Reversion analysis reveals the in vivo immunogenicity of a poorly MHC I-binding cancer neoepitope

Hakimeh Ebrahimi-Nik, Marmar Moussa, Ryan P. Englander, Summit Singhaviranon, Justine Michaux, HuiSong Pak, Hiroko Miyadera, William L. Corwin, Grant L. J. Keller, Adam T. Hagymasi, Tatiana V. Shcheglova, George Coukos, Brian M. Baker, Ion I. Mandoiu, Michal Bassani-Sternberg & Pramod K. Srivastava

High-affinity MHC I-peptide interactions are considered essential for immunogenicity. However, some neo-epitopes with low affinity for MHC I have been reported to elicit CD8 T cell dependent tumor rejection in immunization-challenge studies. Here we show in a mouse model that a neo-epitope that poorly binds to MHC I is able to enhance the immunogenicity of a tumor in the absence of immunization. Fibrosarcoma cells with a naturally occurring mutation are edited to their wild type counterpart; the mutation is then re-introduced in order to obtain a cell line that is genetically identical to the wild type except for the neo-epitope-encoding mutation. Upon transplantation into syngeneic mice, all three cell lines form tumors that are infiltrated with activated T cells. However, lymphocytes from the two tumors that harbor the mutation show significantly stronger transcriptional signatures of cytotoxicity and TCR engagement, and induce greater breadth of TCR reactivity than those of the wild type tumors. Structural modeling of the neo-epitope peptide/MHC I pairs suggests increased hydrophobicity of the neo-epitope surface, consistent with higher TCR reactivity. These results confirm the in vivo immunogenicity of low affinity or ‘non-binding’ epitopes that do not follow the canonical concept of MHC I-peptide recognition.
Antigen presentation by MHC molecules is fundamental to adaptive immunity. In the case of MHC I molecules, such presentation involves a complex series of steps that result in the proteolytic processing of whole or partially synthesized proteins, chaperoning of the peptides through the cytosol and the endoplasmic reticulum, and their rendezvous with MHC I molecules into a tri-molecular MHC I-β2 microglobulin-peptide (pMHC) complex. Based on extensive analyses of peptides recognized by mouse and human T cells against viral antigens, it has been clear that a high affinity (IC50 values <500 nM, but preferably <50 nM) of peptides for MHC I is essential for antigen presentation. This premise has been abundantly validated in its ability to predict the epitopes that can elicit a CD8+ T cell response measurable in vitro.

Subsequent to the advances in our ability to identify somatic mutations in cancers, identification of epitopes that can act as cancer vaccines has become a large area of inquiry. Since affinity of peptides to MHC I has withstood the test of time as a key criterion for predicting immunogenicity, this has been applied to the discovery of cancer neoepitopes as well, and a number of high affinity neoepitopes that elicit tumor rejection as well as CD8 T cell responses measurable in vitro, have been identified. A measurable CD8 response is often considered a valid surrogate for tumor rejection, and several neoepitopes, which have a high affinity for MHC I, and elicit CD8 T cell response have been identified. Indeed, high affinity of a peptide for MHC I has become so entrenched in immunological thought that peptides with a low affinity (IC50 of >500 nM) are routinely excluded from consideration as candidates for vaccines, and are even often referred to as “non binders” to reinforce their irrelevance.

A small number of recent reports have examined the question of immunogenicity of mouse cancer neoepitopes from a vantage point agnostic to peptide-MHC I affinity. Such studies have reported a number of neoepitopes which bind MHC I with low affinity, and mediate CD8-dependent tumor rejection. At the same time, two retrospective human studies analyzing the genomic and clinical outcome data from nearly 7,000 patients with 27 cancer types, have shown that better clinical outcomes and T cell infiltration of tumors are associated with the presence of cancer neoepitopes with low affinities for HLA I molecules, and not with the presence of high affinity HLA I-binding neoepitopes.

Consistent with the lack of association between high affinity of neoepitope to MHC I and anti-tumor activity, all high affinity binding neoepitopes failed to elicit tumor rejection in a mouse model of ovarian cancer. Human clinical trials with high affinity neoepitopes have also failed to elicit significant CD8 T cell responses even when high affinity MHC I binding algorithms were used to predict the immunizing neoepitopes. Such clinical trials have also not shown convincing evidence of anti-tumor activity of the immunizing neoepitopes.

Since the ability of a neoepitope with poor affinity for MHC I to mediate CD8-dependent tumor rejection runs contrary to our dominant conception of MHC I-peptide interaction, it deserves critical scrutiny.

Here we show that the presence or absence of a low affinity MHC I-binding neoepitope in the tumor influences the spontaneous immunogenicity of a tumor in vivo. Upon transplantation, a mouse fibrosarcoma cell line, bearing a mutation known for encoding an MHC I ‘non-binder’ neoepitope, becomes less immunogenic when the sequence is reverted to the wild type allele and regains the original T cell activating capacity when the mutation is re-introduced. These experiments clearly demonstrate that a single MHC I ‘non-binder’ neo-epitope drives spontaneous immunogenicity of the fibrosarcoma and thereby challenge the current view of MHC I affinity determining the tumor immune response.

**Results**

**Definition of the neoepitope Ccd85c\text{MUT}.** The Ccd85c gene encodes a gap junction protein expressed mostly in the brain, colon, lung, kidney and testes in adult mice. The protein has no known oncogenic (driver) function. A non-synonymous (leucine to phenylalanine, Chromosome 12-108221754) somatic SNV in Ccd85c was detected in the BALB/c Meth A fibrosarcoma (Fig. 1a). (See Duan et al. Supplementary Table S1 for a list of all mutations and predicted neoepitopes of the Meth A sarcoma).

