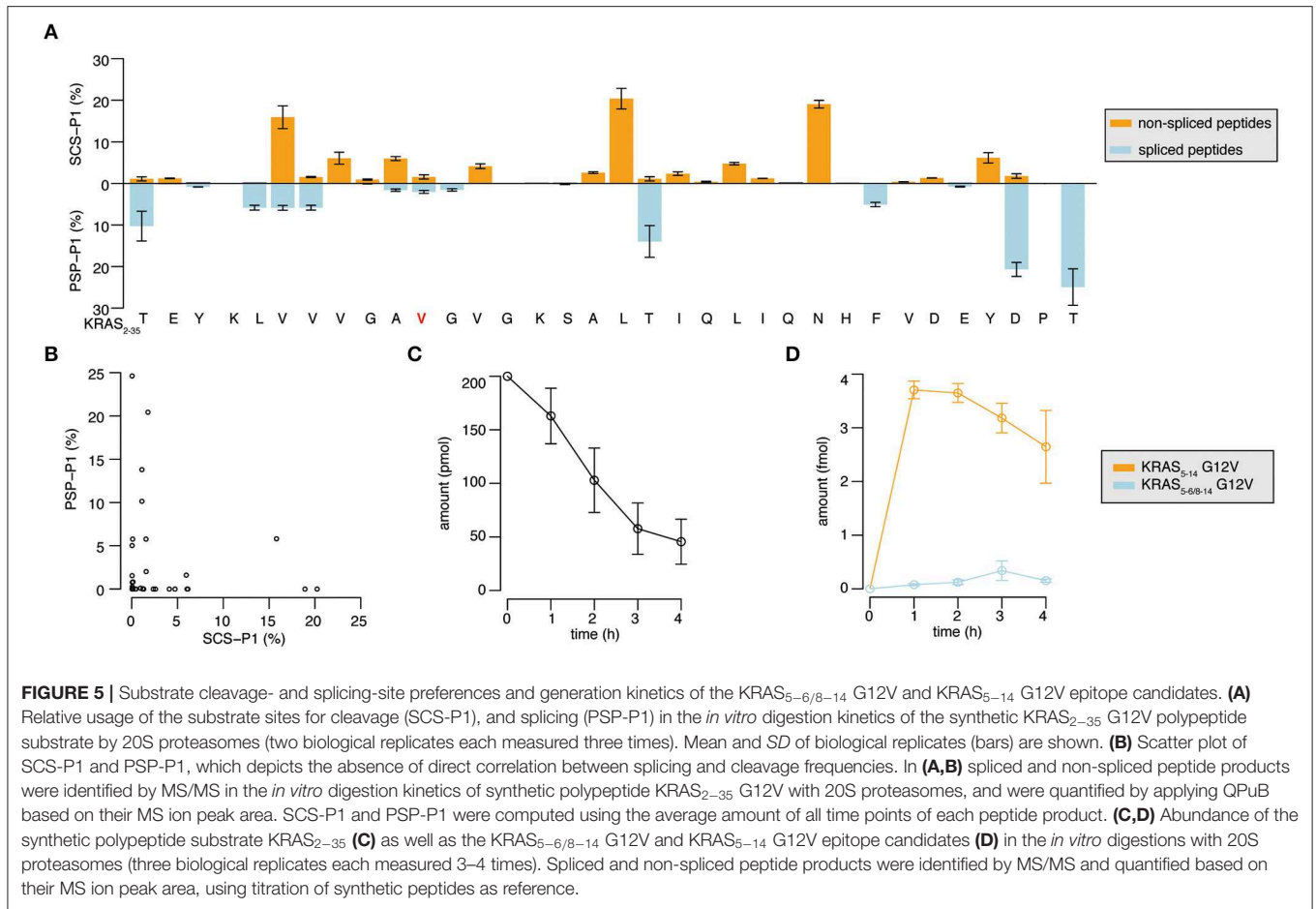
The quantitative analysis of the KRAS<sub>2–35</sub> synthetic substrate degradation (Figure 5C) also showed that the KRAS<sub>5–6/8–14</sub> G12V spliced epitope candidate is produced in amounts smaller than the average amount of spliced peptides (Figures 4E, 5D shall be compared).

### KRAS<sub>5–6/8–14</sub> G12V Spliced Epitope Candidate Is a TAP Substrate and Efficiently Binds HLA-A\*02:01

The production of a peptide by proteasomes is not sufficient alone to incur presentation on the cell surface. There are several other steps in the APP pathway that can direct the peptide fate, such as peptide transport into the ER lumen mediated by TAPs and peptide binding to HLA-A\*02:01 complex. We studied the KRAS<sub>5–6/8–14</sub> G12V [KL][VVGAVGV] spliced epitope candidate in comparison with the KRAS<sub>5–14</sub> G12V

[KLVVVGAVGV] non-spliced epitope candidate. We also extended the study to four control peptides. Two peptides - peptide #1 YLVVVGAVGV and peptide #2 KLVVVGAVGV - shared a large portion of the KRAS<sub>5–6/8–14</sub> G12V and KRAS<sub>5–14</sub> G12V epitope candidate sequences (Figure 6A). The other two control peptides are unrelated peptides predicted to bind to the HLA-A\*02:01 complex (peptides #3 FLHEDLEKI and #4 FLHEDTEKI; see Supplementary Table 2).

To quantify the efficiency of peptide transport into the ER lumen by TAP, we measured the competition between the target peptides and a fluorescent reference peptide for TAP-dependent translocation into free microsomes. The reference peptide has an N-linked glycosylation consensus sequence and peptide glycosylation is used to monitor entry into ER microsomes and as an isolation handle (29). KRAS<sub>5–6/8–14</sub> G12V epitope candidate is efficiently transported by TAPs as the KRAS<sub>5–14</sub> G12V non-spliced epitope candidate is. Peptide #1, which has a K to Y substitution at position 1 as compared to KRAS<sub>5–14</sub> G12V peptide, is transported by TAP as efficiently as the non-spliced epitope candidate. In contrast, peptide #2, which has the removal of residue G in position 6 as compared to KRAS<sub>5–14</sub> G12V peptide and a G to V substitution at position 5 as compared to KRAS<sub>5–6/8–14</sub> G12V peptide, is not competing with the reference peptide and thus ignored by TAP. This suggests a role



of the residue G at the center of the peptides in TAP-mediated transport. The other two control peptides are not substrates for TAP (**Figure 6B**).

