Differential immunopeptidome analysis revealed cancer specific amino acid usage of HLA class-I antigens and novel neoantigens of colorectal cancer

Yuriko Minegishi[1], Kazuma Kiyotani[2], Kensaku Nemoto[2], Yoshikage Inoue[3], Yoshimi Haga[1], Risa Fujii[1], Naomi Saichi[1], Satoshi Nagayama[4], Koji Ueda[1*]

[1] Cancer Proteomics Group, Cancer Precision Medicine Center, Japanese Foundation for Cancer Research, Tokyo, Japan.

[2] Project for Immunogenomics, Cancer Precision Medicine Center, Japanese Foundation for Cancer Research, Tokyo, Japan.

[3] Department of Surgery, Kyoto University, Kyoto, Japan.

[4] Development of gastroenterological surgery, Cancer Institute Hospital of Japanese Foundation for Cancer Research, Tokyo, Japan.

*Corresponding author: Koji Ueda

3-8-31, Ariake, Koto-ku, Tokyo, 135-8550, Japan

Tel.: +81 3 3570 0658, Fax.: +81 3 3570 0644, e-mail: koji.ueda@jfcr.or.jp

Keywords: Immunopeptidomics, HLA peptides, Mass spectrometry, FAIMS, differential ion mobility (DIM), neoantigen, cancer immune therapy

Abstract

Knowing the nature of human leukocyte antigen (HLA) peptides, also called as immunopeptides, is indispensable to realize the cancer precision medicine such as cancer vaccination and the better prediction of efficacy for immunocheckpoint inhibitor (ICI) treatment. Direct interrogation of immunopeptides by mass spectrometry (MS) is of great use for that purpose but the reality in analyses of scarce tissue samples still confronts technical challenges. To elucidate characteristics of HLA class-I immunopeptides specifically presented on colorectal cancer (CRC), the optimized immunopeptide isolation method and differential ion mobility mass spectrometry (DIM-MS) with whole exome sequencing (WES)-based personalized database were established and subjected to differential analysis of tumor or normal tissues of CRC patients. From pilot experiments using 10^8 cells of colon cancer cell line HCT116, total 9,249 unique immunopeptides, including total 11 neoantigens, were identified. Next, approximately 40 mg of tumor or normal regions of CRC tissues were collected from 17 patients and analyzed by our personalized immunopeptidomic technology. As the result, 44,785 unique immunopeptides were profiled, in which 2 neoantigens carrying the mutation KRAS-G12V or CPPED1-R228Q were identified. Interestingly, specific amino acid usage of C-terminus trimming of immunopeptides was found in tumor-exclusive immunopeptides. Thus, our personalized immunopeptidome analysis significantly expands presented antigen knowledgebase and allows direct determination of neoantigens from scarce tissue specimens. This advanced immunopeptidomics holds promise to new outlook of research in immunopeptides both from basic and clinical aspects.

Introduction

The major histocompatibility complex (MHC) governs individual immunity. Under the immunological tolerance, self antigens are not recognized by cytotoxic T cells (CTLs), while non-self antigens derived from alienated cells, such as exogenous pathogen-infected cells or mutation-carrying cancer cells, are recognized and eradicated by CTLs. In cancer immunotherapy, neoantigens carrying cancer specific somatic mutations are presented on cellular surface of cancer cells as non-self antigen and activate the associated repertoire of T cell receptor that will evoke anti-cancer immunity. Recently, numbers of immune checkpoint inhibitors have been established for cancer immunotherapies to reinvigorate the anti-cancer immunity. While the immunotherapies are fully efficacious in patients with certain cancer types such as melanoma, non-small-cell lung and bladder cancer, renal-cell carcinoma, and lymphoma [1], it is also true that there are still obstacles of less efficacy in cancers with less mutational burden such as microsatellite stable (MSS) colon and rectal cancers (CRCs) compare to the microsatellite instable (MSI) CRCs with DNA mismatch-repair deficiency [1, 2]. Predicting the efficacy and choosing the right strategy for ICIs based on the diverse genetic and immunological backgrounds are urgent tasks to grant the benefits by cancer precision immunotherapy. And that can be achieved by the detailed analyses of immunopeptidome under the individual pathological conditions.

Currently, the immunopeptidomics by mass spectrometry (MS) is the only technology to directly interrogate the thousands of HLA-presented antigens (immunopeptides). It is thus a unique technology yet a few concerns have been reported in the immunopeptidomics guidelines as issues to be overcome [3, 4]. According to the report, one of the very disadvantages of MS-based immunopeptidomics from tissue sample is insufficient depth on analysis from a scarce sample, such as endoscopic biopsy, compared to the *in-silico* prediction of immunopeptides based on genomic information. To obtain the robust data by immunopeptidomics such as thousands scale of

immunopeptides, it generally requires more than $1e^8$ of cells for culture cells or more than 1000 ug of net weight for tissue with current immunopeptidomics [5-7]. Furthermore, in order to accomplish the analysis with sufficient depth to identify antigens with tumor-specific mutations (neoantigens), chemical prefractionation have been necessary, which requires more amounts of samples. Regardless of the previous efforts, direct detection of neoantigens from tissue specimens, such as colorectal, liver, or ovarian cancer, have been distressed and unsuccessful [8, 9 10]. However, interrogation of native immunopeptides and neoantigens which vary accordingly to the biological microenvironment comprised of cancer cells, stroma, and blood cells, is indispensable to develop the cancer precision immunotherapy including adoptive T-cell transfer therapy, cancer vaccines, and novel ICIs. Therefore, there is no doubt that it is essential to establish the immunopeptidomics technology with sufficient sensitivity and comprehensiveness.

From that aim, we adopted the differential ion mobility mass spectrometry (DIM-MS) equipped with the high field symmetric wave form ion mobility mass spectrometry (FAIMS) interface for immunopeptidomic analysis. The seamless fractionation of sample at gas-phase by FAIMS secured the sufficient depth of immunopeptidome from the scarce samples. This advanced immunopeptidomics expanded not only the size of immnopeptidome but also the neoantigen identification both from the cellular and the tissue samples. And for the first time, the neoantigens were successfully and directly identified from the clinical tissues of colorectal tumors. Moreover, individual immunopeptidome profiling from normal and tumor tissues revealed the distinct features of C-terminal trimming of immunopeptides. This advanced immunopeptidomics by DIM-MS holds promise to explore a new phase of cancer immunology.

Materials and Methods

Clinical samples and ethical approval

Metastatic colon cancer tissues at pathological stage IV and its adjacent normal tissues were obtained from patients who all had provided written informed consent for this study. This study (ethical committee number 2010-1058) was approved by the ethical committee of the Japanese Foundation for Cancer Research (JFCR). The clinical information of them are listed in **Supplementary Table 1**. All tissue samples were frozen and maintained at least once before the MS sample preparation. Approximately 40 mg in net weight of tissue was dissected from the mass of tumor/ normal tissues respectively and further processed for immunopeptidomics.

Cell culture

The colon cancer cell line HCT-116 was maintained under the general condition. Briefly, the cells were cultured in RPMI1640 media (WAKO) with 10% heat-inactivated fetal bovine serum (FBS), penicillin G and streptomycin. Cells were washed once with PBS before harvest. Desired number of cell pellets were stored at -80 °C until use.

In-house purification of antibody and the IP-beads preparation

A hybridoma clone of anti-panHLA alpha chain antibody (HB95) was obtained from ATCC. Hybridoma cells were first expanded in 10% FBS containing DMEM and then cultured in CELL Line flask (WHEATON) with COS medium (COSMO BIO) for antibody production. After 1-week of cultivation, the culture media was collected and affinity purified by Protein A Fast flow (GE). Small amount of aliquots of eluates were electrophoresed and then stained by GelCode Blue Stain Reagent (Thermo Scientific) to check IgG fractions. The IgG enriched fractions were pooled and then dialyzed to replace the buffer into PBS. The antibody concentration was examined by BCA assay. To confirm the sufficient affinity of our in-house purified W6/32 antibody against HLA molecules, we

tested all batches of W6/32 by general IP and Western blotting by antibody anti-pan HLA class I α-chain molecules (clone EMR8-5, MBL) and β-2-microglobulin (β2M) antibody (PROTEINTECH) in advance. High affinity confirmed W6/32 antibody (800 ug) was then cross linked onto 200 ul of slurry of Protein G Sepharose (GE) by DMP (20 mM) in HEPES (pH 8.4) for 1 hour. After quenching by TBS-T for 1 hour, beads were used for IP or stored at 4 °C before use.

