Personalized immunopeptidomics using differential ion mobility technology enables direct profiling of neoantigens on colorectal cancer tissues

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

Main Text:

Knowing the nature of human leukocyte antigen (HLA)-presented peptides (HLAp) is indispensable to realize the precision medicine [1] [2] while direct identification of HLAp from scarce tissue specimens still confronts technical challenges [3]. Here we optimized the isolation procedures for HLAp and employed the technology of multiple gas phase fractionation differential ion mobility mass spectrometry (DIM-MS) to overcome that limitation. From 1e^{8} cells of HCT116, our method detected 9,249 unique HLAp including 11 peptides with missense mutations (neoantigens). Next, from individual differential immunopeptidome analysis using 40 mg each of colon cancer and adjacent normal tissues (n = 17), 44,785 HLAp including 2 neoantigens, KRAS$_{7\text{-}16}$(G12V) and CPPED1$_{226\text{-}234}$(R228Q), were identified. Furthermore, colon cancer-specific amino acid usage at C-terminus of HLAp was also found. Thus, this technology drastically improves the depth of immunopeptidomic analysis and allows direct determination of clinically important cancer antigens.

Neoantigen, a cancer specific somatic mutation-carrying HLAp, is presented on cellular surface as non-self antigen and activates the associated repertoire of T cells that will evoke anti-cancer immunity [1]. The immune checkpoint inhibitors (ICIs) have been established to invigorate this immune system as cancer immunotherapies [4]. Although immunotherapies are efficacious in patients with certain cancer types [4], there are still obstacles of less efficacy in cancers with less mutational burden such as microsatellite stable (MSS) colorectal cancer (CRC) compare to the microsatellite instable (MSI) CRC with DNA mismatch-repair deficiency [4, 5]. Validating and predicting the efficacy and choosing the right strategy for ICIs based on the background of individual patient are urgent tasks to grant the benefits by cancer precision immunotherapy. For that purpose, the direct scrutiny of HLAp on clinical tissue specimens could provide fundamental

53 information to facilitate them.

54 The immunopeptidomics by mass spectrometry (MS) is the only technology to directly
55 interrogate the HLAp. According to the guidelines for immunopeptidomics[3, 6], one of the
56 disadvantages of MS-based immunopeptidomics is incapability of robust analysis from a
57 scarce sample, such as tiny tissues of endoscopic biopsy, compared to the genomic
58 prediction of HLAp by *in-silico* [7]. To obtain the thousands scale of convincing
59 immunopeptidome with current technology, it generally requires at least $1e^8$ of cultured
60 cells or nearly 1000 μg of tissue sample [8-10] with inevitable chemical prefractionation and
61 this becomes hindrance for the analysis from scarce sample. Actually, direct detection of
62 neoantigens from tissue samples, such as colorectal, liver, or ovarian cancer, have been
63 previously unsuccessful [11, 12 6]. Since the variety of HLAp is dynamic under cancer-
64 associating microenvironments including interferon gamma (IFN-γ) stimulation,
65 comprehending the HLAp under the native pathological environment is crucial for further
66 progress in cancer immunotherapies. Therefore, it is essential to establish more efficient
67 technology that enables the decent immunopeptidomics analysis from scarce tissues.

68 For that purpose, we adopted the differential ion mobility mass spectrometry (DIM-MS)
69 equipped with the interface of high field symmetric wave form ion mobility mass
70 spectrometry (FAIMS). The seamless gas phase fractionation by FAIMS secured the
71 sufficient depth of immunopeptidome from the scarce samples. First, we fully optimized
72 the parameters for DIM-MS to maximize the identification number of HLAp. Especially, a
73 set of 9 compensation voltages (CVs), from -80 to -35 V (see **Methods**), were selected
74 for the best performance of gas phase fractionation of HLAp. Other conditions such as
75 collision mode, detector type, maximum injection time, and scan speed were also fitted
76 accordingly (**Figure 1a, Methods**). By the FAIMS-assisted immunopeptidome analysis
77 using $5e^6$ cells of HCT116, 3,215 HLAp on average were identified, resulting in 1.3-fold
78 increase from the without-FAIMS condition (2,482 HLAp on average, **Fig. 1b**). Accordingly,

79 the number of identified source proteins was also greater in the FAIMS-assisted condition
80 (2,381 proteins on average) than the without-FAIMS condition (2,077 proteins on average,
81 **Fig. 1b**). This result was consistent throughout the experiments and became statistically
82 significant when the sample scale was increased to $1e^7$ cells (3,893 HLAp identification
83 on average) or $1.5e^7$ cells (4,160 HLAp identification on average, **Fig. 1b**). The frequency
84 of neoantigen identification was also enhanced accordingly to the sample scale by FAIMS
85 up to 2.4-fold (**Fig. 1c**). Thus, the DIM-applied gas phase fractionation provided
86 strengthened identification of HLAp as well as neoantigens. These results indicated the
87 potential of DIM-MS in immunopeptidomics from samples that even with an order of
88 magnitude smaller than the most previous reports [8 11 9].

89 From the biologically triplicated analyses using $1e^8$ cells of HCT116, 6,915 HLAp derived
90 from 4,109 source proteins were identified on average for a single analysis, which yielded
91 9,249 unique HLAp from 5,103 unique source proteins in total (**Fig. 1d, Supplementary**
92 **Table 1**). Importantly, 7.3 neoantigens on average were identified per independent
93 experiments, which yielded 11 neoantigens in total (**Fig. 1e**). Among these, 7 were not
94 found in the immune epitope database (IEDB) (http://www.iedb.org/home_v3.php, as of
95 November 27, 2020), hence regarded as newly identified ones (**Table 1**). The distribution
96 of length and the assignment of HLAp against HLA allotypes exhibited typical and
97 reasonable manner in the context of HLA genotype (**Supplementary Figure 1**). These
98 results indicated that the FAIMS-applied DIM-MS system could provide an ideal
99 technological platform for in-depth profiling of HLAp, as well as direct sequencing of
100 neoantigens.

101 We next applied this technology to personalized HLAp profiling analysis of CRC. In this
102 analysis, tumor or adjacent normal regions of CRC tissues (n = 17) were analyzed
103 (**Supplementary Table 2**). Based on the whole exome sequencing (WES) analysis, the
104 mutation-included proteome database was individually constructed and used for database

search analysis of the DIM-MS datasets. In total, 44,785 unique HLAp were identified (6,578 peptides per patient on average) (**Fig. 2a**, **Supplementary Table 3**). Among these, 5,603 HLAp (12.5%) were normal-exclusively identified and 14,052 HLAp (31.4%) were tumor-exclusive, while 25,130 HLAp (56.1%) were commonly detected from both normal and tumor tissues. Thus we succeeded to improve the peptide identification number and quantitative depth of immunopeptidome analysis drastically, compared to the previous reports analyzing solid tumor tissues such as colon[6], liver[12], and ovarian[11] cancers. For the 11,117 source proteins (4,149 proteins per patient on average), 817 (9.7 %) or 1839 proteins (16.5 %) were identified normal-exclusively or tumor-exclusively, respectively (**Fig. 2b**). The individual datasets of immunopeptidomic profiling can be found in **Supplementary figure 2**. The identified HLAp exhibited features of HLA-allotype corresponding features of unsupervised clustering and assignment against the binding motifs (**Supplementary figure 3**). The length of HLAp were variable from patient to patient according to their HLA-allotypes, but the typical dominance of 9-mer in length of HLA class I was common among patients and no statistically significant difference was found between normal and tumor tissue HLAp (**Supplementary figure 4**). Notably, 2 neoantigens were identified from 2 distinct patients' tumor tissues (**Fig. 2c**). The first neoantigen identified from the primary tumor tissue of ID 172 carried a well-known CRC driver mutation, $KRAS_{7\text{-}16}$(G12V). The other one carrying a R228Q mutation of $CPPED1_{226\text{-}234}$, a 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. 2c**, depicted as "Preferred: V" by arrow). On the other hand, the mutation CPPED1-R228Q lost its deleterious amino acid of arginine at position 3 by substitution to glutamine (**Fig.**

131 **2c**, depicted as “Deleterious: P/R” by arrow). Therefore the identified neoantigens
132 exhibited two opposite patterns of gaining affinity against A*11:01 by acquiring a preferred
133 residue or losing a deleterious residue within epitope, compared to the corresponding wild
134 type (WT) sequences. Indeed, the predicted IC_{50} value or $KRAS_{7-16}$(WT), $KRAS_{7-}$
135 $_{16}$(G12V), $CPPED1_{226-234}$(WT), or $CPPED1_{226-234}$(R228Q) was 299.7 nM, 137.3 nM, 729.9
136 nM, or 52.2 nM, respectively. It has been a long-standing challenge to identify neoantigens
137 directly from small clinical tissues, but our personalized immunopeptidomic analysis
138 overcame that task.

139 We next compared HLAp of three groups; HLAp only found in normal tissue (normal-
140 exclusive), HLAp found in both normal and tumor tissues (shared), and HLAp only found
141 in tumor tissue (tumor-exclusive) per patient. If there is a bias in the mechanisms of
142 antigen processing/ presentation between normal and tumor tissues, HLAp should
143 manifest with the characteristics by groups. When we checked the overall usage of amino
144 acid for all 44,785 HLAp, 12 amino acids covered 99.2% of C-terminus usage, in which
145 the lowest frequency was cysteine (0.5%) (**Fig. 2d**). In contrast, tryptic (R, K) and
146 chymotryptic (L, I, V, F, Y, W, A, M) amino acids were more frequently observed for HLAp
147 trimming at the end of C-terminus (pΩ). The usage of amino acid frequency at pΩ was
148 calculated as ratio (%) against subtotal of each groups (normal-exclusive, shared, and
149 tumor-exclusive peptide, respectively) and compared the difference against based on the
150 frequency of the shared-peptide to clarify whether there would be a processing bias shown
151 or not between normal-exclusive and tumor-exclusive groups. Among the 12 pΩ used
152 amino acids, 9 patients carried at least one R-restraint HLA allotypes (A*11:01 and
153 A*33:03) at pΩ, and the usage of R at pΩ exhibited the tendency of reduction in tumor-
154 exclusive HLAp (normal- exclusive vs tumor- exclusive: -0.06 % vs -5.10 % compared to
155 the shared-peptide frequency, p = 0.0975, **Fig. 2e**, [R]). The usage of K exhibited no
156 statistical difference compared to the frequency of shared-peptide (data not shown). The