The mutation is heterozygous and the un-mutated as well as the mutated reads are detected in the transcripts. BALB/c bone marrow derived dendritic cells (BMDCs) pulsed with an 18-mer peptide with the mutant amino acid acid near the center (DPSSTYIRPFETKVKLD) or un-mutated peptide (DPSSTYIRPFLTKVKLD), were used to immunize BALB/c mice. All mice were challenged with the Meth A cells, and tumor rejection was monitored (Fig. 1b upper panel). Immunization of BALB/c mice with the mutated Ccd85c 18-mer elicted potent tumor rejection or control in all mice, while the un-mutated peptide failed to elicit protection (Fig. 1b lower panel). The anti-tumor activity of Ccd85c\text{MUT} was abrogated by depleting the mice of CD8 cells by treating the mice with the anti-CD8 antibody but not by a control antibody during the priming phase as previously described.

Various truncated versions of the 18-mer peptide as indicated in Fig. 1c (and Supplementary Fig. 1) were similarly tested for tumor rejection. A tumor rejection score (TRS) (with a maximum score of 5 indicating near 100% tumor rejection) was used to quantify the extent of tumor rejection as described in Methods. The 18-mer peptide elicited a perfect 5.0 TRS score (Fig. 1c). The most and the least effective peptides along with their TRS scores are shown in Fig. 1d. Since the 10 amino acid peptide YIRPFETKVK was the shortest peptide active in tumor rejection, we consider this the precise epitope.

Evidence of presentation of YIRPFETKVK was sought by analyzing the peptides eluted from MHC I molecules purified from the Meth A cells by mass spectrometry (MS), as described in Methods. No Ccd85c-derived peptides were detected, as expected from the low abundance of expression of this protein. In order to identify the precise peptide derived from the mutant Ccd85c that could be cross-presented by the DCs, BMDCs were pulsed in vitro with the 18-mer peptide as previously described. The BMDCs were extensively washed and MHC I molecules eluted. Targeted-MS analysis of the eluted peptides in the presence of spiked-in heavy labeled synthetic peptides showed the presence of two Ccd85c-derived peptides TYIRPFETKVK and YIRPFETKVK (Fig. 1e). These two peptides detected by cross-presentation of the 18-mer peptide were identical to the two truncated versions of the 18-mer peptide that were observed to be the most effective in tumor rejection (Fig. 1d). These peptides had very low or undetectable predicted as well as measured affinities for K\text{d}, D\text{d} and L\text{d} as shown for K\text{d} in Fig. 1f. With such low affinities, these neoepitopes would normally be considered non-binders.

The 18-mer sequence was queried for the presence of predicted K\text{d}, D\text{d} or L\text{d}-binding peptides. No D\text{d} or L\text{d}-binding peptides were predicted; three peptides were predicted to bind K\text{d} albeit with poor affinity (IC50 values between 692 and 864 nM) (Fig. 1f). Ironically, none of these three peptides were detected by MS among peptides eluted from MHC I of BMDCs pulsed with the 18-mer long peptide.

In order to determine if any peptides within Ccd85c\text{MUT} could be presented by MHC II molecules, we analyzed the interaction of H2-A\text{d} and H2-E\text{d} with TYIRPFETKVK, YIRPFETKVK and IRPFETKVK as well as their wild type counterparts, using a cell-surface density assay. In this assay, β-chains of H2-Ab\text{d} or H2-Eb\text{d} are expressed in fusion with the peptide of
Definition of the precise neoepitope of Ccdc85c\textsuperscript{MUT} that mediates tumor rejection. 

**a** The sequences of the 18-mer wild type and mutant peptides derived from Ccdc85c gene as well as their corresponding allelic fractions (the number of mutant/normal reads divided by the total number of reads [coverage] at a specific genomic position) are shown. 

**b** The top panel shows a schematic diagram of immunization and tumor challenge in BALB/cJ mice. The bottom panel (left) shows tumor growth in BALB/cJ mice immunized with Ccdc85c\textsuperscript{MUT} or Ccdc85c\textsuperscript{WT} and challenged with Meth A as described in Methods. Each line represents tumor growth in a single mouse (\(n=5\) mice per group). AUC for each group is plotted in the panel on the right. Data are presented as mean \(\pm\) SD. 

**c** Several truncated versions of the 18-mer Ccdc85c\textsuperscript{MUT} peptide were tested in tumor rejection assay. BALB/cJ mice were immunized and tumor challenged. Each line represents tumor growth in a single mouse. Although mice were immunized with individual peptides, the data for multiple peptides are grouped into one with the composition of the peptides shown on the right. The tumor rejection data for individual peptides are shown in Supplementary Fig. 1. Tumor rejection score (TRS) for each neoepitope is shown in the yellow box, where five represents a complete tumor protection and zero means no tumor rejection. 

**d** On the left panel, total Area Under the Curve (AUC) scores for each group in B are plotted. Each bar shows the average total AUC score for the indicated group (TRS = 5; \(n=35\), TRS = 4–4.5; \(n=40\), TRS = 3–3.3; \(n=25\), TRS = 2; \(n=30\), TRS = 0.1–1.5; \(n=60\)). Error bars represent standard deviation (SD). The P values corresponding to the comparison of TRS = 0 with TRSs 5.0, 4.0–4.5 and 3.0–3.3 were respectively <0.0001, <0.0001 and 0.0002. P values were calculated using 1-way ANOVA test adjusted for multiple comparisons (Tukey’s multiple comparison test). 

**e** Targeted MS-based detection of TYIRPFETKV and YIRPFETKV among MHC I peptides eluted from BMDCs pulsed with the 18-mer Ccdc85c\textsuperscript{MUT}. Heavy labeled synthetic peptides were spiked into the peptide samples; the labeled amino acid is marked with a bold character and the mutation is in red. Matched peak lists for the “heavy” and “light” ions were extracted and monitored, while only single charge y ions were plotted. See Methods for details. 