Immunoprecipitation of HLA-Class I Complex and immunopeptide purification for immunopeptidomics

Desired cell pellet was lysed by 1 ml of lysis buffer containing [20mM] HEPES, [150mM] NaCl, 1 % NP-40, 0.1% SDS and 10 % Glycerol on ice. For tissue samples, disposable homogenizer Bio Masher System (nippi) was used firstly with 500 ul of lysis buffer for approximately 40 mg of tissue on ice and then add another 500 ul of lysis buffer to make sample volume as 1 ml. Both cells and tissues samples were cleared by centrifugation at 4oC, 15,000 rpm for 15min, the supernatant was collected and conducted immunoprecipitation for 4oC overnight on rotating rotor. For thorough washing, Bio Spin Column (BioRad) was used afterwards. The washing steps were as follows: 1 ml of lysis buffer for 5 times, 1ml of PBS-T for 5 times, 1ml of PBS for 5 times, 1ml of High-salt PBS for twice, 1ml of PBS for 5 times and then 1ml of mass spectrometry-grade water for twice. The trimer of HLA complex was dissociated and eluted by 500 ul of 1 % TFA from PI-beads. Obtained HLA enriched eluate was further processed by tC18 SepPak. For class I HLA peptide, 500ul of 20% acetonitrile (ACN) was firstly used for elution. HLA peptide including eluates were then dried up by evaporator. To elute the other components of HLA complex to confirm the IP-efficiency, 500 ul of 80 % ACN was used for second elution and the obtained eluates were dried up by evaporator as well. Sample purity was assessed by silver staining (Silver Quest).

Culture Cells and the construction of WES-based customized database

A colon cancer cell line, HCT116, was maintained in 10% fetal bovine serum including RPMI1640 in humidified atmosphere at 37 ℃ with 5% CO_2. Customized database for HCT116 was stablished from the WES data deposited in COSMIC Cancer Cell Line Project. Listed somatic mutations were processed by in-house built pipeline previously reported [11, 12]. Obtained mutations were then combined with Swiss-Prot human proteome database (20,181 entries) from UniProt website and used for neoantigen search.

HLA typing, whole exome sequencing and pipe lines for tailored protein database

For clinical tissue samples, genomic DNA was extracted by QIAmp DNA mini Kit (QIAGEN) from normal and tumor tissues. To establish the personalized database for neoantigen identification by immunopeptodomics, whole exome sequencing (WES) was conducted to all samples. In-house pipe line to extract the somatic mutations has been previously reported [12, 13]. Briefly, the genome libraries for WES were established from the DNA samples using xGen Exome Research Panel (IDT, Coralville, IA), according to the manufacturer's instructions. The exons were sequenced as 150 base pairs paired-end reads by a NovaSeq6000 system (Illumina, San Diego, CA). The sequence data obtained were then analyzed to select possible germ line variations and somatic mutations. In short, sequences of the whole exome were compared with a reference human genome (hs37d5), and the possible germ line variants and the somatic mutations were identified. The variants in public databases (>1%) were also excluded. Then cancer somatic mutations were extracted by subtraction of germ line variation identified in normal tissue from tumor tissue respectively. Among those mutations, only the missense mutations were selected and translated into full length of protein and added to the same Swissprot/ Uniprot human database described above as personalized database. The established personalized data base was used for immunopeptidomics both normal and tumor tissues corresponding patient's samples. The OptiType was employed for 4-digit

187 HLA typing from deep sequencing data [14]. The genetic details including mutation burden,
188 oncogenic KRAS mutations and the HLA genotypes of clinical samples can be found in
189 supplementary information (S Table. 1B)
190
191 **Differential ion mobility (DIM) Mass Spectrometry (MS) by High-Field**
192 **Asymmetric-waveform Ion Mobility Spectrometry (FAIMS) for Advanced**
193 **Immunopeptidomics**
194 The sample includes immunopeptides was first trapped by precolumn (C18 Accleim
195 PepMap 100 C18 Trap Cartridge, Thermo Scientific) and then separated by analytical
196 column (Aurora UHPLC Column (C18, 25 cm x 75 ul, 1.6 um FSC, ESI, ion opticks)
197 coupled with a nanospray Flex ion source for electrospray ionization (Thermo Scientific),
198 were used through an Ultimate 3000 RSLC nano HPLC system at a flow rate of 200 nl/
199 min using a linear gradient starting from 2 % solvent B (0.1% Formic acid in Acetnitrile) to
200 28 over 55min, a 1 min hold at 95% Solvent B a prior to a 1 min analytical column
201 equilibration with 2% B. In order to validate the benefit of FAIMS applied DIM-MS in
202 immunopeptidomics, liquid chromatography tandem mass spectrometry (LC-MS/MS)
203 and LC-FAIMS-MS/MS conditions were firstly compared. For LC-FAIMS-MS/MS, the
204 FAIMS-Pro interface (Thermo Fisher Scientific) was installed onto an Orbitrap Fusion
205 Lumos Tribrid mass spectrometer and operated with default parameters except for the
206 compensation voltage (CV) settings for gas-phase fractionation. For LC-MS/MS, we
207 uninstalled FAIMS device from mass spectrometer and performed analyses. Under
208 LC-FAIMS-MS/MS condition, sample was seamlessly fractioned by 3 CVs per single
209 analysis. Total 3 CV sets were used per sample and each CV set includes as follows;
210 set1 (CV= -40, -60, -80 V), set 2 (CV= -35, -50, -70 V) and set 3 (CV= -45, -55, -65 V).
211 Therefore, total 3 analyses (raw data), were acquired per sample. The parameters for
212 mass spectrometry were optimized for immunopeptidomics with small amount of sample.
213 Full MS (rage from 320 to 850 m/z) in the Orbitrap were acquired at resolution of 60 k

followed by a MS^2 acquisition at resolution of 15k. Maximum injection time for full MS scan was 50 ms with an auto gain control (AGC) of $4e^5$. Maximum injection time for MS^2 scan was 100ms with an AGC $1e^4$ followed by top speed MS^2 acquisition by the ion trap. Charge state 2 and 3 were selected for fragmentation by collision-induced dissociation (CID) at rapid scan rate at 30% collision energy. The easy-IC system was used for the internal calibration lock mass during the data acquisition. The HCT116 samples were prepared from $1e^8$ cells and proportionally injected into MS to match the indicated number of cells. For the analyses of $1e^8$ of HCT116 cells and clinical tissue samples, firstly 1/ 20 amount of sample volume was used for analyses of 3 CV sets respectively to check sample conditions, and then from the remaining, 5/ 20 amount of sample volume was used for 3 CV sets respectively. Therefore, total 6 analyses were performed per clinical tissue sample.

The LC/MS raw data and summarized result files are deposited at a public proteomic database, Japan Proteome Standard Repository/Database (jPOST) as follows; HCT116 immunopeptides without FAIMS in JPST001072, HCT116 immunopeptides with FAIMS in JPST001066, Global identification of HCT116 immunopeptides in JPST001068, immunopeptides from normal regions of CRC tissues in JPST001070, and immunopeptides from tumor regions of CRC tissues in JPST001069.

Data analysis by tailored database search for immunopeptidome

Obtained raw files were processed by Proteome Discoverer version 2.4 using Sequest HT engine against corresponding tailored database. The search was set to no-enzyme (unspecific) for the enzymatic designation. The precursor mass tolerance was set to 10 ppm and the fragment mass tolerance to 0.6 Da. The maximum number of missed cleavages was set to zero. Methionine oxidation and cysteine carbamidomethylation were included as dynamic modifications. A false discovery rate (FDR) of 0.01 (Strict) was

used for percolator algorithm to identify highly confident peptides. The output “peptide groups” was exported as a list. Then the peptide sequence was deduplicated for unique peptide sequences only and filtered by the amino acid length from 8 to 15 as a candidate of immunopeptides. To narrow candidate sequences down to the reliable ones, NetMHCpan 4.0[15] was employed to exclude the no-binding peptides from the list. Simultaneously, unsupervised clustering for HLA sequence motifs, Gibbs Cluster 2.0 [16] and Seq2 logo [17] were employed to confirm whether the identified motifs match HLA allotypes accordingly.