157 R and the K are the same tryptic amino acids while exhibited distinct dynamics in normal
158 and tumor tissue. For chymotryptic peptides, only W exhibited statistically significant
159 increase in tumor-exclusive group compared to the shared-peptide (**Fig. 2e**, [W]). The
160 trimming at pΩ of HLAp by L, I, and V which were generally grouped as similar
161 hydrophobic amino acids exhibited inconsistent and non-significant results between
162 normal and tumor-exclusive HLAp (data not shown). Since the usages of these three
163 hydrophobic residues are naturally abundant among the HLAp, it is expected that
164 statistical significance was diminished and was not shown as major differences when
165 compared at the admixture condition of tissue. Thirteen out of 17 patients carry at least
166 one pΩ W-restraint HLA allotypes (A*24:02, B*4402 and B*44:03) and the usage of W at
167 pΩ was -0.45 % in normal-exclusive HLAp and +0.8 % in tumor-exclusive HLAp compare
168 to the shared-peptide group (p = 0.0185) (**Fig. 2e**, [W]). A negative correlation in W usage
169 at pΩ between normal-exclusive and tumor-exclusive HLAp in identical patients indicated
170 the shift of HLAp processing at pΩ in tumor tissue (r = -0.788, p = 0.0014) (**Fig. 2f**). In
171 addition to these two amino acids, the cysteine usage at C-terminus was found to be
172 increased in tumor-exclusive HLAp (+ 0.25 %) than that of normal-exclusive HLAp
173 (0.00 %) compared to the shared-peptide group (p = 0.016) (**Fig. 2e**, [C]). Since there was
174 no clear information available for the binding motif that requires cysteine at pΩ, all 17
175 samples are included for this calculation. It is of note that this distinct cysteine usage
176 between normal and tumor-exclusive peptide may imply the unknown mechanisms of
177 HLAp processing under carcinogenesis and of value for further investigation.
178
179 Through optimization of the protocols for HLAp isolation and the parameters for FAIMS-
180 applied DIM-MS measurement, we here established an ideal technology for
181 immunopeptidome profiling, allowing comprehensive HLAp profiling with clinically-
182 meaningful depth. Indeed, cancer specific signatures of HLAp, such as neoantigens and
183 altered amino acid motifs, were identified from a small amount of CRC tissues. Previously,

184 massive amount of sample was necessary for immunopeptidomic analysis especially from
185 clinical tissue samples [9, 11, 12]. Alternatively, patient-derived cancer cells have to be
186 prepared as primary culture to secure the enough sample amount [13, 14]. The purified
187 culture of patient-derived cancer cells is convenient for in-depth analyses, however, this
188 approach conceived possibility that prolonged cultivation affects the antigen presentation
189 due to the lack of cancer microenvironment. Although a patient-derived xenograft in
190 mouse is also feasible to secure the cancer cell mass, the response against ICIs may
191 differ in host organisms, as shown in the discrepancy result of clinical trial of MEK inhibitor
192 [15] with PD-L1 treatment [16]. Our advanced immunopeptidomics by FAIMS-applied DIM-MS
193 paved the way to identify approximately 5,000 of HLAp from 40 mg of tissue without bias-
194 prone chemical prefractionation procedures [17]. Indeed, since the peptides are fractionated
195 by CVs which uniquely reflect the structural bulkiness of the peptides [18], the majority of
196 HLAp were identified from -60 V to -40 V of CV regardless of HLA allotypes. In addition,
197 rapid transition time for CV stepping (25 ms/ transition) [19] enabled simultaneous
198 measurements with 3 CVs in a single LC/MS run. Eventually, the ratio of the number of
199 peptide spectrum match (PSM) to acquired MS/MS spectra was effectively high, 23.1 ±
200 1.1% on average. Thus, the DIM-MS is considered to be ideal platform for rapid and
201 comprehensive analysis of HLAp.

202 Biosynthesis of HLAp is affected by various external conditions such as stimulation by
203 IFN-γ that increases the number of HLAp by inducing immunoproteasome and modifies
204 the variation of HLAp [6, 31]. However, there was no report showing direct comparison of
205 HLAp profile in cancerous region with that in normal region using individual fresh tumor
206 tissues. In this study, we for the first time demonstrated the detailed difference between
207 HLAp presented on cancer cells and that on normal cells in individual CRC tissues. The
208 total number of identified HLAp was relatively more abundant in tumor tissues than normal
209 tissues although the difference was not significant when normalized by the protein amount

210 (**Supplementary Figure 5**). In the previous *in vitro* experiments under stimulatory
211 condition using IFN-γ, chymotryptic trimming of HLAp was reported to occur frequently.
212 However, among chymotrypsin-targeted amino acids, only tryptophan trimming was
213 statistically increased in the tumor regions of CRC tissues. Regarding trypsin-targeted
214 amino acids, arginine exhibited the tendency of decrease in the tumor regions of CRC
215 tissues, whereas the frequency of lysine at pΩ did not show significant difference. These
216 facts indicated that the cancer-specific biosynthesis of HLAp should not be regulated only
217 by expression level or activity of the trimming enzymes. To fully understand the specific
218 mechanism of HLAp presentation in cancer cells, further large-scaled and HLA allotype-
219 specific HLAp analysis is needed, in combination with comprehensive expression profile
220 of source proteins.

221 It has been suggested that the frequency of neoantigen presentation on solid tumors
222 might be low [6, 12 13], especially in CRC [4 5], because of eradication of neoantigen-presenting
223 cancer cells by anti-tumor immunity [20, 21]. However, we determined 2 neoantigens directly
224 from stage IV CRC tissues in this study. In that respect, the identification of $KRAS_{7-}$
225 $_{16}$(G12V) HLAp in this study is noteworthy because this may overturn the prevailing view
226 that cancer driver mutation-containing neoantigens would be depleted in the later clinical
227 stages of cancer. Since the KRAS-G12D-reactive autologous T cells in lung-metastasized
228 CRC was reported in the previous clinical trial of ICI [22], further identification and utilization
229 of neoantigens containing driver mutations in clinical settings would be anticipated.
230 Another source protein of neoantigen identified in this study, CPPED1, is known as a
231 negative regulator of Akt signaling, which directly dephosphorylates Akt at Ser473 [23-25].
232 Since the amino acid substitution R228Q was found within the phosphatase domain of
233 CPPED1, this may attenuate the phosphatase activity and induce aberrant activation of
234 Akt signaling in CRC cells. Therefore, therapeutic elimination of $CPPED1_{226-234}$(R228Q)-
235 presenting cancer cells could be also considered as an effective strategy to halt

236 progression of CRC.

237 In principle, more mutation burden is considered to be associated with better outcome
238 by ICI treatment [26]. However, in our immunopeptidome dataset, the frequency of
239 neoantigen presentation did not correlated with genomic mutation burden. Indeed, no
240 neoantigen was detected from the tissue specimen derived from the patient ID 119 who
241 carried the highest mutation burden (1,752), even though 6,983 HLAp were identified in
242 total. This fact stresses the importance of taking actually-presented HLAp profiles into
243 account in order to precisely estimate the effect of cancer immunotherapy. These in-depth
244 immunopeptidomic datasets could also lead to interpretation of the complicated biological
245 mechanisms for biogenesis and presentation of antigens.

246 Thus, our FAIMS-applied immunopeptidomic analysis drastically expanded the size of
247 immnopeptidome dataset. Particularly, for the first time, identification of neoantigens from
248 clinical tissue samples was successful. Moreover, individual immunopeptidome profiling
249 of normal and tumor tissues revealed the distinct features of C-terminal trimming of HLAp.
250 Such precision immunopeptidomics holds promise for opening a next phase of cancer
251 immunology with new outlook from both basic and clinical aspects.

252

253

254 **Online Methods**

255 **Clinical samples and ethical approval**

256 Metastatic colon cancer tissues at pathological stage IV and its adjacent normal tissues
257 were obtained from patients who all had provided written informed consent for this study.
258 This study (ethical committee number 2010-1058) was approved by the ethical committee
259 of the Japanese Foundation for Cancer Research (JFCR). The clinical information of them
260 are listed in **Supplementary Table 2**. All tissue samples were frozen and maintained at
261 least once before the MS sample preparation. Approximately 40 mg in net weight of tissue
262 was dissected from the mass of tumor/ normal tissues respectively and further processed
263 for immunopeptidomics.

264

265 **Cell culture**

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

270

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

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

280 our in-house purified W6/32 antibody against HLA molecules, we tested all batches of
281 W6/32 by general IP and Western blotting by antibody anti-pan HLA class I α-chain
282 molecules (clone EMR8-5, MBL) and β-2-microglobulin (β2M) antibody (PROTEINTECH)
283 in advance. High affinity confirmed W6/32 antibody (800 ug) was then cross linked onto
284 200 ul of slurry of Protein G Sepharose (GE) by DMP (20 mM) in HEPES (pH 8.4) for 1
285 hour. After quenching by TBS-T for 1 hour, beads were used for IP or stored at 4 °C before
286 use.
287
288 **Immunoprecipitation of HLA-Class I Complex and HLAp purification for**
289 **immunopeptidomics**
290 Desired cell pellet was lysed by 1 ml of lysis buffer containing [20 mM] HEPES, [150 mM]
291 NaCl, 1% NP-40, 0.1% SDS and 10% Glycerol on ice. For tissue samples, disposable
292 homogenizer Bio Masher System (nippi) was used firstly with 500 µl of lysis buffer for
293 approximately 40 mg of tissue on ice and then add another 500 µl of lysis buffer to make
294 sample volume as 1 ml. Both cells and tissues samples were cleared by centrifugation at
295 4°C, 15,000 rpm for 15min, the supernatant was collected and conducted
296 immunoprecipitation for 4°C overnight on rotating rotor. For thorough washing, Bio Spin
297 Column (BioRad) was used afterwards. The washing steps were as follows: 1 ml of lysis
298 buffer for 5 times, 1 ml of PBS-T for 5 times, 1 ml of PBS for 5 times, 1 ml of high-salt PBS
299 for twice, 1 ml of PBS for 5 times and then 1 ml of mass spectrometry-grade water for
300 twice. The trimer of HLA complex was dissociated and eluted by 500 µl of 1% TFA from
301 PI-beads. Obtained HLA enriched eluate was further processed by tC18 SepPak. For
302 class I HLA peptide, 500 µl of 20% acetonitrile (ACN) was firstly used for elution. HLA
303 peptide including eluates were then dried up by evaporator. To elute the other components
304 of HLA complex to confirm the IP-efficiency, 500 µl of 80% ACN was used for second
305 elution and the obtained eluates were dried up by evaporator as well. Sample purity was

306 assessed by silver staining (Silver Quest).