Once a peptide arrives in the ER lumen, its binding affinity to the specific HLA-I molecule determines whether it will ultimately be presented. Therefore, we measured the binding affinity of the

same six peptides previously tested in the TAP assay and the HLA-A\*02:01 complex in a cell-free system utilizing purified HLA-I molecules. The KRAS<sub>5-6/8-14</sub> G12V epitope candidate was confirmed to efficiently bind the HLA-A\*02:01 complex, in contrast to the KRAS<sub>5-14</sub> G12V non-spliced epitope candidate, which had a measured IC<sub>50</sub> larger than 300 nM (and larger than predicted). The measured and predicted IC<sub>50</sub> of the control peptides was quite similar and the peptides appear to be good binders (Figure 6C).

### Conformation of KRAS<sub>5-6/8-14</sub> G12V and KRAS<sub>5-14</sub> G12V Epitope Candidates Within HLA-A\*02:01 cleft

Once a peptide is bound to HLA-I complexes and presented at the cell surface, it can be recognized by TCRs of CD8<sup>+</sup> T cells. The conformation of the peptide in the HLA-I groove is paramount not only for HLA-I-peptide affinity and stability, but also for the TCR-HLA-I-peptide interaction. To study this aspect, we individually refolded and crystallized HLA-A\*02:01 with spliced epitope candidate KRAS<sub>5-6/8-14</sub> G12V, non-spliced epitope candidate KRAS<sub>5-14</sub> G12V, as well as two control peptides (peptides #1 and #2), in which either the N-terminal residue or one of the central residues was substituted, as compared to the epitope candidates (Figures 7A–K). The crystal structures of these individual HLA-I-peptide complexes were determined at resolutions ranging from 1.4 to 1.58 Å by molecular replacement using PDB ID 5ENW as a search model (Supplementary Table 3). The global superposition of all these four peptide-HLA-I complexes in the peptide binding groove reveals a similar binding orientation with a root mean square deviation value (rmsd) of 0.124 Å. The electron densities for all four peptides are also well-defined over the entire peptide length.

Some structural differences in individual peptide binding were observed when comparing the 9 mer peptides (KRAS<sub>5-6/8-14</sub> G12V [KL][VVGAVGV] and peptide #2 KLVVVAVGV) with the 10 mer peptides (KRAS<sub>5-14</sub> G12V [KLVVVGAVGV] and peptide #1 YLVVVGAVGV). Comparison of peptides with a same length generally only reveals a single amino acid change in a similar orientation or the addition of a side chain, e.g., when V replaces G in peptide #2 as compared to KRAS<sub>5-14</sub> G12V peptide (Figures 7A–D). Specifically, while the N-terminal and C-terminal ends of all four peptides match perfectly, structural superposition reveals conformations in the middle portions of peptide KRAS<sub>5-6/8-14</sub> G12V and peptide #2 unique from the remaining two peptide ligands (KRAS<sub>5-14</sub> G12V and peptide #1). In the latter two cases, the middle portions of the peptides containing P4, P5, and P6 residues bulge out of the binding pockets to accommodate both peptide ends inside the peptide-binding groove of the HLA-I molecule (Figures 7C,D,G,H).

We next evaluated at the detailed interactions between HLA-A\*02:01 and individual peptides. Throughout these interfaces, extensive hydrophobic and hydrogen bonding networking is seen with the majority of peptide residues participating in the contact with HLA-I residues Y7, F9, M45, E63, K66, V67, H70, T73, T80,

L81, Y84, Y99, Y116, T143, K146, W147, V152, Y159, W167, Y171 (Figures 7E–H).

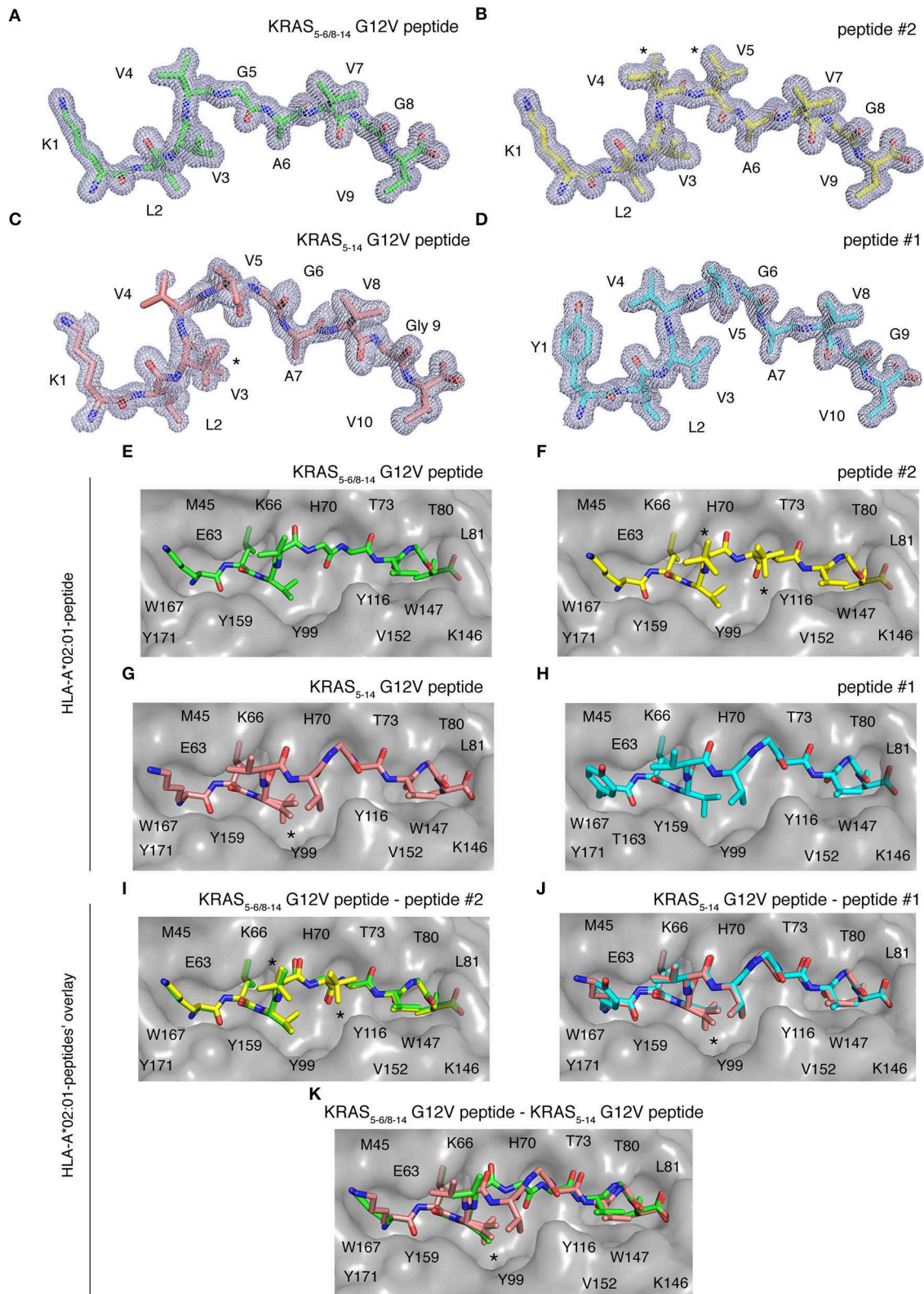
In the HLA-I-peptide #1 complex, the N-terminal P1 Y residue makes hydrophobic contact with T163, which is missing in all the other three peptide complexes and may explain the significantly higher binding affinity of 10 mer peptide #1 compared to 10 mer KRAS<sub>5-14</sub> G12V peptide (Figure 7H). The A and F pockets forming the peptide binding groove of HLA-I are mostly composed of hydrophobic residues and some 12 polar and 21 van der Waals contacts were, throughout the peptide length in all complexes, observed between the peptide moiety and HLA-A\*02:01.