Profiling of Immunopeptides in Normal and Tumor Tissues

In order to interrogate the possible distinctions of normal and tumor-derived immunopeptides, the intersection of immunopeptides between normal and tumor tissues was calculated and draw Venn diagrams by publicly available website (http://bioinformatics.psb.ugent.be/webtools/Venn/). Then classified the immnopeptides into 3 groups; (1) normal-exclusive, (2) shared in both normal and tumor tissues (as “shared-peptide”) and (3) tumor-exclusive ones per patient. The frequency of amino acid usage at position Ω (pΩ), the end position at C-terminus of immunopeptide, was calculated as percentage against the total number of immunopeptide in each group. Based on the frequency of shared-peptide’s amino acid usage at pΩ, the distinction of frequency was calculated by comparing the value between normal-exclusive and shared-peptide or tumor-exclusive and shared-peptide.

Statistical Analysis.

All data were expressed as mean ± standard error (SE). Student’s t-test was used for statistical analyses. Pearson correlation-coefficient was used to assess the correlation. Statistical significance thresholds were determined at $P< 0.05$ (*), $P< 0.01$ (**), $P< 0.001$ (***) values. Pearson correlation-coefficient was used to assess the correlation.

268

269

Results

Differential ion mobility mass spectrometry expands depth of immunopeptidomics

The effectiveness of differential ion mobility mass spectrometry (DIM-MS) in immunopeptidomics was validated by comparing with or without FAIMS interface. For this step, a colon cancer cell line HCT116 was used to optimize the immunopeptidomics-associating parameters to maximize the number of identifications. Finally, 9 compensation voltages (CVs) between -80 to -35 V were selected for DIM and other conditions such as collision mode, detector type, maximum injection time, and scan speed were also fitted to maximize the identification of immunopeptides. Then the efficacy of identification was compared starting from relatively small scale, approximately $5e^6$, compared to the fact that the most of the previous immunopeptidomics prepared MS samples from at least ~e^8 order magnitude to yield a few to several thousands of immunopeptides. By FAIMS-assisted immunopeptidomics, 3215 ± 235 immunopeptides on average were identified from $5e^6$ cells, while 2482 ± 191 peptides were detected in the without-FAIMS condition (**Fig. 2A**, bar graph). Accordingly, the number of identified source protein was also greater in the FAIMS-assisted condition (2381 ± 108 proteins) than the without-FAIMS condition (2077 ± 114 proteins) (**Fig. 2A**, dashed line graph). This result was consistent throughout the experiments and became statistically significant when the number of cells was increased to $1e^7$ (3893 ± 166 immunopeptides) or $1.5e^7$ (4160 ± 163 immunopeptides, **Fig. 2A**). When it comes to neoantigen identification, an average of 3.7 ± 0.7, 3.7 ± 0.3, or 5.7 ± 0.3 neoantigens were identified from $5e^6$, $1e^7$, or $1.5e^7$ cells in the FAIMS-assisted condition, respectively (**Fig. 2B**). Thus, the DIM-based gas phase fractionation provided 27 to 38% enhancement of total identification of immunopeptides and 61 to 235% enhancement of neoantigen identifications. These results indicated the potential of our DIM-MS immunopeptidomics technology even with a small amount of samples that was an order of magnitude smaller than the most previous reports [5 8 6].

From the biologically triplicated analyses using $1e^8$ cells of HCT116, 6915.3 ± 316.3 immunopeptides derived from 4109.7 ± 137.8 source proteins were identified, which yielded 9249 unique immunopeptides and 5103 unique source proteins in total (**Fig. 2C**). Importantly, 7.3 ± 0.3 neoantigens were identified per independent experiments that yielded 11 neoantigens in total (**Fig. 2D**). Among these, 7 were not found in the immune epitope database (IEDB) (http://www.iedb.org/home_v3.php, as of November 27, 2020), hence regarded newly identified ones. And 2 out of 7 newly identified neoantigens, the wild type sequences were not found in IEDB (**Table 1**). The distribution of length and the assignment of immunopeptide against HLA allotypes exhibited typical and reasonable manner in the context of HLA genotype (**Fig. 2E and F**). These results indicated that the FAIMS-applied DIM-MS advanced immunopeptidomics has an advantage to potentiate the immunopeptidome including neoantigens from the scarce samples.

Personalized immunopeptidomic analysis of colorectal cancer tissues

Tumor or adjacent normal regions of CRC tissues (n = 17) were subjected to WES analysis. The identified nucleic acid sequences with somatic mutations were translated into amino acid sequences and utilized for construction of personalized proteome databases. As the result of individual immunopeptidomic analysis using these personalized databases, 44785 unique immunopeptides were identified (6578 peptides per patient on average) (**Fig. 3A**). Among these, 5603 out of 44785 (12.5%) was normal-exclusively identified ones and 14052 (31.4%) was tumor-exclusive, while 25130 (56.1%) was identified from both normal and tumor tissues. This exceeds the previously reported number of immunopeptides from solid tumor tissues such as colon[10], liver[9], and ovarian[8] cancers. For these immunopeptides, 11117 source proteins (4149 proteins per patient on average) were identified (**Fig. 3B**). Among these, 817 out of 11117 (9.7 %) was normal-exclusive ones and 1839 (16.5 %) was tumor-exclusive, while 8431 (75.8 %) was identified from both normal and tumor tissues. The individual datasets of

immunopeptidomic profiling can be found in **Supplementary figure 3**. The identified immunopeptides exhibited features of HLA-allotype corresponding features of unsupervised clustering and assignment against the binding motifs (representative data was shown in **Fig. 3C** and **3D**. Other data of individual peptidomes can be found in **Supplementary figure 4**). The length of immunopeptides were variable from patient to patient according to their HLA-allotypes, but the typical feature of HLA class I immunopeptide length was common in patients, while there was no statistically significant difference in between normal and tumor tissues. Indeed, immunopeptides with 9 amino acid length occupied 74.1 % (ranging 65.3 to 88.3%) of immunopeptides. The 99.7% (ranging 99.1 to 100%) of immunopeptides covered 8 to 12 amino acid in length reflecting the general profile of class I immunopeptide (**Fig. 3E**). Notably, 2 neoantigens were identified from 2 distinct patients' tumor tissues (**Fig. 4A**). The first neoantigen identified from the primary tumor tissue of ID 172 (KRAS, p.7-16) possessed the well-known CRC driver mutation, KRAS-G12V. The other one CPPED1 with R228Q (p.226-234), a mutant of serine/ threonine-protein phosphatase, was identified from ID 261's hepatic metastasized tumor tissue. According to the motif information at IEDB and SYFPEITYHI (http://www.syfpeithi.de/0-Home.htm), the neoantigen carrying KRAS-G12V was predicted to have higher affinity against A*11:01 than wild type sequence due to the preferred amino acid substitution of valine at position 6 within the epitope (**Fig. 4A**, depicted as "Preferred: V" by arrow). On the other hand, the mutation CPPED1-R228Q which lost its deleterious amino acid of arginine at position 3 by substitution to glutamine (**Fig. 4A**, depicted as "Deleterious: P/R" by arrow). The identified neoantigens exhibited of two opposite patterns of gaining affinity against binding motifs, gaining preferred residue or losing deleterious residue within epitope, still, both neoantigens were predicted to have higher affinity against A*11:01 than those of wild type sequences. (KRAS-WT vs KRAS-G12V; 299.7 nM vs 137.3 nM, CPPED1-WT vs CPPED1-R228Q; 729.9 nM vs 52.2 nM). It has been a long-standing challenge to

identify neoantigens directly from tissue samples, but our personalized immunopeptidomic analysis enabled to overcome that task.