307

308 **Culture Cells and the construction of WES-based customized database**

309 A colon cancer cell line, HCT116, was maintained in 10% fetal bovine serum including
310 RPMI1640 in humidified atmosphere at 37°C with 5% CO_2. Customized database for
311 HCT116 was stablished from the WES data deposited in COSMIC Cancer Cell Line
312 Project. Listed somatic mutations were processed by in-house built pipeline previously
313 reported [27, 28]. Obtained mutations were then combined with Swiss-Prot human proteome
314 database (20,181 entries) from UniProt website and used for neoantigen search.

315

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

317 For clinical tissue samples, genomic DNA was extracted by QIAmp DNA mini Kit
318 (QIAGEN) from normal and tumor tissues. To establish the personalized database for
319 neoantigen identification by immunopeptodomics, whole exome sequencing (WES) was
320 conducted to all samples. In-house pipe line to extract the somatic mutations has been
321 previously reported [28, 29]. Briefly, the genome libraries for WES were established from the
322 DNA samples using xGen Exome Research Panel (IDT, Coralville, IA), according to the
323 manufacturer's instructions. The exons were sequenced as 150 base pairs paired-end
324 reads by a NovaSeq6000 system (Illumina, San Diego, CA). The sequence data obtained
325 were then analyzed to select possible germ line variations and somatic mutations. In short,
326 sequences of the whole exome were compared with a reference human genome (hs37d5),
327 and the possible germ line variants and the somatic mutations were identified. The
328 variants in public databases (> 1%) were also excluded. Then cancer somatic mutations
329 were extracted by subtraction of germ line variation identified in normal tissue from tumor
330 tissue respectively. Among those mutations, only the missense mutations were selected
331 and translated into full length of protein and added to the same Swissprot/ Uniprot human

332 database described above as personalized database. The established personalized data
333 base was used for immunopeptidomics both normal and tumor tissues corresponding
334 patient's samples. The OptiType was employed for 4-digit HLA typing from deep
335 sequencing data [30]. The genetic details including mutation burden, oncogenic KRAS
336 mutations and the HLA genotypes of clinical samples can be found in **Supplementary**
337 **Table 2**.
338
339 **Differential ion mobility (DIM) Mass Spectrometry (MS) by High-Field Asymmetric-**
340 **waveform Ion Mobility Spectrometry (FAIMS) for Advanced Immunopeptidomics**
341 The sample includes HLAp was first trapped by precolumn (C18 Accleim PepMap 100
342 C18 Trap Cartridge, Thermo Scientific) and then separated by analytical column (Aurora
343 UHPLC Column (C18, 0.075 × 250 mm, 1.6 um FSC, ESI, ion opticks) coupled with a
344 nanospray Flex ion source for electrospray ionization (Thermo Scientific), were used
345 through an Ultimate 3000 RSLC nano HPLC system at a flow rate of 200 nl/ min using a
346 linear gradient starting from 2% solvent B (0.1% Formic acid in acetonitrile) to 28 over 55
347 min, a 1 min hold at 95% Solvent B a prior to a 1 min analytical column equilibration with
348 2% B. In order to validate the benefit of FAIMS applied DIM-MS in immunopeptidomics,
349 liquid chromatography tandem mass spectrometry (LC-MS/MS) and LC-FAIMS-MS/MS
350 conditions were firstly compared. For LC-FAIMS-MS/MS, the FAIMS-Pro interface
351 (Thermo Fisher Scientific) was installed onto an Orbitrap Fusion Lumos Tribrid mass
352 spectrometer and operated with default parameters except for the compensation voltage
353 (CV) settings for gas-phase fractionation. For LC-MS/MS, we uninstalled FAIMS device
354 from mass spectrometer and performed analyses. Under LC-FAIMS-MS/MS condition,
355 sample was seamlessly fractioned by 3 CVs per single analysis. Total 3 CV sets were
356 used per sample and each CV set includes as follows; set1 (CV = -40, -60, -80 V), set 2
357 (CV = -35, -50, -70 V) and set 3 (CV = -45, -55, -65 V). Therefore, total 3 analyses (raw

358 data), were acquired per sample. The parameters for mass spectrometry were optimized
359 for immunopeptidomics with small amount of sample. Full MS (rage from 320 to 850 m/z)
360 in the Orbitrap were acquired at resolution of 60 k followed by a MS^2 acquisition at
361 resolution of 15k. Maximum injection time for full MS scan was 50 ms with an auto gain
362 control (AGC) of $4e^5$. Maximum injection time for MS^2 scan was 100 ms with an AGC $1e^4$
363 followed by top speed MS^2 acquisition by the ion trap. Charge state 2 and 3 were selected
364 for fragmentation by collision-induced dissociation (CID) at rapid scan rate at 30% collision
365 energy. The easy-IC system was used for the internal calibration lock mass during the
366 data acquisition. The HCT116 samples were prepared from $1e^8$ cells and proportionally
367 injected into MS to match the indicated number of cells. For the analyses of $1e^8$ of HCT116
368 cells and clinical tissue samples, firstly 1/ 20 amount of sample volume was used for
369 analyses of 3 CV sets respectively to check sample conditions, and then from the
370 remaining, 5/ 20 amount of sample volume was used for 3 CV sets respectively. Therefore,
371 total 6 analyses were performed per clinical tissue sample.
372 The LC/MS raw data and summarized result files are deposited at a public proteomic
373 database, Japan Proteome Standard Repository/Database (jPOST) as follows; HCT116
374 HLAp without FAIMS in JPST001072, HCT116 HLAp with FAIMS in JPST001066, Global
375 identification of HCT116 HLAp in JPST001068, HLAp from normal regions of CRC tissues
376 in JPST001070, and HLAp from tumor regions of CRC tissues in JPST001069.
377
378 **Data analysis by tailored database search for immunopeptidome**
379 Obtained raw files were processed by Proteome Discoverer version 2.4 using Sequest
380 HT engine against corresponding tailored database. The search was set to no-enzyme
381 (unspecific) for the enzymatic designation. The precursor mass tolerance was set to 10
382 ppm and the fragment mass tolerance to 0.6 Da. The maximum number of missed
383 cleavages was set to zero. Methionine oxidation and cysteine carbamidomethylation were

384 included as dynamic modifications. A false discovery rate (FDR) of 0.01 (Strict) was used
385 for percolator algorithm to identify highly confident peptides. The output “peptide groups”
386 was exported as a list. Then the peptide sequence was deduplicated for unique peptide
387 sequences only and filtered by the amino acid length from 8 to 15 as a candidate of HLAp.
388 To narrow candidate sequences down to the reliable ones, NetMHCpan 4.0[31] was
389 employed to exclude the no-binding peptides from the list. Simultaneously, unsupervised
390 clustering for HLA sequence motifs, Gibbs Cluster 2.0 [32] and Seq2 logo [33] were employed
391 to confirm whether the identified motifs match HLA allotypes accordingly.
392

393 **Profiling of HLAp in Normal and Tumor Tissues**

394 In order to interrogate the possible distinctions of normal and tumor-derived HLAp, the
395 intersection of HLAp between normal and tumor tissues was calculated and draw Venn
396 diagrams by publicly available website
397 (http://bioinformatics.psb.ugent.be/webtools/Venn/). Then classified the HLAp into 3
398 groups; (1) normal-exclusive, (2) shared in both normal and tumor tissues (as “shared-
399 peptide”) and (3) tumor-exclusive ones per patient. The frequency of amino acid usage at
400 position Ω (pΩ), the end position at C-terminus of HLAp, was calculated as percentage
401 against the total number of HLAp in each group. Based on the frequency of shared-
402 peptide’s amino acid usage at pΩ, the distinction of frequency was calculated by
403 comparing the value between normal-exclusive and shared-peptide or tumor-exclusive
404 and shared-peptide.
405

406 **Statistical Analysis.**

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

410 (***) values. Pearson correlation-coefficient was used to assess the correlation.
411
412

Acknowledgements

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

Author Contributions

Y.M. and K.U. designed the experiments. Y.M., K.K., K.N., Y.H., R.F., and N.S. performed the experiments. Y.I. and S.N. provided the CRC tissue specimens. Y.M., K.K., K.N., Y.H., R.F., and N.S. analyzed and interpreted the data. Y.M. and K.U. wrote the article.

Competing Interests statement

The authors declare no competing interests.

429 **References:**

430

431 1. Jhunjhunwala, S., Hammer, C. & Delamarre, L. Antigen presentation in cancer: insights into
432 tumour immunogenicity and immune evasion. *Nat Rev Cancer* (2021).
433 2. Cancer Genome Atlas, N. Comprehensive molecular characterization of human colon and
434 rectal cancer. *Nature* **487**, 330-337 (2012).
435 3. Caron, E., Aebersold, R., Banaei-Esfahani, A., Chong, C. & Bassani-Sternberg, M. A Case for
436 a Human Immuno-Peptidome Project Consortium. *Immunity* **47**, 203-208 (2017).
437 4. Le, D.T. et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med*
438 **372**, 2509-2520 (2015).
439 5. de Weger, V.A. et al. Clinical effects of adjuvant active specific immunotherapy differ between
440 patients with microsatellite-stable and microsatellite-instable colon cancer. *Clin Cancer Res*
441 **18**, 882-889 (2012).
442 6. Loffler, M.W. et al. Mapping the HLA Ligandome of Colorectal Cancer Reveals an Imprint of
443 Malignant Cell Transformation. *Cancer Res* **78**, 4627-4641 (2018).
444 7. Shukla, S.A. et al. Comprehensive analysis of cancer-associated somatic mutations in class I
445 HLA genes. *Nat Biotechnol* **33**, 1152-1158 (2015).
446 8. Bassani-Sternberg, M. et al. Direct identification of clinically relevant neoepitopes presented
447 on native human melanoma tissue by mass spectrometry. *Nat Commun* **7**, 13404 (2016).
448 9. Chong, C. et al. High-throughput and Sensitive Immunopeptidomics Platform Reveals
449 Profound Interferongamma-Mediated Remodeling of the Human Leukocyte Antigen (HLA)
450 Ligandome. *Mol Cell Proteomics* **17**, 533-548 (2018).
451 10. Liepe, J. et al. A large fraction of HLA class I ligands are proteasome-generated spliced
452 peptides. *Science* **354**, 354-358 (2016).
453 11. Schuster, H. et al. The immunopeptidomic landscape of ovarian carcinomas. *Proc Natl*
454 *Acad Sci U S A* **114**, E9942-E9951 (2017).
455 12. Loffler, M.W. et al. Multi-omics discovery of exome-derived neoantigens in
456 hepatocellular carcinoma. *Genome Med* **11**, 28 (2019).
457 13. Newey, A. et al. Immunopeptidomics of colorectal cancer organoids reveals a sparse
458 HLA class I neoantigen landscape and no increase in neoantigens with interferon or MEK-
459 inhibitor treatment. *J Immunother Cancer* **7**, 309 (2019).
460 14. Demmers, L.C. et al. Single-cell derived tumor organoids display diversity in HLA
461 class I peptide presentation. *Nat Commun* **11**, 5338 (2020).
462 15. Eng, C. et al. Atezolizumab with or without cobimetinib versus regorafenib in
463 previously treated metastatic colorectal cancer (IMblaze370): a multicentre, open-label, phase
464 3, randomised, controlled trial. *Lancet Oncol* **20**, 849-861 (2019).
465 16. Ebert, P.J.R. et al. MAP Kinase Inhibition Promotes T Cell and Anti-tumor Activity in
466 Combination with PD-L1 Checkpoint Blockade. *Immunity* **44**, 609-621 (2016).

