Immune-based mutation classification enables neoantigen prioritization and immune feature discovery in cancer immunotherapy

Peng Bai[1], Yongzheng Li[1], Qiuping Zhou[1], Jiaqi Xia[1], Min Wu[2], Sanny K. Chan[3,4,5], John W. Kappler[3,6,7,8], Yu Zhou[1], Philippa Marrack[3,6,9], Lei Yin[1]

1 State Key Laboratory of Virology, Hubei Key Laboratory of Cell Homeostasis, College of Life Sciences, Wuhan University, Wuhan, 430072, Hubei, China

2 Hubei Key Laboratory of Cell Homeostasis, College of Life Sciences, Wuhan University, Wuhan, 430072, Hubei, China

3 Department of Biomedical Research, National Jewish Health, Denver, CO 80206, USA

4 Department of Pediatrics, University of Colorado Denver School of Medicine, Aurora, CO 80045, USA

5 Division of Pediatric Allergy-Immunology, National Jewish Health, Denver, CO 80206, USA

6 Barbara Davis Center for Childhood Diabetes, University of Colorado, Aurora, CO 80045, USA

7 Department of Immunology and Microbiology, University of Colorado School of Medicine, Aurora, CO 80045, USA

8 Structural Biology and Biochemistry program, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA

9 Department of Biochemistry and Molecular Genetics, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA

Abstract

Evidences have suggested that T cells that target mutation derived neoantigens are the main mediators of many effective cancer immunotherapies. Although algorithms have been used to predict neoantigens, only a handful of those are truly immunogenic. It is unclear which other factors influence neoantigen immunogenicity. Here, we classified clinical human neoantigen/neopeptide data based on their peptide-MHC binding events into three categories. We observed a conserved mutation orientation in anchor mutated neoantigen cohort after classification. By integrating this rule with existing prediction algorithm, we achieved improved performance of neoantigen prioritization. We solved several neoantigen/MHC structures, which showed that neoantigens which follow this rule can not only increase peptide-MHC binding affinity but create new TCR binding features. We also found neoantigen exposed surface area may lead to TCR bias in cancer immunotherapy. These evidences highlighted the value of immune-based classification during neoantigen study and enabled improved efficiency for cancer treatment.

Keywords: cancer immunology; major histocompatibility complex (MHC); vaccine; neoantigen prediction; antigen immunogenicity

Introduction

The immune system can sometimes recognize and destroy cancer cells[1]. $CD8^+$ T cells are a major component in this process by both recognizing and destroying the target cells[2–5]. For T cells to react with tumor cells two components are required: major histocompatibility complex class I and possibly class II proteins (MHCI, MHCII) on the cancer cells and tumor derived antigenic peptides which bound to MHCs. T cells can then recognize such complexes, called pMHC epitopes (the combination of MHC and antigen), on the cancer cells as abnormal cells and target them for destruction. Thus, it is crucial to identify cancer antigens for immunotherapy.

During the last 25 years, great efforts have been made to identify tumor antigens[6]. These tumor antigens can be classified into two broad categories, self-antigens which are seldom expressed in normal tissues or expressed at much higher levels on the cancer cell versus normal tissues (like NY-

ESO-1, MART-1) and nonself-antigens called neoantigens which are derived from somatic mutations in cancer cells[7]. Neoantigens are not presented by normal tissues, hence the immune system is not tolerant to them and views them, correctly, as foreign antigens and responds appropriately[8,9].

Currently, neoantigen identifying methods relies on next-generation sequencing (NGS) to provide formidable identification of cancer mutations, followed by predicting their theoretical binding for the patient's HLA complex[10–14][15,16]. In brief, they compared tumor and normal DNA for identifying non-synonymous mutations and used peptide/MHCI binding prediction algorithms to find mutated peptides which could bind MHC to be potential effective neoantigens. Binding prediction could lighten the burden of immunotherapy by reducing the number of candidate neopeptides[7,10,12,16–19,19–22,22–25]. However, among neoantigens with actual MHC binding, only a small portion of them are immunogenic and have therapeutic effects[26,27]. It is likely that other features beyond MHC binding affinity could affect neoantigen immunogenicity.

It has been hypothesized that the TCR:pMHC ternary binding events may influence neoantigen immunogenicity. Yadav *et al*. and Fritsch *et al*. suggested that the side chain of neoantigen mutation point towards the T-cell receptor (TCR) would be more immunogenicity while Duan *et al*. thought that neoantigen substitutions at MHC anchor positions may be more important[7,18,28]. Recently, Capietto *et al.* also suggested that the MHC binding affinity of neoantigen corresponding wild type peptide is associated with predicting cancer neoantigens, with their mouse model study[29]. However, studies in this area are still limited due to scarce large human neoantigen datasets and neoepitope structures availability. Thus, systematic analyses based on large human neoantigen data and structural understanding of neoantigen immune properties are still needed.

Here, we attempted to determine different features between immunogenic neoantigens and noneffective neopeptides after mutation positional classification of human clinical neoantigens. We found that almost all immunogenic anchor mutated neoantigens in our datasets followed a unique mutation pattern (termed NP rule), rather than other patterns. Combining this rule with existing binding predictor could improve the neoantigen prioritization performance.

To provide structural insights of neoantigens with the NP rule, we solved several pMHC structures which follow this rule: the driver mutation derived KRAS G12D neoantigens in complex with HLA-C*08:02 and the noneffective mouse DPAGT1 peptides in complex with mouse MHC H-2 Kb. We

showed that neoantigens which follow the NP rule could generate new immune features for T cell recognition. The newly generated surface is necessary for anchor mutated cancer peptides to induce T cell response *in vivo*. It also suggested that the antigen exposed surface area might be an additional factor involves in neoantigen immunogenicity.

Results

Classification of cancer neoantigen/neopeptide data by mutation position

To study immune features of clinically relevant neoantigens arising from cancer mutations, we obtained HLA-I immunogenic (positive data, termed neoantigen) and noneffective (negative data, termed neopeptide) human clinical cancer neoantigen/neopeptide data from our NEPdb database (unpublished). These data come from 34 papers which contain neoantigen/neopeptide sequences and relevant HLAs as well as *in vitro* T cell assay reports and clinical tests[5,30,10–12,16,21,31–41,17,42,43,19,44–49,22,50–52].

The Immunogenic Neoantigen Dataset (IND) in this study, contains 128 entries (Fig. 1a; Supplementary Table 1). Neoantigens without T cell assay, minimal length or relevant HLA information were not recorded in the IND datasets. Melanoma and non-small cell lung cancer, with higher mutational burdens and more extensively studies, contained nearly 80% entries in this dataset.[24].

In this study, 11739 noneffective neopeptides with corresponding MHCIs formed the Raw Noneffective Neopeptide Dataset (RNND) (Fig. 1a; Supplementary Table 2). All of these mutant peptides had been tested by T cell assays at least or tested in clinical trials, which were proven to be noneffective. The majority of them are not minimal peptides lengths (MHCI bind peptides that are predominantly 8-11 amino acid in length) known to present by MHCI. As we know, peptide-MHC binding is necessary albeit not sufficient for neoepitope immunogenicity. To process noneffective neopeptides into optimal length (9mer and 10mer) as negative control for further study, we used the NetMHCpan 4.0 algorithm for peptide-MHC binding prediction. This processed dataset is termed Noneffective Neopeptide Dataset (NND) and contains 2883 entries (Fig. 1a; Supplementary Table 3).

Neoantigen with strong binding affinity to MHC can form stable pMHC complex for T cell recognition

107 and therefore induce T cell responses. Thus, neoepitope candidates could be prioritized through
108 prediction algorithms to eliminate those peptides with weak binding affinity to MHCI. While the
109 binding prediction algorithms have used for eliminating candidate neoantigens, the prediction accuracy
110 of neoantigens capable of eliciting efficacious antitumor responses in patients remains quite low[53]. It
111 is likely that features beyond MHC binding affinity involve in neoantigen immunogenicity. Recently,
112 some studies found that the TCR:pMHC ternary binding events may influence immunogenicity[54,55].
113 However, the underlying immunological mechanisms of how the positional mutations affect cancer
114 clinical outcome remains poorly defined.

115 In an attempt to answer above questions, we classified peptides from IND and NND based on the
116 mutation position. To be specific, peptides from IND and NND were categorized into three categories:
117 mutated at an anchor position (binds MHC intensively), MHC-contacting position (contacts MHC but
118 provides less binding affinity) and TCR-contacting position (point towards TCR instead of MHC). The
119 classification was performed by referencing the SYFPEITHI database[56], NetMHCpan antigen
120 binding motif [24] and the Protein Data Bank (PDB) pMHC structures into account (see Methods). The
121 classified information was recorded in Supplementary Table 1 and Supplementary Table 3.

122 We next calculated the percentage of neoantigens in different categories after classification.
123 Immunogenic neoantigens were more likely to mutate at TCR-contacting region (Fig. 1b, c) and were
124 less likely to change at MHC-contacting regions than those noneffective neopeptides (Fig. 1b, c). These
125 results suggested that neopeptides with TCR-contacting position mutation rather MHC-contacting
126 position will preferentially be immunogenic[7,28]. However, the percentage of anchor mutated
127 neoantigens did not show significant differences compared with noneffective neopeptides (Fig.1c). It
128 suggested that classification of anchor mutated peptides is not sufficient to distinguish immunogenic
129 neoantigen from noneffective neoantigen candidates.

130

131 **Anchor mutated neoantigens followed a conservative mutation pattern to acquire**
132 **immunogenicity**

133 We aimed to further characterize the intrinsic immunological properties of anchor mutated neoantigens
134 beyond binding affinity. MHC molecules have many allelic variants with different binding properties of
135 their peptide binding cleft. Thus, the peptides recognized by different MHCs are very diverse, with

136 allele-specific amino acid preferences. To investigate the preferential property of the anchor mutated
137 peptide and its wild type counterparts in the context of HLAs, we set the cut-off frequency to be above
138 10% as HLA preferential amino acids at anchor positions as described previously[57] based on large
139 peptide binding datasets from the IEDB[58] (Supplementary Fig. 1, Supplementary Table 1 and 3).
140 With the definition of "preference", the anchor mutated neopeptide pairs (contain wild-type and mutant
141 peptides) can be divided into 4 groups : non-preferential to non-preferential residues (mutated from
142 wild-type counterparts with non-preferential amino acid to mutation with non-preferential amino acid,
143 within the context of the certain HLA, abbreviated as NN), non-preferential to preferential residues
144 (NP), preferential to non-preferential residues (PN) and preferential to preferential residues (PP) (Fig.
145 1d; Supplementary Table 4). Statistical analysis showed that the frequency of the immunogenic pairs
146 with NP preferential change was significantly higher than those from the noneffective pairs (Fig. 1d,
147 p=8.247e-06). Almost all (26 in 27) of the immunogenic pairs (from IND) appear to be classified into
148 the NP group. In contrast, cancer neopeptides with HLA non-preferential property (the NN or PN
149 groups) or those wild-type counterparts with HLA preferential property (the PN or PP groups) can
150 hardly induce effective anti-tumoral response (Fig. 1e; Supplementary Table 4).

151 Next, we checked the exceptional case (the one which did not follow the NP rule in 27 cases). This
152 neoantigen (MYADM R30W) from the MYADM protein has a mutation at C termini of the peptide that
153 changes Arg to Trp. Of note, both wild type and mutant MYADM peptides can elicit self T cell
154 responses[31]. This evidence suggested that, in this case, wild type peptide can also bind the patient's
155 HLA protein and eliciting T cell autoreactivity *in vivo*. Since both the MYADM neoantigen and its
156 wild-type counterpart were reactive in this patient, this antigen should be excluded for further use.
157 Notwithstanding this exception, our observations suggested the NP rule of anchor mutated neoantigens
158 is a conservative feature to assess T cells for reactivity against neoepitopes.

159 The NP rule of anchor mutated neoantigens thus can be treated as a binary variable (1= true NP, 0=
160 false NP). To assess whether this variable can be used in neoantigen candidate prioritization, we tested
161 the performance of the binary "NP" model and the combination model of the binary "NP" model with
162 NetMHCpan 4.0 Rank% model (termed Com NP+B). Two existing prediction models, the NetMHCpan
163 4.0 (using NetMHCpan 4.0 Rank% score) and the DAI models (differential agretopic index, the
164 difference of predicted binding affinity between the mutated epitope and its unmutated counterpart)
165 were also tested with our data as benchmark comparators. After taking the NP rule into account, the

166 “Com NP+B” model achieved better performance compared with other three models (Fig. 1f,
167 AUC=0.810). When comparing the performance of the four predictors, assessment was also done by
168 50-fold cross-validation (2/3rd random resampling) over the data to check whether the observed
169 difference in average AUC differs significantly (Supplementary Fig. 2). Collectively, we proved that
170 the NP rule can be taken as a predictive feature to improve neoantigen candidate prioritization.