**f** Predicted (by NetMHC4.0) and measured IC\(50\) values of the binding of candidate precise neoepitopes of Ccdc85c\textsuperscript{MUT} to K\(d\). The candidate neoepitopes include those defined by tumor rejection and MS as in panels c and e. The other three candidate neoepitopes were predicted by NetMHC4.0 alone and were not active in tumor rejection. The affinities of the MS/TRS predicted neoepitopes were also measured for D\(d\) and L\(d\); measured affinities were below the level of detection.
interest and the amount of cell-surface MHC II, as a measure of intrinsc stability of MHC II, is quantified in engineered conditions. The measured cell-surface MHC density correlates well with the actual affinity of the peptide to MHC II. No significant difference was observed in binding of Ccdc85c\textsc{MUT} and Ccdc85c\textsc{WT} to H2-IA or H2-IE (Supplementary Table 1 and Supplementary Fig. 2a). Consistent with these findings, depletion of CD4 T cells in mice immunized with Ccdc85c\textsc{MUT} did not lead to reproducible and statistically significant abrogation of tumor rejection (Supplementary Fig. 2b). This was observed when CD4 depletion was carried out during the priming phase alone (as in Supplementary Fig. 2b), or throughout the entire experiment including the effector phase (i.e. post tumor challenge).

**Analysis of immune response against Ccdc85c\textsc{MUT}** BALB/c mice were immunized with 18-mer peptides containing Ccdc85c\textsc{MUT} as well as Ccdc85c\textsc{WT} as described in Methods. Annotation of the clusters was informed by higher expression of Nos2, Mrc1, Il1b, Il7r, Tcf7 and Ccr7 genes and a higher expression of Ifng (Fig. 2b, right panel). The majority of NK Cells (~80%) in the Ccdc85c\textsc{MUT} library were from the NKc(1) cluster which was the more cytokotoxic and active cluster (defined by higher expression of Cd44, Tnfα, and Ifng), while, the fraction of active NK cells in other libraries was about 55%.

To pinpoint differences in T cells of the four libraries, clusters 1 and 5 (activated CD4 and CD8 T cells) were computationally pooled and the expression of cytokotoxicity and other effector function genes were compared between libraries. Proliferating CD4/CD8 T were excluded from further analysis because the gene expression levels in these cells could be influenced by the cell cycle effect prominent in this cluster. Interestingly, Ccdc85c\textsc{MUT} library had the most contribution to the aforementioned pooled cells (31% Ccdc85c\textsc{MUT}, 25% Alms1\textsc{MUT}, 21% Cdr1, 20% Ctr2). Also, the normalized average gene expression (described in Methods) of cytokotoxicity (Gzmbl, Prf1, and Nkg7) and other effector function genes were significantly higher in T cells derived from the Ccdc85c\textsc{MUT} library compared to the control or Alms1\textsc{MUT} libraries. Similarly, T cells of Ccdc85c\textsc{MUT} library had a significantly higher expression of genes involved in TCR engagement (Nrho1 and Ifng). A transcription factor involved in transcription of cytokotoxicity genes, Eomes, had a significantly higher expression in T cells of Ccdc85c\textsc{MUT} library (Fig. 2c).

In the myeloid compartment, nine distinct clusters were identified. These are: macrophage1 (Mφ1), Mφ2 (defined by a moderate expression of Arg1 and lower expression of Cd302, Cd5l, Cd8 and Clqa), monocytic1 (M01), M03 (defined by a lower expression of Ccl8 and a higher expression of Ly6c, Cxcl9, Il1b, H2-Ab1, H2-Dm2, Mmp14, and Cd38), Mφ4 (defined by a higher expression of Nos2, Mrc1, Ifng, Pgf4, Clq1a, Clq1b, and Clq4c), DC1, Mo2 (defined by a higher expression of Ifgα, Tlr7, Tcr, Ace, and Adgre4), neutrophil (Ne) and DC2 (defined by a higher expression of Ccr7, Ccl5, Samsn1, Pgg5s, Gyz, Net1, and Rabgap1p) (Fig. 2b, left panel). The contribution of library Ccdc85c\textsc{MUT} to the most of myeloid clusters was minimal (ranging from 1.6% in Mφ3 to 13.4% in M02). The only exception is for Mφ1 which Ccdc85c\textsc{MUT} library forms ~20% of this cluster (Supplementary Fig. 5).

**Mutation-reversion analysis of CD8 T cell immunogenicity of Ccdc85c\textsc{MUT}** The studies described above examined the immunogenicity of Ccdc85c\textsc{MUT} when administered as a vaccine. We aimed now to analyze the role of Ccdc85c\textsc{MUT} in the immunogenicity of the Meth A tumor itself, and in vivo. The broader objective was to test if a poorly MHC 1-binding neoepitope residing within a tumor influences the CD8 immunogenicity of the tumor. CRISPR-guided gene editing was used to generate two variants of the Meth A. For purpose of this experiment, we refer to the original Meth A tumor with the endogenous Ccdc85c\textsc{MUT} as MUT1. Using CRISPR, the point mutation in Ccdc85c\textsc{MUT} was reversed back to its WT counterpart as described in Methods and
Supplementary Fig. 6; this line with the WT sequence of Ccdc85c is referred to as a Revertant (REV). As shown in Supplementary Fig. 6b, the REV tumor grows significantly faster than the original MUT1 line. The point mutation in Ccdc85cMUT was then reintroduced into the Revertant to generate the line MUT2 which recapitulates the original mutation as in Ccdc85cMUT. Thus, MUT1 and MUT2 are identical tumors (except that MUT1 is heterozygous for the mutation while MUT2 is homozygous for it) and are different from REV only with respect to the mutation in Ccdc85cMUT. Groups of mice were challenged with the three tumors individually (MUT1, REV or MUT2) and their TILs analyzed by scRNA seq. RNA from CD45 + cells (estimated 11,961 cells for the three libraries before QC and 10,265 after QC, and with an average coverage of 61,371 reads per cell and median of 2,137 genes per cell) was sequenced. Data from three libraries were pooled and expression of the top average TF-IDF scoring genes was compared between the three libraries (Fig. 3a). CD8 T cells of all three libraries showed CD8 activation markers including but not limited to: Cd69 and Cd44 as activation markers, LAMP1 (CD107a) as a measure of degranulation and Tbx21 (Tbet), a transcription factor involved in transcription of cytotoxicity-associated genes (Fig. 3b, upper panel). We then...
compared the gene expression patterns of the total immune cell population in TILs of MUT1, REV, and MUT2. The gene expression patterns in TILs of MUT1 and MUT2 showed a higher similarity to each other than to the REV: a simple hierarchical clustering (using Euclidean distance, and complete linkage) of the MUT1, REV and MUT2 libraries represented by the normalized average expression vector of top informative genes (selected by highest average TF-IDF score, see Methods) showed MUT1 and MUT2 are closer to each other (distance 1.097) than to the REV (distance 1.820) (Fig.3a upper panel) with 97% confidence in the hierarchy edge/branch from Rev on one side to MUT1 and MUT2 on the other (see full details in Supplementary Fig. 6).