Tumor-specific usage of C-terminus amino acids in CRC HLAps

To verify this black-boxed physiology, we separated immunopeptides into three groups; immunopeptides only found in normal tissue (normal-exclusive), immunopeptides found in both normal and tumor tissues (shared), and immunopeptides only found in tumor tissue (tumor-exclusive) per patient. This classification concentrates the each immunopeptides into respective groups if there would be a processing or/and presentation bias existed between normal and tumor state. When we checked the overall usage of amino acid for all 44785 immunopeptides, 12 out of 20 amino acids covered 99.2% of C-terminus usage and among these, the lowest frequency was cysteine (0.5%) (**Fig. 5A**). More frequency was observed typical tryptic- (R, K) and chymotryptic- (L, I, V, F, Y, W, A, M) amino acids for immunopeptide trimming at the end of C-terminus (pΩ). Next, we counted the frequency of amino acid usage at pΩ of the immunopeptides per patients. The usage of amino acid frequency at pΩ was calculated as ratio (%) against subtotal of each groups (normal-exclusive, shared, and tumor-exclusive peptide, respectively) and compared the difference against based on the frequency of the shared-peptide to clarify whether there would be a processing bias shown or not between normal-exclusive and tumor-exclusive groups. Among the 12 pΩ used amino acids, there are 9 patients who carries at least one R-restraint HLA allotypes (A*11:01 and A*33:03) at pΩ, and the usage of R at pΩ exhibited the tendency of reduction in tumor-exclusive immunopeptide (normal- exclusive vs tumor- exclusive: -0.06 % vs -5.10 % compared to the shared-peptide frequency, p = 0.0975) but that was not statistically significant (**Fig. 5B**, [R]). While the usage of K exhibited no statistically noteworthy difference compared to the frequency of shared-peptide (data not shown). The R and the K are the same tryptic amino acids while exhibited distinct dynamics in

normal and tumor tissue. For chymotryptic peptides, only the W exhibited the statistically significantly increased in tumor-exclusive group than normal-exclusive group compare to the shared-peptide (**Fig. 5B**, [W]). The trimming at pΩ of immunopeptides by L, I and V, generally grouped as similar hydrophobic amino acids, exhibited inconsistent and non-significant results between normal and tumor-exclusive immunopeptides (data not shown). Since the usages of these three hydrophobic residues are naturally abundant among the immunopeptides, it is expected that statistical significance was diminished and was not shown as major differences when compared at the admixture condition of tissue. Thirteen out of 17 patients carry at least one pΩ W-restraint HLA allotypes (A*24:02, B*4402 and B*44:03) and the usage of W at pΩ was -0.45 % in normal-exclusive immunopeptide and +0.8 % in tumor-exclusive immunopeptides compare to the shared-peptide group (*p-value* = 0.0185) (**Fig. 5B**, [W]). There was a negative correlation in W usage at pΩ between normal-exclusive and tumor-exclusive groups indicated the shift of immunopeptide processing at pΩ in patient (r = -0.788, p = 0.0014) (**Fig. 5C**). In addition to these two amino acids, the cysteine usage at C-terminus was found to be increased in tumor-exclusive immunopeptides (+ 0.25 %) than that of normal-exclusive immuopeptide (0.00 %) compared to the shared-peptide group (p = 0.016) (**Fig. 5B**, [C]). Since there was no clear information available for the binding motif that requires cysteine at pΩ, all 17 samples are included for this calculation. It is of note that this distinct cysteine usage between normal and tumor-exclusive peptide may imply the unknown mechanisms of immunopeptide processing and of value to investigate by analyzing more number and variable tissue samples.

Discussion

In this study, advanced immunopeptidomics by FAIMS-applied DIM-MS which enabled direct identification from scarce sample was established. The very proposition of this study was to challenge and overcome the technological limitations of current immunopeptidomics. From this aspect, the methodology was thoroughly optimized for simultaneous gas-phase fractionations to analyze the immunopeptides from limited samples such as biopsy-sized clinical tissues.

In the past, massive amount of sample was necessary for MS-based immunopeptidomics especially from tissues [6, 8, 9]. Otherwise, as an alternative, patient-derived cancer cells were isolated for culture expansion to secure the enough number of cancer cells [18, 19]. The purified culture of patient-derived cancer cells is convenient for in-depth analyses of immunopeptidomics, however, it may not be denied that the antigen presentation can be modified during the isolated long-time cultivation due to the lack of cancer microenvironment. To mimic more tissue-like environment, the cancer organoids were previously reported as well. While establishing multiple organoids from every clinical samples may be too laborious and time-consuming task considering the fact that it takes at least a few months to establish. Another option to secure the cancer cell mass is patient-derived xenograft models while the response against ICI candidates may differ in host organisms of micronenvironment that previously shown in the discrepancy result of clinical trials of MEK inhibitor [20] with PD-L1 treatment from mouse xenograft models [21]. Instead of increasing the amount of sample amount, the database search for immunopeptides from non-canonical coding regions and the spliced immunopeptides have been also reported as another option to explore the variation of immunopeptides. The former, the gigantic database established from both exon-coding and non-coding sequences of individuals is used to identify immunopeptides[22]. The latter employs so called *de novo* sequencing to identify immunopeptides directly from the MS spectra without referring database [7, 23]. By these methodologies, the size of

immunopeptidome was indeed expanded and the new insights of the source and the processing of immunopeptides were given, while the hardships remain in terms of the accessibility to these methods and the versatility of identified neoantigens. The necessity to establish the gigantic database for non-canonical sequence by both genomics (from DNA) and transcriptomics (from RNA) individually may become another burden and that requires again more sample amount. For the spliced immunopeptides, it seems more likely to be by-chance product at that moment of processing, compared to that derived from somatic mutations of exonic origin. Therefore, it may be more difficult to find the shared antigens among patients. Although these approaches have its own advantages in biology of immunopeptides, these both non-canonical derived- and spliced immunopeptides have one common concern that insufficient information about the individual differences. Investigating the neoantigens based on the somatic mutations encoded within exon sequences from well-established cancer genetics [24, 25] has more expectations to become profit for clinical practice, i.e., the identification of shared neoantigens from known pan-cancer oncogenic mutations for versatile cancer immunotherapy. And the close scrutiny of actually presented neoantigens across tumor tissues will enlighten us the optimal epitopes for cancer vaccination as well as to relieve the labor of screening the repertoire of autologous/ engineered T cells for adoptive therapy from the validation of costful numerous number of synthesized candidate peptides and tandem mini genes [26].

Here, advanced immunopeptidomics by FAIMS-applied DIM-MS paved the way to make the identification of approximately 5000 (4917.9 on average/ sample) of immunopeptides possible from 40 mg of tissue. In the sample preparation for MS analysis, the chemical pre-fractionations based on the chemical properties such as hydrophobicity and the ionic bonding preference, are inevitable and routinely employed to expand the depth of MS analysis. They surely expand the depth of analysis while irresistibly requires more sample amount. It has also been reported that the pre-fraction creates a bias in

identifiable immunopeptides in the context of distinct chemical properties of immunopeptides governed by the diverse binding motifs of each HLA-allotypes[27]. This suggested that as long as the chemical pre-fractionation employed, we never cover both chemical properties and there left certain immunopeptides abandoned. There are still uncertainties remain in the consensus sequence of binding motif in each HLA-allotypes and the combination of HLA-allotype in individuals is complex. Under these circumstances, it is particularly difficult to determine in advance which chemical pre-fractionation would be the best option for diverse samples. By FAIMS, the peptides are fractionated by CV which uniquely defined by the structural bulkiness of the peptide [28]. So, every peptide has distinct ideal CV to be identified. Indeed, during the optimization, we noticed the different preference of CVs in distinct cell lines for immunopeptidomics. For example, TE-11, an esophagus cancer cells, exhibited more weak CVs from -60v to -35v preferred distribution compared to the HCT116 which exhibited rather broad CV distribution from -80v to -35v (data not shown). Although the back and forth difference of preferred CVs can be seen accordingly to each HLA-allotypes, the majority of immunopeptides were mainly identified from -60v to -40v of CV. So, the multi-CV with rapid scanning covers the wide and sufficient range of ideal CVs that led to robust immunopeptidome regardless of diverse CV preferences. Thus, the best virtue of DIM-MS with seamless gas-phase fractionation is the feasibility of broad identification of immunopeptides independent from the elusive chemical properties and that is also ideal for the unbiased analysis of scarce samples. The progress in transition time for CV stepping (25 ms/ transition) [29] enabled multi-experiments of 3 CVs per single analysis. The ratio from the number of acquired MS/MS spectra to peptide spectrum match (PSM) was 23.1±1.1% on average indicated the efficacious data acquisition by advanced immunopeptidomics of global immunopeptide identification.

During the analyses of HCT116, on average, we identified more than 7 neoantigens from $1e^8$ scale of HCT116 and three independent experiments resulted in total 11

neoantigen identification. Five of those were identified in all three independent experiments constantly and 4 were less constant. Although there are multiple possibilities that defines the fluctuation of neoantigen identification, e.g., particularly low abundance of source protein of less constant neoantigens compare to the constant ones, unraveling the processing and the presentation mechanisms of constant and not constant neoantigens may be interesting from the aspect of how to boost the neoantigen presentation to reinforce the efficacy of ICI treatment.

