

# Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity

Fei Duan,<sup>1</sup> Jorge Duitama,<sup>2</sup> Sahar Al Seesi,<sup>2</sup> Cory M. Ayres,<sup>3</sup> Steven A. Corcelli,<sup>3</sup> Arpita P. Pawashe,<sup>1</sup> Tatiana Blanchard,<sup>1</sup> David McMahan,<sup>1</sup> John Sidney,<sup>4</sup> Alessandro Sette,<sup>4</sup> Brian M. Baker,<sup>3</sup> Ion I. Mandoiu,<sup>2</sup> and Pramod K. Srivastava<sup>1</sup>

<sup>1</sup>Department of Immunology and Carole and Ray Neag Comprehensive Cancer Center, University of Connecticut School of Medicine, Farmington, CT 06030

<sup>2</sup>Department of Computer Science and Engineering, University of Connecticut, Storrs, CT 06269

<sup>3</sup>Department of Chemistry and Biochemistry and Harper Cancer Research Institute, University of Notre Dame, Notre Dame, IN 46556

<sup>4</sup>LaJolla Institute of Allergy and Immunology, La Jolla, CA 92037

**The mutational repertoire of cancers creates the neoepitopes that make cancers immunogenic. Here, we introduce two novel tools that identify, with relatively high accuracy, the small proportion of neoepitopes (among the hundreds of potential neoepitopes) that protect the host through an antitumor T cell response. The two tools consist of (a) the numerical difference in NetMHC scores between the mutated sequences and their unmutated counterparts, termed the differential agretopic index, and (b) the conformational stability of the MHC I–peptide interaction. Mechanistically, these tools identify neoepitopes that are mutated to create new anchor residues for MHC binding, and render the overall peptide more rigid. Surprisingly, the protective neoepitopes identified here elicit CD8-dependent immunity, even though their affinity for K<sup>d</sup> is orders of magnitude lower than the 500-nM threshold considered reasonable for such interactions. These results greatly expand the universe of target cancer antigens and identify new tools for human cancer immunotherapy.**

## CORRESPONDENCE

Pramod Srivastava:  
Srivastava@uchc.edu  
OR  
Ion I. Mandoiu:  
ion@enr.uconn.edu

Abbreviations used: DAI, differential agretopicity index; RMSD, root mean square deviation; RMSF, root mean square fluctuation; SNV, single-nucleotide variant.

The idea that neoepitopes created by random mutations in tumor cells, termed as individually specific tumor antigens or unique antigens, are responsible for the immunogenicity of tumors has been around for over 20 yr (Srivastava, 1993). There has been strong experimental evidence for their existence and activity in murine and human tumors (Duan et al., 2009; van der Bruggen et al., 2013), and mathematic modeling has predicted the existence of tens to hundreds of neoepitopes in individual human tumors (Srivastava and Srivastava 2009). The recent revolution in high-throughput DNA sequencing and accompanying bioinformatics approaches has finally made it possible to actually identify the individually specific neoepitopes in individual

cancers. Using this methodology, Segal et al. (2008) showed that human breast and colon cancers harbor tens of putative mutational neoepitopes; Rajasagi et al. (2014) have also published similar data for chronic myelogenous leukemia and other human cancers. A genomic/bioinformatic approach to identify such epitopes in a mouse melanoma also led to identification of hundreds of epitopes (Castle et al., 2012). A similar approach led to identification of a neoepitope in a methylcholanthrene-induced sarcoma in an immunocompromised mouse; transplantation of this tumor into an immune-competent animal led to epitope-dependent tumor regression (Matsushita et al., 2012). In human studies, association of favorable clinical

F. Duan, J. Duitama, and S. Al Seesi contributed equally to this paper.

J. Duitama's present address is Agrobiodiversity Research Area, International Center for Tropical Agriculture (CIAT), 6713 Cali, Colombia.

© 2014 Duan et al. This article is distributed under the terms of an Attribution-Noncommercial-Share Alike-No Mirror Sites license for the first six months after the publication date (see <http://www.rupress.org/terms>). After six months it is available under a Creative Commons License (Attribution-Noncommercial-Share Alike 3.0 Unported license, as described at <http://creativecommons.org/licenses/by-nc-sa/3.0/>).

course of disease with a dominant immune response to mutated neopeptides has been demonstrated (Lennerz et al., 2005; Lu et al., 2013; van Rooij et al., 2013). These growing numbers of studies strongly suggest that the host immune response to mutant neopeptides plays the dominant role in protection of the host from tumor growth.