467 17. Demmers, L.C., Heck, A.J.R. & Wu, W. Pre-fractionation Extends but also Creates a
468 Bias in the Detectable HLA Class Iota Ligandome. *J Proteome Res* **18**, 1634-1643 (2019).
469 18. Purves, R.W. et al. Using gas modifiers to significantly improve sensitivity and
470 selectivity in a cylindrical FAIMS device. *J Am Soc Mass Spectrom* **25**, 1274-1284 (2014).
471 19. Hebert, A.S. et al. Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid
472 Orbitrap Mass Spectrometer. *Anal Chem* **90**, 9529-9537 (2018).
473 20. Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G. & Hacohen, N. Molecular and genetic
474 properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48-61 (2015).
475 21. Davoli, T., Uno, H., Wooten, E.C. & Elledge, S.J. Tumor aneuploidy correlates with
476 markers of immune evasion and with reduced response to immunotherapy. *Science* **355**
477 (2017).
478 22. Tran, E. et al. T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. *N Engl J*
479 *Med* **375**, 2255-2262 (2016).
480 23. Zhuo, D.X. et al. CSTP1, a novel protein phosphatase, blocks cell cycle, promotes cell
481 apoptosis, and suppresses tumor growth of bladder cancer by directly dephosphorylating Akt
482 at Ser473 site. *PLoS One* **8**, e65679 (2013).
483 24. Zhuo, D., Wu, Y., Luo, J., Deng, L. & Niu, X. CSTP1 inhibits IL-6 expression through
484 targeting Akt/FoxO3a signaling pathway in bladder cancer cells. *Exp Cell Res* **380**, 80-89
485 (2019).
486 25. Yang, R. et al. Combined Transcriptome and Proteome Analysis of Immortalized
487 Human Keratinocytes Expressing Human Papillomavirus 16 (HPV16) Oncogenes Reveals
488 Novel Key Factors and Networks in HPV-Induced Carcinogenesis. *mSphere* **4** (2019).
489 26. Hugo, W. et al. Genomic and Transcriptomic Features of Response to Anti-PD-1
490 Therapy in Metastatic Melanoma. *Cell* **165**, 35-44 (2016).
491 27. Kato, T. et al. Effective screening of T cells recognizing neoantigens and construction
492 of T-cell receptor-engineered T cells. *Oncotarget* **9**, 11009-11019 (2018).
493 28. Kiyotani, K., Chan, H.T. & Nakamura, Y. Immunopharmacogenomics towards
494 personalized cancer immunotherapy targeting neoantigens. *Cancer Sci* **109**, 542-549 (2018).
495 29. Choudhury, N.J. et al. Low T-cell Receptor Diversity, High Somatic Mutation Burden,
496 and High Neoantigen Load as Predictors of Clinical Outcome in Muscle-invasive Bladder
497 Cancer. *Eur Urol Focus* **2**, 445-452 (2016).
498 30. Szolek, A. et al. OptiType: precision HLA typing from next-generation sequencing
499 data. *Bioinformatics* **30**, 3310-3316 (2014).
500 31. Jurtz, V. et al. NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions
501 Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol* **199**, 3360-3368
502 (2017).
503 32. Andreatta, M., Alvarez, B. & Nielsen, M. GibbsCluster: unsupervised clustering and
504 alignment of peptide sequences. *Nucleic Acids Res* **45**, W458-W463 (2017).
505 33. Thomsen, M.C. & Nielsen, M. Seq2Logo: a method for construction and visualization

506 of amino acid binding motifs and sequence profiles including sequence weighting, pseudo
507 counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids*
508 *Res* **40**, W281-287 (2012).

509

510

511

512

513 **Figure Legends**

514 **Figure 1. Schematic workflow and data validation of personalized**
515 **immunopeptidome analysis assisted by differential ion mobility mass spectrometry**
516 **(DIM-MS). a.** A proteogenomic workflow for personalized immunopeptidome analysis is
517 shown. After constructing the personalized proteome database in which cancer-specific
518 mutated amino acid sequences were involved, the DIM-MS analysis was performed for
519 isolated HLAp. The FAIMS-Pro interface was coupled to mass spectrometer for seamless
520 gas-phase fractionation to enable the DIM-MS. CV; compensation voltage. **b.** The number
521 of identified HLAp from HCT116 cells was displayed with bars and the left y-axis (white
522 bars; without FAIMS, black bars; with FAIMS). The number of identified source proteins
523 was shown with the dashed line and the right y-axis. * $p < 0.05$, # $p < 0.1$. **c.** The average
524 number of identified neoantigens from the same experiments of b was shown (white bars;
525 without FAIMS, black bars; with FAIMS). * $p < 0.05$, ** $p < 0.01$. **d.** The number of identified
526 HLAp from three independent analyses using $1e^8$ of HCT116 cells was shown with bars
527 and the left y-axis. White bars; the number of HLAp from respective analyses, gray bar;
528 an average number of HLAp from 3 independent analyses, and dark gray bar; total
529 number of non-redundant identified HLAp. The number of source proteins were depicted
530 in dashed line and the right y-axis. **e.** The number of neoantigens identified from the same
531 experiments as d was shown.

532

533 **Figure 2. Personalized differential immunopeptidome analysis using cancer tissues**
534 **and adjacent normal tissues of colorectal cancer (n = 17). a.** An average of total
535 number of HLAp identified per patient (HLAp/ Patient) and the total number of unique
536 HLAp from all patients (Total Unique HLAp) were displayed in bar graphs with respective
537 numerals. **b.** An average of total number of source proteins identified per patient (Protein/
538 Patient) and the total number of unique HLAp from all patients (Total Unique Protein) were

539 displayed in bar graphs with respective numerals. Full information for 17 patients can be
540 found in **Supplementary Table 3**. **c.** The MS/MS spectra for $KRAS_{7-16}$(G12V) and
541 $CPPED1_{226-234}$(R228Q) neoantigens were shown. Possible causes of enhanced affinity
542 by mutations were also depicted with motif sequence of A*11:01. **d.** A pie chart depicts the
543 frequency of amino acid usage at the C-terminus (pΩ) of HLAp. **e.** The difference of amino
544 acid usage (%) at pΩ in normal-exclusive and tumor-exclusive HLAp compared to the
545 shared-HLAp. **f.** The dot plot demonstrates negative correlation of tryptophan (W) usage
546 at pΩ in normal-exclusive and tumor-exclusive HLAp.
547
548
549

550 **Table 1. Neoantigens identified from HCT116 by immunopeptidomic analysis**

551

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	

552 The position of mutation is highlighted by underline. The affinity prediction by NetMHC4.0

553 was shown. The number of times identified out of 3 independent MS analyses was shown.

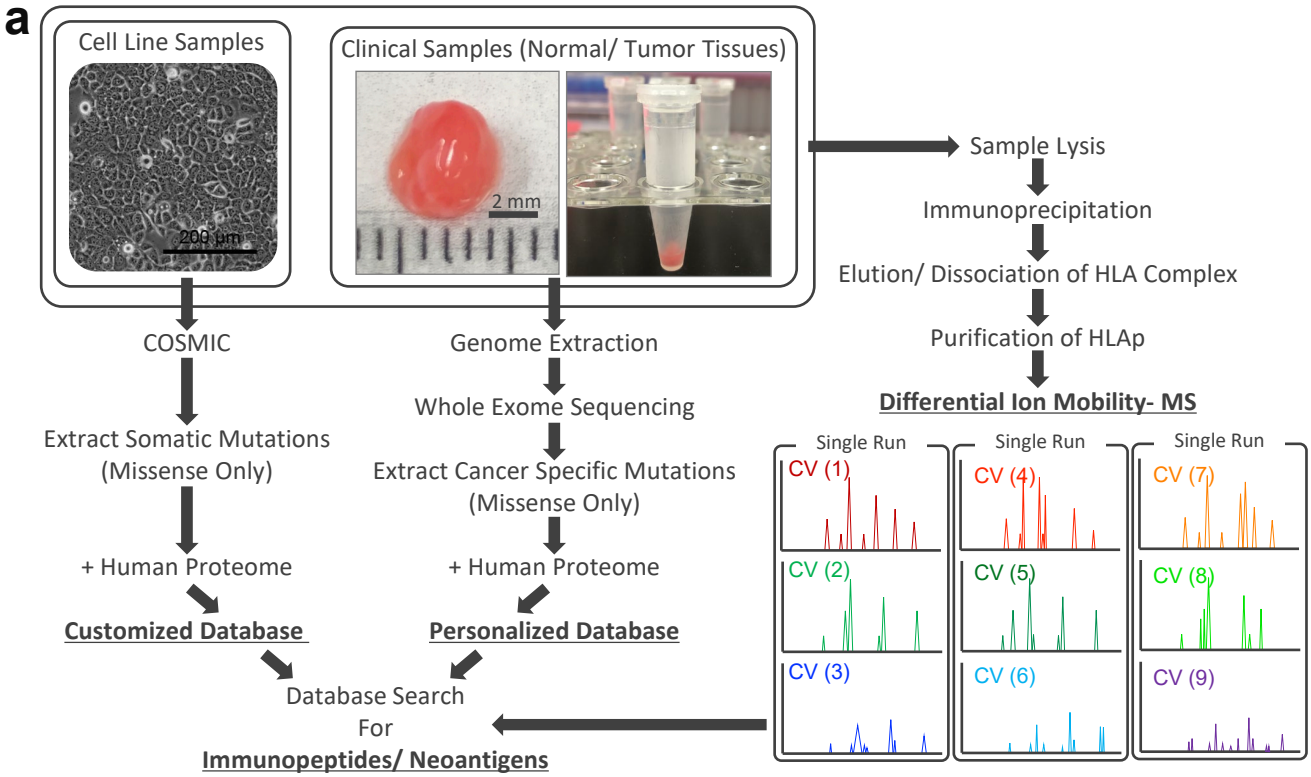
554 * Newly identified neoantigens.