In Fig. 3a bottom panel, where genes with variability among the three libraries are juxtaposed, it is clear that the difference between REV and MUT1 and MUT2 is more pronounced than the difference between MUT1 and MUT2.

To identify differences in CD8 T cells of the three libraries, the RNA sequencing data of the combined libraries were clustered based on the gene expression pattern of each cell type as described in Methods. Annotation of the clusters was informed by TCGA data. A simple hierarchical clustering of MUT1, REV and MUT2 libraries, based on the normalized average expression vector of top informative genes (selected by the highest average TF-IDF score) where, the Y value reflects the distance between clusters. The bottom panel represents the violin heat map plots of the top average TF-IDF scoring gene expression for the three libraries (expression of 26 genes is shown). Significant difference in distance of the REV versus MUT1 and MUT2 in the hierarchy is indicated by asterisk (please refer to Sup Fig. 6 for more details).

To identify differences in CD8 T cells of the three libraries, the RNA sequencing data of the combined libraries were clustered based on the gene expression pattern of each cell type as described in Methods. Annotation of the clusters was informed by TCGA data. A simple hierarchical clustering of MUT1, REV and MUT2 libraries, based on the normalized average expression vector of top informative genes (selected by the highest average TF-IDF score, see Methods) showed MUT1 and MUT2 are closer to each other (distance 1.097) than to the REV (distance 1.820) (Fig.3a upper panel) with 97% confidence in the hierarchy edge/branch from Rev on one side to MUT1 and MUT2 on the other (see full details in Supplementary Fig. 6).

In Fig. 3a bottom panel, where genes with variability among the three libraries are juxtaposed, it is clear that the difference between REV and MUT1 and MUT2 is more pronounced than the difference between MUT1 and MUT2.
NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-021-26646-5
ARTICLE

Fig. 4 Models of peptide/MHC I complexes indicate structural and physical correlates with immunogenicity. a For the tumor rejecting TYIRPETFVKV neoepitope, the leucine-to-phenylalanine substitution at position 6 is predicted to increase hydrophobic solvent accessibility by $17 \text{Å}^2$, with the aromatic phenylalanine ring partially exposed for interactions with T cell receptors. An overlay of the neoepitope and its wild type counterpart demonstrates the substantial differences between the wild type peptide and neoepitope. b For the tumor rejecting YIRPFETKV neoepitope, with the leucine-to-phenylalanine substitution at position 5, the modeling predicts structural alterations in exposed side chains in response to the mutation, as well as a reduction in exposed hydrophobic solvent accessible surface area of $23 \text{Å}^2$.

by both differentially expressed genes (DE Genes) and per cluster highly expressed genes identified by the TF-IDF analysis. T cells were re-clustered into 7 clusters by unsupervised clustering as described in Methods and enriched CD8 T cells populations were further analyzed (Supplementary Fig. 8a, b). Some cytotoxicity and effector function genes (Tnf (TNF), Gzm a (Granzyme a) and Ifng (Interferon)) had similar expression pattern in CD8 T cells of all three libraries; however, the normalized average gene expression of other cytotoxicity genes (Fasl (Fas ligand), Gzm b (Granzyme b) and Prf1 (Perforin1)) as well as other effector function genes (Pdcd1 (PD1) and Tbx21 (Tbet)) were significantly higher in CD8 T cells derived from MUT1 and MUT2 libraries compared to the REV library ($P$ value < 0.001). Similarly, CD8 T cells of MUT1 and MUT2 libraries had a significantly higher expression than the REV library, of early response genes which are involved in TCR engagement ($Nra4a1$, $Nra4a2$, and $Nra4a3$, $P$ value < 0.001) (Fig. 3b, bottom panel).

T cell receptors (TCRs) in the TILs of MUT1, REV, and MUT2 tumors. T cell receptors (TCRs) in the TILs of the three libraries were characterized using Grouping of Lymphocyte Interactions by Paratope Hotspots (GLIPH) analysis that groups together the TCRs into specificity groups based on the global and local similarities of the CDR3 regions of the TCRs. Based on the GLIPH algorithm, 40–42.9% of all distinct clonotypes contributed to forming a network/similarity-based specificity groups in each of the libraries, while the rest were standalone clonotypes (with no similarity to other clonotypes). The similar percentage of network-based specificity groups (40–42.9%) in all three libraries was expected because of the existence of other mutations (except Ccdc85cMUT) in all the three libraries. To further analyze the networks, we performed Louvain graph-based clustering of the specificity networks and calculated the modularity scores of the identified communities for each of the libraries (score of zero means the communities are the same and score of one refers to a perfect separation between communities). The modularity of a network with respect to its division into communities measures how well separated (diverse) the different nodes (clonotypes) forming the communities are from each other (see Methods). In the TILs, the TCR clonotypes that form communities/specificity groups are almost identical in frequency (42.9% for REV, and 42% and 40% for MUT1 and MUT2). However, the average modularity score of the communities/specificity groups including the most frequent (expanded) clonotypes is 0.53 for the REV, 0.73 for MUT1 and 0.77 for MUT2, indicating lower diversity of TCR clonotype in the TILs of REV than those of MUT1 and MUT2.