171

172 **Anchor mutated neoantigens can generate new surface for T cell recognition**

173 Our results above indicated that the NP rule is a conservative feature in the immunogenic anchor
174 mutated neoantigens. To understand how this rule forms the basis for acquired immunogenicity of
175 anchor mutated neoantigens, we attempted to solve the structures of the typical neoepitopes: HLA-
176 C*08:02 (C08) in complex with KRAS mutation derived neoantigens form a successful clinical case,
177 reported by Tran *et al*[10]. In this case, objective regression of metastases was observed in a metastatic
178 colorectal cancer patient after the infusion of HLA-C*08:02-restricted tumor-infiltrating lymphocytes
179 (TILs) targeting KRAS G12D neoantigens. The KRAS G12D mutant 9mer peptide GA<u>D</u>GVGKSA and
180 10mer peptide GA<u>D</u>GVGKSAL (both mutated at position 3 from glycine to aspartic acid) can stimulate
181 autologous HLA-C*08:02 reactive T cells, while their wild-type counterparts cannot[10]. These two
182 neoantigens can be categorized in the anchor mutated group (Supplementary Table 1 and reference[59])
183 and follow the NP rule.

184 First, we attempted to get the protein complexes of HLA-C*08:02 (C08) in complex with four
185 peptides: wild type KRAS 9mer GAGGVGKSA (wt9m); wild type KRAS 10mer peptide
186 GAGGVGKSAL (wt10m); mutant KRAS G12D 9mer peptide GA<u>D</u>GVGKSA (mut9m, mutation site
187 is indicated with underline) and mutant KRAS G12D 10mer peptide GA<u>D</u>GVGKSAL (mut10m). C08-
188 mut9m and C08-mut10m complexes were successfully refolded. However, C08-wt9m complex and
189 C08-wt10mer complex failed to refold even with ten-fold increase in concentration of the peptides.
190 This result confirmed that the mutation from glycine to aspartic acid of these neoantigens is crucial to
191 stabilize pMHC complex.

192 We next solved the C08-mut9m complex at 2.4 angstroms (Å) and the C08-mut10m complex at 1.9 Å
193 (Supplementary Table 5). The electron density at the peptide region was unambiguous (Fig. 2a, b). The
194 C08-mut9m complex and the C08-mut10m complex showed a conserved conformation except the

195 peptide region (Fig.2c, d). The smaller residues such as alanine and serine are preferably selected at
196 peptide P1 and P2 position by the peptide binding motif of HLA-C*08:02 (Fig.2e, Supplementary Fig.
197 1). This is due to the narrow cleft formed by several C*08:02 aromatic residues (Tyr7, Phe33, Tyr67,
198 Tyr99, Tyr59, Tyr171, Tyr159 and Trp167) which limits the size of the P1 and P2 side chains that can
199 bind at that site (Fig.2f).

200 Generally, peptides use anchor positions at P2 and PΩ to occupy the B and F pockets of HLA class I
201 molecules. However, the structures of the C08-mut9m and C08-mut10m, followed by the peptide
202 binding analysis, revealed that KRAS G12D 9mer and 10mer peptides use unconventional P3 and PΩ
203 sites as anchors to form stable complexes with HLA-C*08:02 (Fig. 2g, h; Supplementary Fig. 1). The
204 B pocket of HLA-C*08:02 bind P3D via interactions with Arg97 and Arg156. The P3D side chain of
205 C08-mut9m also forms an intra-chain hydrogen bond with P4G, while this hydrogen bond is absent in
206 the C08-mut10m structure. Interestingly, it is apparent from the C08-mut9m structure that the anchor
207 residue P3D could also provide TCR accessible surface by partially exposing of its charged side chain
208 (Supplementary Fig. 3). This unusual phenomenon suggested that anchor mutation of some neoantigens
209 can not only improve pMHC binding force but could also provide additional accessible surface for
210 TCR interaction, under certain conditions, and therefore change the total strength of the TCR:pMHC
211 complex.

212 At the PΩ anchor position, although the 9mer and 10mer peptides occupied the same F pocket with
213 their PΩ residues, the P10L side chain from the 10mer peptide is buried more deeply than P9A in the
214 9mer structure (Fig. 2g, h). Meanwhile, instead of the upward residue P8S in mut9m peptide, the
215 mut10m peptide uses P8S as an auxiliary anchor to bind the HLA protein via hydrogen bonds to
216 Glu152 and Arg156. These interactions from the mut10m squeezed the P4 to P7 residues together with
217 residue P3D (Fig. 2h). Although the mut9m and the mut10m have similar sequences with only one
218 amino acid different at PΩ termini, the surface details are largely different after they engage with HLA-
219 *08;02. We postulated that these different features from the mut9m and mut10m neoepitopes may
220 respectively modulate the effective T cell recognition and activate different T cell repertoires (discussed
221 below).

222 Noneffective anchor mutated neoantigens could not generate more differentiated TCR binding surface
223 than their wild-type analogues, as they buried the anchor mutations into MHC pocket and can hardly

224 contact with TCRs. However, structural analysis of immunogenic anchor mutated neoantigens
225 suggested that these neoantigens can not only generate new immune features, but even create novel
226 neoepitope surface. Thus, the necessity and conservation of NP rule in anchor mutated neoantigen
227 might be explained by the generation of novel surface from mutant peptide, rather than the wild-type
228 analogue, which could enable the boosting of neoepitope-specific response from T cells.

229

230 **Structural of non-therapeutic mouse neoantigen DPAGT1 V213L in complex with H-2 Kb**

231 We next determined two structures of non-immunogenic neopeptides with anchor mutation. Yadav *et al*
232 described a neopeptide that can be presented by MHC but showed non-immunogenic property *in*
233 *vivo*[7]. This mouse DPAGT1 V213L neopeptide contains a mutated C-terminal anchor residue that
234 falls into the "preferential to preferential residues (PP)" group, with the changing of valine(V) to
235 leucine(L).

236 Soluble mouse H-2 Kb in complex with mutant DPAGT1 V213L 8mer peptide (SIIVFNLL, termed
237 mut8mL) and the wild type 8mer counterpart (SIIVFNLV, termed wt8mV) were separately expressed,
238 refolded and purified for crystallization trials. Crystal diffraction data of Kb-wt8mV and Kb-mut8mL
239 were processed to 2.4 Å and 2.5 Å resolution respectively (Supplementary Table 5) and provided
240 electron density for each peptide (Fig. 3a).

241 The overall structure of the Kb-wt8mV complex closely resembles that of Kb-mut8mL with the
242 exception of a slight difference at PΩ (P8) (Fig. 3b). The C-terminal PΩ residue acted as an anchor in
243 both the Kb-wt8mV and the Kb-mut8mL complexes (Fig. 3b). Moreover, both PΩ valine and leucine
244 were preferably selected by H-2 Kb (Fig. 3c). Both of the two PΩ residues formed hydrogen bonds
245 with Kb Asp77, Tyr84, Thr143 and Lys146 (Fig. 3d, e). Although the side chain of leucine in the
246 mut8mL peptide inserted more deeply into Kb than valine in the wt8mV peptide because of its longer
247 side chain, these two peptides did not provide different TCR binding surface with relevant MHCs (Fig.
248 3b, d, e). These findings indicated that neopeptides with the "PP" rule cannot readily change the
249 binding surface and therefore immunogenicity. The non-effectiveness across the neopeptides with PP
250 rule might be explained by the pre-existence of the wild-type peptide-MHC complex in thymus, which
251 can lead to negative selection of potential neopeptide restricted T cell repertoires. Considering that
252 peptide-MHC binding is necessary for neoepitope immunogenicity, we did not further discuss the

253 situation of those neopeptides with the “PN” or the “NN” rule.

254

255 **Neoantigen exposed surface areas may affect T cell selection in cancer immunotherapy**

256 Peptide antigen can form stable complexes with HLA by lying the peptide chain into the HLA antigen-
257 binding cleft. Some antigens, called “featureless” antigens, have relatively small exposed surface areas
258 (ESA) form side chains pointing towards the T cell receptor when they fill the peptide-binding cleft of
259 HLAs[60]. Studies have indicated that featureless epitope are more likely to select relatively narrow
260 TCR repertoires than epitopes with large exposed features *in vivo*[61]. We next examined the ESA
261 features of the C08-mut9m and the C08-mut10m structures, by employing the PDBePISA server. Of
262 note, the ESA of two representative T cell epitopes were also calculated as benchmarks. One is the
263 HLA-A2–M1, a viral antigen “M1” ($M1_{58-66}$ from the IAV) in complex with HLA-A*02:01(A2-M1, in
264 Fig. 4a), which considered as a featureless epitope[62]. In contrast, a viral epitope HLA-A2-RT, which
265 has a “reverse transcriptase peptide” ($RT_{468-476}$ from HIV) in complex with HLA-A*02:01(called A2-
266 RT, in Fig. 4a), is considered as largely exposed epitope[63]. After calculation, we found that the
267 mut9m neoantigen has the smallest peptide ESA at 240 $Å^2$, even less than the well-known featureless
268 M1 peptide (251 $Å^2$, in Fig. 4a-c). However, the mut10m neoantigen has relatively large ESA at 317
269 $Å^2$, which is comparable with the typical largely exposed antigen RT (330 $Å^2$, in Fig. 4a-c). These data
270 suggested that the mut9m provides relatively less SEA than canonical T cell antigens. We thus
271 postulated that the specific TCRs for C08-mut9m may be constrained in patients because of the
272 featureless area available for specific recognition.

273 Studies have suggested that narrow TCR repertoires can recognize featureless epitope, because of the
274 lack of TCR recognition modes[61,62,64]. To investigate the diversity of KRAS G12D neoantigen
275 specific TCRs in clinical cases, we examined the TCR sequences of the restricted T cell repertoire
276 targeting the C08-mut9m neoepitope (Fig. 4d, patients 3995 and 4095, both expressing HLA-
277 C*08:02)[10,44]. Patient 3995 received ACT targeting KRAS G12D neoantigen and did not response.
278 The transferred RK5 T cell repertoire (T cells expressed the RK5 TCR) was identified to recognize
279 C08-mut9m neoepitope. Patient 4095 received a similar treatment and observed objective tumor
280 regressions. In patient 4095, the RK1, RK3, RK4 T cell repertoires were verified to recognize the C08-
281 mut9m while the RK2 recognized the C08-mut10m. All four T cell repertoires (RK1, RK3, RK4, RK5)

282 with C08-mut9m restriction were identified biased usage with a public TCR pair (TRAV4/TRBV5-1)
283 across different patients. The length and sequence of these TCRα chains was highly restricted, with the
284 same TCR-V region and "CLVGDxDQAGTALIF" CDR3α motif among the four TCRs (Fig. 4d). The
285 TCRβ chains was also restricted at TCR-V region but showed differences at CDR3β regions. Generally,
286 the CDR1 and CDR2 loops of TCR can recognize the two conserved α-helixes on the MHC, whereas
287 the CDR3 loops mainly interact with the exposed peptide. Moreover, the CDR3β had proved to be the
288 main factor that determines TCR bias (compared with CDR3α) in many cases, due to the greater
289 sequence diversity in TCR repertoire and the extensive contact of peptide region[64]. However, in this
290 case, similar CDR1, CRR2 and CDR3α sequences with C08-mut9m restriction were identified
291 consistently from different patients. We thus speculation that these public CDR1, CRR2 and CDR3α
292 regions are important in the C08-mut9m recognition, rather than the CDR3β regions[65]. In contrast,
293 the C08-mut10m neoepitope did not observe dominant public TCRs in patient 4095 or across different
294 individuals. Collectively, these data suggested that the featureless mut9m neoepitope can be recognized
295 by T cells with public TCRs across different patients.