The opportunity to identify a vast number of putative neopeptides from individual human tumors creates a corresponding problem: how does one differentiate and identify actual tumor protective neopeptides from among the large number of putative neopeptides identified *in silico*? The problem is daunting in scale because an examination of the tumor transcriptomes and their comparison with normal exomes in the TCGA database shows that many tumors harbor hundreds of putative neopeptides. Presumably, only a small fraction of these virtual neopeptides are immunoprotective against cancer.

This question has been addressed before in viral systems. Assarsson et al. (2007) performed a systematic analysis of the putative and real epitopes of the vaccinia virus. This study revealed the magnitude of the problem: starting from all possible 9–10 amino acid peptides encoded by the vaccinia genome, only 2.5% are high-affinity binders to a given HLA allele. Of the high-affinity binders, half elicit a CD8 response. Of these, only 15% are naturally processed and presented. Finally, they observed little correlation between the dominance of an epitope with HLA-peptide affinity, HLA-peptide stability, TCR avidity, or the quantity of processed epitope. Thus, without the benefit of information from T cell responses, one would be unable to start from the vaccinia genome and identify useful epitopes.

The problem is orders of magnitude more complex for identifying useful epitopes from cancer genomes because the mammalian genome is considerably larger than that of vaccinia. Moreover, viral genomes are entirely nonself, whereas the cancer genomes are mutated-self; hence the neopeptides may be cross-reactive with self, and tolerance or suppression mechanisms are highly likely to come into play.

Here, we make the first systematic analysis of the transcriptomes and CD8 immunomes of tumors, and seek to understand the rules that govern the immunogenicity and tumor-protective ability of mutation-generated neopeptides. This effort has led to surprising observations regarding MHC I-peptide interactions that distinguish the recognition of neopeptides from that of viral epitopes, and a recognition that the proportion of putative mutational neopeptides that is translatable *in vivo* is far smaller than the corresponding proportion for viral systems.

## RESULTS

### From transcriptome to immunome

A methylcholanthrene-induced fibrosarcoma, CMS5 (BALB/*c*, *d* haplotype) was used as the primary workhorse, and the results were cross-tested to varying degrees with another chemically induced (Meth A) mouse tumor, as well as several human cancers. The CMS5 sarcoma is well characterized, as is the primary host in which this tumor first arose. It is a

progressively growing tumor, which is lethal in a syngeneic host. CTLs against a CMS5 line have led to the identification of a single immunogenic and tumor-protective neopeptide (Ikeda et al., 1997; Hanson et al., 2000).

Transcriptome sequencing was chosen over genome or exome sequencing to identify mutations specifically in the genes expressed in CMS5 and Meth A. Broadly speaking, the cDNA sequences obtained were compared with the normal mouse sequences, and single-nucleotide variants (SNVs) were identified (Table 1). This pipeline is named Epi-Seq. The SNVs were analyzed for their potential to generate MHC I-restricted epitopes of the murine H-2 K, D, or L alleles using the NetMHC algorithm. The complete list of epitopes was filtered based on default NetMHC 3.0 PWM peptide binding score thresholds for weak binders of 8.72, 8.08, and 8.19 for K<sup>d</sup>, D<sup>d</sup>, and L<sup>d</sup>, respectively. Using these thresholds, CMS5 and Meth A were observed to harbor 112 and 823 potential epitopes, respectively (see Table S1 for the total output of the pipeline for both tumors); the difference in the number of epitopes identified between these two lines is a reflection of the depth to which their transcriptomes were sequenced (Table 1). The putative neopeptides are randomly distributed over the entire genome.

We made an attempt to determine if the numbers of MHC I-restricted neopeptides in these mouse tumors are within the range expected in actual primary human tumors. An analysis of exome sequences of several human melanomas and their comparison with corresponding normal sequences (Wei et al., 2011) through our bioinformatic pipeline reveals

**Table 1.** Single nucleotide variants and predicted epitopes of tumor lines as deduced from transcriptome sequencing and bioinformatic analyses

Mouse strain	BALB/ <i>c</i>	
	Meth A	CMS5
<b>Tumor type</b>		
RNA-Seq reads (million)	105.8	23.4
Genome mapped	75%	54%
Transcriptome mapped	83%	59%
HardMerge mapped	65%	48%
After PCR amplification filter	18%	22%
HardMerge and filtered mapped bases (Gb)	1.15	0.24
High-quality heterozygous SNVs in CCDS exons <sup>a</sup>	1,528	208
Tumor specific	1,504	191
Non-synonymous	77.1%	78.5%
Missense	1,096	146
Nonsense	63	4
No-stop	1	-
NetMHC predicted epitopes <sup>b</sup>	823	112
H2 K <sup>d</sup> -restricted	203	15
H2 D <sup>d</sup> -restricted	328	58
H2 I <sup>d</sup> -restricted	292	39

<sup>a</sup>The number of mutations identified depends on the sequencing depth.

