Abstract

Cancer genome sequencing has enabled the rapid identification of the complete repertoire of coding sequence mutations within a patient’s tumor and facilitated their use as personalized immunogens. Although a variety of techniques are available to assist in the selection of mutation-defined epitopes to be included within the tumor vaccine, the ability of the peptide to bind to patient MHC is a key gateway to peptide presentation. With advances in the accuracy of predictive algorithms for MHC class I binding, choosing epitopes on the basis of predicted affinity provides a rapid and unbiased approach to epitope prioritization. We show herein the retrospective application of a prediction algorithm to a large set of bona fide T cell–defined mutated human tumor antigens that induced immune responses, most of which were associated with tumor regression or long-term disease stability. The results support the application of this approach for epitope selection and reveal informative features of these naturally occurring epitopes to aid in epitope prioritization for use in tumor vaccines. Cancer Immunol Res; 2(6); 522–9. ©2014 AACR.

Introduction

We and others (1–3) have suggested that the vast number of personal, tumor-specific mutations found in the genome of patients with cancer provides a rich source of unique immunogens (“neoantigens”) for use in tumor vaccination strategies. These tumor neoantigens are attractive as vaccine targets as they are expected to bypass the immune-dampening effects of central tolerance and because their expression is exquisitely tumor specific. To bring this highly personalized treatment approach to patients with cancer, one crucial challenge is the choice of which of the many possible personal mutated epitopes to incorporate in the vaccine.

Cancer genomes vary widely in the number of total and coding sequence mutations depending on the tumor type (4). The five most common tumor types in the United States (prostate, breast, lung, colon, and melanoma) harbor an average of 25 to 500 nonsynonymous coding sequence mutations. Vaccination approaches that use irradiated whole-tumor cells or various forms of cell lysates (5–11) have attempted to capture all such neoantigens (in addition to native tumor-associated antigens). Although these strategies seem comprehensive and have resulted in clinical benefit in some cases, they do not favor any particular T-cell immunogens. Hence, potentially highly effective immunogens may be drowned out within the vast sea of immunologically irrelevant antigens. Such complete antigen preparations are similar to the endogenous presentation of the tumor cell to the immune system and lack the “pharmacologic specificity” of a rationally designed vaccine.

A more selective but still comprehensive approach for neoantigens could be envisioned by using every identified coding mutation as a separate immunogen. Although possibly feasible from the technical standpoint—especially for tumors with a low mutation load—the dilution of the potent immunogens is likely to reduce their effectiveness, and, thus, a more discriminating approach to identify the most effective subset seems advisable.

Potential Strategies to Identify and Prioritize Mutated Antigens

Multiple biochemical and biologic techniques are available that can help prioritize candidate mutated antigens for inclusion in tumor vaccines.

Mass spectrometry

Great strides in the fields of mass spectrometry (MS) and associated computational algorithms have enabled the characterization of the MHC-displayed “ligandome” (12, 13). This approach can be used to test whether a mutated peptide (or a native tumor-associated peptide) is displayed by tumor cells. This information is important as the peptide–MHC complexes are the substrates recognized by T-cell receptors (TCR). However, the approach is limited technically by insufficient amounts of tumor tissue and conceptually by the observation that few peptide-bound MHC targets are needed for an effective T-cell response (14, 15). As a result, many useful but less abundant targets on the cell may be bypassed in favor of those that are less potent but more highly represented.
**Ex vivo T-cell assays**

Peripheral blood mononuclear cells or tumor-infiltrating lymphocytes can be tested in antigen-specific *ex vivo* assays to identify neoantigens that stimulate existing T-cell populations. This strategy would be expected to reveal the patient’s natural response to neoantigens (2, 16). However, routine clinical application of *ex vivo* assays is costly and technically challenging, given the number of neoantigen mutations (requiring rapid preparation of many stimulatory immunogens), the requirement of MHC-matched antigen-presenting cells for some of these assays, and the relative insensitivity of these techniques. Most importantly, using *ex vivo* assays as a filter for neoantigen selection limits the spectrum of T-cell reactivity to existing T-cell responses. In patients with clinically evident tumors, this would restrict the selected neoantigen repertoire to the existing and possibly ineffective T-cell responses. It is currently unknown whether enhancement of an ongoing T-cell response or generation of *de novo* responses is clinically relevant for an effective tumor vaccine. Other biologic assays, such as *in vitro* or *in vivo* immunization of a humanized mouse, could also be considered, but such assays are likewise technically challenging, costly, and conceptually limited.