296 Previous reports have shown that the constrained TCR repertoires are associated with poor efficient to
297 control viral infection[66,67]. However, the correlation between TCR bias and clinical outcome in
298 cancer treatment is unclear. In an exploratory analysis, we examined the TCR bias and clinical
299 performance in this case to address above question. The C08-mut9m restricted T cell repertoires with
300 public TCR usage (RK1, RK3 RK4 and RK5) did not dominant presence in TILs or not show long-
301 term persistence after infusion (Fig. 4d, detected after 39 and 266 days). However, the RK2 T cells with
302 C08-mut10m restriction showed dominant persistent abilities both in tumor infiltrating stage and after
303 cell transfer for 266 days. While we did not observe direct correlation between TCR diversity and
304 clinical outcome due to the limitations of clinical data, we still observed short-lived T cell persistence
305 in the presence of the featureless C08-mut9m neoepitope. This phenomenon might be explained due to
306 the existence of cross-reactivity with self antigen and thereby leads to the suppression by regulatory
307 effectors *in vivo*. We thus postulated that, outcomes of patients who receive more diverse adoptive T
308 cells tended to be better than patients who receive constrained T cells in cancer immunotherapy.

309

310 **Discussion**

311 How the T cells recognize neoantigen as “non-self” is an important question in cancer immunotherapy.
312 In contrast to conventional pathogenic peptides that may totally different from self peptides, the
313 neoantigens are single amino acid altered peptides compare with self peptides. Thus, differences
314 between neoantigens and their wild-type counterparts and how neoantigen-MHC binding events
315 involve in forming new TCR binding surface should be elucidated in neoantigen immunogenicity
316 studies.

317 Efforts have devoted to discuss the complexity of neoantigen immunogenicity with their binding
318 events. Different pieces of evidence led to different conclusions[68]. Fritsch *et al*. and Yadav *et al*.
319 suggest that neoantigen are more commonly mutated at TCR-contacting position, however, Duan
320 *et al*. thought that neoantigen substitutions at anchor may be more immunogenic[7,18,28]. Further,
321 using mouse model, Capietto *et al*. showed that where the mutation is at an anchor residue, increased
322 affinity relative to the corresponding wild-type peptide can influence neoantigen immunogenicity[29].
323 To elucidate the questions above, we assigned human neoantigen/neopeptide data into three different
324 categories: mutations at TCR-contacting positions, mutations at MHC-contacting regions but not
325 anchor positions and mutations at anchor positions. In our study, mutations occur more frequently at
326 TCR-contacting positions rather than at MHC-contacting position in immunogenic dataset. It is
327 possible that mutation characteristics, amino acid contact potentials and force-dependent interactions
328 may affect the interactions in the TCR-contacting group[69,70]. For anchor mutated group, we showed
329 that the NP rule in anchor mutated neoantigen is a more pervasive element of immunogenicity than
330 previously understood. Also, we found the NP rule combined with binding predictor (NetMHCpan 4.0)
331 that could contribute to prioritize neoantigen candidates. In a sense, the NP rule could be taken as a
332 reflection of antigen binding property. This binary indicator can provide a direct understanding of the
333 MHC binding difference between mutant peptides and their wild type counterparts. However, the
334 uniqueness of the NP rule was not fully determined, since the test depends on how comprehensive the
335 validation database is. More analyses to understand the NP rule are still needed with the data
336 increasing.

337 We next performed structural studies to understand the NP rule. The molecular insight of anchor
338 mutated neoantigens provided evidences that they can generate new surface and features for T cell
339 recognition but the wild-type cannot. In contrast, the anchor mutated neopeptides with PP rule reveal
340 low immunogenicity in clinical treatment, reflecting that the most neoepitope restricted T cells might

341 have been removed by negative selection. It also suggested that neoantigen exposed surface area (ESA)
342 might be a factor to influence TCR diversity and clinical outcome based on our analysis. More
343 experimental data and neoantigen-MHC structures are needed to fully understand the relevance of ESA
344 and TCR diversity.

345 Our study showed three possible neoantigen binding models within the context of MHC
346 (Supplementary Fig. 4). Model A represents the situation which mutation occurs at the TCR-contacting
347 region and therefore directly towards to T cells. Model B represents the situation in which mutation
348 occurs at MHC-contacting region (not including anchor sites) and therefore might be immunogenically
349 irrelevant. Model C represents the situation which the mutation occurs at an anchor position. Anchor
350 mutations may not change TCR-contacting surface but instead lead to de novo presentation. If the wild
351 type allele prevented presentation in the thymus, self-reactive T cells would not have been selected
352 against.

353 KRAS G12D mutations is indicative of poor prognosis with negative/poor response to standard cancer
354 treatment. It is one of the most infamous driver mutations target leads to oncogenesis[71]. The
355 neoantigen KRAS G12D in complex with HLA-C*08:02 is a typical case of a human driver mutation
356 derived neoantigen-MHC structure which has been linked to clinical benefits. With these structures,
357 further research could be undertaken to heighten the immunogenicity and stability of KRAS G12D-
358 C*08:02 neoepitope, by taking modification of agonist peptides or by screening non-natural synthetic
359 epitopes[72,73]. Alternatively, our structures could be used to design artificial receptors that bind
360 mutant KRAS peptides based on synthetic biology means. Our findings reveal that immune-based
361 classification is essential for neoantigen immunogenicity study The NP rule, which found in anchor
362 mutated neoantigens, can be used to prioritize neoantigens. Further structural analyses indicated that
363 newly generated surface as well as the ESA in neoantigens affected T cell activation and clinical
364 outcome, suggesting that these factors could be considered in neoantigen discovery and design of
365 future clinical trials.

366

367 **Materials and Methods**

368 **Noneffective neopeptides dataset (NND) generation**

369 In order to generate a minimal noneffective neopeptide dataset, binding prediction was done by
370 NetMHCpan 4.0 Server[24] based on the data from noneffective neopeptide dataset (NND). Different
371 length of mutant peptides from NND were input in NetMHCpan 4.0 Server with custom python scripts.
372 The mutant peptides in 9- or 10-mer length with recommend cut-off (affinity < 500 nM) based on
373 NetMHCpan 4.0 Server were considered to be binders and recorded in the Supplementary Table 3.

374

375 **Differential agretopic index (DAI) score calculation**

376 The calculation of DAI was described previously[74]. Briefly, peptide binding affinity with HLA was
377 predicted for mutant and wildtype peptide pairs by NetMHCpan 4.0. The DAI score of each neoantigen
378 pair was calculated by subtraction of its predicted IC50 binding affinity from the corresponding wild
379 type counterparts.

380

381 **Receiver operating characteristic (ROC) curve generation**

382 The AUC value (the area under the ROC Curve) was calculated based on the different predictors listed.
383 The ROC curve was plotted from the false positive rate (FPR) and true positive rate (TPR) values
384 calculated by varying the cut-off value (separating the predicted positive from the predicted negative)
385 from high to low.

386

387 **Anchor, MHC-contacting and TCR-contacting positions determination**

388 Identification of the anchor positions were based on the SYFPEITHI online database[56] and manually
389 defined from NetMHCpan binding motif results[24]. Anchors were defined for each allele with the
390 SYFPEITHI database definition and highest information content in NetMHCpan binding motif record.
391 The anchor position for each entry was cross-validated based on solved HLA structures from Protein
392 Data Bank (PDB). Thus, the combination of these tools can provide anchor position information for
393 most HLA alleles in our datasets. We recorded the cross-validated result as the “consensus anchor”.
394 The anchor position information was also recorded in IND and NND. Alleles without relevant
395 information above were recorded as “null” in the tables.

Peptide MHC-contacting and TCR-contacting positions was determined based on solved 9mer peptide-MHC complex. Briefly, positions from peptides which prove to be non-anchor position can be divided into MHC-contacting and TCR-contacting positions. With the pMHC structural model from PDB, the position on peptide which contact MHC was treated as MHC-contacting position. Contrary to MHC-contacting position, TCR-contacting position often harbor a residue with a side chain that points toward outside from the pMHC complex and may contact with TCR. The MHC contact and TCR contact region information was recorded in IND and NND. Alleles without relevant structure information above were recorded as "null" in the tables.

Peptide library and preferential HLA anchor position determination

The nonameric peptide library of 30 HLA alleles was obtained from IEDB database[58]. Sequence logos were generated base using the sequence logo generator[75]. Threshold for preferential amino acids at anchor position of each HLA alleles was set to include and above 10% based on the data from nonameric peptide library. Preferential information of amino acids at anchor position for each HLA allele was recorded in Supplementary Table 1 and 3.

Combined NP+Binding (Con NP+B) prediction model building

To combined the binary NP rule with existing prediction methods NetMHCpan 4.0, we took a logistic regression algorithm to model the prediction of immunogenicity. Analyses were performed using the R build-in function glm(), as below:

$$\text{immunogenicity} \sim \text{NP} + \text{Binding prediction}.$$

Anchor mutated neoantigen data was selected to train the model. After fitting, the performance of this model was shown by ROC curve.

To further test this model, we used 50 times resampling. Random resampling of the data (2/3rd resampling) was used for training. The AUC values were calculated by plotROC package. After iteration, the differences of AUC between four models were measured using paired t test in R software.

423 **Protein Expression, refolding and purification**

424 Inclusion bodies of HLA heavy chains and β2M was expressed as described previously[76]. Briefly,
425 The DNA encoding MHC heavy chain (HLA-C*08:02, HLA-C*05:01 and H-2 Kb) and light chain
426 (human β2M and mouse β2M) were synthesized (Idobio) and cloned into pET-22(b) vector (Novagen).
427 The vectors were transformed into the E. coli strain BL21 DE3 (Novagen). Transformants were
428 selected fromand selected on Lurian broth (LB) agar plates containing ampicillin. A single colony was
429 selected and cultured in LB fluid medium with the antibiotics listed above at 37°C. Upon reaching an
430 optical density OD600 of 0.6, expression was induced with the addition of 1mM IPTG. Incubation
431 continued at 37°C for 5h. The cells were harvested by centrifugation and then resuspended in PBS
432 buffer with 1 mM PMSF at 4°C. The cells were lysed, and the lysate was clarified by centrifugation at
433 10,000 g to collect inclusion bodies. Inclusion bodies were harvested and solubilized in 20 mM Tris
434 (Vetec) pH 8.0, 8 M urea (Vetec), 1 mM EDTA (BBI life sciences), 1 mM DTT (Sinopharm chemical
435 reagent) and 0.2 mM PMSF (Sinopharm chemical reagent).

436 Refolding was performed in the presence of MHC heavy chain, β2M and peptides as described
437 previously[77]. Briefly, the resolubilized heavy chain (60 mg each) and light chain (25 mg each) in the
438 presence of the corresponding peptide were added into 1 liter of refolding buffer [100 mM Tris (pH
439 8.4), 0.5 mM oxidized glutathione (BBI life sciences), 5 mM reduced glutathione (BBI life sciences),
440 400mM L-arginine (Vetec), 2mM EDTA (BBI life sciences)]. After 48h of refolding, the 1 L mixture
441 was transferred into dialysis bag (Spectra) and dialyzed against 15 liters of 10 mM Tris buffer (pH 8.0)
442 at 4°C for 24h.

443 Refolded proteins were purified by anion exchange chromatography with Q Sepharose HP (GE
444 Healthcare) column then Mono Q column (GE Healthcare) and concentrated by tangential flow
445 filtration using Amicon Ultra centrifugal filters (Merck). For desalination and purification, samples
446 were loaded onto a Superdex 200 increase 10/300 GL column (GE Healthcare) for size exclusion
447 chromatography. Chromatography was taken with BioLogic DuoFlow system (Bio-rad) at a flow rate
448 of 1 ml/min. Peak analysis was performed using the ASTRA software package (BioLogic
449 Chromatography Systems).

450

451 **Crystallization, data collection, and processing**

452 Purified pMHC complex were concentrated to 10 mg/ml for crystallization trials prior to screening
453 using a series of kits from Hampton Research. Protein complex were crystalized by sitting drop vapor
454 diffusion technique at 4 °C. Single crystals of C08-mut9m and C08-mut10m were obtained in the
455 condition of 0.2 M ammonium acetate, 0.1 M HEPES (pH 6.5), 25% w/v polyethylene glycol 3,350.
456 For the H-2 Kb complex, single crystal of Kb-8mV and Kb-8mL complex were obtained when 4% v/v
457 Tacsimate (pH 6.0), 12% w/v Polyethylene glycol 3,350 was used as the reservoir buffer.