Not only the neoantigen processing, but the global immunopeptides can be affected by the external conditions such as stimulation by cytokines. The IFN-γ induces so called immune-proteasomal degradation intracellularly to produce different pattern of C-terminus trimming of immunopeptides compare to the constitutive-proteasomal degradation. More recently, this immunoproteasome expression has shown to associate better response of ICI treatment in melanoma [30]. In the melanoma cells, the population of immunopeptides which significantly increased by IFN-γ stimulation exhibited more potential to activate cytotoxic T cells than stably presented immunopeptides. While in CRC-organoids, although the IFN-γ treatment drastically increased the amount of HLA complex on cellular surface, the variation of immunopeptides were not commensurately expanded[18]. In this study, the number of identified immunopeptides seemed apparently abundant in tumor tissues than normal tissues while there was no statistically significant difference between the number of identified immunopeptide when normalized by the protein amount (**Supplementary Figure 1**). Not only the number of immunopeptides, it has previously shown that the variation of immunopeptides could be modified by proteasomal conditions [6, 31]. Recently, it is getting more clearer that the proteasomal dynamics affects immunopeptidome largely and thus the population of immunopeptides could be shifted accordingly. To date, there are three distinct proteasomes are reported and in the field of oncology, the inflammatory proteasome got an attention because of its association with interferon gamma (IFN-γ). The proteasome subunits alter its

combination under inflammatory conditions such as IFN-γ stimulation which is often associated with cancer lesions. The combination of proteasome induced by IFN-γ, called immunoproteasome, produces distinct property of immunopeptides because of its altered catalytic subunit from caspase-like catalytic subunit to chymotryptic-like catalytic ones. Therefore, the more chymotryptic-trimmed immunopeptides are produced and presented under IFN-γ stimulation. And the expression status of immunoperoteasome has been reported as a prediction marker for better ICIs response in melanoma. [30] The immunopeptidome with or without IFN-γ stimulation by various types of cancer cell lines have been previously reported, while there was no report of direct comparison of immunopeptidome from tissue level to clarify how much the immunopeptide presentation could be differ under cancer microenvironment. Further classification of immunopeptides into normal or tumor- exclusive as well as shared population was revealed in identical patients, the shift of immunopeptides population but was not that clear as previously reported in *in vitro* samples. By *in vitro* samples, more chymotryptic-type trimming will be increased by IFN-γ condition assuming an inflammatory condition in tissue. In tissue-based immunopeptidome, unlike the *in vitro*-based analyses, only the tryptophan (W) exhibited the significant increase in tumor- exclusive immunopeptides. Arginine and lysine both were in the group of trypsin-cleaved amino acids and often used at the same residues, only the arginine exhibited the tendency of decrease. Intriguingly, the cysteine-end immunopeptides were statistically increased in tumor-exclusive immunopeptides. As such, the visibility of the shift of immunopeptide by C-terminus trimming is different between the cellular and tissue samples. This may because, at least in part, the admixture condition of cancer cells in tissues. The meanings should be validated and the mechanisms behind would further be elucidated for the unknown biology of immunopeptide processing.

The scarce presentation of neoantigens in solid tumors including colon and liver cancers compare to the ICI-responsive cancers have been previously suggested [9, 10].

Actually, even by the in-depth immunopeptidome from 5 patient-derived organoids, a highly purified cancer cell mass, of CRC identified only 3 neoantigens [18]. For that reasons, the number of neoantigens with cancer driver mutation are believed to be extremely rare because the number of neoantigens could be further depleted as a result of cancer immunity response that eradicates neoantigen presenting cancer cells from tissues [32, 33] with disagreement [34]. Therefore, the consensus is not clear yet but not only in the CRC, identification of the neoantigen with major cancer driver mutation has been believed to be unlikely. Apparently, the number of neoantigens presented in CRC did not seem to be so abundant in this study, while we successfully identified 2 neoantigens directly from tumor tissues from stage IV CRCs. It is of worthy to investigate the early stage of CRC whether it presents more number of neoantigens. The identification of KRAS-G12V is noteworthy because this may overturn the myth of depleted neoantigen with cancer driver mutation. Based on the fact that the previous report of clinical trials of cancer immunotherapy which validated the efficacy of KRAS-G12D reactive autologous T cells restrained by HLA-A*11:01 in lung metastasized CRC[35], the identification of immunopeptides that carry KRAS-G12V and other oncogenic KRAS mutations from various cancers will be anticipated to develop the KRAS-targeted ICI. And the CPPED1 has been reported as cancer associated molecule. The CPPED1 that dephosphorylates AKT family kinase and reported to have the function of tumor suppression by inhibiting IL-6 expression through the axis of Akt/FoxO3 signal pathway that downregulates STAT3 in bladder cancer cells[36, 37]. And the Downregulation of CPPED1 in human papilloma virus (HPV)-associated cancers. [38] Further investigation of immunopeptidome in tissues of pathologically early stages or even in the pre-cancerous state will debunk the current controversial situation of depletion in driver gene-derived neoantigens and those contributions in cancer immunity.

Unlike the ICI-responsive cancers, it's been getting gradually clearer that for some reasons, neoantigen presentation is suppressed in CRC. The microsatellite stable (MSS)

CRC receives less profit by ICI treatments compare to the microsatellite instable (MSI) ones [1, 2]. And together, it has been reported that more mutation burden associates with the better outcome by ICI treatment [39]. This is based on the notion/ theory that more mutation burden evokes more possibility of neoantigen presentation. In this study, on the contrary, no neoantigen was identified in the tumor sample of ID 119 carrying the highest mutation burden (1752) with total 6983 immunopeptides. While the neoantigens were identified from the tumor sample of ID 172 carrying 123 missense mutations with total 5661 immunopeptides and the ID 261 carrying less abundant mutation burden (69 missense mutations) with 7437 immunopeptides. Even the ID 119 which carries equivalent mutation burden and the sufficient depth of immunopeptidome as HCT116, the neoantigen identification was not available so far. To reinvigorate the cancer immunity in the ICI non-responsive cancers like MSS-CRC, the agents that potentially boost the antigen presentation and TCR activation are of great interest for next generation of ICIs.

Currently, the STING agonist, one of the candidates that invigorates the cancer immunity, are now in the phase 3 clinical trials [40]. The STING agonist activates IRF3 to upregulate the type I interferon and cytokines that boosts cancer immunity via the axis of CD4+ T cell activation that governed by HLA class II immunopeptides[41]. On the other hand, the oncogenic KRAS (KRAS-G12D) has been recently reported to suppress IRF2 which is indispensable for IFN-γ secretion [42]. The IRF2 thus activate the axis of CD8+ T cell activation governed by class I immunopeptides. The physiological mechanism(s) how KRAS-G12D possibly down-regulates IRF2 is still unclear, while the cancer patients with this is oncogenic mutation exhibited statistically significantly lower expression of IRF2 together with negative prognosis in several cancers [42]. It is intriguing to evaluate the efficacy of STING agonists against oncogenic KRAS carrying tumors from the view of immunopeptidomics so that can tell us how much and the balance of cell-mediated immunity and the humoral immunity play roles in cancer immunity. And this can be validated by immunopeptidomics from tissue samples as one clear solution.

591 The FAIMS-applied advanced immunopeptidomics hold promise to search the shared
592 neoantigens as well as the to validate the efficacy of agents for cancer immunotherapy by
593 directly assessing the actually presented immunopeptides.
594

595 **Acknowledgements**

596 This work was supported by “the Development of Technology for Patient Stratification
597 Biomarker Discovery” of the Japan Agency for Medical Research and Development
598 (20ae0101074s0302) and “Grant-in-Aid for Scientific Research (C)” of Japan Society for
599 the Promotion of Science (20K05759).

600

Figure Legends

Figure 1. Schematic diagram of personalized immunopeptidome analysis assisted by differential ion mobility mass spectrometry (DIM-MS)

Advanced immunopeptidomics optimized for the analyses from small amount of cellular or tissue samples together with the tailored database established from whole exome sequencing for neoantigen identification. The FAIMS-Pro interface was coupled to mass spectrometer for seamless gas-phase fractionation to enable the DIM-MS.