**In silico prediction of peptide–MHC binding**

Generation of an immune response to any mutated peptide sequence and recognition of tumor cells containing that peptide depend critically on the ability of the patient’s MHC molecules to effectively bind to the mutated peptide and present it to a T cell. Advanced algorithms using neural network-based learning approaches have been developed to capitalize on large amounts of data describing peptides that bind with different strengths to a wide variety of class I MHC molecules (17). These algorithms allow rapid *in silico* prediction of peptide-binding strength to patient-specific MHC alleles, and potentially enable a more rapid and less restrictive approach to filter the list of candidate neoantigens from sequencing data. Using results from next-generation DNA sequencing, we have evaluated the binding for more than 100 different predicted peptides to understand the boundaries of the accuracy of prediction by these algorithms (18). To link this *in silico* analysis to potentially clinically and biologically relevant observations, we present here an analysis of 40 neoantigens previously identified as CD8⁺ T-cell targets in the literature.

The Predicted Binding Characteristics of Tumor Neoepitopes Recognized by T Cells in Patients with Antitumor Immunity

We have conducted an extensive search of the literature, including recent reviews on neoantigens (1, 2) from PubMed, and the most comprehensive list of cancer vaccine antigens compiled by the Cancer Research Institute (19), identifying reports of spontaneous CD8⁺ T-cell responses in patients with cancer in whom the target epitopes were discovered subsequently. To avoid bias of the results, reports of vaccinations with known epitopes or of selected searches for single T-cell epitopes (such as for an immune response to a known mutated oncogene) were not included. Multiple reports of spontaneous CD8⁺ T-cell epitopes were identified, and, remarkably, in each case following an unbiased search for the dominant T-cell epitope, the target epitope was a neoantigen. Two thirds of the patients in these reports experienced significant partial or complete tumor regression or long-term stable disease, either spontaneously or following therapy.

As shown in Table 1, 31 of these 40 neoepitopes were identified in an unbiased manner based on cDNA expression cloning or MHC–peptide elution, while the remaining nine were found on the basis of genomic mutation and epitope-binding predictions. These neoantigens resulted from 35 missense mutations and five frameshift mutations (that led to novel open reading frames, neoORFs) and are restricted by 11 different HLA alleles, representing both common and less common alleles, as expected from sampling of the population at large. Approximately 80% of these are somatic mutations found exclusively in the tumors of individual patients. The remaining alterations are polymorphic loci within hematopoietically restricted minor histocompatibility antigens (mHAg) identified following hematopoietic stem cell transplantation for blood malignancies. In almost every case, the mutated peptide was significantly (>100×) more potent than the cognate native peptide in the induction of T-cell IFN-γ production or cytotoxicity. These examples represent seven different cancer types (non–small cell lung cancer, melanoma, renal cell carcinoma, bladder cancer, B-cell acute lymphoblastic leukemia, multiple myeloma, and chronic lymphocytic leukemia).

Because these neoepitopes are associated with biologic responses, they provide an ideal set of sequences for retrospective peptide affinity predictions to "reverse engineer" predictable characteristics of effective epitopes. For this analysis, we used the netMHCpan2.4 algorithm (Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark; www.dtu.dk; ref. 20). NetMHCpan is an artificial neural network–trained algorithm with an extensive training dataset (17), including 43 HLA-A and HLA-B alleles, representing approximately 90% and 60%, respectively, of the allelic population distribution, with more than 1,000 members each in the training set. NetMHCpan was determined to be one of the most accurate predictive algorithms in a 2012 competition (21). We applied this algorithm to individually predict MHC binding for all possible tiled peptides containing the mutated or the corresponding unmutilated residues of these observed spontaneous epitopes to determine:

- whether the naturally recognized epitopes would have been predicted;
- the predicted affinity of each mutated epitope; and
- the predicted affinity of each cognate native epitope (focusing on only the missense mutations).