Using TCR-Seq analysis, top ten clonally expanded CD8 T cells were computationally pooled and further analyzed for gene expression patterns of their cytotoxic and effector functions. The normalized average gene expression of cytotoxicity-associated genes (Fasl or Fas ligand, Gzm b or Granzyme b, Prf1 or Perforin1, Nkg7 or Protein NK7) as well as other effector function genes (Tbx21 or Tbet, Pdcd1 or PD1 and Ifng or Interferon) were significantly higher in the clonally expanded CD8 T cells derived from MUT1 and MUT2 libraries compared to the REV library ($F$as $P$ value < 0.001, Gzm $b$ or $P$ value < 0.001, Prf1 or $P$ value < 0.001 Nkg7 or $P$ value < 0.001, Tbx21 or $P$ value < 0.001, Pdcd1 or $P$ value < 0.001 and Ifng or $P$ value < 0.001). Similarly, the top 10 clonally expanded CD8 T cells of MUT1 and MUT2 libraries had a significantly higher expression of early response genes which are involved in TCR engagement ($Nra4a1$ or NUR/77 $P$ value < 0.001, Nra4a2, and NUR-related factor 1 $P$ value < 0.001 and Nra4a3 or Orphan nuclear receptor TEC $P$ value < 0.001) (Fig. 3c). Interestingly, Ifng and Nkg7 which had a similar expression pattern in the pooled CD8 T cells of all three libraries (Supplementary Fig. 8c), had significantly higher expression in the top 10 clonally expanded CD8 T cells derived from MUT1 and MUT2 libraries compared to the REV library.

Molecular modeling of Ccdc85cMUT. To gain insight into how the leucine to phenylalanine mutation leads to immunogenic epitopes, we modeled the structures of the 11-mer TYIRPFETKV neoepitope and the 10-mer YIRPFETKV neoepitope bound to $K^b$. We modeled each corresponding WT peptide as well, to assess possible changes resulting from the mutation and thus infer how the neoepitopes might differ from self. We used
the same stochastic, template-based modeling procedure previously applied to murine neoepitopes. For the TYIRPFETKV 11-mer, the phenylalanine at position 6 is predicted to extend up from peptide near the MHC I α2 helix, increasing the amount of expressed hydrophobic surface 5% over the wild type peptide and potentially allowing the aromatic phenylalanine to interact with T cell receptors (Fig. 4a). Other than the side chain replacement, no conformational changes are predicted to occur in the peptide. For the YIRPFETKV 10-mer, the new phenylalanine at position 5 is predicted to pack between the peptide and the α2 helix, in this case reducing exposed hydrophobic surface area (Fig. 4b). Subtle structural changes are predicted for the exposed side chains at positions 6 and 9, which could be suggestive of changes not captured by static structural modeling, such as changes in peptide flexibility that lead to altered TCR recognition.

Discussion

A high binding affinity of peptides to MHC I is generally considered essential for immunogenicity. However, some reports with cancer neoepitopes show that even peptides with very low affinities for MHC I elicit CD8-dependent tumor rejection. These reports have used immunization with peptides to demonstrate immunogenicity. Here, we have asked and addressed if low affinity neoepitopes actually influence the natural immunogenicity of a tumor in vivo in the absence of artificial immunization. The answer is a clear affirmative. Using CRISPR to edit the cancer genome, our results show that introduction of a single point mutation into the Meth A tumor results in strong transcripntomic signatures of TCR engagement and cytotoxic functions in the CD8 T cells infiltrating the tumor. Extinction of this mutation eliminates that signature. Remarkably, the Ccdc85cMUT neoepitopes used here have very low affinities (IC50 values of 1,434 and 39,661 nM) for Kd. These results have been obtained during the examination of the natural growth of a tumor in the absence of any immunization and indicate that the low affinity MHC I-binding neoepitopes have a functional role in the immunogenicity of a tumor in vivo.

These findings are the most detailed yet, on the activity of a neoepitope that would be considered a non-MHC I binding epitope. Under the canonical view of MHC I–peptide interaction, epitopes with such low affinities are typically considered to be non-immunogenic and are routinely eliminated from further study. Our results show that such non-canonical neoepitopes indeed behave in a manner similar to the traditional high affinity MHC I-binding epitopes, and in ignoring them, we run the risk of ignoring a significant proportion of the cancer immunome. Studies with several thousand cancer patients with a wide array of cancers have also noted the strong correlation between the presence of low affinity neoepitopes and good clinical outcomes.

Methods

Purification of MHC I eluted peptides. MHC I peptides were immunoinaffinity purified as described before. MethA cells or BMDCs were lysed and MHC I molecules were immuno-affinity purified from cleared lysates with Hib antibodies cross-linked to Protein A-Sepharose 4B beads at 4°C. MHC I complexes and the bound peptides were eluted with 1% trifluoroacetic acid (TFA). Eluates containing MHC I molecules were loaded in pre-conditioned Sep-Pak tC18 96-well plates and light counterpart for visualization of the MS/MS spectra. The assessment of proper alignment of transitions between heavy and endogenous peptides. Raw data were processed and analyzed by Skyline (MacCoss Lab, Skyline v19.1.0.193, Seattle, USA). An ion mass tolerance of 0.055 m/z was used to extract fragment ion chromatogram. Peptides with precursor’s charge state z ≤ 3+ and fragment ion with z ≥ 2+ were used to monitor and control transitions corresponding to the peptide sequence and monitored by both fragment ions and fragment transitions with z = 1–4. We then enabled synchronization of isotopes for a proper alignment of transitions between heavy and endogenous peptides. Raw data were converted into Mascot generic format (mgf) by MConvert (Proteowizard, Palo Alto, CA 94304, USA) in order to extract matched peak lists for heavy and light counterparts for visualization of the MS/MS spectra. The achievement of MS/MS matching was done by plabel (Version 2.4.0.8, pfind studio, Sci. Ac., China) and Skyline.