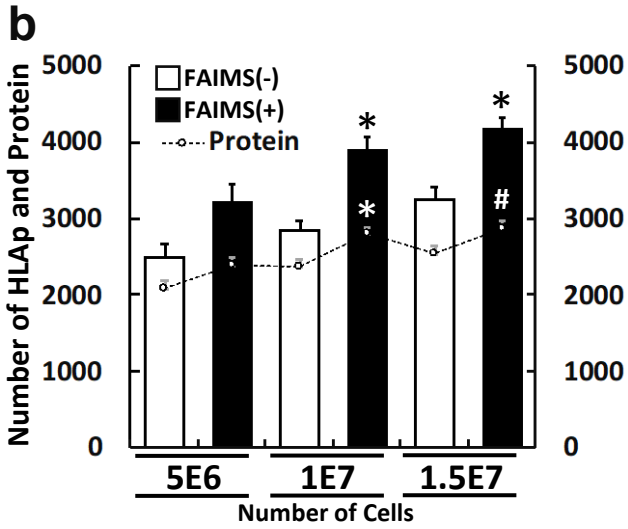
555 # There was no epitope information of the corresponding wild type sequence at IEDB.

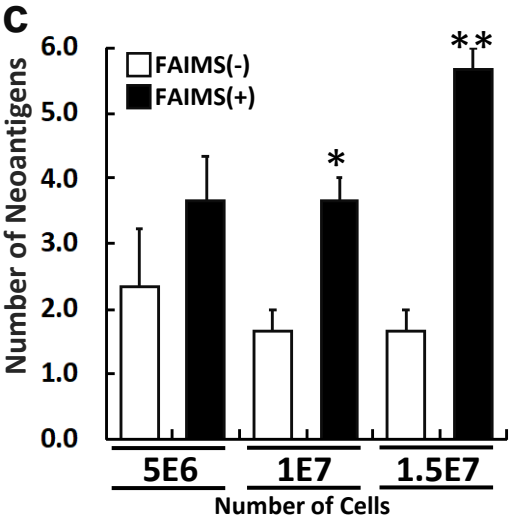
556

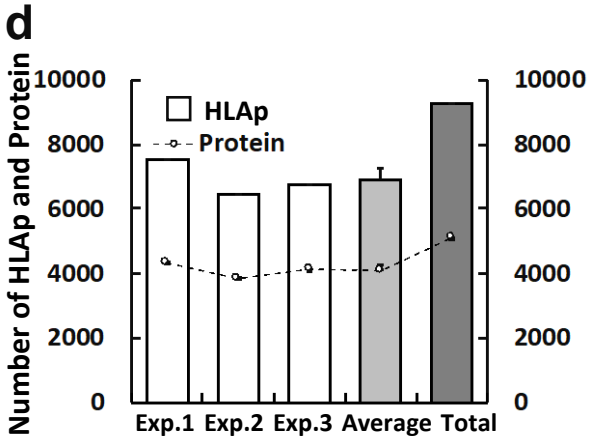
557

Figure 1


a
Cell Line Samples
200 µm
Clinical Samples (Normal/ Tumor Tissues)
2 mm
Sample Lysis
Immunoprecipitation
Elution/ Dissociation of HLA Complex
Purification of HLAp
Differential Ion Mobility- MS
Single Run
CV (1)
CV (2)
CV (3)
CV (4)
CV (5)
CV (6)
CV (7)
CV (8)
CV (9)
COSMIC
Extract Somatic Mutations
(Missense Only)
+ Human Proteome
Customized Database
Genome Extraction
Whole Exome Sequencing
Extract Cancer Specific Mutations
(Missense Only)
+ Human Proteome
Personalized Database
Database Search
For
Immunopeptides/ Neoantigens


b
Number of HLAp and Protein
FAIMS(-)
FAIMS(+)
Protein
5000
4000
3000
2000
1000
0
*
*
*
#
5E6
1E7
1.5E7
Number of Cells


c
Number of Neoantigens
FAIMS(-)
FAIMS(+)
6.0
5.0
4.0
3.0
2.0
1.0
0.0
*
**
5E6
1E7
1.5E7
Number of Cells


d
Number of HLAp and Protein
HLAp
Protein
10000
8000
6000
4000
2000
0
Exp.1
Exp.2
Exp.3
Average
Total

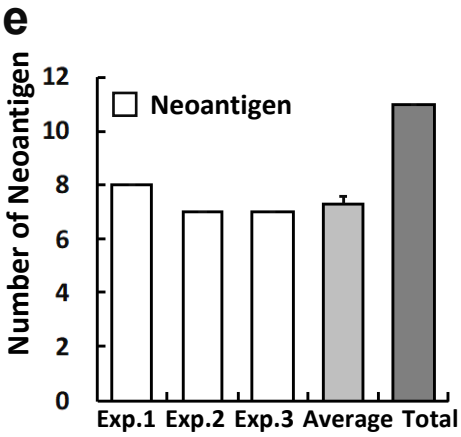
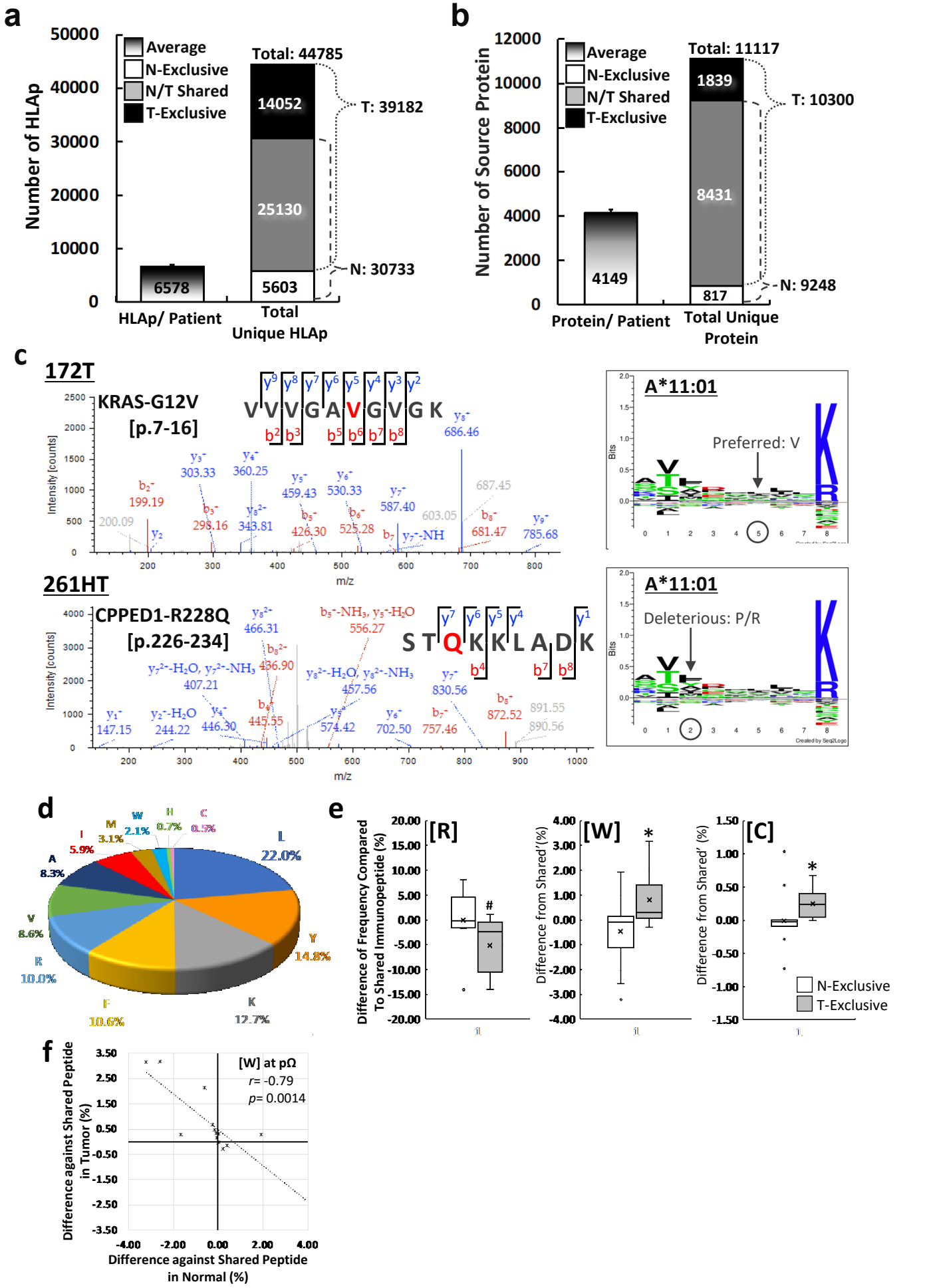
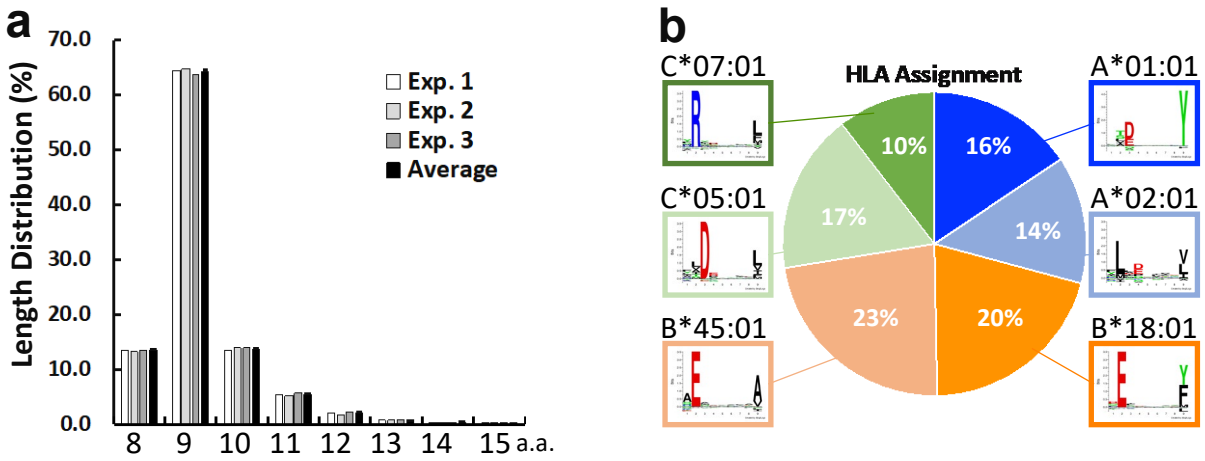

e
Number of Neoantigen
Neoantigen
12
10
8
6
4
2
0
Exp.1
Exp.2
Exp.3
Average
Total