While the HLA-I interaction interface seems to be conserved in both spliced peptide KRAS<sub>5-6/8-14</sub> G12V and non-spliced peptide KRAS<sub>5-14</sub> G12V, the binding affinity of KRAS<sub>5-6/8-14</sub> G12V peptide toward HLA-A\*02:01 is higher compared to the KRAS<sub>5-14</sub> G12V peptide (Figure 6C). Hence, to understand this differential affinity of these peptides for HLA-A\*02:01, we compared the crystal structures of KRAS<sub>5-6/8-14</sub> G12V and KRAS<sub>5-14</sub> G12V peptides and modified variants (peptide #1 and #2) bound to HLA-A\*02:01 complexes in more detail (Figures 7I–K). The superpositions of either KRAS<sub>5-6/8-14</sub> G12V peptide and peptide #2 or KRAS<sub>5-14</sub> G12V peptide and peptide #1 do not show any relevant differences (Figures 7I,J).

In contrast, although the structural superposition of KRAS<sub>5-6/8-14</sub> G12V and KRAS<sub>5-14</sub> G12V peptides bound to HLA-A\*02:01 molecules reveals a similar type of HLA-I interaction network at their N-terminal and C-terminal regions, their structural arrangements deviate in their middle portions. Due to this, the spliced peptide KRAS<sub>5-6/8-14</sub> G12V makes several unique interactions with HLA-A\*02:01. Firstly, in the structure of HLA-A\*02:01 complexed with spliced peptide KRAS<sub>5-6/8-14</sub> G12V, the P6 A residue makes both hydrogen bonding and van der Waals contacts with the side chain of T73 residue of HLA-I, whereas the P6 G residue of KRAS<sub>5-14</sub> G12V is not in contact with HLA-A\*02:01 and its P7 A residue maintains only hydrophobic interactions with T73.

Another difference between both these complexes is at their C-termini. In the HLA-I-KRAS<sub>5-14</sub> G12V peptide complex, the HLA-A\*02:01 K146 residue adopts a different orientation, due to which it interacts with only the terminal PΩ residue. In the HLA-I-KRAS<sub>5-6/8-14</sub> G12V complex, the amino group of K146 forms a hydrogen bond with both the carbonyl oxygen of the PΩ-1 residue and the terminal PΩ residue. Furthermore, though KRAS<sub>5-14</sub> G12V is longer [10 amino acids, compared to KRAS<sub>5-6/8-14</sub> G12V (9 amino acids)] and reorganizes its central region in the HLA binding groove, this structural rearrangement does not favor any additional contacts with HLA-A\*02:01. From our structural analysis, we can predict that the higher affinity of the spliced KRAS<sub>5-6/8-14</sub> G12V peptide, compared to KRAS<sub>5-14</sub> G12V peptide, might be due to these two additional hydrogen bonding contacts between the spliced peptides P6 A residue and the T73 of HLA-A\*02:01, as well as the spliced peptides PΩ-1 residue and the K146 residue of the HLA-A\*02:01 molecule (Figure 7K).





**FIGURE 7 |** HLA-A\*02:01-KRAS G12V peptides binding mode. Binding mode of spliced epitope candidate KRAS<sub>5-6/8-14</sub> G12V [KL][VGVAVGV], non-spliced epitope candidate KRAS<sub>5-14</sub> G12V [KLWVGVAVGV], peptides #1 YLWVGVAVGV and #2 KLWVAVGV to HLA-A\*02:01 complex. **(A-D)** 2Fo-Fc electron density map contoured at 1σ for KRAS<sub>5-6/8-14</sub> G12V peptide **(A)**, peptide #2 **(B)**, non-spliced peptide KRAS<sub>5-14</sub> G12V **(C)**, and peptide #1 **(D)**. **(E-H)** Binding of spliced peptide

(Continued)



**FIGURE 7** | KRAS<sub>5-6/8-14</sub> G12V (**E**; green sticks), peptide #2 (**F**; yellow sticks), non-spliced peptide KRAS<sub>5-14</sub> G12V (**G**; brown sticks), and peptide #1 (**H**; cyan sticks) to HLA-A\*02:01 protein (gray molecular surface). (**I-K**) Overlay of KRAS<sub>5-6/8-14</sub> G12V and peptide #2 binding to HLA-A\*02:01 (**I**), KRAS<sub>5-14</sub> G12V peptide and peptide #1 (**J**), as well as KRAS<sub>5-6/8-14</sub> G12V and KRAS<sub>5-14</sub> G12V peptides (**K**) binding to HLA-A\*02:01 molecule. All peptides are shown as sticks and in (**A-D**), peptide residues are labeled with one-letter amino acid codes. In (**E-K**), the residues of HLA-A\*02:01 that are extended in the peptide binding interface are labeled with single-letter amino acid codes. In (**B,F,I**), \* indicates the alternate conformations for residues V4 and V5 of peptide #2. In (**C,G,J-K**), \* indicates the alternate conformations for residues V3 of KRAS<sub>5-14</sub> G12V peptide.