458 Crystals were transferred to crystallization buffer containing 20% (w/v) glycerol and flash-cooled in
459 liquid nitrogen immediately. The diffraction data were collected at the Shanghai Synchrotron Radiation
460 Facility (Shanghai, China) on beam line BL17U1/BL18U1/BL19U1, and processed using the iMosflm
461 program[78]. Data reduction was performed with Aimless and Pointless in the CCP4 software
462 suite[79]. All structures were determined by molecular replacement using Phaser[80]. The models from
463 the molecular replacement were built using the COOT (Crystallographic Object-Oriented Toolkit)
464 program[81] and subsequently subjected to refinement using Phenix software[82]. Data collection,
465 processing, and refinement statistics are summarized in (Supplementary Table 5). All the structural
466 figures were prepared using PyMOL (http://www.pymol.org) program. The atomic coordinates and
467 structure factors for the reported crystal structures have been deposited on the Protein Data Bank (PDB;
468 http://www.rcsb.org/pdb/).

469

470 **Acknowledgements**

471 We thank the staff from BL17U1/BL18U1/BL19U1 beamline of National Center for Protein Sciences
472 Shanghai (NCPSS) at Shanghai Synchrotron Radiation Facility, for assistance during crystal data
473 collection. We would like to thank Dr. Eric Tran (Providence Cancer Institute, USA) for his helpful
474 comments and thank Chang-yi Ma (The Chinese University of Hong Kong, China) for advices on data
475 analysis. This research was funded by the National Institutes of Health Grants AI018785 and AI135374
476 (to P.M.); the National Natural Science Foundation of China 31870728 and 31470738 (to L.Y.); the
477 National Basic Research Program of China 2014CB910103 (to L.Y.); the Science Foundation of
478 Wuhan University 2042016kf0169 (to L.Y.).

479

480 **Author contributions**

481 P.B., L.Y., and P.M. designed all the experiments; P.B. performed all experiments; P.B. collected
482 neoantigen data; P.B., L.Y., P.M., J.W.K., M.W., Y.L., Y.Z., S.K.C., and J.X. analyzed neoantigen data;
483 P.B., and Q.Z. determined crystal structures; P.B., L.Y., and P.W. analyzed crystal structures; P.B., L.Y.,
484 P.M., J.W.K., and S.K.C. wrote the manuscript.

485

486 **Competing interests**

487 The authors declare that they have no conflict of interest.

488

489 **References**

490 [1] Vesely MD, Kershaw MH, Schreiber RD, Smyth MJ. Natural Innate and Adaptive Immunity to
491 Cancer. Annu Rev Immunol 2011;29:235–71. https://doi.org/10.1146/annurev-immunol-
492 031210-101324.
493 [2] Sykulev Y, Joo M, Vturina I, Tsomides TJ, Eisen HN. Evidence that a Single Peptide–MHC
494 Complex on a Target Cell Can Elicit a Cytolytic T Cell Response. Immunity 1996;4:565–71.
495 https://doi.org/10.1016/S1074-7613(00)80483-5.
496 [3] Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJM, Robert L, et al. PD-1 blockade
497 induces responses by inhibiting adaptive immune resistance. Nature 2014;515:568–71.
498 https://doi.org/10.1038/nature13954.
499 [4] Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. Science 2015;348:69–
500 74. https://doi.org/10.1126/science.aaa4971.
501 [5] Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational
502 landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science
503 2015;348:124–8. https://doi.org/10.1126/science.aaa1348.
504 [6] Wang R-F, Wang HY. Immune targets and neoantigens for cancer immunotherapy and
505 precision medicine. Cell Res 2017;27:11–37. https://doi.org/10.1038/cr.2016.155.
506 [7] Yadav M, Jhunjhunwala S, Phung QT, Lupardus P, Tanguay J, Bumbaca S, et al. Predicting
507 immunogenic tumour mutations by combining mass spectrometry and exome sequencing.
508 Nature 2014;515:572–6. https://doi.org/10.1038/nature14001.
509 [8] Coulie PG, Van den Eynde BJ, van der Bruggen P, Boon T. Tumour antigens recognized by T
510 lymphocytes: at the core of cancer immunotherapy. Nat Rev Cancer 2014;14:135–46.
511 https://doi.org/10.1038/nrc3670.
512 [9] Gascoigne NRJ, Rybakin V, Acuto O, Brzostek J. TCR Signal Strength and T Cell Development.
513 Annu Rev Cell Dev Biol 2016;32:327–48. https://doi.org/10.1146/annurev-cellbio-111315-
514 125324.
515 [10] Tran E, Robbins PF, Lu Y-C, Prickett TD, Gartner JJ, Jia L, et al. T-Cell Transfer Therapy Targeting

516 Mutant KRAS in Cancer. N Engl J Med 2016;375:2255–62.
517 https://doi.org/10.1056/NEJMoa1609279.
518 [11] Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal
519 neoantigen vaccine for patients with melanoma. Nature 2017;547:217–21.
520 https://doi.org/10.1038/nature22991.
521 [12] Sahin U, Derhovanessian E, Miller M, Kloke B-P, Simon P, Löwer M, et al. Personalized RNA
522 mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. Nature
523 2017;547:222–6. https://doi.org/10.1038/nature23003.
524 [13] Melief CJM. Precision T-cell therapy targets tumours. Nature 2017;547:165–7.
525 https://doi.org/10.1038/nature23093.
526 [14] Chen F, Zou Z, Du J, Su S, Shao J, Meng F, et al. Neoantigen identification strategies enable
527 personalized immunotherapy in refractory solid tumors. J Clin Invest 2019;129:2056–70.
528 https://doi.org/10.1172/JCI99538.
529 [15] Matsushita H, Vesely MD, Koboldt DC, Rickert CG, Uppaluri R, Magrini VJ, et al. Cancer exome
530 analysis reveals a T-cell-dependent mechanism of cancer immunoediting. Nature
531 2012;482:400–4. https://doi.org/10.1038/nature10755.
532 [16] Robbins PF, Lu Y-C, El-Gamil M, Li YF, Gross C, Gartner J, et al. Mining exomic sequencing data
533 to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. Nat
534 Med 2013;19:747–52. https://doi.org/10.1038/nm.3161.
535 [17] van Rooij N, van Buuren MM, Philips D, Velds A, Toebes M, Heemskerk B, et al. Tumor Exome
536 Analysis Reveals Neoantigen-Specific T-Cell Reactivity in an Ipilimumab-Responsive
537 Melanoma. J Clin Oncol 2013;31:e439–42. https://doi.org/10.1200/JCO.2012.47.7521.
538 [18] Duan F, Duitama J, Al Seesi S, Ayres CM, Corcelli SA, Pawashe AP, et al. Genomic and
539 bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer
540 immunogenicity. J Exp Med 2014;211:2231–48. https://doi.org/10.1084/jem.20141308.
541 [19] Cohen CJ, Gartner JJ, Horovitz-Fried M, Shamalov K, Trebska-McGowan K, Bliskovsky VV, et
542 al. Isolation of neoantigen-specific T cells from tumor and peripheral lymphocytes. J Clin
543 Invest 2015;125:3981–91. https://doi.org/10.1172/JCI82416.
544 [20] Carreno BM, Magrini V, Becker-Hapak M, Kaabinejadian S, Hundal J, Petti AA, et al. A
545 dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific
546 T cells. Science 2015;348:803. https://doi.org/10.1126/science.aaa3828.
547 [21] Stronen E, Toebes M, Kelderman S, van Buuren MM, Yang W, van Rooij N, et al. Targeting of
548 cancer neoantigens with donor-derived T cell receptor repertoires. Science 2016;352:1337–
549 41. https://doi.org/10.1126/science.aaf2288.
550 [22] Anagnostou V, Smith KN, Forde PM, Niknafs N, Bhattacharya R, White J, et al. Evolution of
551 Neoantigen Landscape during Immune Checkpoint Blockade in Non–Small Cell Lung Cancer.
552 Cancer Discov 2017;7:264–76. https://doi.org/10.1158/2159-8290.CD-16-0828.
553 [23] Zacharakis N, Chinnasamy H, Black M, Xu H, Lu Y-C, Zheng Z, et al. Immune recognition of
554 somatic mutations leading to complete durable regression in metastatic breast cancer. Nat
555 Med 2018;24:724–30. https://doi.org/10.1038/s41591-018-0040-8.
556 [24] Jurtz VI, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan 4.0: Improved
557 peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding
558 affinity data 2017. https://doi.org/10.1101/149518.
559 [25] Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks:

application to the MHC class I system. Bioinformatics 2016;32:511–7. https://doi.org/10.1093/bioinformatics/btv639.

[26] Vitiello A, Zanetti M. Neoantigen prediction and the need for validation. Nat Biotechnol 2017;35:815–7. https://doi.org/10.1038/nbt.3932.

[27] Bulik-Sullivan B, Busby J, Palmer CD, Davis MJ, Murphy T, Clark A, et al. Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification. Nat Biotechnol 2019;37:55–63. https://doi.org/10.1038/nbt.4313.

[28] Fritsch EF, Rajasagi M, Ott PA, Brusic V, Hacohen N, Wu CJ. HLA-Binding Properties of Tumor Neoepitopes in Humans. Cancer Immunol Res 2014;2:522–9. https://doi.org/10.1158/2326-6066.CIR-13-0227.

[29] Capietto A-H, Jhunjhunwala S, Pollock SB, Lupardus P, Wong J, Hänsch L, et al. Mutation position is an important determinant for predicting cancer neoantigens. J Exp Med 2020;217. https://doi.org/10.1084/jem.20190179.

[30] Hogan KT, Eisinger DP, Cupp SB, Lekstrom KJ, Deacon DD, Shabanowitz J, et al. The Peptide Recognized by HLA-A68.2-restricted, Squamous Cell Carcinoma of the Lung-specific Cytotoxic T Lymphocytes Is Derived from a Mutated Elongation Factor 2 Gene. Cancer Res 1998;58:5144.

[31] McGranahan N, Furness AJS, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science 2016;351:1463–9. https://doi.org/10.1126/science.aaf1490.

[32] Robbins PF. A mutated beta-catenin gene encodes a melanoma-specific antigen recognized by tumor infiltrating lymphocytes. J Exp Med 1996;183:1185–92. https://doi.org/10.1084/jem.183.3.1185.

[33] Zorn E, Hercend T. A natural cytotoxic T cell response in a spontaneously regressing human melanoma targets a neoantigen resulting from a somatic point mutation. Eur J Immunol 1999;29:592–601. https://doi.org/10.1002/(SICI)1521-4141(199902)29:02<592::AID-IMMU592>3.0.CO;2-2.

[34] Chiari R, Foury F, De Plaen E, Baurain J-F, Thonnard J, G. Coulie P. Two Antigens Recognized by Autologous Cytolytic T Lymphocytes on a Melanoma Result from a Single Point Mutation in an Essential Housekeeping Gene. Cancer Res 1999;59:5785.

[35] Echchakir H, Mami-Chouaib F, Vergnon I, Baurain J-F, Karanikas V, Chouaib S, et al. A Point Mutation in the α-Actinin-4 Gene Generates an Antigenic Peptide Recognized by Autologous Cytolytic T Lymphocytes on a Human Lung Carcinoma. Cancer Res 2001;61:4078.

[36] Guéguen M, Patard J-J, Gaugler B, Brasseur F, Renauld J-C, Van Cangh PJ, et al. An Antigen Recognized by Autologous CTLs on a Human Bladder Carcinoma. J Immunol 1998;160:6188.

[37] Zhou J, Dudley ME, Rosenberg SA, Robbins PF. Persistence of Multiple Tumor-Specific T-Cell Clones Is Associated with Complete Tumor Regression in a Melanoma Patient Receiving Adoptive Cell Transfer Therapy: J Immunother 2005;28:53–62. https://doi.org/10.1097/00002371-200501000-00007.

[38] Sensi M, Nicolini G, Zanon M, Colombo C, Molla A, Bersani I, et al. Immunogenicity without Immunoselection: A Mutant but Functional Antioxidant Enzyme Retained in a Human Metastatic Melanoma and Targeted by CD8+ T Cells with a Memory Phenotype. Cancer Res 2005:10.