<sup>b</sup>Based on default NetMHC 3.0 PWM peptide binding score thresholds for weak binders, of 8.72, 8.08, and 8.19 for K<sup>d</sup>, D<sup>d</sup>, and L<sup>d</sup> alleles, respectively.

**Table 2.** Tumor-specific polymorphisms and epitopes for human melanomas

Sample	Somatic mutations <sup>a</sup>						NetMHC predicted epitopes <sup>b</sup>			
	Synonymous/nonsynonymous in UCSC annotation		Subset of nonsynonymous mutations within CCDS coding regions				Total	A0101	A0201	A0301
	Synonymous	Nonsynonymous	Total	Nonsynonymous	Missense	Non-sense				
01T	160	304	292	291	272	12	473	26	262	185
05T	56	115	106	106	100	6	175	12	116	47
09T	39	83	81	80	76	4	131	5	77	49
12T	427	741	706	705	656	49	1193	50	657	486
18T	91	190	179	179	168	11	286	15	144	127
22T	69	126	116	115	108	7	181	20	94	67
24T	163	397	381	379	358	21	625	26	263	336
35T	13	34	32	32	31	1	68	1	44	23
43T	68	94	91	91	86	5	175	4	90	81
51T	51	136	126	126	117	9	229	7	148	74
60T	67	129	121	120	112	8	209	14	91	104
91T	99	215	209	209	196	13	329	18	176	135
93T	54	130	125	124	116	8	184	6	105	73
96T	68	118	112	112	103	9	192	10	101	81

<sup>a</sup>Calculated from mutation data published by Wei et al. (2011), and using the epitope prediction step of the Epi-Seq pipeline to call the epitopes based on the mutation data.

<sup>b</sup>Most common HLA alleles were chosen according to the published frequencies of such alleles (Cao et al., 2001).

hundreds of putative neoepitopes per melanoma (Table 2). Similarly, the list of mutations derived from transcriptome sequencing of 14 human prostate cancers and normal tissues (Ren et al., 2012), led to identification of a median of 14 putative epitopes (range, 2–82) for the common HLA alleles (Table 3). The smaller number of neoepitopes in prostate cancers is related to the fact that they harbor relatively smaller

numbers of mutations as compared with melanomas (Vogelstein et al., 2013). The published data for the B16 melanoma line of spontaneous origin also reveal the presence of >100 MHC I-restricted neoepitopes. These measurements indicate that regardless of their murine or human origin, and regardless of etiology, tumors harbor a significant number of candidate MHC I-restricted neoepitopes.

**Table 3.** Tumor-specific polymorphisms and epitopes for human prostate cancers

Tumor sample	Somatic mutations <sup>a</sup>						NetMHC predicted epitopes			
	Synonymous/nonsynonymous in UCSC annotation		Subset of nonsynonymous mutations within CCDS coding regions				Total	A1101	A2402	A0201
	Synonymous	Nonsynonymous	Total	Heterozygous	Missense	Non-sense				
1T	7	10	10	10	9	1	9	4	0	5
2T	4	3	2	2	2	0	4	4	0	0
3T	7	4	3	3	3	0	10	5	1	4
4T	5	3	3	3	3	0	2	2	0	0
5T	4	7	7	7	7	0	18	12	2	4
6T	3	5	5	5	4	1	3	1	0	2
7T	2	3	2	2	2	0	2	1	1	0
8T	20	44	40	40	38	2	82	42	7	33
9T	13	22	18	18	16	2	30	13	0	17
10T	20	32	29	29	28	1	36	20	3	13
11T	6	13	9	9	9	0	17	8	2	7
12T	3	11	10	10	10	0	28	13	5	10
13T	2	5	5	5	5	0	9	2	3	4
14T	19	32	29	29	24	5	53	25	8	20

<sup>a</sup>Calculated from mutation data published by Ren et al. (2012), and using the epitope prediction step of the Epi-Seq pipeline to call the epitopes based on the mutation data.

**Heterogeneity of neoepitopes**

Meth A cells were cloned and 30 distinct clones were tested for four SNVs picked at random (Tnp3, Nfkb1, Prp31, and Psg17). To our surprise, all 30 clones harbored three of the four SNVs and fourth SNV was detected in 29/30 clones. We attribute this apparent lack of antigenic heterogeneity to the relatively shallow depth of sequencing. It is also possible that cancer cell lines show less antigenic heterogeneity than primary tumors. Most importantly, these results suggest it is possible to use a relatively shallow sequencing as a methodology to identify the neoepitopes that are the most broadly distributed among cancer cells.