Figure 2


a
Number of HLAp
Average
N-Exclusive
N/T Shared
T-Exclusive
Total: 44785
14052
25130
5603
6578
T: 39182
N: 30733
HLAp/ Patient
Total Unique HLAp
b
Number of Source Protein
Total: 11117
1839
8431
817
4149
T: 10300
N: 9248
Protein/ Patient
Total Unique Protein
c
172T
KRAS-G12V [p.7-16]
V V V G A V G V G K
Intensity [counts]
m/z
A*11:01
Preferred: V
261HT
CPPED1-R228Q [p.226-234]
S T Q K K L A D K
Deleterious: P/R
d
L 22.0%
Y 14.8%
K 12.7%
F 10.6%
R 10.0%
V 8.6%
A 8.3%
I 5.9%
M 3.1%
W 2.1%
H 0.7%
C 0.5%
e
[R]
[W]
[C]
Difference of Frequency Compared To Shared Immunopeptide (%)
Difference from Shared' (%)
N-Exclusive
T-Exclusive
f
[W] at pΩ
r= -0.79
p= 0.0014
Difference against Shared Peptide in Tumor (%)
Difference against Shared Peptide in Normal (%)

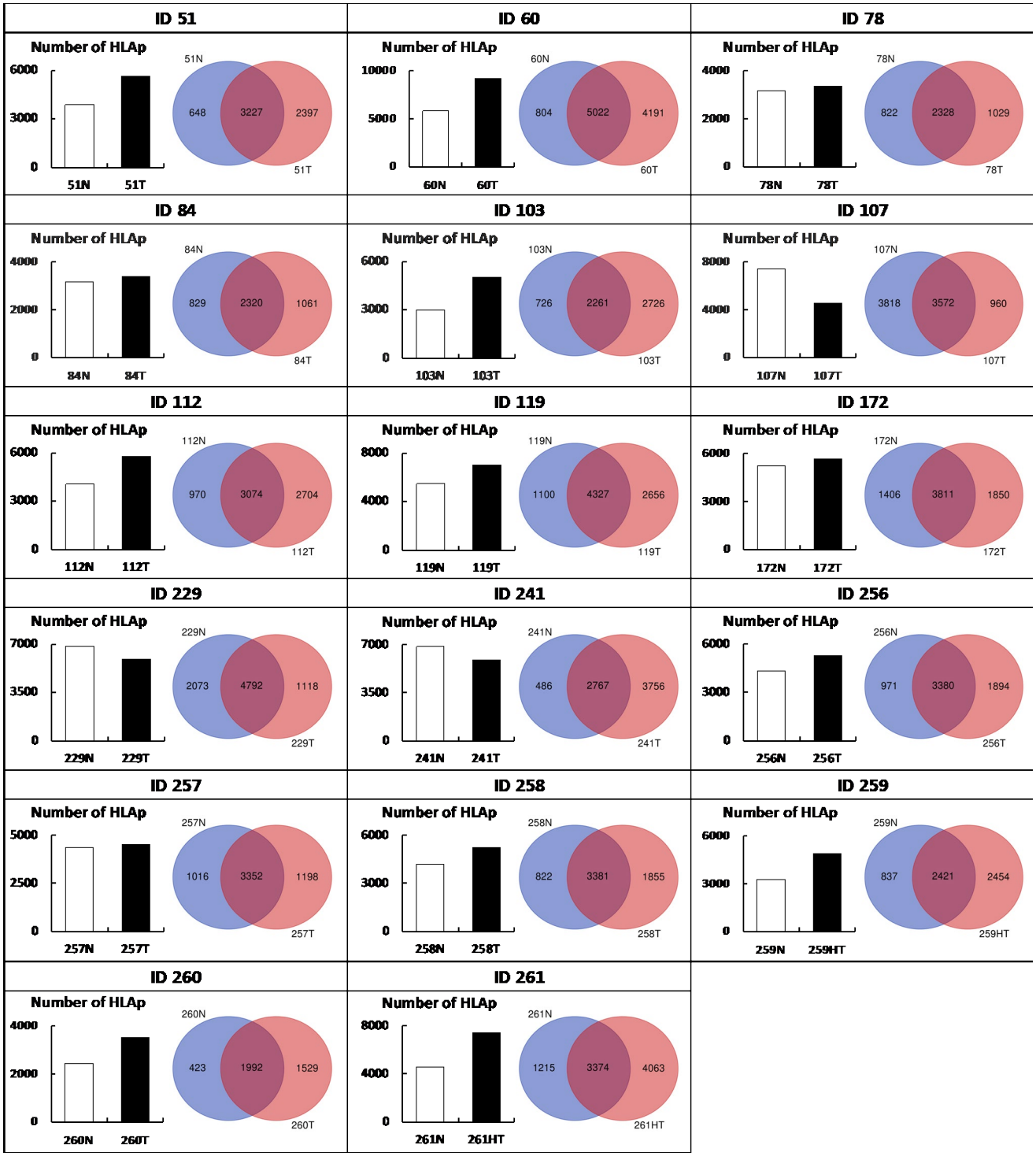
Supplementary Figure 1



Supplementary Figure 1. Characteristics of immunopeptidomics and neoantigen identification from HCT116 cells

a. The length of HLAp identified from HCT116 in Figure 2d was counted. **b.** Representative image of unsupervised HLA motif clustering and the population of HLA assignment against genotype-matched HLA allotypes.

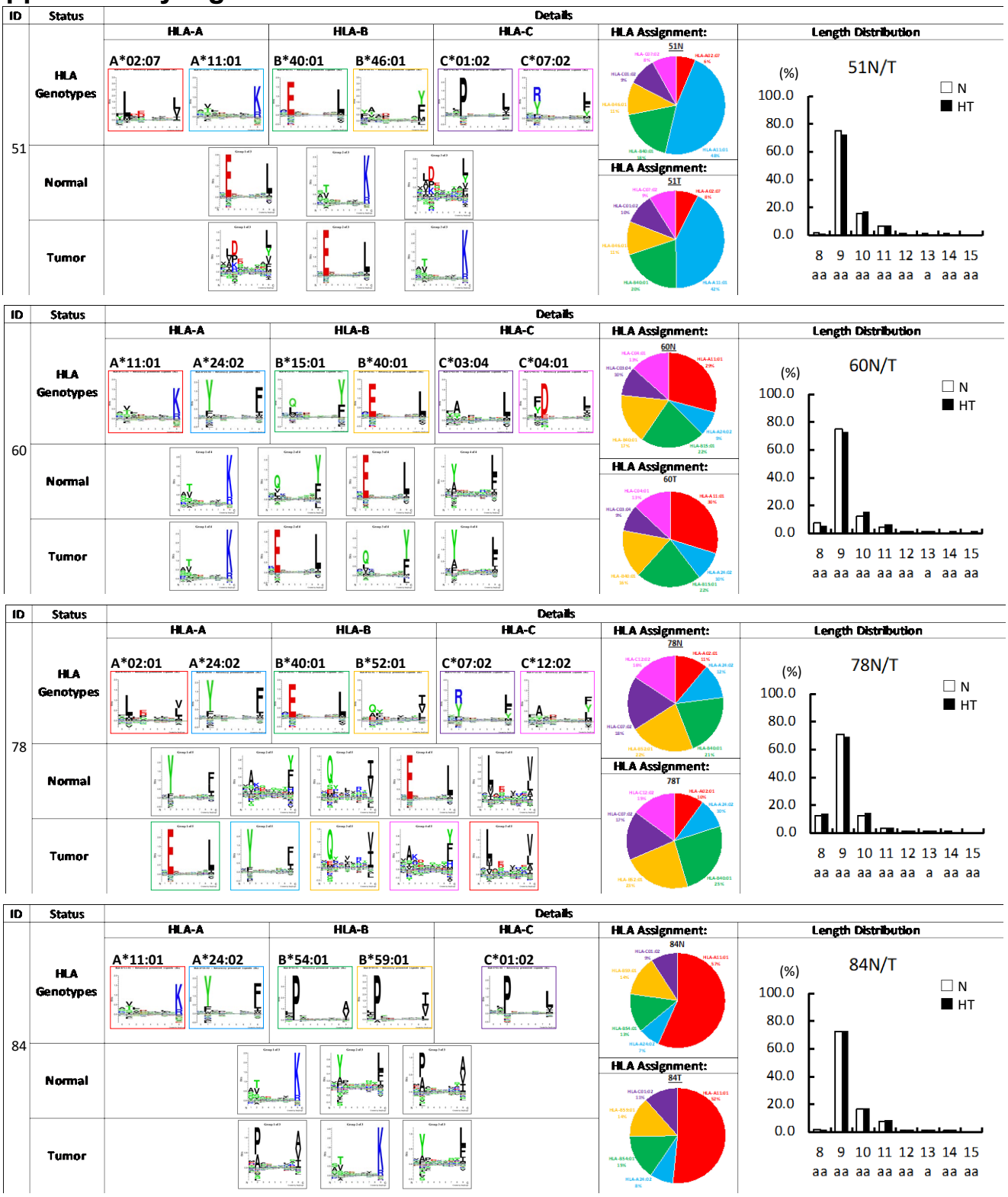
Supplementary Figure 2



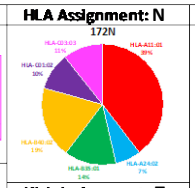
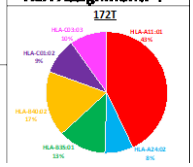
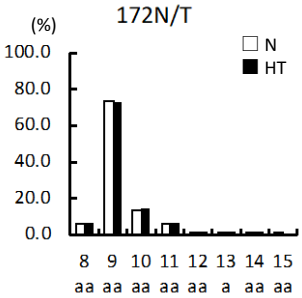
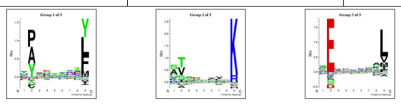
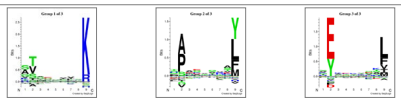
Supplementary Figure 2. Number of HLAp per patients and Venn diagrams