[39] Lennerz V, Fatho M, Gentilini C, Frye RA, Lifke A, Ferel D, et al. The response of autologous T

604 cells to a human melanoma is dominated by mutated neoantigens. Proc Natl Acad Sci
605 2005;102:16013–8. https://doi.org/10.1073/pnas.0500090102.
606 [40] Takenoyama M, Baurain J-F, Yasuda M, So T, Sugaya M, Hanagiri T, et al. A point mutation in
607 the NFYC gene generates an antigenic peptide recognized by autologous cytolytic T
608 lymphocytes on a human squamous cell lung carcinoma: NF-YC Mutated Peptide-Reactive
609 CTL in NSCLC. Int J Cancer 2006;118:1992–7. https://doi.org/10.1002/ijc.21594.
610 [41] Ito D, Visus C, Hoffmann TK, Balz V, Bier H, Appella E, et al. Immunological characterization
611 of missense mutations occurring within cytotoxic T cell-defined p53 epitopes in HLA-
612 A*0201+ squamous cell carcinomas of the head and neck. Int J Cancer 2007;120:2618–24.
613 https://doi.org/10.1002/ijc.22584.
614 [42] Wick DA, Webb JR, Nielsen JS, Martin SD, Kroeger DR, Milne K, et al. Surveillance of the Tumor
615 Mutanome by T Cells during Progression from Primary to Recurrent Ovarian Cancer. Clin
616 Cancer Res 2014;20:1125–34. https://doi.org/10.1158/1078-0432.CCR-13-2147.
617 [43] Lu Y-C, Yao X, Crystal JS, Li YF, El-Gamil M, Gross C, et al. Efficient Identification of Mutated
618 Cancer Antigens Recognized by T Cells Associated with Durable Tumor Regressions. Clin
619 Cancer Res 2014;20:3401–10. https://doi.org/10.1158/1078-0432.CCR-14-0433.
620 [44] Tran E, Ahmadzadeh M, Lu Y-C, Gros A, Turcotte S, Robbins PF, et al. Immunogenicity of
621 somatic mutations in human gastrointestinal cancers. Science 2015;350:1387–90.
622 https://doi.org/10.1126/science.aad1253.
623 [45] Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic Basis for
624 Clinical Response to CTLA-4 Blockade in Melanoma. N Engl J Med 2014;371:2189–99.
625 https://doi.org/10.1056/NEJMoa1406498.
626 [46] Kalaora S, Wolf Y, Feferman T, Barnea E, Greenstein E, Reshef D, et al. Combined Analysis of
627 Antigen Presentation and T-cell Recognition Reveals Restricted Immune Responses in
628 Melanoma. Cancer Discov 2018;8:1366–75. https://doi.org/10.1158/2159-8290.CD-17-1418.
629 [47] Gros A, Parkhurst MR, Tran E, Pasetto A, Robbins PF, Ilyas S, et al. Prospective identification
630 of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. Nat Med
631 2016;22:433–8. https://doi.org/10.1038/nm.4051.
632 [48] Prickett TD, Crystal JS, Cohen CJ, Pasetto A, Parkhurst MR, Gartner JJ, et al. Durable Complete
633 Response from Metastatic Melanoma after Transfer of Autologous T Cells Recognizing 10
634 Mutated Tumor Antigens. Cancer Immunol Res 2016;4:669–78.
635 https://doi.org/10.1158/2326-6066.CIR-15-0215.
636 [49] Parkhurst M, Gros A, Pasetto A, Prickett T, Crystal JS, Robbins P, et al. Isolation of T-Cell
637 Receptors Specifically Reactive with Mutated Tumor-Associated Antigens from Tumor-
638 Infiltrating Lymphocytes Based on CD137 Expression. Clin Cancer Res 2017;23:2491–505.
639 https://doi.org/10.1158/1078-0432.CCR-16-2680.
640 [50] Stevanović S, Pasetto A, Helman SR, Gartner JJ, Prickett TD, Howie B, et al. Landscape of
641 immunogenic tumor antigens in successful immunotherapy of virally induced epithelial
642 cancer. Science 2017;356:200–5. https://doi.org/10.1126/science.aak9510.
643 [51] Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch repair deficiency
644 predicts response of solid tumors to PD-1 blockade. Science 2017;357:409–13.
645 https://doi.org/10.1126/science.aan6733.
646 [52] Deniger DC, Pasetto A, Robbins PF, Gartner JJ, Prickett TD, Paria BC, et al. T-cell Responses to
647 *TP53* "Hotspot" Mutations and Unique Neoantigens Expressed by Human Ovarian Cancers.

648 Clin Cancer Res 2018;24:5562–73. https://doi.org/10.1158/1078-0432.CCR-18-0573.
649 [53] The problem with neoantigen prediction. Nat Biotechnol 2017;35:97–97.
650 https://doi.org/10.1038/nbt.3800.
651 [54] Chen DS, Mellman I. Elements of cancer immunity and the cancer–immune set point. Nature
652 2017;541:321–30. https://doi.org/10.1038/nature21349.
653 [55] Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor
654 immunotherapy. Nat Rev Cancer 2019;19:133–50. https://doi.org/10.1038/s41568-019-
655 0116-x.
656 [56] Rammensee H-G, Bachmann J, Emmerich NPN, Bachor OA, Stevanović S. SYFPEITHI: database
657 for MHC ligands and peptide motifs. Immunogenetics 1999;50:213–9.
658 https://doi.org/10.1007/s002510050595.
659 [57] Falk K, Rötzschke O, Stevanovié S, Jung G, Rammensee H-G. Allele-specific motifs revealed by
660 sequencing of self-peptides eluted from MHC molecules. Nature 1991;351:290–6.
661 https://doi.org/10.1038/351290a0.
662 [58] Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The Immune Epitope
663 Database (IEDB): 2018 update. Nucleic Acids Res 2019;47:D339–43.
664 https://doi.org/10.1093/nar/gky1006.
665 [59] Di Marco M, Schuster H, Backert L, Ghosh M, Rammensee H-G, Stevanović S. Unveiling the
666 Peptide Motifs of HLA-C and HLA-G from Naturally Presented Peptides and Generation of
667 Binding Prediction Matrices. J Immunol 2017:ji1700938.
668 https://doi.org/10.4049/jimmunol.1700938.
669 [60] Davis MM. The problem of plain vanilla peptides. Nat Immunol 2003;4:649–50.
670 https://doi.org/10.1038/ni0703-649.
671 [61] Song I, Gil A, Mishra R, Ghersi D, Selin LK, Stern LJ. Broad TCR repertoire and diverse structural
672 solutions for recognition of an immunodominant CD8+ T cell epitope. Nat Struct Mol Biol
673 2017;24:395–406. https://doi.org/10.1038/nsmb.3383.
674 [62] Stewart-Jones GBE, McMichael AJ, Bell JI, Stuart DI, Jones EY. A structural basis for
675 immunodominant human T cell receptor recognition. Nat Immunol 2003;4:7.
676 [63] Celie PHN, Toebes M, Rodenko B, Ovaa H, Perrakis A, Schumacher TNM. UV-Induced Ligand
677 Exchange in MHC Class I Protein Crystals. J Am Chem Soc 2009;131:12298–304.
678 https://doi.org/10.1021/ja9037559.
679 [64] Turner SJ, Doherty PC, McCluskey J, Rossjohn J. Structural determinants of T-cell receptor bias
680 in immunity. Nat Rev Immunol 2006;6:883–94. https://doi.org/10.1038/nri1977.
681 [65] Rudolph MG, Stanfield RL, Wilson IA. HOW TCRS BIND MHCS, PEPTIDES, AND CORECEPTORS.
682 Annu Rev Immunol 2006;24:419–66.
683 https://doi.org/10.1146/annurev.immunol.23.021704.115658.
684 [66] Wang GC, Dash P, McCullers JA, Doherty PC, Thomas PG. T Cell Receptor αβ Diversity
685 Inversely Correlates with Pathogen-Specific Antibody Levels in Human Cytomegalovirus
686 Infection. Sci Transl Med 2012;4:128ra42. https://doi.org/10.1126/scitranslmed.3003647.
687 [67] Yager EJ, Ahmed M, Lanzer K, Randall TD, Woodland DL, Blackman MA. Age-associated
688 decline in T cell repertoire diversity leads to holes in the repertoire and impaired immunity
689 to influenza virus. J Exp Med 2008;205:711–23. https://doi.org/10.1084/jem.20071140.
690 [68] Yarchoan M, Johnson BA, Lutz ER, Laheru DA, Jaffee EM. Targeting neoantigens to augment
691 antitumour immunity. Nat Rev Cancer 2017;17:209–22.

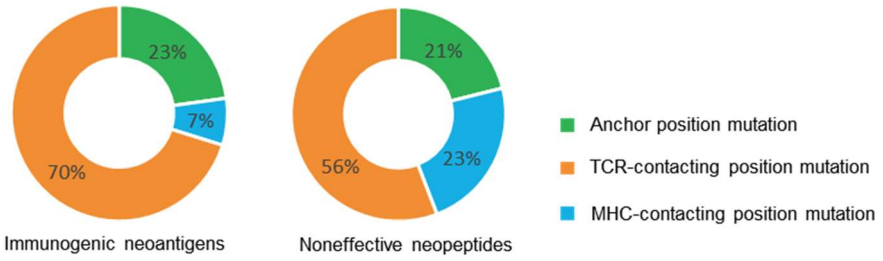
692 https://doi.org/10.1038/nrc.2016.154.
693 [69] Kawashima S, Pokarowski P, Pokarowska M, Kolinski A, Katayama T, Kanehisa M. AAindex:
694 amino acid index database, progress report 2008. Nucleic Acids Res 2007;36:D202–5.
695 https://doi.org/10.1093/nar/gkm998.
696 [70] Sibener LV, Fernandes RA, Kolawole EM, Carbone CB, Liu F, McAffee D, et al. Isolation of a
697 Structural Mechanism for Uncoupling T Cell Receptor Signaling from Peptide-MHC Binding.
698 Cell 2018;174:672-687.e27. https://doi.org/10.1016/j.cell.2018.06.017.
699 [71] June CH. Drugging the Undruggable Ras — Immunotherapy to the Rescue? N Engl J Med
700 2016;375:2286–9. https://doi.org/10.1056/NEJMe1612215.
701 [72] Chen J-L, Stewart-Jones G, Bossi G, Lissin NM, Wooldridge L, Choi EML, et al. Structural and
702 kinetic basis for heightened immunogenicity of T cell vaccines. J Exp Med 2005;201:1243–55.
703 https://doi.org/10.1084/jem.20042323.
704 [73] Miles JJ, Tan MP, Dolton G, Edwards ESJ, Galloway SAE, Laugel B, et al. Peptide mimic for
705 influenza vaccination using nonnatural combinatorial chemistry. J Clin Invest 2018;128:1569–
706 80. https://doi.org/10.1172/JCI91512.
707 [74] Ghorani E, Rosenthal R, McGranahan N, Reading JL, Lynch M, Peggs KS, et al. Differential
708 binding affinity of mutated peptides for MHC class I is a predictor of survival in advanced lung
709 cancer and melanoma. Ann Oncol 2018;29:271–9. https://doi.org/10.1093/annonc/mdx687.
710 [75] Thomsen MCF, Nielsen M. Seq2Logo: a method for construction and visualization of amino
711 acid binding motifs and sequence profiles including sequence weighting, pseudo counts and
712 two-sided representation of amino acid enrichment and depletion. Nucleic Acids Res
713 2012;40:W281–7. https://doi.org/10.1093/nar/gks469.
714 [76] Yin L, Huseby E, Scott-Browne J, Rubtsova K, Pinilla C, Crawford F, et al. A Single T Cell
715 Receptor Bound to Major Histocompatibility Complex Class I and Class II Glycoproteins
716 Reveals Switchable TCR Conformers. Immunity 2011;35:23–33.
717 https://doi.org/10.1016/j.immuni.2011.04.017.
718 [77] Macdonald WA, Chen Z, Gras S, Archbold JK, Tynan FE, Clements CS, et al. T Cell
719 Allorecognition via Molecular Mimicry. Immunity 2009;31:897–908.
720 https://doi.org/10.1016/j.immuni.2009.09.025.
721 [78] Battye TGG, Kontogiannis L, Johnson O, Powell HR, Leslie AGW. *iMOSFLM* : a new graphical
722 interface for diffraction-image processing with *MOSFLM*. Acta Crystallogr D Biol Crystallogr
723 2011;67:271–81. https://doi.org/10.1107/S0907444910048675.
724 [79] Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the *CCP*
725 4 suite and current developments. Acta Crystallogr D Biol Crystallogr 2011;67:235–42.
726 https://doi.org/10.1107/S0907444910045749.
727 [80] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. *Phaser*
728 crystallographic software. J Appl Crystallogr 2007;40:658–74.
729 https://doi.org/10.1107/S0021889807021206.
730 [81] Emsley P, Cowtan K. *Coot* : model-building tools for molecular graphics. Acta Crystallogr D
731 Biol Crystallogr 2004;60:2126–32. https://doi.org/10.1107/S0907444904019158.
732 [82] Adams PD, Grosse-Kunstleve RW, Hung L-W, Ioerger TR, McCoy AJ, Moriarty NW, et al.
733 *PHENIX* : building new software for automated crystallographic structure determination.
734 Acta Crystallogr D Biol Crystallogr 2002;58:1948–54.
735 https://doi.org/10.1107/S0907444902016657.
736