**Immunogenicity of neoepitopes identified in silico**

To reduce the complexity of analyses, attention was directed toward the 218 K<sup>d</sup>-restricted epitopes (for Meth A and CMS5 combined) from a total list of 935 for all three alleles (Table 1). All the mutations used for immunological analyses were confirmed individually by Sanger sequencing.

The neoepitopes were ranked in descending order of their NetMHC scores for K<sup>d</sup> binding. The top 7 neoepitopes from CMS5 and the top 11 from Meth A are shown in Table 4, which also shows the NetMHC score of the WT peptide corresponding to the neoepitope. Peptides corresponding to these 18 putative neoepitopes and their WT counterparts were synthesized, and the affinity of all peptides for K<sup>d</sup> (IC<sub>50</sub> values) was determined experimentally as described in Materials and

methods. (NetMHC scores and experimentally determined IC<sub>50</sub> values were significantly correlated; r = -0.44, two-tailed Student's *t* test; P < 0.001). 10 of 18 top-ranked neoepitopes (56%) bound K<sup>d</sup> with an affinity of 500 nM or better, and 7/18 (39%) bound with an affinity of 100 nM or better.

All 18 peptides were used to immunize BALB/c mice. The draining LNs of immunized mice were harvested 1 wk after the single immunization, and the cells were stimulated in vitro for 16 h without any added peptide, the mutant peptide used for immunization, or the corresponding WT peptide. The CD8<sup>+</sup> cells were analyzed for activation (CD44<sup>+</sup>) and effector function (intracellular IFN-γ<sup>+</sup>; see Materials and methods for details of FACS analysis.) All possible patterns of immunoreactivity were observed (Fig. 1 A): no immune response (12/18), a mutant peptide-specific, i.e., tumor-specific immune response (5/18), and a cross-reactive response between the mutant and corresponding WT peptides (1/18). Altogether, 6/18 or 33% of the neoepitope candidates identified in silico actually elicited functional effector T cells in vivo. When analyzed by an IFN-γ ELISpot assay, four additional neoepitopes showed immunogenicity, bringing the total to 10/18 or ~56%.

Of the 10 peptides with a K<sup>d</sup>-binding affinity of 500 nM or better, 7 or 70% were determined to be immunogenic experimentally. Of the 7 peptides with a K<sup>d</sup>-binding affinity of 100 nM or better, 4 or 57% were immunogenic. Three

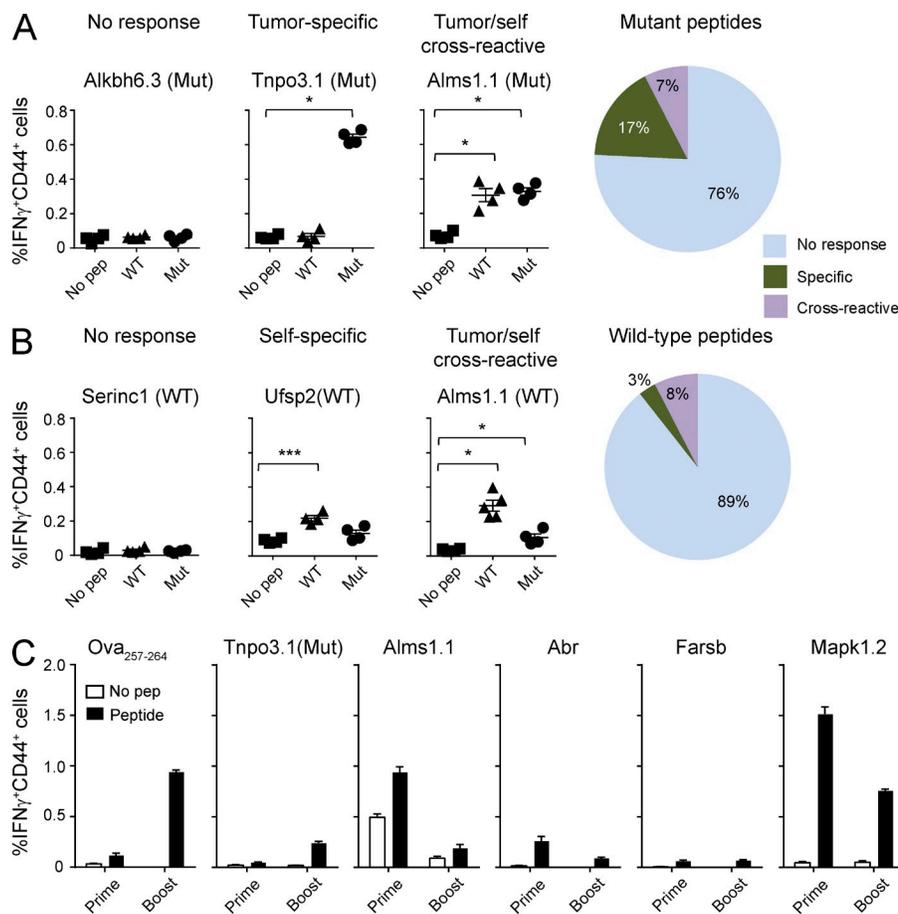
**Table 4.** CMS5 and Meth A epitopes with highest NetMHC PWM scores.