**Functional neoepitopes are correctly predicted by a class I MHC–peptide binding algorithm**

Thirty-one of the epitopes shown in Table 1 were identified by *ex vivo* T-cell reactivity or MS and did not use genomic sequence or binding prediction information as a component of their identification. For all but one of these 31 epitopes, we found that the reported epitope was the peptide with the

---

**Tumor Neoepitopes in Humans**

*Published OnlineFirst March 3, 2014; DOI: 10.1158/2326-6066.CIR-13-0227*
Table 1. Biologic features and predicted binding affinities of neoantigen-directed T-cell responses in humans

<table>
<thead>
<tr>
<th>Group</th>
<th>Gene (Ref.)</th>
<th>Identification approach</th>
<th>Favorable clinical response</th>
<th>MUT&gt;&gt;&gt;NAT T-cell response detected (approach)</th>
<th>HLA allele</th>
<th>Mutated epitope (native allele)</th>
<th>Predicted binding affinity (IC50 nmol/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ECGF-1 (ref. 32; miHAg)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (y)</td>
<td>B'-07:02</td>
<td>RH-KRRPLAL (R)</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>ME-1 (31)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (y)</td>
<td>A'-02:01</td>
<td>FLK-ERMGEV (A)</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>PLEKHM2 (16)</td>
<td>WES</td>
<td>Yes</td>
<td>Yes (y)</td>
<td>A'-01:01</td>
<td>LTDRLFTCY (H)</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>FNDC3B (18)</td>
<td>WES</td>
<td>Yes</td>
<td>≤(10×)</td>
<td>A'-02:01</td>
<td>VMSHPAPPP (L)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>PRDX5 (27)</td>
<td>cDNA</td>
<td>NR</td>
<td>Yes (C)</td>
<td>A'-02:01</td>
<td>LLLDDLVLIS (S)</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>MATN2 (16)</td>
<td>WES</td>
<td>Yes</td>
<td>Yes (y)</td>
<td>A'-02:01</td>
<td>KTLTSSAFV (E)</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>DDX21 (33)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (C)</td>
<td>A'-01:01</td>
<td>EALTQPRR (S)</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>RBAF (29)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (C)</td>
<td>B'-07:02</td>
<td>RPHEPESAF (G)</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>GAS7 (16, 34)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (C)</td>
<td>A'-02:01</td>
<td>SLADEAYV (H)</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>ATR (35)</td>
<td>WES</td>
<td>Yes</td>
<td>≤(10×)</td>
<td>A'-02:01</td>
<td>KLYEYPKK (S)</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>SIRT2 (29)</td>
<td>cDNA</td>
<td>Yes</td>
<td>≤C (10×)</td>
<td>A'-03:01</td>
<td>KIFSEVTLP (P)</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>EF2 (36)</td>
<td>MS</td>
<td>NR</td>
<td>Yes (C)</td>
<td>A'-02:01</td>
<td>ETQVSSNY (E)</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>KIAA0223 (HA-1; ref. 37; miHAg)</td>
<td>MS</td>
<td>Yes (severe)</td>
<td>NR</td>
<td>A'-02:01</td>