For the data shown in Fig. 1d, targeted MS-based detection of TYIRPFETKV and YIRPFETKV among MHC I peptides eluted from BMDCs pulsed with the 18-mer Ccdc85cMUT heavy labeled synthetic peptides were spiked into the peptide samples; the labeled amino acid is marked with a bold character and the mutation is in red. Matched peak lists for the “heavy” and “light” ions were extracted and monitored, while only single charge y ions were plotted. First, the absence of “light” peptide and the presence of the “heavy” peptide were confirmed by Parallel Reaction Monitoring (PRM). The identity control was performed through tandem mass spectrometry of the both peptide samples (upper left and lower left, respectively). Then, the co-elution of the synthetic “heavy” and endogenous “light” fragment ions was measured by PRM in Cdd85cMUT pulsed BMDC MHC I–peptides. Figures were edited to improve resolution and readability.

MHC II peptide binding analysis. MHC II–peptide binding analysis was conducted by cell-surface density assay with modifications. NIH3T3 cells that stably express H2-Astd or -Ea1 heavy labeled through retroviral transduction of cells with pmxS-puro containing the full-length H2-Astd or -Ea1 with a C-terminal Strep-tag II (IBA GmbH), using packaging cell PLAT-E22,23. The stable cells were obtained by selection with puromycin (5 μg/ml) for two weeks. The expression constructs for the β subunit were designed to contain the signal peptide for HLA-DQβ1*0602, followed by the peptide via linkers and the mature region of the β subunit (H2-Aβ11 or H2-Eβ12), with a C-terminal 6x His-tag. The construct was inserted into pMXs-IG that contains IRES and GFP tag. The construct was inserted into pMXs-IG that contains IRES and GFP.
downstream of H2-Ab1d or H2-Eb1d. The stable H2-Aa1d or H2-Ea1d cells were transduced with retrovirus particles containing H2-Ab1d/peptide/pMXs-Ig or IA-1β1-2-IIIa peptide/pMXs-Ig. Cell surface expression of H2-Ab1d and cytoplasmic expression of GFP in GFP+ MHC+ cells were measured by flow cytometry 48 h after the transduction, using the following antibodies: H2-Ab, anti-mouse I-Ad mAb with the dilution of 1:10, 20 μl per sample (39-10-8, BioLegend) or mouse IgG3 isotype control with the dilution of 1:100, 10 μl per sample (m078-3, clone 6A3, Medical & Biological Laboratories Co. Ltd.) and goat anti-mouse IgG-PE with the dilution of 1:20, 15 μl per sample (sc-3767, Santa Cruz Biotechnology, Inc.); H2-A, anti-I-Eβ mAb with the dilution of 1:40, 10 μl per sample (115002, 14-4-8, Thermo Fisher) or mouse IgG2a isotype control with the dilution of 1:10, 10 μl per sample (Genetex, 6H13, Medical & Biological Laboratories Co. Ltd.) and goat F(ab′)2 Anti-mouse Ig-PE with the dilution of 1:20, 20 μl per sample (1012-09, Southern Biotechnology Associates Inc.). The ratio of MHC MFI to GFP MFI (MHC/GFP) for each MHC II-peptide combination was calculated and normalized to the MHC/GFP for respective MHC II-allele-Gp9. On each assay date, MHC/GFP for each MHC II-peptide combination was measured for three or four wells and their average was determined. The assay was repeated twice. Data were collected with SA3800 (Sony Imaging Products & Solutions Inc.) and analyzed using FCS Express 6 software (6.06.0022, De Novo Software, CA). The double-stranded DNA oligonucleotides encoding the signal sequence and peptide were synthesized (Genewiz Japan). The NIH3T3 cell line was obtained from the RIKEN Bioresource Center.

Mice and tumors. BALB/c mice (6–8 week-old females, stock # 000651) were purchased from the Jackson Laboratory and maintained in our specific pathogen-free mouse facilities under ethical approval from the Institutional Animal Care and Use Committee of the University of Connecticut School of Medicine. Experimental and control mice were kept in separate cages. Twelve light/12 dark cycle was used for mice housing. The temperature of the mice room and cages were kept around 75 °F. The humidity of the housing was 40–60%. Mice were euthanized by inhalation of carbon dioxide. Mth A cells that have been in our lab since 1988 were originally obtained from Lloy J Old. Mth A cells were passaged in ascites and were determined to be free from mycoplasma contamination.

Analysis of tumor growth. Area under the curve (AUC) is used as a tool to measure tumor growth23,24. Briefly, AUC was calculated by selecting “Curves & Regression” and then “Area under curve” from the “analyze” tool, using the Prism 5.0 (GraphPad). A tumor rejection score (TRS) has been utilized for reporting the proportion of mice which reject tumors completely or near completely in response to vaccination peptide. A mouse is scored asTRS of 5 indicates a tumor rejection in 100% of the mice, and a TRS score of 0 indicates tumor rejection in no mice. The values between 0 and 5 are allocated based on the proportion of mice rejecting a tumor.

Immunization. Fifty microliter of TiterMax (CytRx Corporation) or Day 7 granulocyte macrophage colony-stimulating factor-derived BMDCs (GM-CSF-BMDCs) were polymerized with 40 μg of the peptide per well. The peptide-polymerized BMDCs were used to immunize a single mouse. Immunizations were done twice, one week apart, and the mice were challenged with live tumor cells one week after the last immunization (Fig. 1b, upper panel). All immunizations were performed in the presence of CTLA4 blockade, using the IgG2a isotype (Clone: 9D9, Bio X Cell), administered with the second immunization and every 3 days after tumor challenge. Peptides were synthesized by JPT Peptide Technologies.

Bone marrow derived dendritic cells (BMDCs). Bone marrow cells (2–3 million/10 cm2 bacteriological Petri dishes) of 6- to 8-week-old mice were cultured in complete RPMI supplemented with 20 ng/mL recombinant murine GM-CSF (Peprotech) and incubated at 37 °C for 7 days to generate GM-CSF-BMDCs.