Figure 2. Validation of advanced immunopeptidomics and neoantigen identification from HCT116 cells

A. Composite graph of the number of identified immunopeptide (HLAp) in bar graph (white bars = without FAIMS, black bars = with FAIMS) and the number of source proteins in dashed line graph. The index showed the approximate number of cells injected for the single MS analysis. * $p < 0.05$, # $p < 0.1$ B. The number of neoantigens identified from the same experiments of A. White bars= without FAIMS, black bars= with FIAMS. * $p < 0.05$, ** $p < 0.01$ C. Composite graph of the number of identified immunopeptide (HLAp) from three independent analyses approximately from $1e^8$ of cells. White bars = the number of immunopeptides from respective analyses, gray bar= an average number of immunopeptides from 3 independent analyses, dark gray bar = total number of identified immunopeptides. The number of source proteins were depicted in dashed line graph. D. The number of neoantigens identified from the same experiments as C. E. Length of immunopeptides identified. The length of immunopepties identified in C was counted. F. Representative image of unsupervised HLA motif clustering and the population of HLA assignment against genotype-matched HLA allotypes.

Figure 3. Results of personalized immunopeptidome analysis for CRC tissue samples

The results of immunopeptidome from clinical tissue samples (total 34 samples) of colon cancer (n=17) together with adjacent normal tissues (n=17). A. An average of total number of immunopeptides identified per patient (HLAp/ Patient) and the total number of unique immunopeptides from all patients (Total Unique HLAp) were displayed in bar graphs with respective numerals. B. An average of total number of source proteins identified per patient (Protein/ Patient) and the total number of unique immunopeptides from all patients (Total Unique Protein) were displayed in bar graphs with respective numerals. C. Representative image of the information of patient-specific HLA-allotypes and unsupervised HLA motif clustering by GibbsCluster2.0 for corresponding normal and tumor immunopeptidome. Full information for 17 patients can be found in **Supplementary Table 1**. D. Representative pie chart of immunopeptide assignment for HLA allotypes according to patient specific HLA genotypes. E. Length distribution of immunopeptides shown in A.

Figure 4. Mass spectrometric identification of neoantigens from CRC tissue samples

A. Mass spectra of identified neoantigens from colon cancer tissues. From ID 172, the oncogenic KRAS-G12V carrying neoantigen was identified. Possible cause of enhanced affinity by mutation was depicted with motif sequence of A*11:01. The CPPED1-R228Q was identically shown. B. Table for the details of identified neoantigens. The binding motif and affinity was calculated by NetMHC4.0.

Figure 5. Distinct features of c-terminus trimming between normal or tumor immunopeptides in CRC tissues.

A. A pie chart depicts the frequency of amino acid usage at the C-terminus (pΩ) of immunopeptide. More than 99.5% of immunopeptide identified from colon cancer tissues were included. B. The difference of amino acid usage (%) at pΩ in normal-exclusive and

655 tumor-exclusive immunopeptides compared to the shared-immunopeptide. C. A plot
656 depicted the negative correlation of tryptophan (W) usage at pΩ in normal-exclusive and
657 tumor-exclusive immunopeptides.
658
659
660

Table 1. Neoantigens identified from HCT116 by immunopeptidomic analysis

Mutation	Neoantigen Sequence	Assigned HLA	Affinity [nM]	Detection	First Report (*) No WT in IEDB (#)
AGO2-H336Y	QEQK**Y**TYLP	B4501	980.7	1/3	*
CHMP7-A324T	QTDQMVFN**T**Y	A0101	33.48	3/3	
FNBP4-K318E	EEE**E**KGVAA	B4501	83.24	3/3	*
NAPA-A181V	K**V**IDIYEQV	A0201	50.07	3/3	*
NR1D1-G39D	YSDNSN**D**SF	A0101/ C0701	28.34/ 5.89	3/3	
GAPDH-I69T	AENGKLV**T**N	B4501	963.6	3/3	*, #
IQGAP-S1070T	VLEDKVL**T**V	A0201	35.1	1/3	*
RBBP7-N17D	EERVI**D**EEY	B1801	194.48	1/3	
NOTCH2-C41S	NEGM**S**VTY	B1801	6.2	2/3	*, #
PDP1-N379D	**D**EYTKFIPP	B1801	158.49	1/3	*
UQCRB-N88K	EEE**K**FYLEP	B4501	728.67	1/3	

The position of mutation is highlighted by underline. The affinity prediction by NetMHC4.0 was shown. The number of times identified out of 3 independent MS analyses was shown.

* Newly identified neoantigens.

There was no epitope information of the corresponding wild type sequence at IEDB.

References:

1. Le, D.T. et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* **372**, 2509-2520 (2015).
2. de Weger, V.A. et al. Clinical effects of adjuvant active specific immunotherapy differ between patients with microsatellite-stable and microsatellite-instable colon cancer. *Clin Cancer Res* **18**, 882-889 (2012).
3. Caron, E., Aebersold, R., Banaei-Esfahani, A., Chong, C. & Bassani-Sternberg, M. A Case for a Human Immuno-Peptidome Project Consortium. *Immunity* **47**, 203-208 (2017).
4. Lill, J.R. et al. Minimal Information About an Immuno-Peptidomics Experiment (MIAIPE). *Proteomics* **18**, e1800110 (2018).
5. Bassani-Sternberg, M. et al. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun* **7**, 13404 (2016).
6. Chong, C. et al. High-throughput and Sensitive Immunopeptidomics Platform Reveals Profound Interferongamma-Mediated Remodeling of the Human Leukocyte Antigen (HLA) Ligandome. *Mol Cell Proteomics* **17**, 533-548 (2018).
7. Liepe, J. et al. A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science* **354**, 354-358 (2016).
8. Schuster, H. et al. The immunopeptidomic landscape of ovarian carcinomas. *Proc Natl Acad Sci U S A* **114**, E9942-E9951 (2017).
9. Loffler, M.W. et al. Multi-omics discovery of exome-derived neoantigens in hepatocellular carcinoma. *Genome Med* **11**, 28 (2019).
10. Loffler, M.W. et al. Mapping the HLA Ligandome of Colorectal Cancer Reveals an Imprint of Malignant Cell Transformation. *Cancer Res* **78**, 4627-4641 (2018).
11. Kato, T. et al. Effective screening of T cells recognizing neoantigens and construction of T-cell receptor-engineered T cells. *Oncotarget* **9**, 11009-11019 (2018).
12. Kiyotani, K., Chan, H.T. & Nakamura, Y. Immunopharmacogenomics towards personalized cancer immunotherapy targeting neoantigens. *Cancer Sci* **109**, 542-549 (2018).
13. Choudhury, N.J. et al. Low T-cell Receptor Diversity, High Somatic Mutation Burden, and High Neoantigen Load as Predictors of Clinical Outcome in Muscle-invasive Bladder Cancer. *Eur Urol Focus* **2**, 445-452 (2016).
14. Szolek, A. et al. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* **30**, 3310-3316 (2014).
15. Jurtz, V. et al. NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol* **199**, 3360-3368 (2017).
16. Andreatta, M., Alvarez, B. & Nielsen, M. GibbsCluster: unsupervised clustering and alignment of peptide sequences. *Nucleic Acids Res* **45**, W458-W463 (2017).

17. Thomsen, M.C. & Nielsen, M. Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res* **40**, W281-287 (2012).
18. Newey, A. et al. Immunopeptidomics of colorectal cancer organoids reveals a sparse HLA class I neoantigen landscape and no increase in neoantigens with interferon or MEK-inhibitor treatment. *J Immunother Cancer* **7**, 309 (2019).
19. Demmers, L.C. et al. Single-cell derived tumor organoids display diversity in HLA class I peptide presentation. *Nat Commun* **11**, 5338 (2020).
20. Eng, C. et al. Atezolizumab with or without cobimetinib versus regorafenib in previously treated metastatic colorectal cancer (IMblaze370): a multicentre, open-label, phase 3, randomised, controlled trial. *Lancet Oncol* **20**, 849-861 (2019).
21. Ebert, P.J.R. et al. MAP Kinase Inhibition Promotes T Cell and Anti-tumor Activity in Combination with PD-L1 Checkpoint Blockade. *Immunity* **44**, 609-621 (2016).
22. Chong, C. et al. Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nat Commun* **11**, 1293 (2020).
23. Faridi, P. et al. A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands. *Sci Immunol* **3** (2018).
24. Cancer Genome Atlas, N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-337 (2012).
25. Dienstmann, R. et al. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat Rev Cancer* **17**, 79-92 (2017).
26. Tran, E. et al. Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science* **350**, 1387-1390 (2015).
27. Demmers, L.C., Heck, A.J.R. & Wu, W. Pre-fractionation Extends but also Creates a Bias in the Detectable HLA Class Iota Ligandome. *J Proteome Res* **18**, 1634-1643 (2019).
28. Purves, R.W. et al. Using gas modifiers to significantly improve sensitivity and selectivity in a cylindrical FAIMS device. *J Am Soc Mass Spectrom* **25**, 1274-1284 (2014).
29. Hebert, A.S. et al. Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Anal Chem* **90**, 9529-9537 (2018).
30. Kalaora, S. et al. Immunoproteasome expression is associated with better prognosis and response to checkpoint therapies in melanoma. *Nat Commun* **11**, 896 (2020).
31. Murata, S., Takahama, Y., Kasahara, M. & Tanaka, K. The immunoproteasome and thymoproteasome: functions, evolution and human disease. *Nat Immunol* **19**, 923-931 (2018).
32. Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48-61 (2015).
33. Davoli, T., Uno, H., Wooten, E.C. & Elledge, S.J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* **355** (2017).