## Potential Recognition of KRAS<sub>5-6/8-14</sub> G12V and KRAS<sub>5-14</sub> G12V Epitope Candidates Within HLA-A\*02:01 Cleft by Different TCRs

Once the peptide binds to a HLA-I molecule, it gets displayed for TCR recognition, which can then induce effective immune responses. Using structure as a tool, we tried to determine the mode of TCR-HLA-I-peptide interaction. Our evaluation of HLA-I-KRAS<sub>5-6/8-14</sub> G12V and HLA-I-KRAS<sub>5-14</sub> G12V peptide complexes provides a link to the potential cross recognition by a given CD8<sup>+</sup> T cell clone. The middle portion of both spliced and non-spliced epitope candidates containing P4 and P5 residues does not make ample contacts with the HLA-A\*02:01 molecule but has limited flexibility in the crystal structure, otherwise this would not have been solved in the structure (**Figures 8A,B**). In the HLA-A\*02:01-KRAS<sub>5-14</sub> G12V peptide complex, the middle portion that bulges out from the binding groove makes it more accessible for TCR recognition (**Figure 8B**). The side chain of P4 V and P5 V residues are facing in an upward direction and can be easily accommodated into the binding pocket located over the central peptide, formed by the most structurally diverse CDR loops, CDR3 $\alpha$  and CDR3 $\beta$  of the TCR. Similarly, in the HLA-I-KRAS<sub>5-6/8-14</sub> G12V complex, the side chain of the P4 V residue is available to mediate hydrophobic contact for TCR recognition. Also, in both complexes there is a possibility of hydrogen bonding interactions between the main chain carbonyl and amide groups of P4 and P5 residues with the TCR (**Figures 8A-C**). Hence, our analysis of the crystal structure suggests that both the spliced KRAS<sub>5-6/8-14</sub> G12V and the non-spliced KRAS<sub>5-14</sub> G12V epitope candidates can be contacted by the same TCR at their P4 site, thereby promoting cross reactivity.

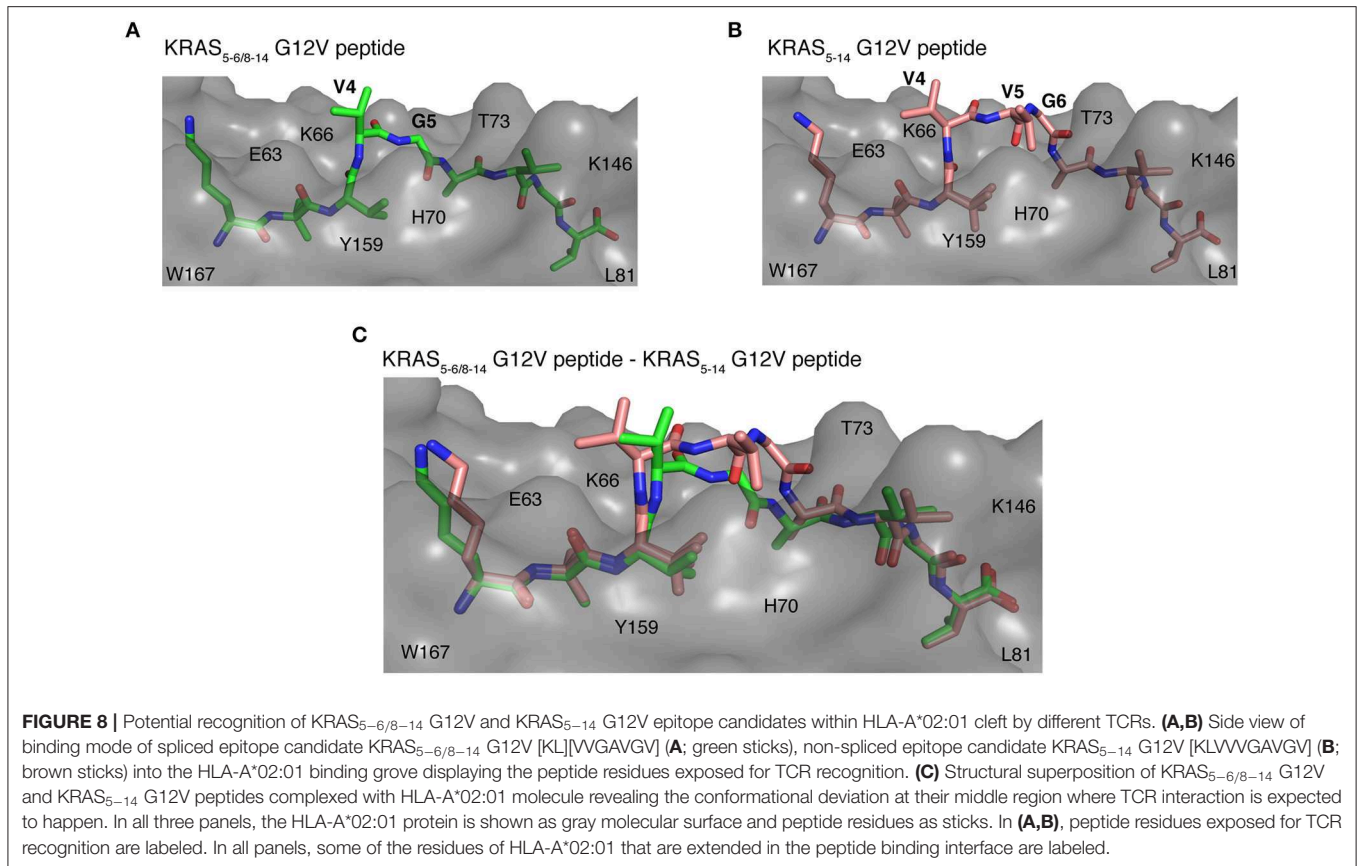
On the other hand, even though some potential TCR cross reactivity exists toward both spliced and non-spliced epitope candidates, the structural superposition of both peptide complexes revealed deviation in their peptide conformation at the region where the TCR interaction is expected to happen (**Figure 8C**). Hence, depending on the direction that the TCR encounters in the HLA-I-peptide complex, there might be a definite possibility of having TCRs that exhibit preference or exclusive binding toward either the non-spliced or the spliced epitope, rather than recognizing both of them. As the KRAS<sub>5-6/8-14</sub> G12V possess three peptide residues, P4 V, P5 V, and P6 G, that can mediate both hydrogen bonding and van der Waals contacts with a TCR, whilst the spliced peptide contains only P4 V and P5 G residues; hence, we can speculate that TCRs more likely will have selectivity and specificity for one of the two epitope candidates.