Figure 1

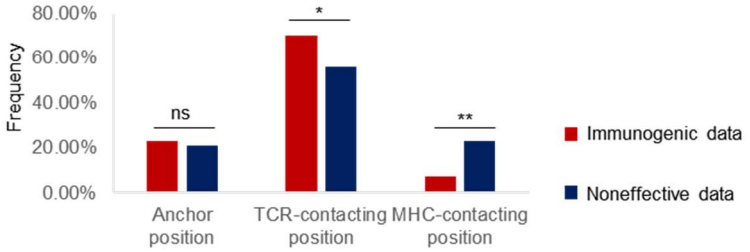
a

	Immunogenic Neoantigen Dataset (IND)	Raw Noneffective Neopeptide Dataset (RNND)	Noneffective Neoantigen Dataset (NND)
Entries	128	11,739	2,883
Anchor Mutated Peptides	29	N	428
Tumor Types	15	11	10
HLA-A	14	15	15
HLA-B	14	20	17
HLA-C	7	15	13
Genes	105	1,929	1,081
References	34	13	11

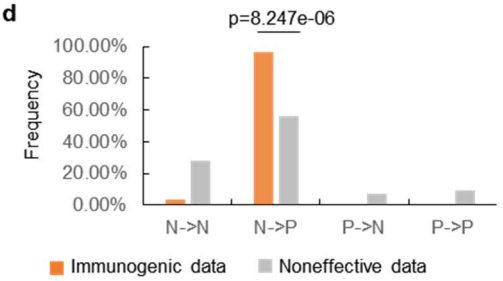
b



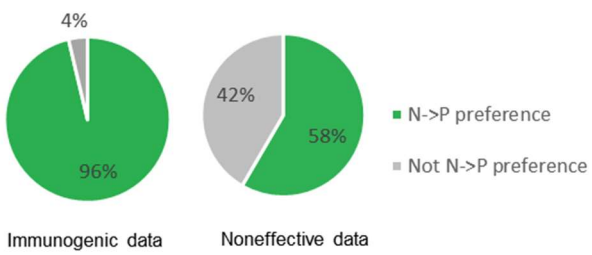
c



d



e



f

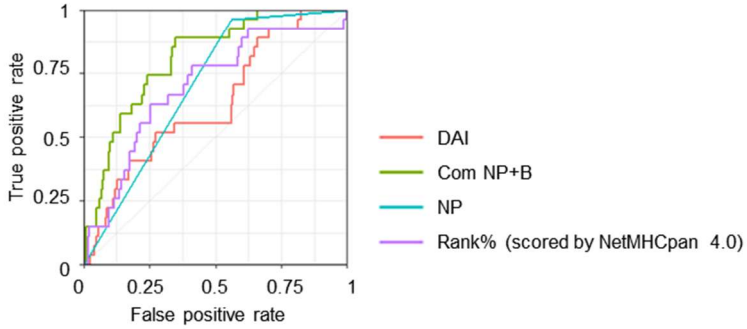


Figure 2

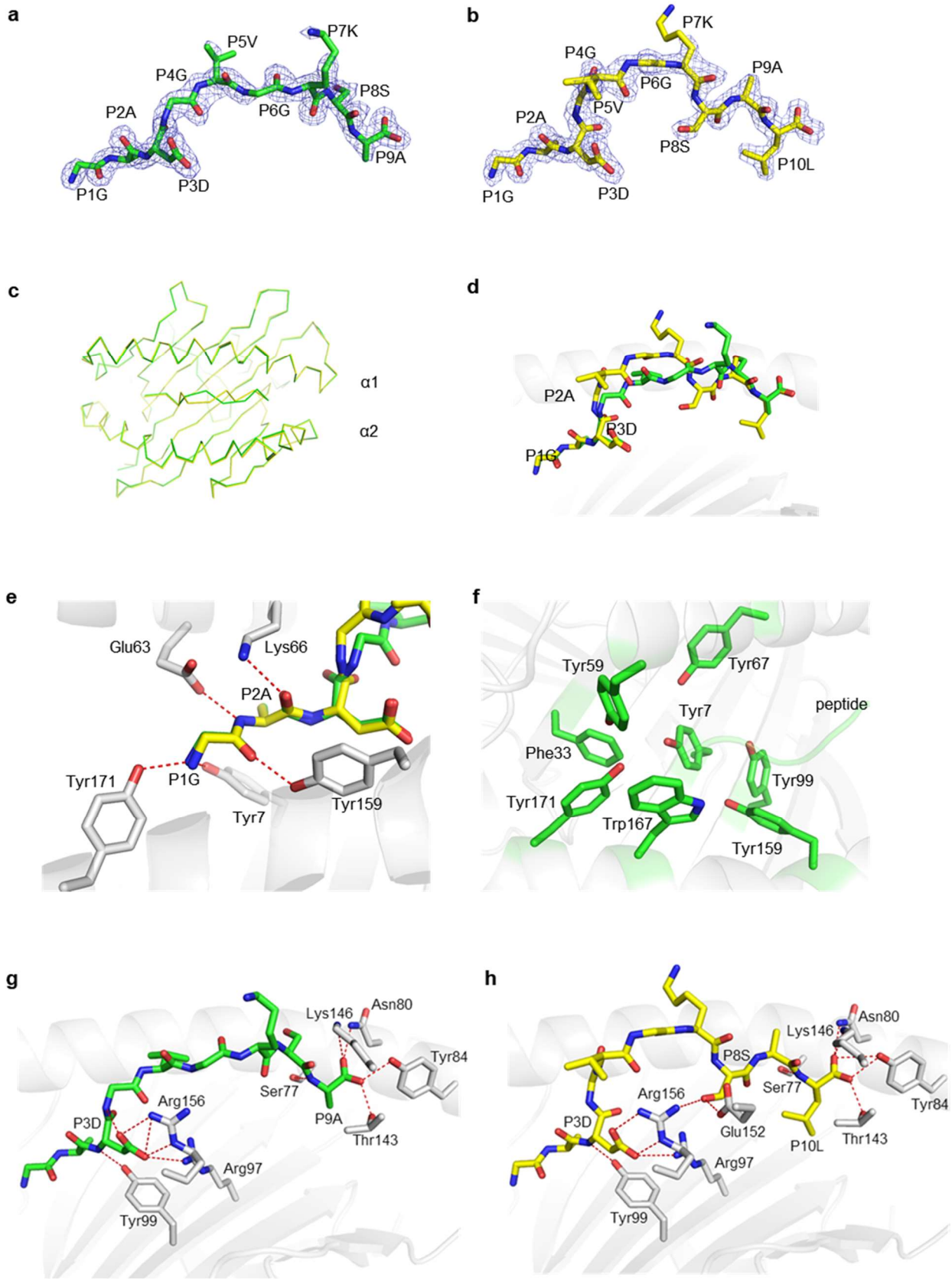

a
P7K
P5V
P4G
P8S
P2A
P6G
P9A
P1G
P3D
b
P7K
P4G
P9A
P6G
P5V
P2A
P8S
P10L
P1G
P3D
c
α1
α2
d
P2A
P3D
P1G
e
Glu63
Lys66
P2A
Tyr171
P1G
Tyr7
Tyr159
f
Tyr59
Tyr67
Tyr7
peptide
Phe33
Tyr99
Tyr171
Trp167
Tyr159
g
Lys146
Asn80
Tyr84
Ser77
Arg156
P3D
P9A
Thr143
Arg97
Tyr99
h
Asn80
Lys146
P8S
Ser77
Arg156
Tyr84
P3D
Glu152
P10L
Thr143
Arg97
Tyr99

Figure 3

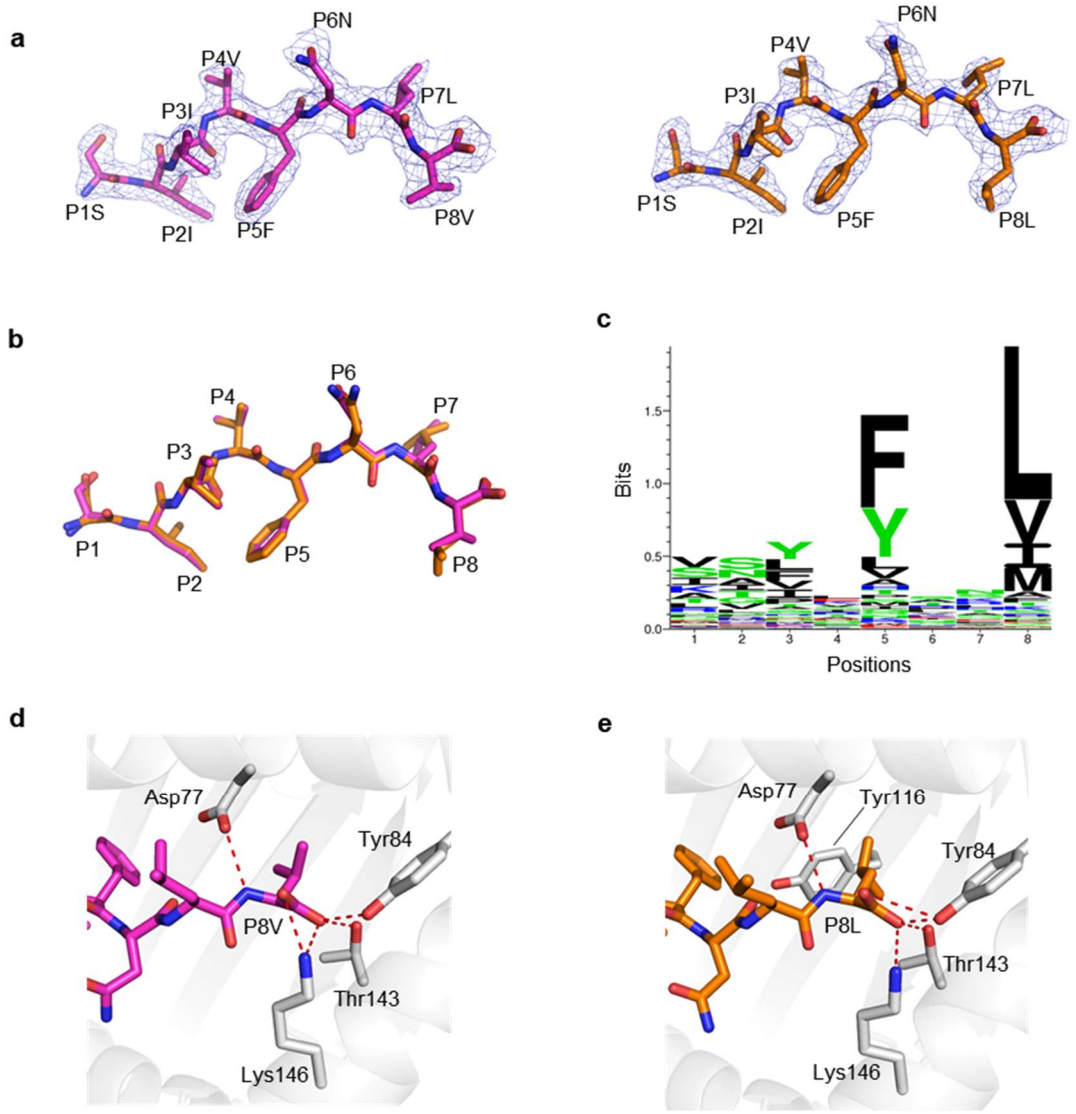
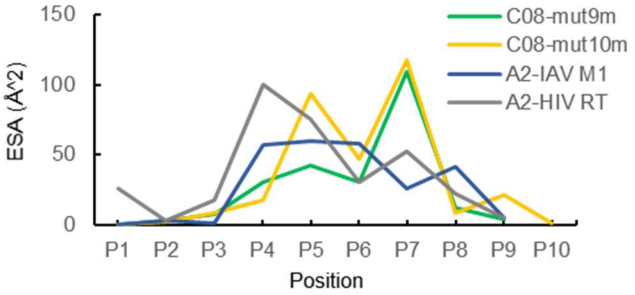

a
P6N
P4V
P3I
P7L
P1S
P2I
P5F
P8V
P6N
P4V
P3I
P7L
P1S
P2I
P5F
P8L
b
P6
P4
P7
P3
P1
P5
P8
P2
c
Bits
1.5
1.0
0.5
0.0
1
2
3
4
5
6
7
8
Positions
d
Asp77
Tyr84
P8V
Thr143
Lys146
e
Asp77
Tyr116
Tyr84
P8L
Thr143
Lys146