34. Van den Eynden, J., Jimenez-Sanchez, A., Miller, M.L. & Larsson, E. Lack of detectable neoantigen depletion signals in the untreated cancer genome. *Nat Genet* **51**, 1741-1748 (2019).
35. Tran, E. et al. T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. *N Engl J Med* **375**, 2255-2262 (2016).
36. Zhuo, D.X. et al. CSTP1, a novel protein phosphatase, blocks cell cycle, promotes cell apoptosis, and suppresses tumor growth of bladder cancer by directly dephosphorylating Akt at Ser473 site. *PLoS One* **8**, e65679 (2013).
37. Zhuo, D., Wu, Y., Luo, J., Deng, L. & Niu, X. CSTP1 inhibits IL-6 expression through targeting Akt/FoxO3a signaling pathway in bladder cancer cells. *Exp Cell Res* **380**, 80-89 (2019).
38. Yang, R. et al. Combined Transcriptome and Proteome Analysis of Immortalized Human Keratinocytes Expressing Human Papillomavirus 16 (HPV16) Oncogenes Reveals Novel Key Factors and Networks in HPV-Induced Carcinogenesis. *mSphere* **4** (2019).
39. Hugo, W. et al. Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* **165**, 35-44 (2016).
40. Le Naour, J., Zitvogel, L., Galluzzi, L., Vacchelli, E. & Kroemer, G. Trial watch: STING agonists in cancer therapy. *Oncoimmunology* **9**, 1777624 (2020).
41. Du, H., Xu, T. & Cui, M. cGAS-STING signaling in cancer immunity and immunotherapy. *Biomed Pharmacother* **133**, 110972 (2020).
42. Liao, W. et al. KRAS-IRF2 Axis Drives Immune Suppression and Immune Therapy Resistance in Colorectal Cancer. *Cancer Cell* **35**, 559-572 e557 (2019).