Fluorescence activated cell sorting. The antibody specific for Fixable Viability Dye eFluor® 780 (65-0885-14, dilution of 1:1000/sample) and CD45-PE (103106, clone: 30-F11, dilution of 1:50/sample) were purchased from eBioscience and Biolegend, respectively. Mouse FcR blocking reagent (130-092-575, dilution of 1:2000, Thermo Fisher) or mouse IgG2a isotype control with the dilution of 1:10, 20 μl per sample was used to block FcR interactions by paratope hotspots (GLIPH) algorithm from Glanville et al.19. GLIPH searches for global and local motif CDR3 similarity in TCR CDR regions with high contact probability. Each specificity group is analyzed in GLIPH for enrichment (of motif clusters). Global similarity measures CDR3 differing by up to one amino acid and local similarity measures the shared enriched CDR3 amino acid motifs with 10x fold-enrichment and probability < 0.001. Supplementary Tables 2–4 show the enriched CDR3 motifs of TCRs from TILs of MUT1, REV and MUT2 libraries.

Modularity score (as defined in igraph R package). The modularity of a graph with respect to some division (or vertex types) measures how good the division is, or how separated are the different vertex types from each other. It is defined as
\[
Q = \frac{1}{2m} \sum_{i,j} (A_{ij} - \frac{k_i k_j}{2m}) \delta(c(i),c(j)),
\]
here is the number of edges, \(A_{ij}\) is the element of the A adjacency matrix in row i and column j, \(k_i\) is the degree of i, \(k_j\) is the degree of j, \(c(i)\) is the type (or component) of i, \(c(j)\) that of j, the sum goes over all i and j pairs of vertices, and \(\delta(c(i),c(j))\) is 1 if \(c(i) = c(j)\) and 0 otherwise.

Clustered regularly interspaced short palindromic repeats (CRISPR). A guide RNA was designed that included the C -> A Gilead mutation in its seed region (Supplementary Table 5). The seed region refers to the 8–12 nucleotides proximal to the PAM; mutations in this region significantly limit Cas9’s ability to cleave target DNA. A single stranded oligodeoxynucleotide (donor ssODN) template was designed that contained 50-base pairs of homology to the endogenous sequence on either side of the target base (Supplementary Table 5). The donor ssODN was resuspended in TE buffer and aliquoted and stored at −20 °C. Prior to

Sample data analysis. Samples from the libraries were analyzed using the SC1 tool available at sc1.engr.uc conn.edu. Pre-processing quality control was conducted to exclude outliers and low quality cells based on the data distribution, from 20422 cells from 10x pipeline 15925 cells met our QC criteria (cells with over 30000 total UMIs or expressing less than 500 genes or over 5000 genes, with higher than 10% mitochondrial genes or less than 5% ribosomal genes were excluded from the analysis).

Cells were then clustered using Ward’s Hierarchical Agglomerative Clustering algorithm using the top average TF-IDF genes as features26, after log2(x+1) transformation of the data. Clusters were annotated based on one-versus-all differential expression analysis between clusters, determined by a p < 0.01 and absolute value of Log fold change > 1.
transfection, Cas9 ribonucleoprotein complexes (Cas9 RNP) were formed by incubating 10 µg of Alt-R HiFi Cas9 Nuclease v3 (IDT) with 8 µg of sgRNA in 0.3 M NaCl for 30 min at room temperature. To produce revertant clones, Cas9 RNP complexes as previously described and were mixed with 200 pmol of donor ssODN and delivered into 10⁵ MethA cells via electroporation with a Lonza 4D Nucleofector X kit using program DS-150 and Cell Line Solution SG (Lonza) according to the manufacturer’s protocols. Two days post-delivery, cells were split via limiting dilution into single-cell clones, allowed to expand for 21 days, and genotyped with PCR and Sanger sequencing at the Cdc25c mutation locus in a 96-well plate. Clones that were successfully edited by CRIPR, were identified by Sanger sequencing.

### Structural modeling of wild type/neoepitope peptide-MHC pairs.

Structural modeling of the 9-mer, 10-mer, and 11-mer wild type and neoepitope peptide/MHC pairs was conducted as previously described. Briefly, modeling utilized Rosetta and the re2015 energy function. The structures used as templates for modeling were PDB STG2⁹ for the 9-mer peptides and 5GSV³¹ for the 10-mer and 11-mer peptides. The templates were energy minimized via Rosetta FastRelax. As there was no 11-mer peptide/MHC structure containing H-2Kd as of November 2019, the 11-mer was approximated by interpolating a glycine between residues 5 and 6 of the template. Subsequently, the desired peptide sequence was introduced via mutation of the template peptide. Structures were first modeled with a low resolution centroid kinematic closure protocol then with a high resolution classical molecular dynamics protocol. To sufficiently sample the available conformational space, we modeled 10,000 decoys for each peptide/MHC. The lowest scoring decoy of each was retained for further analysis. Root-mean-square deviation of atomic positions (RMSD) of peptide common or backbone heavy atoms between wild type and mutant peptides was calculated and models were inspected visually for differences in structural features. Solvent-accessible surface areas were calculated in Rosetta using a probe radius of 1.4 Å.

### Statistical analysis.

P values for comparisons of MHC/GFP MFI and AUC scores were calculated using t-test and 1-way ANOVA test, respectively, adjusted for multiple comparisons. P values were adjusted for multiple comparison by False Discovery Rate method or “Dunnnet’s multiple comparison test” or “Tukey’s multiple comparison test”. P < 0.05 was considered statistically significant. Differential expression (DE) analysis is done by performing t-test to compare clusters/libraries. The t-test uses the Welch (or Satterthwaite) approximation with 0.95 confidence interval by calling the t-test available in R stats package. Results of the Log, fold change and the P value from the analysis are provided with a 1.5-fold change cutoff and 0.05 for P value.

### Reporting summary.

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability.

Single Cell RNA-Seq and TCR-Seq data generated in this study have been deposited in the GEO database under accession code GSE171100 and in the Supplementary Information Data file. There are no restrictions on data availability. Source data are provided with this paper.