<td>VNLDDILEA (R)</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>GAPDH (34)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (C)</td>
<td>A'-02:01</td>
<td>GIEQUHV (M)</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>BCL2A1 (ref. 38; miHAg)a</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (C)</td>
<td>A'-24:02</td>
<td>DYLQVLIQ (C)</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>HPS70 (39)</td>
<td>cDNA</td>
<td>NR</td>
<td>NR (C)</td>
<td>A'-02:01</td>
<td>SIFGDEY (F)</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>ACTININ (30)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (C)</td>
<td>A'-02:01</td>
<td>FASNQVGV (K)</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>CDK12 (16)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (y)</td>
<td>A'-02:01</td>
<td>CILQKLT (E)</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>KIAA1440 (40)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (y)</td>
<td>A'-01:01</td>
<td>QTAGEDL (T)</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>HAUS3 (16)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (y)</td>
<td>A'-02:01</td>
<td>ILNAMAKI (T)</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>BCL2A1 (ref. 38; miHAg)b</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (C)</td>
<td>B'-44:03</td>
<td>KEFEDDIWN (G)</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>PPP1R2B (16, 28)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (C)</td>
<td>A'-01:01</td>
<td>YTDHQCQY (P)</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>HB-1 (ref. 41; miHAg)</td>
<td>cDNA</td>
<td>NR</td>
<td>Yes (C)</td>
<td>B'-44:03</td>
<td>EKEKRSLiH (Y)</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>MUM-2 (42)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (C)</td>
<td>B'-24:02</td>
<td>SSSLRRGSLDSY (R)</td>
<td>184</td>
</tr>
<tr>
<td></td>
<td>KIAA0205 (43)</td>
<td>cDNA</td>
<td>NR</td>
<td>Yes (C)</td>
<td>B'-44:03</td>
<td>AEPDQOTW (N)</td>
<td>258</td>
</tr>
<tr>
<td></td>
<td>GPNMB (29)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (C)</td>
<td>A'-03:01</td>
<td>TLGQWLLTTP (G)</td>
<td>282</td>
</tr>
<tr>
<td>2</td>
<td>CSNK1A1 (16)</td>
<td>WES</td>
<td>Yes</td>
<td>Yes (y)</td>
<td>A'-02:01</td>
<td>GLQGDIYL (S)</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>CLPP (44)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (C)</td>
<td>A'-02:01</td>
<td>ILDKVLYL (P)</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>C1TP8B (45)</td>
<td>cDNA</td>
<td>Yes (?)</td>
<td>Yes (C)</td>
<td>A'-24:02</td>
<td>SYLDSGIF (S)</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>SNRP116 (29)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (C)</td>
<td>A'-03:01</td>
<td>KILDAWAQ (E)</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>OS9 (46)</td>
<td>cDNA</td>
<td>NR</td>
<td>Yes (C)</td>
<td>B'-44:03</td>
<td>KELEGILL (P)</td>
<td>60</td>
</tr>
</tbody>
</table>