Gene	Mut/WT sequence	Mut/WT Score	Measured IC <sub>50</sub> for K <sup>d</sup> (Mut/WT)	ICS	ELISpot	Tumor rejection
CMS5 Epitopes with highest NetMHC scores						
Ssx2ip	CYAK(v/L)KEQL	14.5/14.1	26/3.2	-	-	-
Mapk1.1	(q/K)YIHSANVL	13.2/12.4	57/0.2	-	-	-
Farsb	HY(v/L)HIIESKPL	13/13.6	423/52	+	-	-
Ncoa3	(h/Q)YLQYKQEDL	11.5/11.7	2162/54074	-	-	-
Mapk1.2	(q/K)YIHSANV	11.4/10.7	2135/295	-	-	-
Mapk1.3	LYQILRGL(q/K)YI	11.3/11	110/333	-	+++++	-
Serinc1	NYLLSLVAV(m/V)L	11/10.2	2679/20861	-	+++++	-
Meth A Epitopes with the highest NetMHC scores						
Usp12	SY(l/R)VFPL	14.2/12.3	6835/1155	-	+++++	-
Tfdp1	QYSGS(w/R)VETPV	14.2/15.3	- <sup>a</sup> /603	-	-	-
Ufsp2	HYINM(i/S)LPi	14.2/14.5	0.23/+++ <sup>b</sup>	+	+++++	-
Apc	AYCETCWE(l/W)	14/8.1	23/60	+	+++++	-
Hspg2	SY(l/Q)LGSGEARL	14/14.4	2623/79	-	-	-
Ccdc85c	TYIRP(f/L)ETKV	13.5/13.1	6155/118	-	-	-
Pacs2	HYLS(s/A)ILRL	13.4/12.7	41/1269	-	-	-
Alms1.1	(l/S)YLDSKSDTTV	13.3/15.2	79/16	+	-	-
Alms1.2	YYVPLKRVV(l/S)	13.3/7.3	421/1485	-	++	-
Ckap5	K(y/D)MSMLEERI	13.2/1.7	17/7686	+	-	-
Abr	GYFVSKAKT(s/R)V	13.1/12.8	958/570	+	-	-

<sup>a</sup>IC<sub>50</sub> > 70,000 nM

<sup>b</sup>++++, IC<sub>50</sub> < 0.1 nM

Note for ELISpot results: 1-9 spots/10<sup>6</sup> CD8 = +; 10-20 spots/10<sup>6</sup> CD8 = ++; 21-50 spots/10<sup>6</sup> CD8 = +++; 51-100 spots/10<sup>6</sup> CD8 = ++++; >100 spots/10<sup>6</sup> CD8 = +++++.



**Figure 1. Immunogenicity of epitopes generated by point mutations.** Mice were immunized in the footpad with indicated mutant peptides. 1 wk later, draining LNs were harvested and their cells were stimulated overnight *in vitro* without peptides (*No pep*) or with mutated (*Mut*) peptides or their unmutated (*WT*) counterparts. Surface CD44 and intracellular IFN- $\gamma$ <sup>+</sup> cells were counted on 20,000-gated CD8<sup>+</sup> cells. (A) Representative examples of mutant peptides that elicited no response, tumor-specific (i.e., mutant peptide specific) response, or tumor/self-cross-reactive response. The right pie chart shows the total percentage of each type of T cell response elicited by mutated peptides from Meth A ( $n = 39$ ) and CMS5 ( $n = 27$ ). (B) Representative examples of unmutated counterparts of selected mutant peptides that elicited no response, unmutated peptide-specific response, or cross-reactive functional CD8 response (as in A). The right pie chart shows the percentage of type of T cell response elicited by unmutated peptides from both tumors. (C) Mice were immunized 100  $\mu$ g of the indicated peptide once (Prime), or twice (Boost) with a 29-d interval. 7 d after the last immunization, draining popliteal LNs were harvested for intracellular cytokine assay. Lymphocytes were stimulated, or without stimulation, with 10  $\mu$ g/ml cognate peptides for 16 h and stained for CD4, CD8 and CD44, followed by permeabilization and staining for IFN- $\gamma$ . Shown are the percent CD44<sup>+</sup> IFN- $\gamma$ <sup>+</sup>

cells of total CD8<sup>+</sup> cells stimulated with peptides and without peptide stimulation (*No pep*). Error bars represent SEM. \*,  $P < 0.05$ ; \*\*\*,  $P < 0.001$ . Immunogenicity of each peptide was tested in two to four mice each and the experiments were performed between four and six times. See Fig. S2 for FACS gating strategy and representative primary data.

peptides with a K<sup>d</sup>-binding affinity of 500 nM or worse were immunogenic.