## DISCUSSION

Epitope discovery is an essential first step for antigen-targeted immunotherapies against cancer, infection and some autoimmune diseases. In the last decade, several studies proposed strategies to achieve this, especially in light of anti-cancer immunotherapies. The majority of these studies identify epitope candidates in HLA-I immunopeptidomes eluted from cells. Although this strategy guarantees that the identified epitope is presented at the cell surface, it cannot include all targetable epitopes because of its relative low sensitivity (51). TCRs can still be considered more sensitive than MS-based methods and can sense even a few epitope molecules bound to HLA-I molecules to trigger cytotoxic responses. There are several examples of epitopes that were not identified by analytical methods based on HLA-I immunopeptidomes of cells, but were well-recognized by specific CTLs. The pipeline that we proposed here tries to circumvent this problem. While starting from a large number of theoretical epitope candidates, the pipeline narrows them down to a few selected candidates step by step. One of the advantages of our strategy is that its sequential steps could be exchanged and adapted to the specific requirements of a given application. For instance, in this study we developed a model to rank antigens by their potential over-representation in HLA-I immunopeptidomes considering four protein features, without including any cell-specific assays, such as transcriptome or intracellular proteome analysis. If such data was available, our pipeline could use more complex algorithms, such as that published by Pearson et al. (37), and likely reach a more in-depth antigen selection.

The same principle of flexible structure and interchangeable steps could be applied to the “*in vitro* selection” section of our pipeline. In this study, we tested *in vitro* three steps of the HLA-I APP pathway: proteasome-mediated generation; TAP-mediated transport into the ER lumen; and the efficient binding to the selected HLA-I variant. While some epitopes may be presented by HLA-I molecules in a proteasome- and TAP-independent fashion, the majority of HLA-I-restricted epitopes depends on these two steps.

Efficient binding to the selected HLA-I molecule is, on the contrary, mandatory. However, although a threshold of 500 nM would capture ~85% of all HLA-I-bound peptides (52, 53), it is still an open question what the optimal IC<sub>50</sub> threshold is to define a “good epitope target” for ATT. The most determining factor could be the off-rate of peptide binding, a feature that we likely determine indirectly via IC<sub>50</sub>, because poor and good peptides have been reported to have similar on-rates but different off-rates (54).



Of course, the “*in vitro* selection” section of our pipeline could enlist other APP steps such as tapasin-dependency, cytosolic peptidase and ERAP trimming, etc. (2, 55), which could be selectively chosen based on tumor features and the known APP pathway of the target antigen.

Our pipeline identifies epitope candidates which shall further be validated by isolating specific CTL clones and their TCRs and using them to confirm that the epitope candidates are produced *in cellulo* and eventually *in vivo*. There are several strategies to this end. For example, the KRAS<sub>5-6/8-14</sub> spliced epitope candidate here identified has been validated in collaboration with Blankenstein et al. (56). Specific TCRs have been isolated from humanized ABAbDII mice (57), cloned into expressing vectors and transduced into human PBMCs. Transduced human CD8<sup>+</sup> T cells selectively recognized the KRAS<sub>5-6/8-14</sub> spliced epitope. They also recognized human cancer cell lines expressing KRAS G12V antigen and HLA-A\*02:01 complex and release IFN $\gamma$ . They do not recognize a cancer cell line expressing the wild type KRAS G12 protein and the HLA-A\*02:01 complex. These outcomes validated the KRAS<sub>5-6/8-14</sub> spliced epitope candidate as a genuine epitope (56). We tried to identify the KRAS<sub>5-6/8-14</sub> G12V spliced and the KRAS<sub>5-14</sub> G12V non-spliced epitope candidates through the MS measurement of the HLA-I immunopeptidomes of the SW480 pancreatic adenocarcinoma cell line, which expresses the HLA-A\*02:01 complex and the KRAS G12V mutated protein (42). Despite CTL clones could recognize both epitope candidates presented

by cancer cell lines (42, 56), none of the two peptides was identified in the SW480-derived HLA-I immunopeptidomes (data not shown), thereby confirming the usefulness of the pipeline described here.

Another advantage of our pipeline is its ability to select and identify proteasome-generated spliced epitope candidates, which we and other groups have found to represent a sizeable pool of immunologically relevant epitopes, especially within the framework of anti-cancer immunotherapies (3, 11, 12, 20, 21, 44, 58). The PCPS reaction was shown to generate a large number and a significant amount of spliced peptides in the *in vitro* processing of KRAS<sub>2-35</sub> neoantigen by 20S proteasomes. This suggests that we might have previously underestimated PCPS frequency in the *in vitro* proteasome digestions, likely due to the low MS sensitivity available at that time (23).