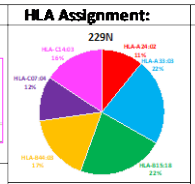
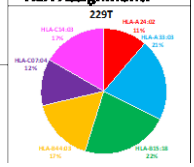
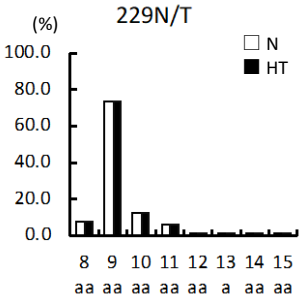
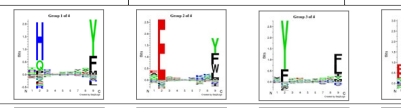
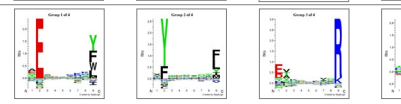
The details of identified number of HLAp in each normal and tumor tissue samples and the overlap of HLAp in identical patients were shown. Venn diagram does not reflect the sample size in this figure.

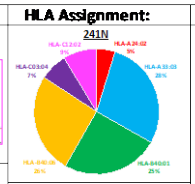
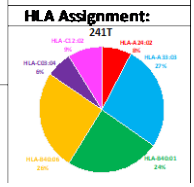
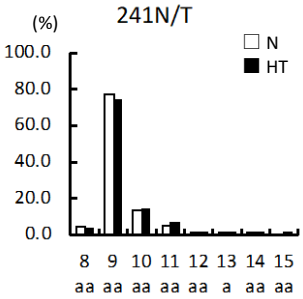
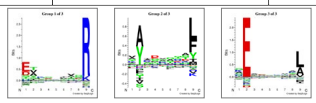
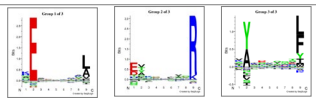
Supplementary Figure 3

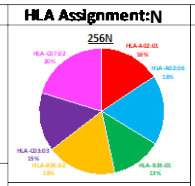
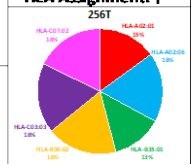
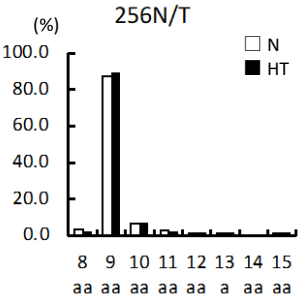
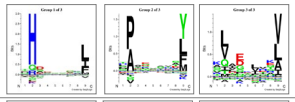
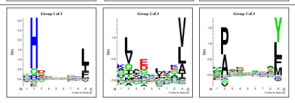


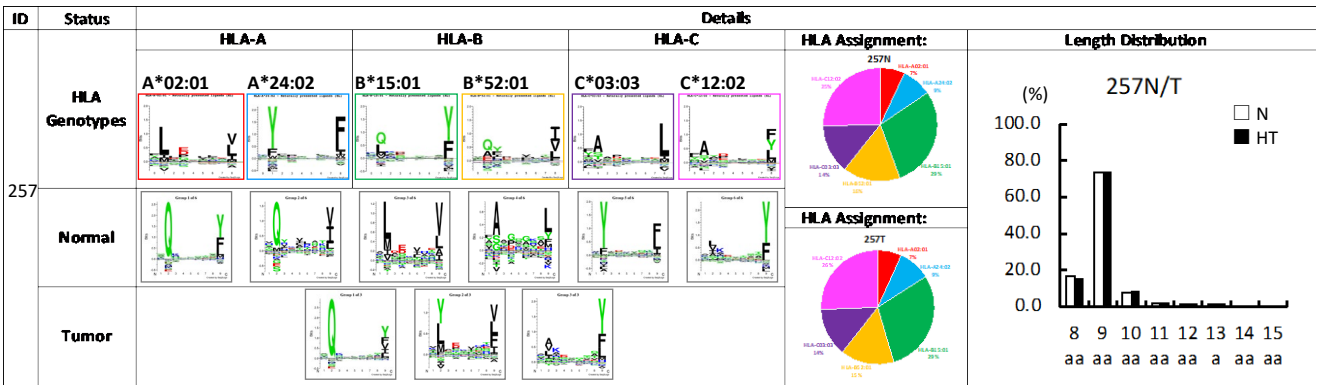
| ID | Status | Details | | |

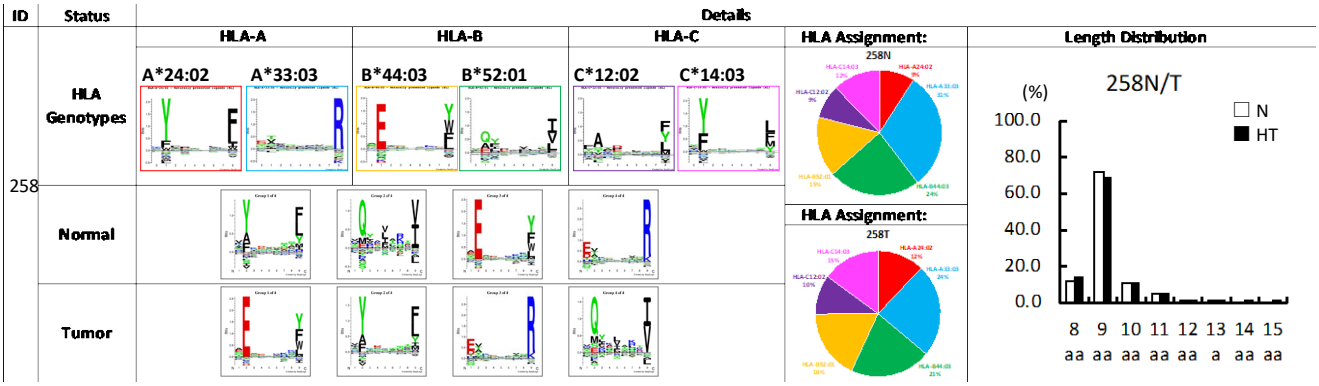
ID	Status	Details					
		HLA-A	HLA-B	HLA-C	HLA Assignment: N	HLA Assignment: T	Length Distribution
172	HLA Genotypes	A*11:01 A*24:02	B*35:01 B*40:02	C*01:02 C*03:03			
	Normal						
	Tumor						

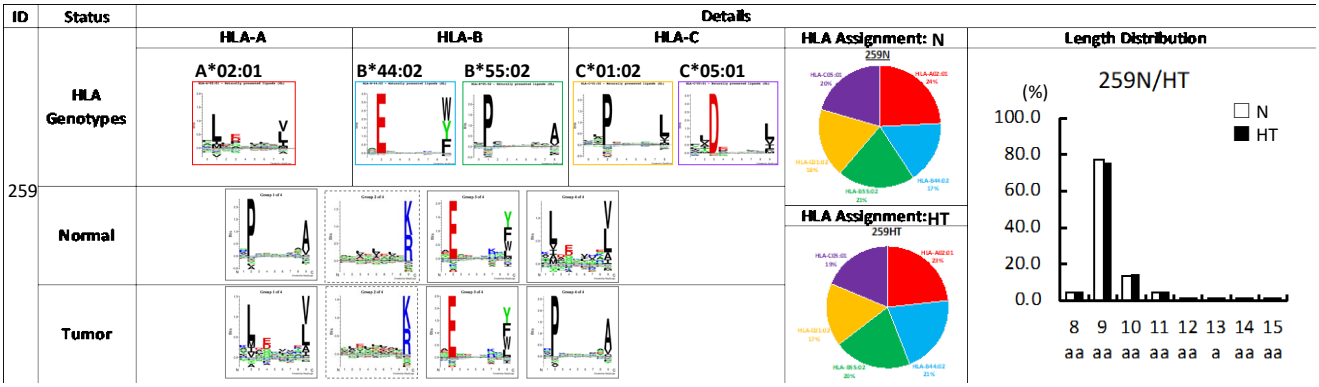
ID	Status	Details					
		HLA-A	HLA-B	HLA-C	HLA Assignment:	HLA Assignment:	Length Distribution
229	HLA Genotypes	A*24:02 A*33:03	B*15:08 B*44:03	C*07:04 C*14:03			
	Normal						
	Tumor						

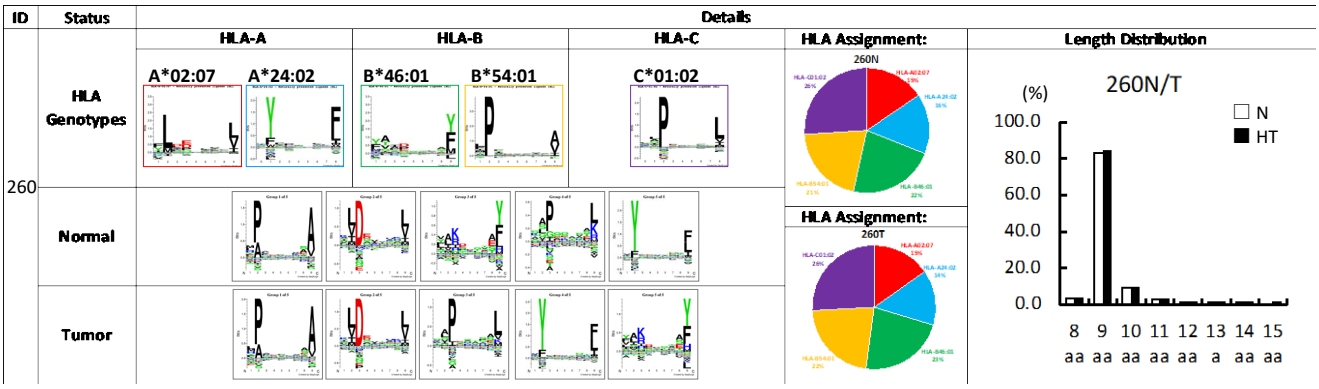
ID	Status	Details					
		HLA-A	HLA-B	HLA-C	HLA Assignment:	HLA Assignment:	Length Distribution
241	HLA Genotypes	A*24:02 A*33:03	B*40:01 B*40:06	C*03:04 C*12:02			
	Normal						
	Tumor						