### Lack of immunogenicity of WT peptides

While testing for immunogenicity of neopeptides, their WT counterparts were similarly tested (Fig. 1 B). Surprisingly, as with the mutant neopeptides (Fig. 1 A), all possible patterns of immunoreactivity were observed (Fig. 1 B): no immune response (11/18), a WT peptide-specific immune response (5/18), and a cross-reactive response between the mutant and corresponding WT peptides (2/18). Altogether, 7/18 or 39% of the WT counterparts of neopeptide candidates identified *in silico* elicited functional effector T cells *in vivo* (Alms1.1, Alms1.2, Abr, Ccdc85c, Farsb, Mapk1.2, and Ufsp2; Table 4).

This was still a surprisingly large proportion of self-reactive peptides in view of the strong role of negative selection in sculpting of the T cell repertoire. We tested the possibility that the WT peptides that are immunogenic are functionally tolerated, even though they show immunogenicity after a 16-h stimulation *in vitro*. We reasoned that if a small proportion of low affinity autoreactive T cells escaped negative selection, they may still be clonally expanded to a degree upon immunization

with a self-peptide, but that such an expansion would prove self-limiting. Hence, naive mice were immunized with the peptides, followed by a second immunization, and the responses in naive mice, once-immunized mice, and twice-immunized mice were compared. The ovalbumin-derived K<sup>b</sup>-binding epitope SIINFEKL was used as a positive control in C57BL/6 mice, and indeed the magnitude of the anti-SIINFEKL CD8 response was amplified by a second immunization (Fig. 1 C). This same phenomenon was observed with the mutant neopeptide Tnpo3 in BALB/c mice. However, second immunization with 4/4 WT peptides tested did not elicit an amplification of the response. Indeed responses detected after the second immunization were significantly diminished as compared with the response after the first immunization. By this stringent criterion, not a single WT peptide was observed to be immunogenic.

### Lack of immunoprotective activity of the strongest K<sup>d</sup>-binding neopeptides

All 18 neopeptides (Table 4) were tested for their ability to elicit tumor rejection of CMS5 or Meth A. BALB/c mice were immunized with the individual peptides and were

challenged with the appropriate tumor 1 wk after the last immunization. None of the peptides elicited tumor rejection (Fig. 2 A). One of the CMS5 neoepitopes FarsB shows significant protection from tumor growth in this panel; however, this result could not be reproduced unambiguously, leading us to assign this as a negative neoepitope with respect to protection from tumor growth. Interestingly, one of the neoepitopes identified by us Mapk1 (listed as Mapk1.1, 1.2, and 1.3; Table 4), was also identified by Ikeda et al. (1997) in the CMS5 sarcoma as a tumor rejection antigen. Immunization with it does not elicit tumor rejection in our hands, just as it did not in the original paper. The authors of the original paper noted that “IL-12 treatment was essential to show antitumor immunity in this system, because mice vaccinated with 9m-pulsed spleen cells in the absence of exogenous IL-12 showed no resistance to CMS5 challenge.” This is an unintended validation of our pipeline and also highlights the fact that we have used a very stringent tumor rejection assay in our analyses.

### Differential agretopicity

From the aforementioned results, it is evident that the NetMHC score is not a valuable predictor of immunogenicity or tumor rejection. A close examination of the data in Table 4 suggests an underlying possibility: for each neoepitope, both the neoepitope and its WT counterpart have similar NetMHC scores characteristic of high-affinity peptide binding. Moreover, examining the experimental  $IC_{50}$  values, 4/7 WT peptides of CMS5 and 7/11 WT peptides of Meth A have stronger affinity for  $K^d$  than the mutant peptides. Thus, unless the mutations alter TCR contacts or the structural properties of the peptides in the  $K^d$  binding groove, T cells potentially reactive to the neoepitopes may have been centrally deleted or peripherally tolerized.