Figure 1

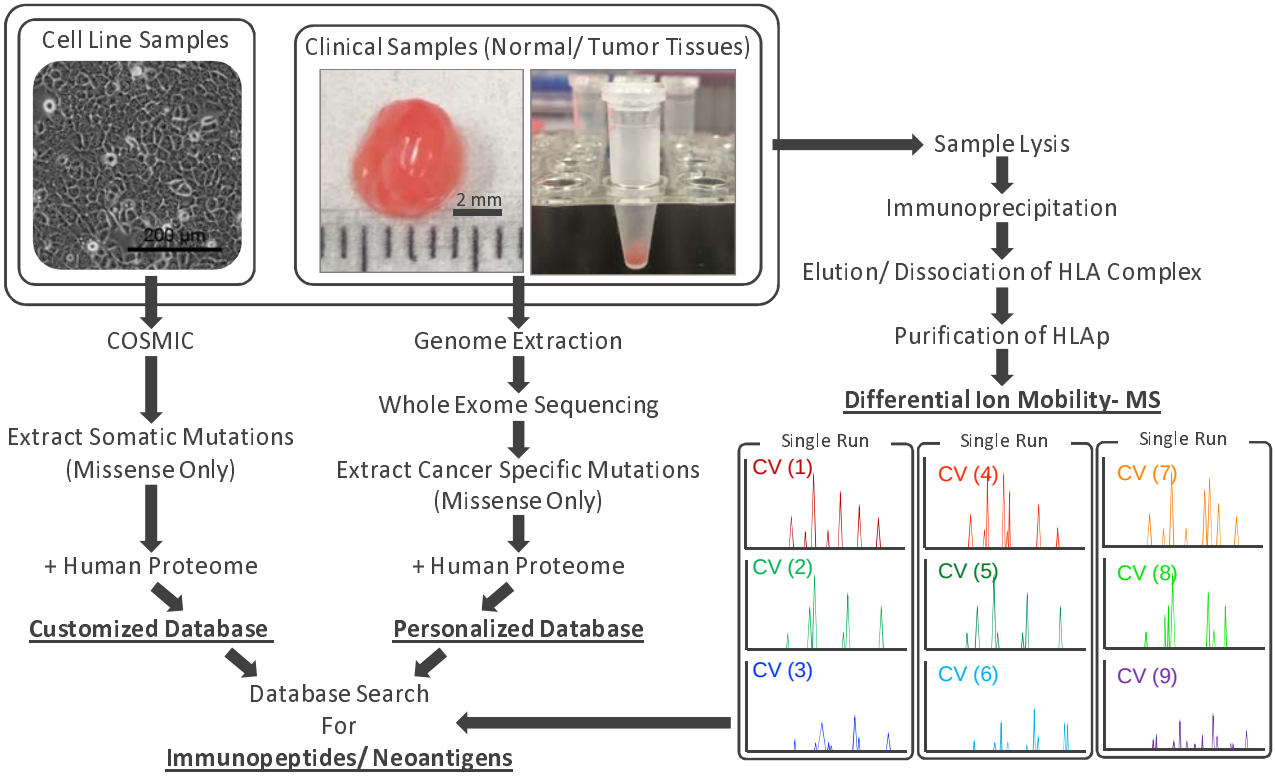


Figure 2

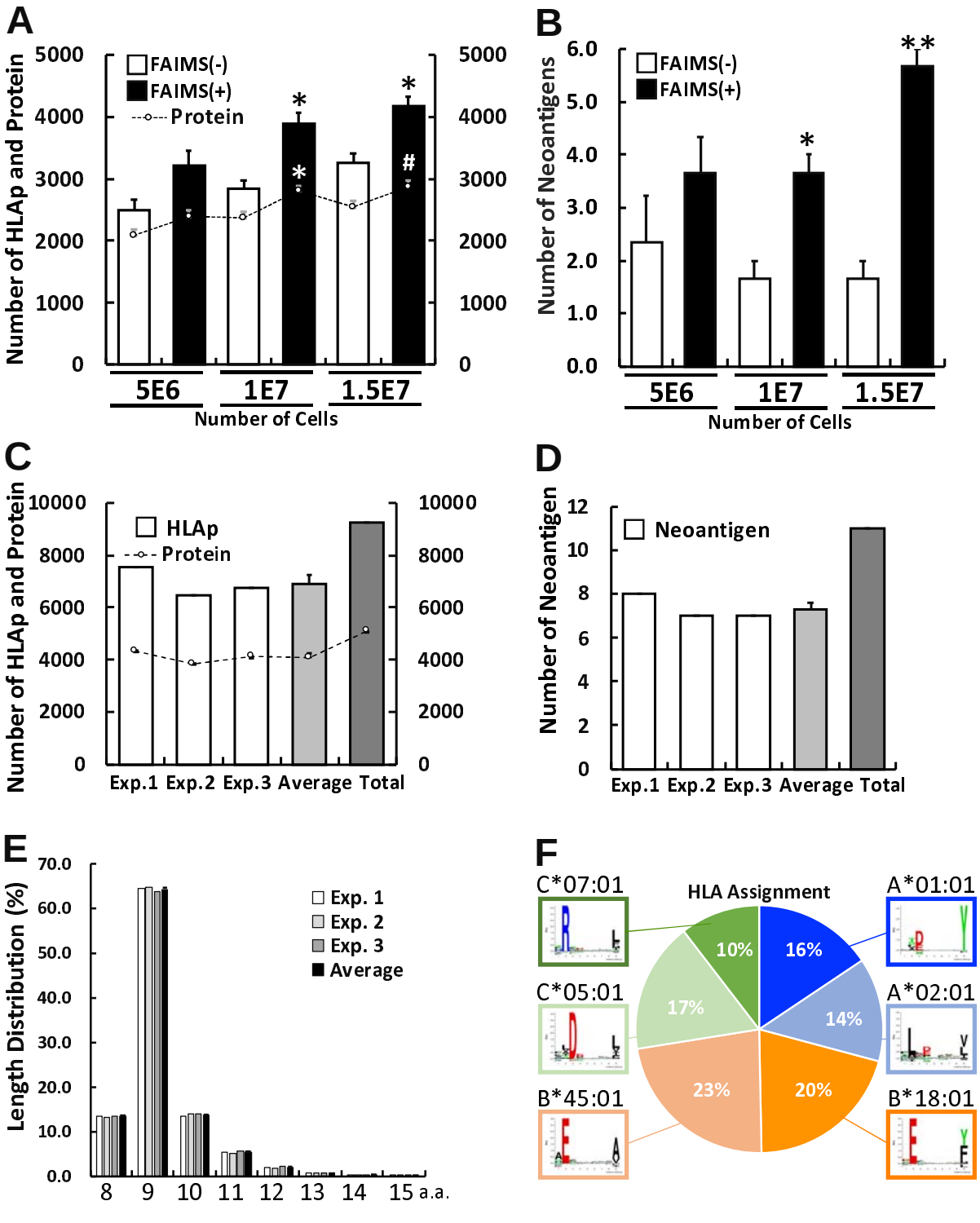


Figure 3

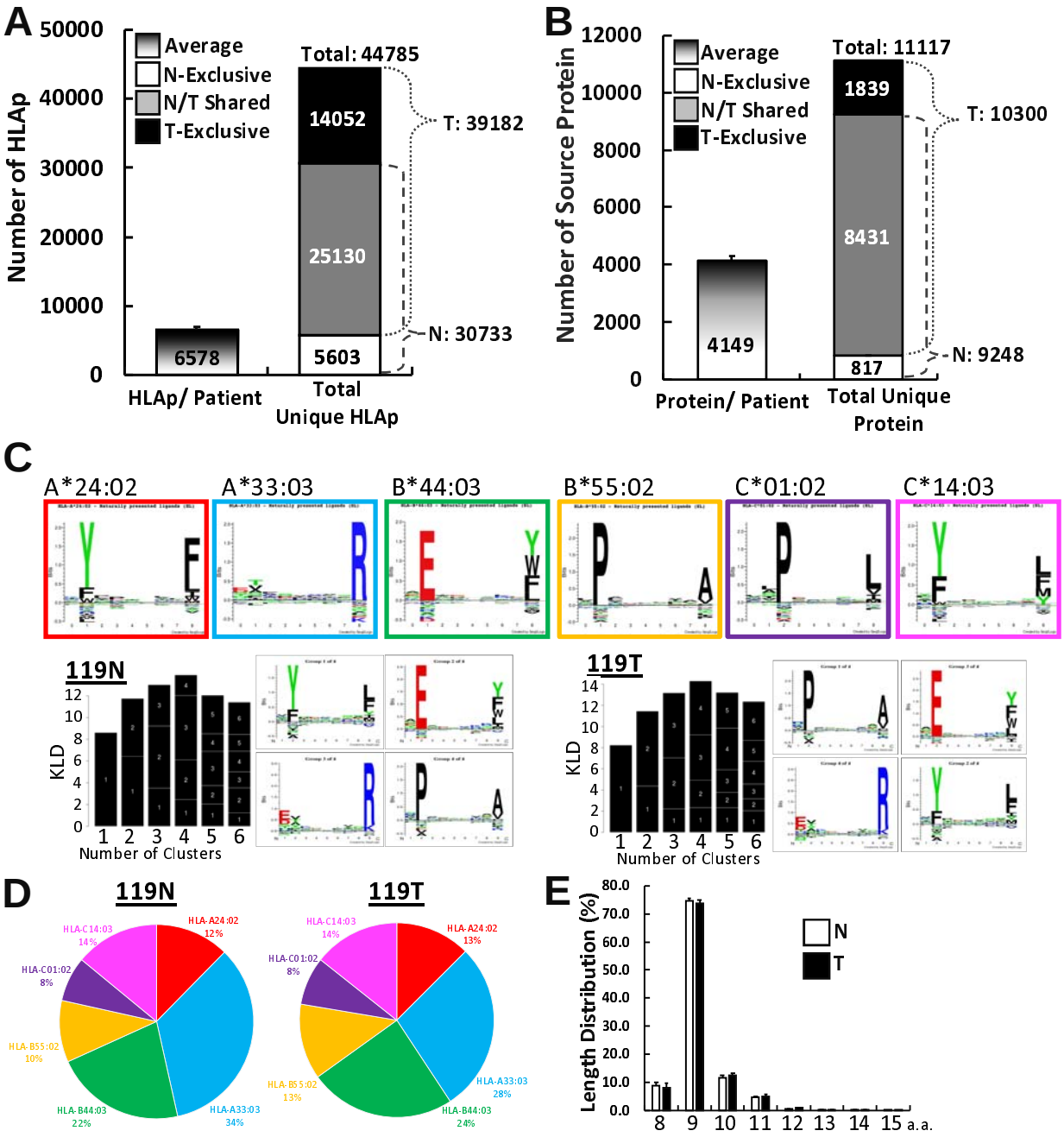
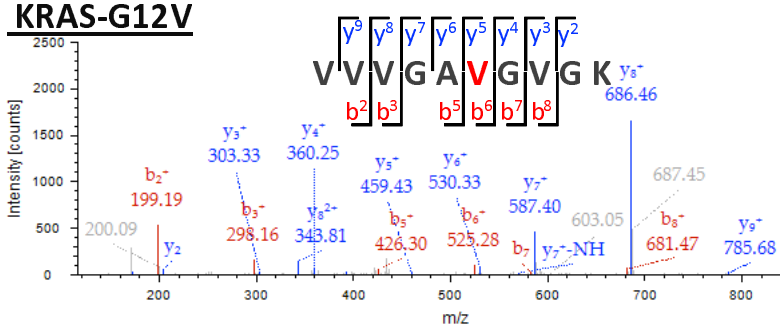
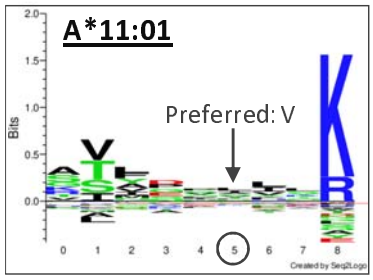


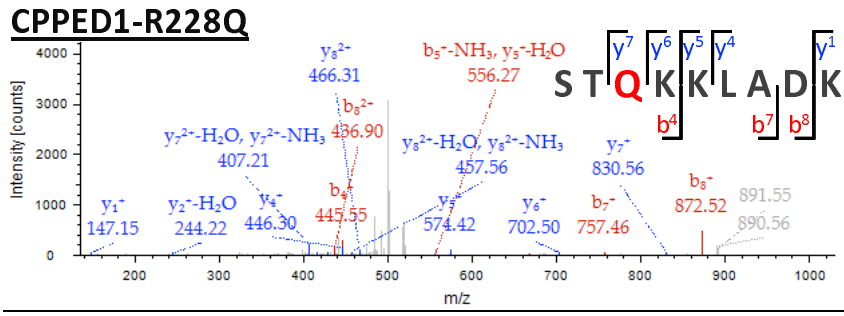
Figure 4

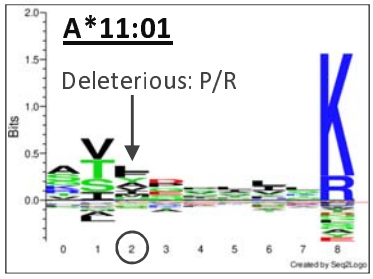
A **KRAS-G12V**





CPPED1-R228Q





B

Sample	Mutation	Position	Sequence	Assigned Allele	Affinity [nM]
172T	KRAS-G12V	[7-16]	VVVGAVGVGK	**A1101**	137.3
261HT	CPPED1-R228Q	[226-234]	STQKKLADK	**A1101**	52.2

Figure 5