Figure 4

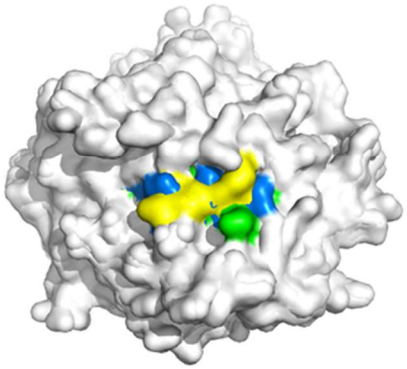
a

pMHC name	Peptide	PESA (Å^2)
C08-mut9m	GADGVGKSA	239.89
C08-mut10m	GADGVGKSAL	316.90
A2-IAV M1	GILGFVFTL	250.96
A2-HIV RT	ILKEPVHGV	329.88

b



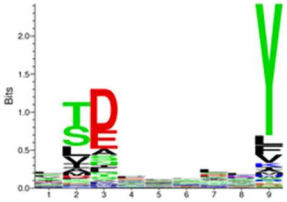
c



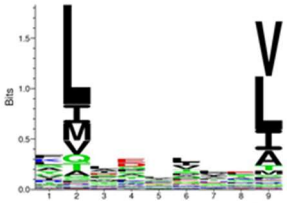
d

Patient ID	TCR Pair Name	TCRα/TCRβ	TCRα/TCRβ CDR3 AA sequence	Rank in Tumor Sample*	Rank in Infusion Product	Rank in Blood
4095	RK1	TRAV4/ TRBV5-6	CLVGDMDQAGTALIF/ CASSLGEGRVDGYTF	20/26/33	1	ND (d+266)
4095	RK2	TRAV12-2/ TRBV10-2	CAAAMDSSYKLIF/ CASSDPGTEAFF	1/1/2	2	1 (d+266)
4095	RK3	TRAV4/ TRBV5-6	CLVGDRDQAGTALIF/ CASSFGQSSTYGYTF	ND	3	5 (d+266)
4095	RK4	TRAV4/ TRBV5-6	CLVGDMDQAGTALIF/ CASSLGRASNQPQHF	917/1589/2243	45	1784 (d+266)
3995	RK5	TRAV4/ TRBV5-6	CLVGDMDQAGTALIF/ CASSLGQTNYGYTF	527	2699	ND (d+39)

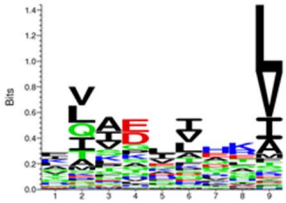
Supplementary Figure 1



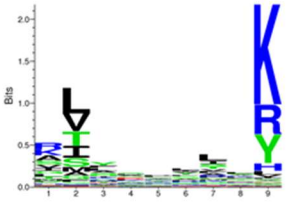
HLA-A*01:01 (n=4661)



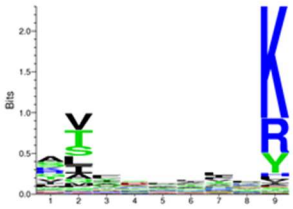
HLA-A*02:01 (n=21172)



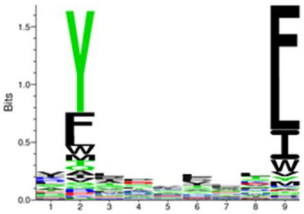
HLA-A*02:05 (n=252)



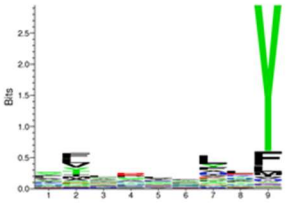
HLA-A*03:01 (n=7487)



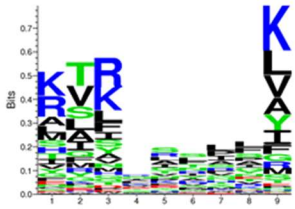
HLA-A*11:01 (n=4850)



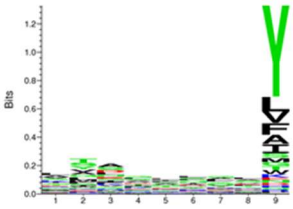
HLA-A*24:02 (n=6683)



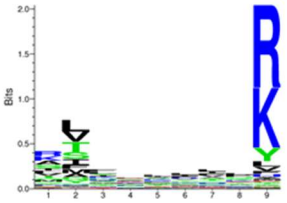
HLA-A*29:02 (n=6009)



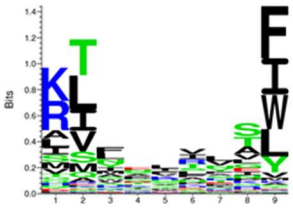
HLA-A*30:01 (n=810)



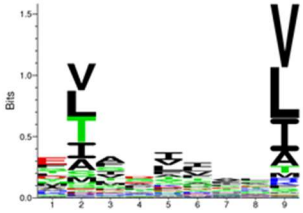
HLA-A*30:02 (n=819)



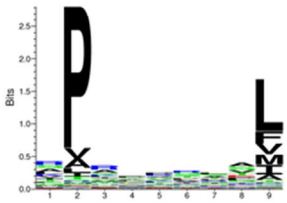
HLA-A*31:01 (n=2000)



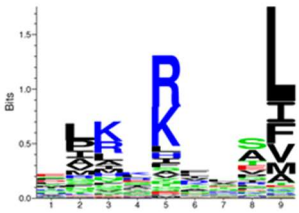
HLA-A*32:01 (n=878)



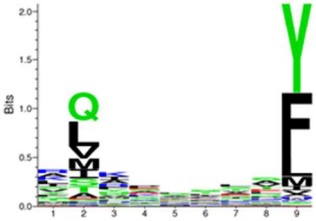
HLA-A*68:02 (n=2995)



HLA-B*07:02 (n=9573)

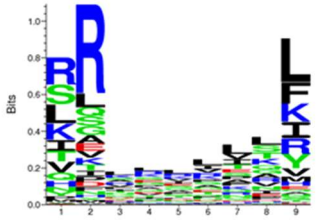


HLA-B*08:01 (n=3955)

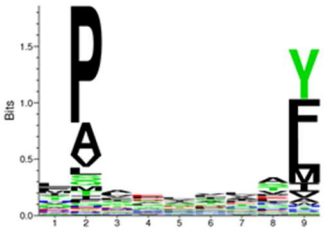


HLA-B*15:01 (n=8349)

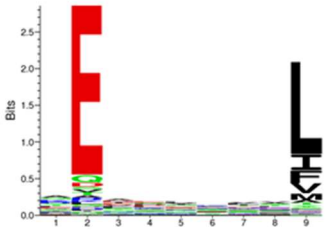
(continued on next page)



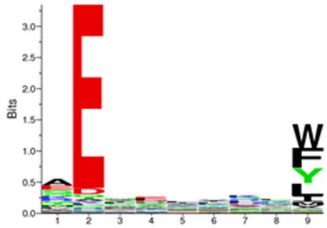
HLA-B*27:05 (n=13590)



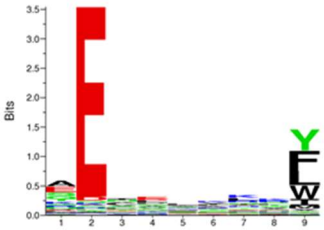
HLA-B*35:01 (n=3023)



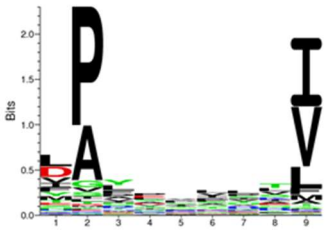
HLA-B*40:01 (n=2148)



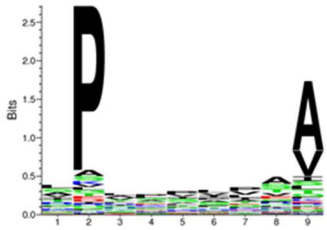
HLA-B*44:02 (n=3531)



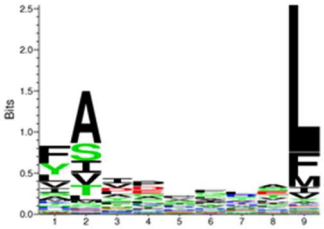
HLA-B*44:03 (n=5166)



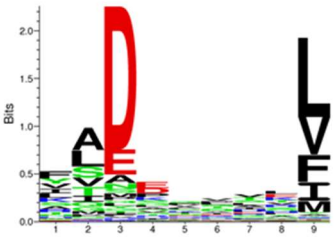
HLA-B*51:01 (n=2869)



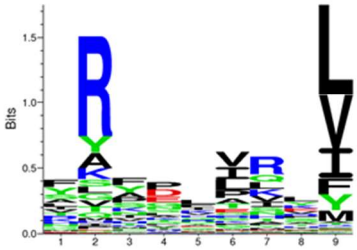
HLA-B*56:01 (n=398)



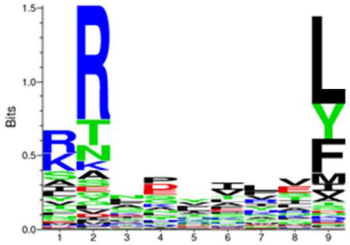
HLA-C*03:03 (n=1020)



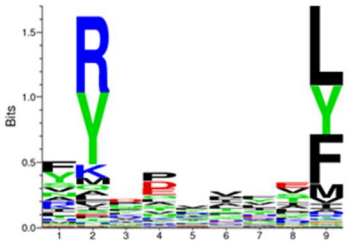
HLA-C*05:01 (n=3251)



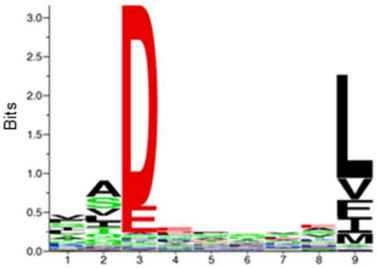
HLA-C*06:02 (n=2069)



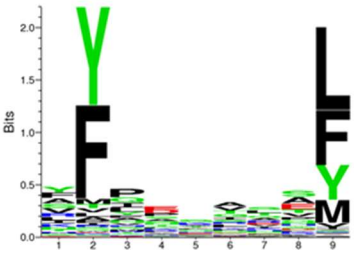
HLA-C*07:01 (n=389)



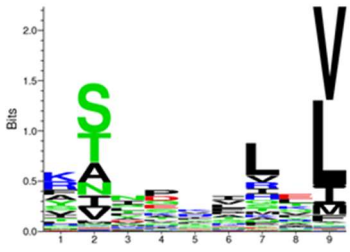
HLA-C*07:02 (n=980)



HLA-C*08:02 (n=2321)

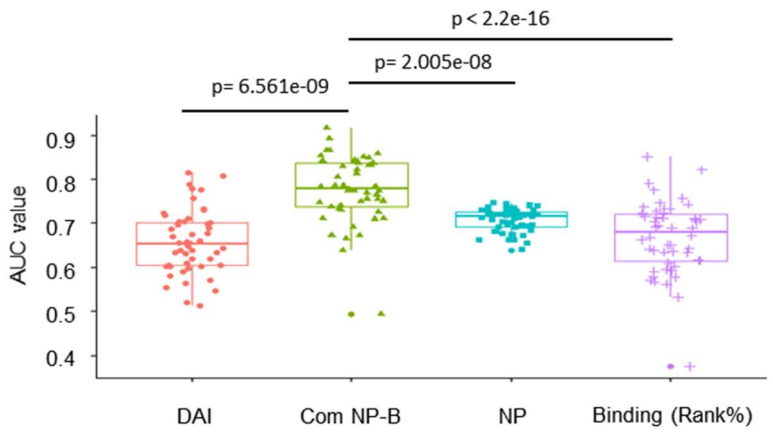


HLA-C*14:02 (n=1899)