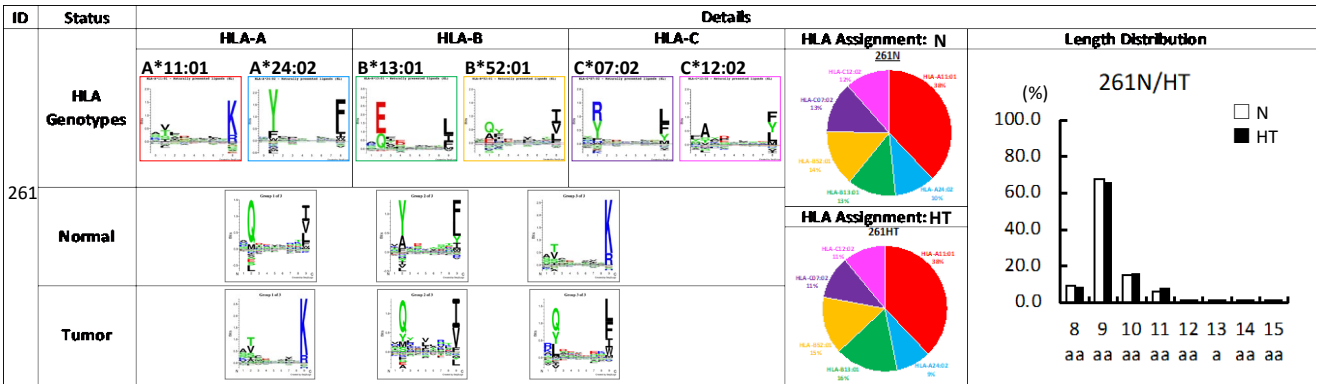
ID	Status	Details					
		HLA-A	HLA-B	HLA-C	HLA Assignment:N	HLA Assignment: T	Length Distribution
256	HLA Genotypes	A*02:01# A*02:06	B*35:01 B*38:02	C*03:03 C*07:02			
	Normal						
	Tumor						


ID
Status
Details
HLA-A
HLA-B
HLA-C
HLA Assignment:
Length Distribution
257
HLA Genotypes
A*02:01
A*24:02
B*15:01
B*52:01
C*03:03
C*12:02
Normal
Tumor
257N
257T
257N/T
(%)
N
HT
100.0
80.0
60.0
40.0
20.0
0.0
8 aa
9 aa
10 aa
11 aa
12 aa
13 a
14 aa
15 aa


ID
Status
Details
HLA-A
HLA-B
HLA-C
HLA Assignment:
Length Distribution
258
HLA Genotypes
A*24:02
A*33:03
B*44:03
B*52:01
C*12:02
C*14:03
Normal
Tumor
258N
258T
258N/T
(%)
N
HT
100.0
80.0
60.0
40.0
20.0
0.0
8 aa
9 aa
10 aa
11 aa
12 aa
13 a
14 aa
15 aa


ID
Status
Details
HLA-A
HLA-B
HLA-C
HLA Assignment: N
Length Distribution
259
HLA Genotypes
A*02:01
B*44:02
B*55:02
C*01:02
C*05:01
Normal
Tumor
259N
HLA Assignment:HT
259HT
259N/HT
(%)
N
HT
100.0
80.0
60.0
40.0
20.0
0.0
8 aa
9 aa
10 aa
11 aa
12 aa
13 a
14 aa
15 aa

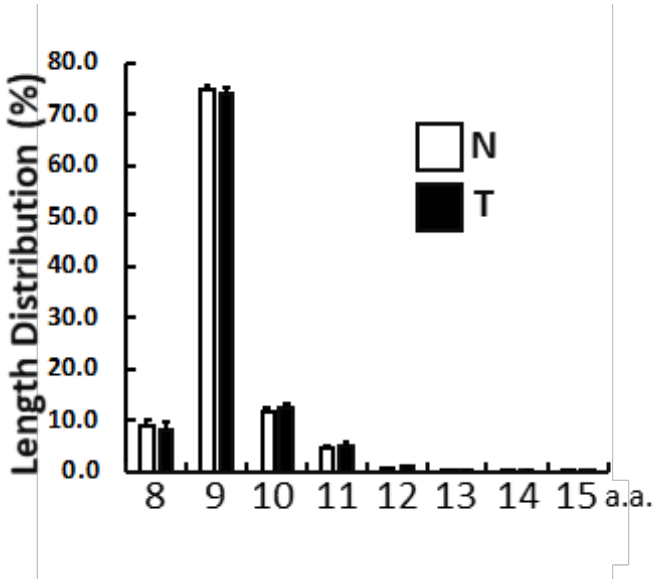

ID
Status
Details
HLA-A
HLA-B
HLA-C
HLA Assignment:
Length Distribution
260
HLA Genotypes
A*02:07
A*24:02
B*46:01
B*54:01
C*01:02
Normal
Tumor
260N
HLA Assignment:
260T
260N/T
(%)
N
HT
100.0
80.0
60.0
40.0
20.0
0.0
8 aa
9 aa
10 aa
11 aa
12 aa
13 a
14 aa
15 aa



Supplementary Figure 3. Individual dataset for HLA allotypes and identified HLAp

HLA allotype and its motif sequence, unsupervised clustering of actually identified HLAp, HLAp assignment to corresponding HLA allotypes, and the length distribution of HLAp were shown individually. The details of identified HLAp in each patient was validated by using Gibbsclustering2.0, NetMHCpan. The length of HLAp were compared between normal and tumor tissues in identical patients.

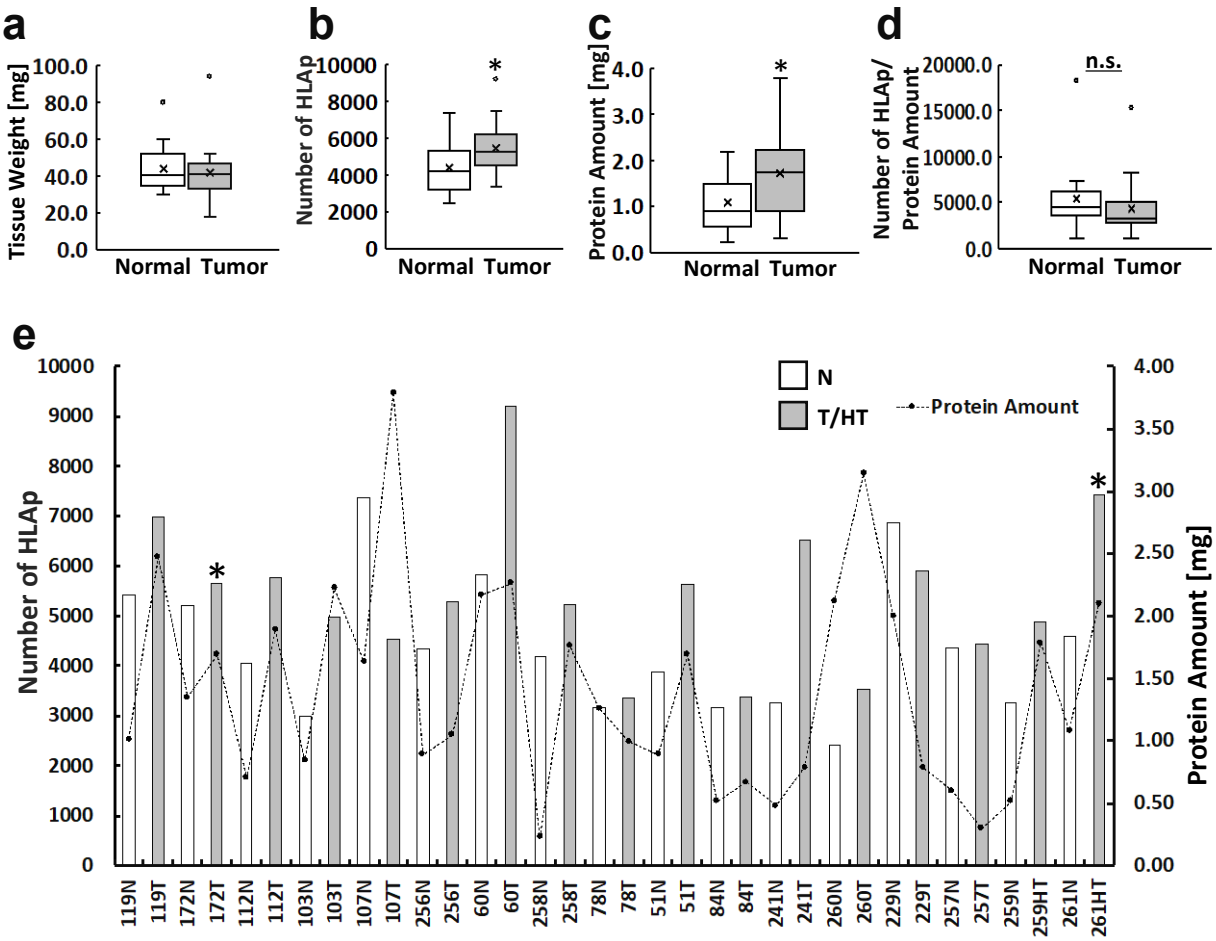
Supplementary Figure 4



Supplementary Figure 4. Comparison of personalized immunopeptidome analysis for CRC tissue samples

Length distribution of HLAp shown in Figure 2a.

Supplementary Figure 5



Supplementary Figure 5. Sample conditions and their correlation with identified HLAp

Box plots depict **a.** net weight of tissues, **b.** total number of identified HLAp, **c.** protein amount included in lysate and **d.** the number of identified HLAp normalized by protein amount. There was no statistically significant increase of HLAp presentation in tumor tissues compare to the normal tissue when the identified number of HLAp was normalized by protein amount in the starting sample of lysate. **e.** A composite graph of depicts the identified number of HLAp in each samples (white bars indicates normal tissues, gray bars indicates tumor tissues) and the protein amount used. The identified number indicates the variation of HLAp and is vary from sample to samples. Asterisk indicates the sample in which neoantigen was identified respectively.