The benefit of including these unconventional epitopes in our pipeline is evident. The recurrent KRAS G12V mutation can be efficiently presented by HLA-A\*02:01 complexes only through spliced peptides. TCRs specific to this spliced epitope candidate could be used to treat around 15–20% of colorectal cancer and pancreatic adenocarcinoma by ATT. According to our analysis, the G12V mutation promotes not only the binding affinity of the KRAS<sub>5-6/8-14</sub> G12V spliced epitope candidate to HLA-A\*02:01 complex, but also the splicing reaction at that site, since we did not identify the KRAS<sub>5-6/8-14</sub> G12 spliced peptide in the *in vitro* digestion of wild type KRAS<sub>2-35</sub>

G12 by 20S proteasomes. The KRAS<sub>5-6/8-14</sub> G12V sequence [KL][VVGAVGV] cannot be generated by any other human protein by peptide hydrolysis or *cis* peptide splicing (data not shown), thereby defining it as a unique neopeptide. In the KRAS<sub>2-35</sub> G12V polypeptide digestion, we also identified the non-spliced epitope candidate KRAS<sub>5-14</sub> G12V, which was shown to be recognized by PBMCs of pancreatic adenocarcinoma patients (42). For a cancer-targeted strategy, it would be informative to perform *in vitro* digestions using proteasome isoforms recapitulating those present in the target cancer, since they vary from tumor to tumor with implications for the quantity (and perhaps the quality) of peptide produced (4, 5, 27). Both KRAS<sub>5-6/8-14</sub> G12V and KRAS<sub>5-14</sub> G12V epitope candidates are efficiently transported by TAP into the ER lumen, which is then no bottleneck. The non-spliced epitope candidate, however, binds HLA-A\*02:01 less efficiently than the spliced epitope candidate with an IC<sub>50</sub> larger than 300 nM, which could be higher than the binding affinity currently suggested for immunodominant epitopes. According to analysis of the HLA-A\*02:01-peptide crystal structures, the two epitope candidates differ in the region exposed to TCRs. Therefore, we would expect that unique CD8<sup>+</sup> T cell clones could recognize these, although cross-reactive TCRs cannot be excluded. In the case of cross-reactivity, the immunodominance of the spliced epitope over the non-spliced epitope might be favored by the higher binding affinity to HLA-A\*02:01 complex. We do not have enough information about the other steps of their APP pathways, including production by cancer-associated proteasome isoforms, to conclude about presentation in cancer patients. Yet, our pipeline allows identification of potential new neopeptides derived from peptide splicing that are unique for a driver in oncogenesis, KRAS G12V. Such epitopes could be critical in new vaccination approaches for the related tumors.

## DATA AVAILABILITY STATEMENT

All datasets generated for this study are either included in the article/**Supplementary Material** or deposited in online repositories, as clarified in the Materials and Methods section.

## REFERENCES

- Weissman AM, Shabek N, Ciechanover A. The predator becomes the prey: regulating the ubiquitin system by ubiquitylation and degradation. *Nat Rev Mol Cell Biol.* (2011) 12:605–20. doi: 10.1038/nrm3173
- Rock KL, Reits E, Neefjes J. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends Immunol.* (2016) 37:724–37. doi: 10.1016/j.it.2016.08.010
- Mishto M, Liepe J. Post-Translational peptide splicing and T cell responses. *Trends Immunol.* (2017) 38:904–15. doi: 10.1016/j.it.2017.07.011
- Guillaume B, Chapiro J, Stroobant V, Colau D, Van Holle B, Parvizi G, et al. Two abundant proteasome subtypes that uniquely process some antigens presented by HLA class I molecules. *Proc Natl Acad Sci USA.* (2010) 107:18599–604. doi: 10.1073/pnas.1009778107
- Guillaume B, Stroobant V, Bousquet-Dubouch MP, Colau D, Chapiro J, Parmentier N, et al. Analysis of the processing of seven human tumor antigens by intermediate proteasomes. *J Immunol.* (2012) 189:3538–47. doi: 10.4049/jimmunol.1103213
- Solomon H, Brauning B, Fainer I, Ben-Nissan G, Rabani S, Goldfinger N, et al. Post-translational regulation of p53 function through 20S proteasome-mediated cleavage. *Cell Death Differ.* (2017) 24:2187–98. doi: 10.1038/cdd.2017.139
- Olshina MA, Ben-Nissan G, Sharon M. Functional regulation of proteins by 20S proteasome proteolytic processing. *Cell Cycle.* (2018) 17:393–4. doi: 10.1080/15384101.2017.1414682
- Hanada K, Yewdell JW, Yang JC. Immune recognition of a human renal cancer antigen through post-translational protein splicing. *Nature.* (2004) 427:252–6. doi: 10.1038/nature02240
- Vigneron N, Stroobant V, Chapiro J, Ooms A, Degiovanni G, Morel S, et al. An antigenic peptide produced by peptide splicing in the proteasome. *Science.* (2004) 304:587–90. doi: 10.1126/science.1095522

## AUTHOR CONTRIBUTIONS

MM and JL developed the whole project, performed and supervised the data analysis and data generation, and wrote the manuscript. DZ supervised the crystal structure analysis and wrote the cognate text. JS, JN, and AS supervised part of the analysis, wrote, and edited the text. AM, GY, AB, and RC performed and analyzed part of the experiments and proofread the manuscript. SH and DP carried out QPUB analysis and proofread the manuscript. HU supervised MS measurement and optimization.

## FUNDING

This study has been in part supported by: (i) NIH to AS (R21AI134127), (ii) Cancer Research UK King's Health Partners Center at King's College London (Development Fund 2018), KCL-Monash Collaborative Award 2018 and MPI-BPC collaboration agreement 2018 to MM; the initial design of the study was conceptualized and the spliced epitope was identified by MM, while appointed at Charité – Universitätsmedizin Berlin with a contract financially supported by the BIH grant awarded to P.M. Kloetzel (BIH, CRG1-TP1), and by JL, whilst she was supported by NC3Rs through a David Sainsbury Fellowship (NC/K001949/1) at Imperial College London. SH and AM are supported by the International Max Planck Research School (IMPRS) for Genome Science.