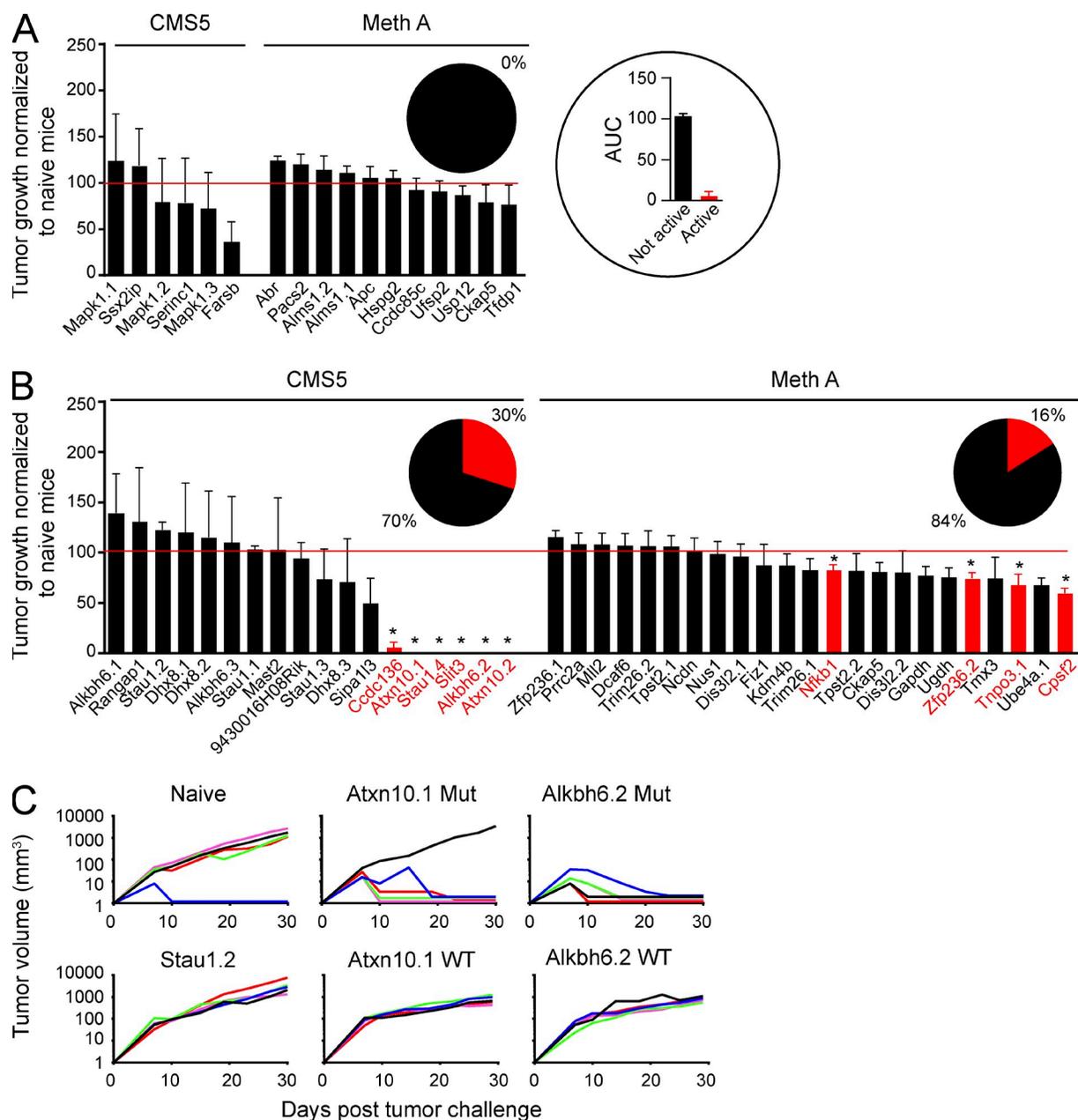
We therefore created a new algorithm wherein the NetMHC scores of the unmutated counterparts of the predicted mutated epitopes were taken into consideration by subtracting them from the corresponding NetMHC scores of the mutated epitopes. We refer to this new property of an epitope as the differential agretopicity index (DAI; agretopicity is the ability of a peptide to act as an agretope), and we expect it to reflect the degree to which the peptide-binding determinants of the neoepitopes differ from those of their WT counterparts. In the search for the rules for immunogenicity of viral or other clearly nonself-epitopes, such a parameter is not necessary, nor possible, and has therefore never been sought. The putative epitopes were ranked on basis of the DAI (Table 5). A review of this DAI-ranked list for both tumors shows curiously that all the neoepitopes in this new ranking are mutated at one of the two primary anchor residues at position 2 or the C terminus (we did not identify any neoepitope with changes at both anchor residues.) Consistent with the  $K^d$  preferences gleaned from structural analyses (Mitaksov and Fremont, 2006), all of these mutations involve aspartic acid to tyrosine at position 2 or proline/arginine

to leucine at the C terminus. Although the DAI was not crafted with this intent, it is a perfectly reasonable outcome in hindsight.