(Continued on the following page)
### Table 1. Biologic features and predicted binding affinities of neoantigen-directed T-cell responses in humans (Cont’d)

<table>
<thead>
<tr>
<th>Group</th>
<th>Gene (Ref.)</th>
<th>Identification approach&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Favorable clinical response</th>
<th>MUT&gt;&gt;&gt;NAT T-cell response detected (approach)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>HLA allele</th>
<th>Mutated epitope (native allele)</th>
<th>Predicted binding affinity (IC&lt;sub&gt;50&lt;/sub&gt; nmol/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>MYH2 (47)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Yes (C)</td>
<td>A’03:01</td>
<td>KINKNPKYK (E)</td>
<td>4.960</td>
</tr>
<tr>
<td></td>
<td>MART-2 (48)</td>
<td>cDNA</td>
<td>Yes (weak)</td>
<td>Yes (C)</td>
<td>A’01:01</td>
<td>FLEGNEVKGTY (G)</td>
<td>4.504</td>
</tr>
<tr>
<td>neoORF</td>
<td>NFYC (49)</td>
<td>cDNA</td>
<td>NR</td>
<td>Yes (C)</td>
<td>B’52:01</td>
<td>AQQTKEV (Q)</td>
<td>5.701</td>
</tr>
<tr>
<td></td>
<td>CDK4 (50)</td>
<td>cDNA</td>
<td>NR</td>
<td>Yes (C)</td>
<td>A’02:01</td>
<td>ACDPHSGFHVFV (R)</td>
<td>25.222</td>
</tr>
<tr>
<td>neoORF</td>
<td>pARF14-ORF3 (51)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Not relevant</td>
<td>A’11:01</td>
<td>AVCPWTVLIR</td>
<td>Not relevant</td>
</tr>
<tr>
<td></td>
<td>HAM5-1 (ref. 52; miHAg)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Not relevant</td>
<td>B’44:03</td>
<td>MEIFREVSHF</td>
<td>Not relevant</td>
</tr>
<tr>
<td></td>
<td>PANE-1 (ref. 53; miHAg)</td>
<td>cDNA</td>
<td>NR</td>
<td>Not relevant</td>
<td>A’03:01</td>
<td>RWDLPPGLK</td>
<td>Not relevant</td>
</tr>
<tr>
<td></td>
<td>MUM1 (54)</td>
<td>cDNA</td>
<td>Yes</td>
<td>Not relevant</td>
<td>B’44:02</td>
<td>(L)EEKLMLVF&lt;sup&gt;c&lt;/sup&gt; (S)</td>
<td>409 (434)</td>
</tr>
<tr>
<td>neoORF</td>
<td>P2&gt;5 (ref. 55; miHAg)</td>
<td>Linkage</td>
<td>Yes</td>
<td>Not relevant</td>
<td>B’07:02</td>
<td>TPHNQRQVC</td>
<td>Not relevant</td>
</tr>
</tbody>
</table>

**NOTE:** Predicted affinity highlighting: green type, strong/moderate binder (IC<sub>50</sub> ≤ 150 nmol/L); blue type, weak binder (IC<sub>50</sub> between 150 and 500 nmol/L); and black type, nonbinder (IC<sub>50</sub> ≥ 500 nmol/L).

Bold, identification by binding predictions and either genomic sequence or genetic linkage. Underlined values: comparable experimentally measured binding affinity of mutated and native peptides were reported.

Abbreviations: cDNA, cDNA library expression cloning; NR, not reported; WES, whole-exome sequencing.

<sup>a</sup>Two separate CTL clones specific for peptides restricted by distinct MHC alleles (A2402 and B4403) were identified. Each peptide spanned a distinct missense mutation in the BCL2A1 gene.

<sup>b</sup>C, cytotoxicity; γ, IFNγ stimulation; T, CD107a stimulation.

<sup>c</sup>Special case of a missense, partial neoORF.
strongest predicted MHC-binding affinity among the tiled peptides containing the mutation. The only exception was a MUM1-derived 10mer containing an additional leucine at the N-terminus that had a slightly better predicted affinity (IC\textsubscript{50} 409 nmol/L) than the observed 9mer (IC\textsubscript{50} 434 nmol/L). We conclude that the MHC–peptide binding prediction algorithm netMHCpan consistently predicts the naturally recognized tumor neoepitope from all of the possible epitopes harboring a specific mutation.

Most functional neoepitopes have strong to moderate predicted IC\textsubscript{50}

Twenty of 31 (65%) of the naturally recognized missense and neoORF epitopes had predicted IC\textsubscript{50} < 50 nmol/L (strong binders) and three of 31 (10%) had a predicted IC\textsubscript{50} between 50 and 150 nmol/L (moderate binders). Thus, 75% of the dominant T-cell clones isolated from the naturally occurring T-cell populations recognize an epitope with a strong or moderate predicted affinity (IC\textsubscript{50} < 150 nmol/L) for the patient’s MHC allele. Because an unbiased functional assay (cytolysis, IFN\gamma production, or MS) was the critical test used to identify each of the stimulating peptides in these 31 examples, it is unlikely that there was an experimental bias toward the identification of epitopes with higher predicted affinity. Four of 31 naturally recognized peptides were predicted to be “weak” binding peptides (IC\textsubscript{50} between 150 and 500 nmol/L), indicating that a total of 27 of 31 (87%) of the naturally occurring epitopes would have been considered as binding peptides (IC\textsubscript{50} < 500 nmol/L affinity) using netMHCpan.