## ACKNOWLEDGMENTS

We thank P. Henklein, P. Kunert, B. Brecht-Jachan (Charité Berlin), and the Peptide Synthesis Facility of the Charité, as well as O. Jahn (Max Planck Institute for Experimental Medicine) for peptide synthesis. We thank W. Scott (KCL) for proofreading the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.02572/full#supplementary-material>



10. 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. doi: 10.1126/science.aaf4384
11. 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:eaar3947. doi: 10.1126/sciimmunol.aar3947
12. 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. doi: 10.1158/2326-6066.CIR-18-0424
13. 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. doi: 10.1074/mcp.RA118.000877
14. Rolfs Z, Solntsev SK, Shortreed MR, Frey BL, Smith LM. Global identification of post-translationally spliced peptides with neo-fusion. *J Proteome Res*. (2018) 18, 349–58. doi: 10.1021/acs.jproteome.8b00651
15. Faridi P, Li C, Ramarathinam SH, Illing PT, Mifsud NA, Ayala R, et al. Response to Comment on “A subset of HLA-I peptides are not genomically templated: evidence for cis- and trans-spliced peptide ligands”. *Sci Immunol*. (2019) 4:eaaw8457. doi: 10.1126/sciimmunol.aaw8457
16. Rolfs Z, Muller M, Shortreed MR, Smith LM, Bassani-Sternberg M. Comment on “A subset of HLA-I peptides are not genomically templated: evidence for cis- and trans-spliced peptide ligands”. *Sci Immunol*. (2019) 4:eaaw1622. doi: 10.1126/sciimmunol.aaw1622
17. Ebstein F, Textoris-Taube K, Keller C, Golnik R, Vigneron N, Van Den Eynde BJ, et al. Proteasomes generate spliced epitopes by two different mechanisms and as efficiently as non-spliced epitopes. *Sci Rep*. (2016) 6:24032. doi: 10.1038/srep24032
18. Gonzalez-Duque S, Azoury ME, Colli ML, Afonso G, Turatsinze JV, Nigi L, et al. Conventional and neo-antigenic peptides presented by beta cells are targeted by circulating naive CD8+ T cells in type 1 diabetic and healthy donors. *Cell Metab*. (2018) 28:946–60.e6. doi: 10.1016/j.cmet.2018.07.007
19. Platteel ACM, Liepe J, Textoris-Taube K, Keller C, Henklein P, Schalkwijk HH, et al. Multi-level strategy for identifying proteasome-catalyzed spliced epitopes targeted by CD8+ T cells during bacterial infection. *Cell Rep*. (2017) 20:1242–53. doi: 10.1016/j.celrep.2017.07.026
20. Robbins PF, El-Gamil M, Kawakami Y, Stevens E, Yannelli JR, Rosenberg SA. Recognition of tyrosinase by tumor-infiltrating lymphocytes from a patient responding to immunotherapy. *Cancer Res*. (1994) 54:3124–6.
21. Dalet A, Robbins PF, Stroobant V, Vigneron N, Li YF, El-Gamil M, et al. An antigenic peptide produced by reverse splicing and double asparagine deamidation. *Proc Natl Acad Sci USA*. (2011) 108:E323–31. doi: 10.1073/pnas.1101892108
22. Dalet A, Vigneron N, Stroobant V, Hanada K, Van Den Eynde BJ. Splicing of distant peptide fragments occurs in the proteasome by transpeptidation and produces the spliced antigenic peptide derived from fibroblast growth factor-5. *J Immunol*. (2010) 184:3016–24. doi: 10.4049/jimmunol.0901277
23. Mishto M, Goede A, Taube KT, Keller C, Janek K, Henklein P, et al. Driving forces of proteasome-catalyzed peptide splicing in yeast and humans. *Mol Cell Proteomics*. (2012) 11:1008–23. doi: 10.1074/mcp.M112.020164
24. Berkers CR, De Jong A, Schuurman KG, Linnemann C, Meiring HD, Janssen L, et al. Definition of proteasomal peptide splicing rules for high-efficiency spliced peptide presentation by MHC class I molecules. *J Immunol*. (2015) 195:4085–95. doi: 10.4049/jimmunol.1402455
25. Liepe J, Ovaas H, Mishto M. Why do proteases mess up with antigen presentation by re-shuffling antigen sequences? *Curr Opin Immunol*. (2018) 52:81–6. doi: 10.1016/j.coi.2018.04.016
26. Mishto M, Raza ML, De Biase D, Ravizza T, Vasuri F, Martucci M, et al. The immunoproteasome beta5i subunit is a key contributor to ictogenesis in a rat model of chronic epilepsy. *Brain Behav Immun*. (2015) 49:188–96. doi: 10.1016/j.bbi.2015.05.007
27. Mishto M, Liepe J, Textoris-Taube K, Keller C, Henklein P, Weberruss M, et al. Proteasome isoforms exhibit only quantitative differences in cleavage and epitope generation. *Eur J Immunol*. (2014) 44:3508–21. doi: 10.1002/eji.201444902
28. Liepe J, Mishto M, Textoris-Taube K, Janek K, Keller C, Henklein P, et al. The 20S proteasome splicing activity discovered by spliceMet. *PLoS Comput Biol*. (2010) 6:e1000830. doi: 10.1371/journal.pcbi.1000830
29. Neefjes JJ, Momburg F, Hammerling GJ. Selective and ATP-dependent translocation of peptides by the MHC-encoded transporter. *Science*. (1993) 261:769–71. doi: 10.1126/science.8342042
30. Nielsen M, Andreatta M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med*. (2016) 8:33. doi: 10.1186/s13073-016-0288-x
31. Remesh SG, Andreatta M, Ying G, Kaever T, Nielsen M, McMurtrey C, et al. Unconventional peptide presentation by major histocompatibility complex (MHC) class I Allele HLA-A\*02:01: BREAKING CONFINEMENT. *J Biol Chem*. (2017) 292:5262–70. doi: 10.1074/jbc.M117.776542
32. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. (2010) 66:486–501. doi: 10.1107/S0907444910007493
33. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*. (1997) 53:240–55. doi: 10.1107/S0907444996012255
34. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res*. (2019) 47:D442–50. doi: 10.1093/nar/gky1106
35. Bassani-Sternberg M, Pletscher-Frankild S, Jensen LJ, Mann M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol Cell Proteomics*. (2015) 14:658–73. doi: 10.1074/mcp.M114.042812
36. Mommen GP, Frese CK, Meiring HD, Van Gaans-Van Den Brink J, De Jong AP, Van Els CA, et al. Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (ETHCD). *Proc Natl Acad Sci USA*. (2014) 111:4507–12. doi: 10.1073/pnas.1321458111
37. Pearson H, Daouda T, Granados DP, Durette C, Bonneil E, Courcelles M, et al. MHC class I-associated peptides derive from selective regions of the human genome. *J Clin Invest*. (2016) 126:4690–701. doi: 10.1172/JCI88590
38. Prior IA, Lewis PD, Mattos C. A comprehensive survey of Ras mutations in cancer. *Cancer Res*. (2012) 72:2457–67. doi: 10.1158/0008-5472.CAN-11-2612
39. Shin SM, Choi DK, Jung K, Bae J, Kim JS, Park SW, et al. Antibody targeting intracellular oncogenic Ras mutants exerts anti-tumour effects after systemic administration. *Nat Commun*. (2017) 8:15090. doi: 10.1038/ncomms15090
40. Wang QJ, Yu Z, Griffith K, Hanada K, Restifo NP, Yang JC. Identification of T-cell receptors targeting KRAS-mutated human tumors. *Cancer Immunol Res*. (2016) 4:204–14. doi: 10.1158/2326-6066.CIR-15-0188
41. Tran E, Robbins PF, Lu YC, Prickett TD, Gartner JJ, Jia L, et al. T-cell transfer therapy targeting mutant KRAS in cancer. *N Engl J Med*. (2016) 375:2255–62. doi: 10.1056/NEJMoa1609279
42. Kubuschok B, Neumann F, Breit R, Sester M, Schormann C, Wagner C, et al. Naturally occurring T-cell response against mutated p21 ras oncoprotein in pancreatic cancer. *Clin Cancer Res*. (2006) 12:1365–72. doi: 10.1158/1078-0432.CCR-05-1672
43. Chapiro J, Claverol S, Piette F, Ma W, Stroobant V, Guillaume B, et al. Destructive cleavage of antigenic peptides either by the immunoproteasome or by the standard proteasome results in differential antigen presentation. *J Immunol*. (2006) 176:1053–61. doi: 10.4049/jimmunol.176.2.1053
44. Warren EH, Vigneron NJ, Gavin MA, Coulie PG, Stroobant V, Dalet A, et al. An antigen produced by splicing of noncontiguous peptides in the reverse order. *Science*. (2006) 313:1444–7. doi: 10.1126/science.1130660
45. Deol P, Zaiss DM, Monaco JJ, Sijts AJ. Rates of processing determine the immunogenicity of immunoproteasome-generated epitopes. *J Immunol*. (2007) 178:7557–62. doi: 10.4049/jimmunol.178.12.7557
46. Tenzer S, Wee E, Burgevin A, Stewart-Jones G, Friis L, Lamberth K, et al. Antigen processing influences HIV-specific cytotoxic T lymphocyte immunodominance. *Nat Immunol*. (2009) 10:636–46. doi: 10.1038/ni.1728
47. Dalet A, Stroobant V, Vigneron N, Van Den Eynde BJ. Differences in the production of spliced antigenic peptides by the standard proteasome and the immunoproteasome. *Eur J Immunol*. (2011) 41:39–46. doi: 10.1002/eji.201040750
48. Zanker D, Waithman J, Yewdell JW, Chen W. Mixed proteasomes function to increase viral peptide diversity and broaden antiviral CD8+ T cell responses. *J Immunol*. (2013) 191:52–9. doi: 10.4049/jimmunol.1300802