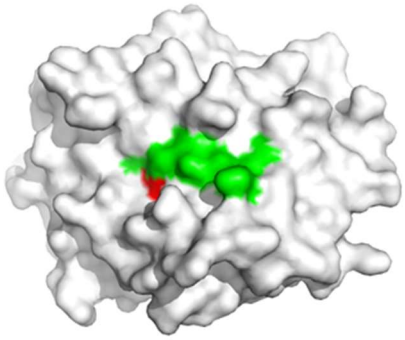


HLA-C*15:02 (n=1747)

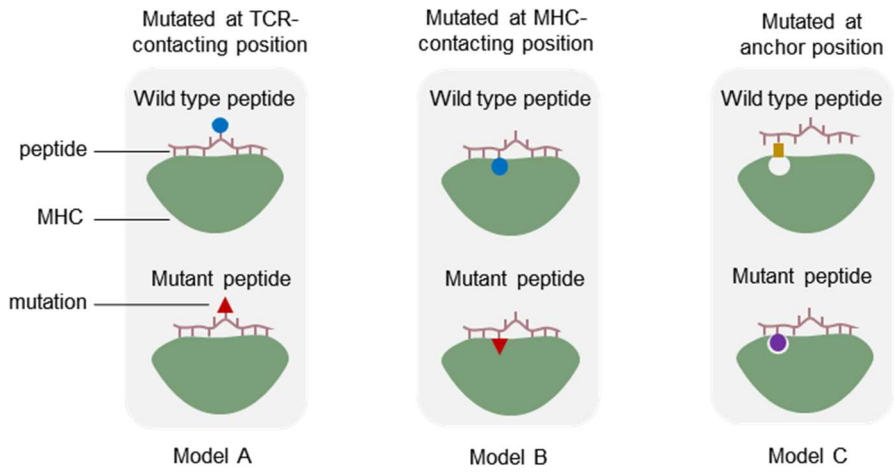
Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4



Supplementary Table 1. Immunogenic Neopeptide Dataset (IND)

Supplementary Table 2. Raw Noneffective Neopeptide Dataset (RNND)

Supplementary Table 3. Noneffective Neopeptide Dataset (NND)

（see in separated EXCEL files)

Supplementary Table 4. Classification of anchor mutated neoantigens based on MHC preference

	N->N frequency	N->P frequency	P->N frequency	P->P frequency	Number of peptide
Anchor mutated neoantigens (immunogenic)	3.7% (1)	96.3% (26)	0% (0)	0% (0)	27
Anchor mutated neopeptides (noneffective)	27.76% (118)	56% (238)	6.82% (29)	9.41% (40)	425

Data come from Immunogenic Neoantigen Dataset (IND) and Noneffective Neopeptide Dataset (NND). Peptides defined with anchor position were calculated in this analysis.

Supplementary Table 5. Data collection and refinement statistics (molecular replacement).

	C08-mut9m	C08-mut10m	Kb-8mV	Kb-8mL
Resolution range	42.55 - 1.9 (1.968 - 1.9)	51.13 - 1.9 (1.968 - 1.9)	57.76 - 2.4 (2.486 - 2.4)	68.67 - 2.5 (2.589 - 2.5)
Space group	P 21 21 21	P 21 21 21	C 1 2 1	C 1 2 1
Unit cell	68.2796 75.9199 78.0298 90 90 90	68.9397 76.24 78.6999 90 90 90	107.24 69.42 70.4397 90 103.79 90	106.63 67.9699 70.73 90 103.86 90
Total reflections	64768 (6323)	66521 (6583)	37661 (3734)	33008 (3336)
Unique reflections	32450 (3174)	33289 (3294)	19013 (1886)	16646 (1679)
Multiplicity	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)
Completeness (%)	0.96 (0.99)	1.00 (1.00)	0.96 (0.96)	0.97 (0.99)
Mean I/sigma(I)	7.15 (4.02)	9.26 (5.12)	14.38 (4.80)	11.60 (2.85)
Wilson B-factor	10.39	20.99	38.48	46.21
R-merge	0.04092 (0.1168)	0.03331 (0.1176)	0.02281 (0.1159)	0.02767 (0.1982)
R-meas	0.05787 (0.1652)	0.04711 (0.1663)	0.03225 (0.1639)	0.03913 (0.2803)
CC1/2	0.995 (0.963)	0.996 (0.956)	0.999 (0.979)	0.999 (0.944)
CC*	0.999 (0.99)	0.999 (0.989)	1 (0.995)	1 (0.985)
Reflections used in refinement	31350 (2603)	33285 (3294)	18981 (1879)	16622 (1674)
Reflections used for R-free	1936 (155)	2001 (198)	1899 (188)	1663 (167)
R-work	0.2406	0.174	0.2194	0.2153
R-free	0.2872	0.2044	0.263	0.2934
CC(work)	0.933	0.96	0.943	0.958
CC(free)	0.897	0.946	0.896	0.902
Number of non-hydrogen atoms	3641	3542	3152	3093
macromolecules	3127	3139	3044	3045
Protein residues	381	382	373	373
RMS(bonds)	0.008	0.007	0.01	0.01
RMS(angles)	0.93	0.93	1.08	1.19
Ramachandran favored (%)	96	99	94	94
Ramachandran allowed (%)	3.2	1.3	5.3	5.1
Ramachandran outliers (%)	0.53	0	0.28	0.85
Rotamer outliers (%)	2.1	1.8	4.3	2.8
Clashscore	8.39	4.91	18.38	15.2
Average B-factor	17.27	25.67	50.83	59.08
Macromolecules	16.15	24.62	50.95	59.2
solvent	24.05	33.81	47.71	51.52

Statistics for the highest-resolution shell are shown in parentheses.

Figure Legends

Figure 1. Analyses of three datasets and immune-based data classification of Immunogenic neoantigen dataset (IND) and Noneffective neopeptides dataset (NND).

a, Overview of the IND, RNND, and NND datasets.

b, Donut plots of percentage of 9mer peptide mutations categorized based on three different mutation locations: anchor mutation, MHC-contacting position and TCR-contacting position (immunogenic data, n =51; noneffective data, n=763).

c, Nonameric peptides mutation distribution of three different groups (mutated at anchor mutation, MHC-contacting position and TCR-contacting position) from IND and NND. The frequency of mutation distribution at TCR-contacting position and MHC-contacting position showed significant difference (TCR-contacting position, p=0.0378; MHC-contacting position, P=0.0027. n (immunogenic) =51; n (noneffective) =763. Fisher's exact test). The percentage of mutation distribution at anchor position did not show significant difference between immunogenic and noneffective data (ns, non-significant).

d, The frequency between anchor mutated IND and NND with significant difference in subgroup of non-preferential to preferential residues (N to P) change (p=8.247e-06, n=27 (immunogenic); n=425 (noneffective), Fisher's exact test). There was no difference between residue changes from wild type neopeptides to different types mutants: non-preferential to non-preferential residues (N to N), preferential to non-preferential residues (P to N) and preferential to preferential residues (P to P)) of anchor mutated immunogenic data (from IND) and non-immunogenic data (from NND).

e, Pie charts represented the percentage of NP group and not NP group of anchor mutated datasets (n (immunogenic) = 27; n (noneffective) =425).

f, Receiver operator characteristic (ROC) curve showed the performance of different predictors (DAI score, binary NP rule, binding prediction (Rank% scored by NetMHCpan 4.0), combination of NP rule + binding prediction (Com NP+B)) with anchor mutated data (data from IND and NND datasets, n (immunogenic) = 27; n (noneffective) =425). The AUC (Area Under the ROC Curve) was calculated for each predictive rule (AUC_{DAI}= 0.632; $AUC_{NP\ rule}$ =0.701; $AUC_{com\ NP\text{-}B}$ =0.810; $AUC_{Rank\%}$=0.698).

Figure 2. Structural comparison of C08-mut9m and C08-mut10m complexes.

a-b, Unambiguous 2Fo-Fc electron density maps of (**a**) KRAS G12D 9mer (GADGVGKSA, green) and (**b**)10mer (GADGVGKSAL, yellow) peptides from solved pMHC complex structures. The underlined amino acids represented the mutation in the peptides.

c, Overlay of Cα traces (C08-mut9m, green; C08-mut10m, yellow). Differences in peptide conformation were observed.

d, Overlay of the KRAS G12D 9mer and 10mer peptides in the MHC binding groove.

e, Polar interactions at P1G and P2A positions within HLA-C*08:02 molecule showed in grey, 9mer peptide showed in green and 10mer peptide showed in yellow.

f, Aromatic residues (green) from HLA-C*08:02 accommodating P1 and P2 residues of KRAS G12D 9mer peptide.

g, The P3D and P9A side chains of mut9m peptide interact with HLA-C*08:02.

h, The P3D, P8S and P10LA side chains of mut10m peptide interact with HLA-C*08:02.

Figure 3. H-2 Kb presented DPAGT V213L wild type peptide wt8mV and mutant peptide mut8mL in a similar manner

a, Unambiguous 2Fo-Fc electron density maps of DPAGT1 wild type 8mer peptide (wt8mV peptide SIIVFNLV, magenta) and mutant 8mer peptide (mut8mL peptide SIIVFNLL, orange) from solved structures.

b, Overlay of the DPAGT1 wt8mV (magenta) and mutant mut8mL (orange) peptides.

c, Sequence logo based on the data from IEDB showing amino acid preferences for 8mer peptides bound to H-2 Kb. The peptide library was obtained from IEDB (n=4141).

d, Anchor residue P8V of wt8mV peptide interacts with H-2 Kb (grey).

e, Anchor residue P8L of mut8mLpeptide interacts with H-2 Kb (grey).

Figure 4. Different peptide presentation patterns and peptide exposed surface areas (PESA) between C08-mut9m and C08-mut10m

a, Peptide exposed surface areas (PESA) of 4 pHLAs. PESA of A2-IAV M1 complex was calculated based on 2VLL (PDB ID). PESA of A2-HIV RT complex was calculated based on 2X4U (PDB ID).

b, Exposed surface areas (ESA) of individual residues at each position of four peptides within HLAs.

c, Mut10m exhibits a relatively large exposed surface area compared to A2-IAV M1 (the featureless benchmark), C08-mut9m and C08-mut10m. HLA backbone surface was shown in grey. IAV M1 (blue), mut9m (green) and mut10m (yellow) peptide surface in different colors.

d, Different T cell fate of KRAS G12D neoantigen restricted T cells reported by previous clinical studies. Rank in tumor sample represents the rank of restricted T cells in patient TILs before cell therapy. Rank in infusion products represents the rank of restricted T cells in ex vivo cell transfer products before cell transfer. Rank in blood represents the rank of restricted T cells in patients' peripheral blood after cell transfer (d+ represents the day after cell transfer). The rank of T cells with same TCR pair was validated in one or three different metastatic tumor fragments from the patient. *, tested in three fragments; †, tested in one fragment; ND, not detected.

Supplementary Figure Legends

Supplementary Figure 1. 9mer peptides binding profile of 30 HLA alleles.

Sequence logos frequent amino acid binding profiles for 9mer peptides bound to each of 30 HLA alleles generated from peptide-binding matrices using the Seq2Logo. Peptide libraries were obtained from IEDB. The sample size of each allele was shown on the figure.

Supplementary Figure 2. The AUC values testing using a 50-fold cross-validation.

The 50-fold cross-validation (2/3rd random resampling) within the exploration set. Differences of AUC values were determined using paired T-test.

Supplementary Figure 3. Additional TCR accessible surface of mut9m neoantigen in complex with HLA-C*08:02.

P3D of C08-mut9m can provide additional TCR accessible surface. HLA-C*08:02 (grey), 9mer peptide (green) and P3D side chain (red) showed in different colors.

Supplementary Figure 4. The topology of different neoantigen-MHC binding models

Model A demonstrates the neoantigens with mutations at TCR-contacting positions that are presented by MHC. The blue circle represents wild type residues. The red triangle represents the mutant residues. Model B demonstrates the neoantigens with mutations at MHC-contacting positions. The red triangle represents the mutant residues compared to the wild type that would affect binding specificity. Model C demonstrates the peptides which have mutations at anchor position and are

presented by MHC. The yellow rectangle represents wild type residues which cannot bind into the anchor position in the MHC peptide presentation groves and thus are not presented. The purple circle represents mutated residues which are preferential selected and bind better at the anchor position.