Conversely, only four of 31 naturally recognized peptides were predicted by netMHCpan to be nonbinders (IC\textsubscript{50} > 500 nmol/L); they may be false negatives from the prediction algorithm or may represent low affinity yet functional epitopes. Although these alternatives cannot be distinguished on the basis of the available data, three observations are relevant. First, for the three epitopes arising from missense mutations (MART-2, NFYC, and CDK4), cytolytic activity was preferentially induced by the mutated peptide and not by the native peptide at a range of peptide concentrations (1–10 nmol/L) comparable with those observed with more strongly predicted binding peptides. Second, for the Arg → Cys CDK4 mutation, the highly oxidizable sulfur residue may contribute serendipitously to MHC binding as a “pseudo-” anchor residue that could not have been accounted for by the prediction algorithms. Finally, T cells recognizing the fourth epitope (the mHAg P2 × 5) represented as much as 1.6% of all circulating T cells following the therapeutic infusion of donor lymphocytes. Results from these 40 examples dataset suggest that there are limitations to the capability of predictive algorithms and that up to 15% of target T-cell epitopes may be missed by the prediction algorithms.

Most of the cognate native peptides are predicted to bind MHC equally to the mutated peptides

In addition to analyzing MHC binding to the mutated epitopes, we also compared the predicted affinities of the cognate native epitopes corresponding to all 35 missense epitopes in Table 1 and identified three distinct classes. The predominant class (26 of 35 or 74%; group 1) was composed of native/mutated pairs that were predicted to bind with comparable affinity [with 23 of 26 showing strong to moderate predicted binding (IC\textsubscript{50} < 150 nmol/L) and the remaining three showing weak binding (IC\textsubscript{50} between 150 and 500 nmol/L)]. Despite comparable predicted binding, in almost all cases, the mutated peptide had been found to be significantly more potent in stimulating T cells than the native peptide. A smaller group (6 of 35 or 17%; group 2) showed low predicted binding for the native epitope and strong binding for the mutated epitope, directly correlating with the differential T-cell responses to the mutated and native peptides. Finally, in a minority of cases both mutated and native peptides were predicted to be nonbinding (3 of 35 or 9%; group 3). We note that each group in Table 1 comprises multiple HLA with no apparent bias in representation. Although the existence of the group 1 and 2 epitopes is not surprising, the predominance of the group 1 epitopes (containing 74% of all missense epitopes and four times more abundant than group 2) with comparable affinities for both the mutated and native peptides is unexpected.

Discussion

The MHC-bound peptide can be considered as a double-sided “key,” which must fit both the MHC and the TCR “locks” to stimulate an immune response and for subsequent target-cell cytolysis (Fig. 1A). Sequence-specific binding of peptides to the MHC molecule is highly dependent on the interactions of the peptide side chains at particular positions (“anchors”) along the length of the peptide with chemical moieties defined by the polymorphic residues that constitute the MHC-binding pocket (22–24): hence, predictive calculations are sequence dependent (25). Furthermore, analysis of these critical MHC-binding positions and residues over a wide range of MHC alleles shows that only a few positions of the peptide are anchor positions and only a few amino acids at the anchor positions of the peptide contribute to binding in a positive manner (26). On the other side of the “key,” TCR recognition of the peptide–MHC complex gains specificity from the ordered presentation of the other face of the peptide conferred by the anchoring residues.

For the majority of the missense mutations, both the native and the mutated peptides were predicted to be binding peptides (group 1; Fig. 1B). This observation is almost certainly a consequence of the mutations affecting the region of the peptide “key” that is involved in TCR recognition. In all but two of these 26 examples, the mutation was in a nonanchor position (as identified by the online tool provided at http://www.syfpeithi.de/; ref. 26). In the two nonconforming examples (PLEKHM2 and KIAA11440), a second anchor residue was already present in the native peptide. Other investigators have also reported mutations with equivalent affinity predictions for the native and the mutated peptide pairs (16, 27). Our broader analysis suggests that such mutant epitopes are a common phenomenon. Only a minority of the missense mutations were found in group 2 characterized by nonbinding of the native peptide. All of the group 2 examples, except for MYOSIN, were
mutations to preferred anchor residues at critical anchor positions.