Received: 30 November 2020; Accepted: 27 September 2021; Published online: 05 November 2021

### References.

1. Blum, J. S., Wearsch, P. A. & Cresswell, P. Pathways of antigen processing. Annu. Rev. Immunol. 31, 443–473 (2013).

2. Sette, A. et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. J. Immunol. (Baltim., Md.: 1950) 153, 5586–5592 (1994).

3. Assarsson, E. et al. A quantitative analysis of the variables affecting the repertoire of T cell specificities recognized after Vaccinia virus infection. J. Immunol. 178, 7890–7991 (2007).

4. Castle, J. C. et al. Exploiting the mutanome for tumor vaccination. Cancer Res. 72, 1081–1091 (2012).

5. Gubin, M. M. et al. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. Nature 515, 577–581 (2014).

6. Yadav, M. et al. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. Nature 515, 572 (2014).

7. Capitello, A. H. et al. Mutation position is an important determinant for predicting cancer neoantigens. J. Exp. Med. 217, https://doi.org/10.1084/jem.20190179 (2020).

8. Liu, S. et al. Efficient identification of neoantigen-specific T-cell responses in advanced human ovarian cancer. J. Immunother. Cancer 7, 156 (2019).

9. Koşaloğlu-Yalçın, Z. et al. Predicting T cell recognition of MHC class I restricted neoepitopes. Oncoimmunology 7, e1492508 (2018).

10. Duan, F. et al. Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict antitumor immunogenicity. J. Exp. Med. 211, 2231–2248 (2014).

11. Ebrahimi-Nik, H. et al. Mass spectrometry-driven exploration reveals nuances of neoepitope-driven tumor rejection. JCI Insight 4, https://doi.org/10.1172/JCI91215 (2019).

12. Glorani, E. et al. Differential binding affinity of mutated peptides for MHC class I is a predictor of survival in advanced lung cancer and melanoma. Ann. Oncol.: Off. J. Eur. Soc. Med. Oncol. 29, 271–279 (2018).

13. Rech, A. J. et al. Tumor immunity and survival as a function of alternative neoepitopes in human cancer. Cancer Immunol Res. https://doi.org/10.1158/ 3272-6066.cr-17-0559 (2020).

14. Martin, S. D. et al. Low mutation burden in ovarian cancer may limit the utility of neoantigen-targeted vaccines. PLoS ONE 11, e0155189 (2016).

15. Ott, A. P. et al. An immunogenic personal neoantigen vaccine for patients with melanoma. Nature 547, 217–221 (2017).

16. Keskin, D. B. et al. Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma. Nature 565, 234–239 (2019).

17. Zaidi, N. et al. Role of in silico structural modeling in predicting immunogenic neoepitopes for cancer vaccine development. JCI Insight 5, https://doi.org/10.1172/jci.insight.136991 (2020).

18. Miyadera, H., Ohashi, J., Lermark, Å., Kitamura, T. & Tokunaga, K. Cell-surface MHC density profiling reveals instability of autoimmunity-associated HLA. J. Clin. Investig. 125, 275–291 (2015).

19. Glanville, J. et al. Identifying specificity groups in the T cell receptor repertoire. Nature 547, 94–98 (2017).

20. Richman, L. P., Vanderheide, R. H. & Rech, A. J. Neoantigen dissimilarity to the self-proteome predicts immunogenicity and response to immune checkpoint blockade. Cell Syst. 9, 375–382 e374 (2019).

21. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367–1372 (2008).

22. Kitamura, T. et al. Retrovirus-mediated gene transfer and expression cloning: powerful tools in functional genomics. Exp. Hematol. 31, 1007–1014 (2003).

23. Mortira, S., Kojima, T. & Kitamura, T. Plat-E: an efficient and stable system for transient packaging of retroviruses. Gene Ther. 7, 1063–1066 (2000).

24. Duan, F. et al. Area under the curve as a tool to measure kinetics of tumor growth in experimental animals. J. Immunol. Methods 382, 224–228 (2012).

25. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 8, 14049 (2017).

26. Mousa, M. & Mánodoi, I. J. Single cell RNA-seq data clustering using TF-IDF based methods. BMC Genomics 19, 569 (2018).

27. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. Bioinformatics 26, 689–691 (2010).

28. Doo, R. & Baker, D. Macromolecular modeling with rosetta. Annu. Rev. Biochem. 77, 363–382 (2008).

29. Park, H. et al. Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. J. Chem. Theory Comput. 12, 6201–6212 (2016).

30. Frey, B. F. et al. Effects of cross-presentation, antigen processing, and peptide binding in HIV evasion of T cell immunity. J. Immunol. (Baltim., Md.: 1950) 200, 1853–1864 (2016).

31. Liu, W. J. et al. Protective T cell responses featured by concordant recognition of middle east respiratory syndrome coronavirus-derived CD8+ T cell epitopes and host MHC. J. Immunol. 198, 873 (2017).

32. Bradley, P., Misura, K. M. S. & Baker, D. Toward high-resolution de novo structure prediction for small proteins. Science 309, 1688 (2005).

33. Mandell, D. J., Coutsias, E. A. & Kortemme, T. Sub-angstrom accuracy in structure prediction for small proteins. Science 309, 1853 (2000).
Author contributions
P.S. and H.E.-N. conceptualized and designed the research, analyzed the data and wrote the manuscript. H.E.-N., R.E., W.C., H.M. and A.H. did the experiments. I.M. guided the analysis of RNA-seq data. M.M. designed and developed the methods for data analysis of single cell RNA-Seq. M.M., S.S. and T.S. analyzed the RNA-seq data. B.B. and G.K. performed the modeling of the peptides and MHC I. M.B.-S., J.M., H.S.P. and G.C. performed immunopeptidomics and MS analyses and interpreted the MS data.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-26646-5.

Correspondence and requests for materials should be addressed to Pramod K. Srivastava.

Peer review information Nature Communications thanks Neeha Zaidi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.