49. Michaux A, Larrieu P, Stroobant V, Fonteneau JF, Jotereau F, Van Den Eynde BJ, et al. A spliced antigenic peptide comprising a single spliced amino acid is produced in the proteasome by reverse splicing of a longer peptide fragment followed by trimming. *J Immunol.* (2014) 192:1962–71. doi: 10.4049/jimmunol.1302032
50. Platteel AC, Mishto M, Textoris-Taube K, Keller C, Liepe J, Busch DH, et al. CD8 T cells of *Listeria monocytogenes*-infected mice recognize both linear and spliced proteasome products. *Eur J Immunol.* (2016) 44:3543–59. doi: 10.1002/eji.201545989
51. Loffler MW, Mohr C, Bichmann L, Freudenmann LK, Walzer M, Schroeder CM, et al. Multi-omics discovery of exome-derived neoantigens in hepatocellular carcinoma. *Genome Med.* (2019) 11:28. doi: 10.1186/s13073-019-0636-8
52. Sette A, Vitiello A, Rehman B, Fowler P, Nayarsina R, Kast WM, et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J Immunol.* (1994) 153:5586–92.
53. Paul S, Weiskopf D, Angelo MA, Sidney J, Peters B, Sette A. HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity. *J Immunol.* (2013) 191:5831–9. doi: 10.4049/jimmunol.1302101
54. Garstka MA, Fish A, Celie PH, Joosten RP, Janssen GM, Berlin I, et al. The first step of peptide selection in antigen presentation by MHC class I molecules. *Proc Natl Acad Sci USA.* (2015) 112:1505–10. doi: 10.1073/pnas.1416543112
55. Boulanger DSM, Eccleston RC, Phillips A, Coveney PV, Elliott T, Dalchau N. A mechanistic model for predicting cell surface presentation of competing peptides by MHC class I molecules. *Front Immunol.* (2018) 9:1538. doi: 10.3389/fimmu.2018.01538
56. Blankenstein T, Willimsky G, Liepe J, Kloetzel PM, Mishto M, Beier C. T cell receptors for tumor specific proteasome splice variants and uses thereof. *European Patent* (2019). PCT/EP2019/050027. Pub. No. WO/2019/129892
57. Li LP, Lampert JC, Chen X, Leita C, Popovic J, Muller W, et al. Transgenic mice with a diverse human T cell antigen receptor repertoire. *Nat Med.* (2010) 16:1029–34. doi: 10.1038/nm.2197
58. Bonnet D, Warren EH, Greenberg PD, Dick JE, Riddell SR. CD8(+) minor histocompatibility antigen-specific cytotoxic T lymphocyte clones eliminate human acute myeloid leukemia stem cells. *Proc Natl Acad Sci USA.* (1999) 96:8639–44. doi: 10.1073/pnas.96.15.8639

**Conflict of Interest:** MM and JL are co-inventors of the spliced epitope and specific TCRs protected by the patent PCT/EP2019/050027.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Mishto, Mansurkhodzhaev, Ying, Bitra, Cordfunke, Henze, Paul, Sidney, Urlaub, Neefjes, Sette, Zajonc and Liepe. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.