Although the majority of the naturally occurring tumor epitopes were derived from the corresponding native peptides predicted to bind MHC, the vast majority (>98%) of the native human peptidome is not predicted to contain peptides that are binding epitopes of human MHC (our unpublished analysis using netMHCpan). Random mutational events that convert a nonbinding peptide (the vastly predominant target) to a binding peptide (group 2) are expected to be rare because they require mutation to one of only a few specific amino acids at a small number of anchor positions. For most MHC molecules, there are only one or two important anchor positions, and usually only two or three amino acids at those positions promote binding. Conversely, nonanchor positions are three to four times more abundant than anchor positions, and most mutations to native binding epitopes in these nonanchor positions would maintain MHC binding (group 1). This simple probabilistic explanation may be sufficient to account for the predominance of the observed group 1 epitopes (derived from the vastly underrepresented class of native peptides that are predicted to bind MHC). Alternatively, more complex explanations may be required. For example, aspects of central immune tolerance that are currently not well understood may cause the extent TCR repertoire to more effectively respond to peptides presenting a surface chemically distinct from any native peptide (group 1) than to peptides that more efficiently present an otherwise native peptide surface (group 2).

We did not observe weak native peptide binders that converted to strong/moderate mutated binders or strong/moderate native peptide binders that converted to weak mutated peptide binders. Although this may reflect the relatively limited dataset we used, it could also be that such upgrading or downgrading of binding involves anchor residue changes that moderately increase/decrease binding affinity but retain similar chemical structure of the nonanchor residues available for TCR recognition. In these scenarios, because the native peptides could bind to MHC, central immune tolerance may have effectively deleted cells with the reactive TCR, rendering both native and mutated peptides nonimmunogenic.

From the perspective of vaccine efficacy, we propose that mutations resulting in either group 1 or 2 binders should be considered as acceptable for use as immunogens, as both types of mutated neoepitopes have been found in vivo in patients with cancer with spontaneous tumor regressions and in long-term cancer survivors. Notably, in long-term survivors, T cells specific to mutated tumor epitopes from both groups 1 and 2 have been found to persist over many years (28, 29).

From the perspective of safety, there have been no reports of immune-mediated toxicities (except for the expected occurrence of GVHD as a result of responses against miHAg) despite the observation that for most of the mutations, the cognate native peptide was predicted and experimentally demonstrated in some cases (18, 27, 30, 31), to bind MHC as well as the mutated peptide (group 1). Importantly, in almost all cases the mutant peptide was shown to be more potent than
the native peptide in stimulating T-cell cytotoxicity or IFNγ production. The absence of autoimmune toxicity in these patients fits with the model that T cells reactive to the native epitope were eliminated by central immune tolerance and that T cells reactive to the mutated epitope do not cross-react to the native epitope as the mutation exclusively affects the TCR binding region.

In conclusion, in silico peptide-binding predictions provide a useful and rapidly deployable tool to capture the types of immunogens that are naturally observed in patients with cancer, many of whom experienced tumor regression and sometimes long-term tumor control. Moreover, results from our retrospective prediction study reveal features of these epitopes to further guide inclusion as immunogens in vaccines. We have recently initiated a clinical study using personalized neoantigen epitopes identified by whole-exome sequencing and prioritized by MHC-binding predictions in which we will carefully monitor the immune response to each mutation (NCT01970358).

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

References
7. Randazzo M, Temess P, Opelz G, Kleist C. Immunization with neoepitope-based vaccines was financially supported by the Blavatnik Family Foundation and the NIH (NHLBI:5 R01 HL103532-03; NCI-8R01CA155010-02). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
19. van der Bruggen P, Sto khoembt V, Vigneron N, Van den Eynde BJ.