The top DAI-ranking 20 epitopes of CMS5 and 28 epitopes of Meth A were tested in tumor rejection assays. The results (Fig. 2 B) show that 6/20 or 30% CMS5 epitopes and 4/28 or 14% Meth A epitopes showed statistically significant tumor protective immunogenicity. The six CMS5 neoepitopes are particularly impressive in that they elicited near complete or complete protection from a lethal tumor challenge. Fig. 2 C shows representative tumor rejection curves of the two protective and one unprotective CMS5 epitope from Table 5. Corresponding WT peptides did not mediate any tumor rejection. Detailed data on tumor rejection elicited by a Meth A neoepitope Tnp3 are shown in Fig. 5. Statistical comparison of the NetMHC alone versus the DAI algorithms in predicting antitumor protective immunity, using one-tailed Fisher's exact test, shows the DAI to be far superior (0/18 vs. 10/48, for NetMHC and DAI, respectively, one-sided Fisher's exact test,  $P = 0.031$ ).

Although the DAI algorithm yielded a far richer harvest of tumor-protective epitopes than the reliance on the highest NetMHC or MHC-binding scores, most epitopes identified by DAI still fail to elicit tumor protection. Further, the DAI-ranked neoepitopes (Table 5) that elicit protection from tumor growth are not necessarily the highest ranking in DAI. Clearly, other properties of the putative epitopes (see Discussion) contribute to the tumor rejection activity of individual neoepitopes. Regardless of its imperfection, the DAI is a statistically significant and novel predictor of tumor-protective immunogenicity, and more importantly, permits a dissection of the other potential criteria for antitumor immunogenicity *in vivo*.

The DAI algorithm also uncovers a new paradox of fundamental significance. All the six neoepitopes that elicit protection against CMS5 have NetMHC scores between 6.8 and 1.2, which are well below 8.72, the PWM peptide binding score threshold for weak binders for  $K^d$  (NetMHC3.0). Consistent with that observation, their measured  $IC_{50}$  values for binding to  $K^d$  are  $>70,000$  nM (for all except Slit3, for which is  $IC_{50}$  is  $\sim 50,000$  nM). This observation is surprising because epitopes are typically considered good MHC I-binders if they have an  $IC_{50}$  value of  $<100$  nM, or at least  $<500$  nM. The  $IC_{50}$  values for the CMS5 neoepitopes are so high (i.e., their binding to  $K^d$  is so poor) that these neoepitopes would never be considered suitable candidates for being epitopes based simply on their  $K^d$ -binding characteristics. For Meth A as well, all the four neoepitopes that elicit tumor immunity have a NetMHC below the 8.72 threshold for weak binders for  $K^d$ ; the measured  $K^d$ -binding affinities of only two of the four Meth A neoepitopes are  $<100$  nM. To explore the possibility that these neoepitopes may be binding to another allele,  $D^d$  or  $L^d$ , the six tumor-protection eliciting neoepitopes for CMS5 were tested for binding to these two alleles by direct peptide-binding studies; none of the peptides showed significant binding (unpublished



**Figure 2. Landscape of protective tumor immunity elicited by tumor-specific peptides.** (A) Tumor-protective activity of the mutated epitopes with top NetMHC or DAI scores for CMS5 and Meth A. (A and B), Mice were immunized with peptides that had the highest NetMHC scores (A) or top DAI scores (B) and challenged with live tumor cells, and tumor growth monitored. Area under the curve (AUC; Duan et al., 2012) for each individual tumor growth curve was calculated and normalized by setting the naive group to a value of 100, shown by a horizontal red line. Bars corresponding to peptides that show statistically significant tumor-protective immunogenicity are filled in red, and indicated by an asterisk ( $P = 0.015\text{--}0.03$ ). The peptides are arranged in order of increasing antitumor activity. The pie charts show the percentage of neopeptides tested that did not (black) and did (red) elicit protection from tumor challenge. (C) Examples of tumor growth curves in untreated mice (naive) and mice immunized with indicated mutant peptides from CMS5 or their WT counterparts. Stau1.2 is a representative neopeptide that does not elicit protective immunity, whereas Atxn10.1 and Alkbh6.2 are representative neopeptides that do. Each line shows the kinetics of tumor growth in a single mouse. The experiments were performed three times.

data). Because the requirement for MHC I binding to be  $<500$  nM is so well established (Assarsson et al., 2007), the possibility that the peptides identified as being potent in eliciting protection from tumor challenge (Fig. 2, B and C) may do so through nonimmunological means was investigated.

#### CD8 dependence of tumor protection

Mice immunized with five of the six active neopeptides of CMS5 (Alkbh6.2, Slit3, Atxn10.1, Atxn10.2, and Ccdc136) were tested for immunogenicity; as shown in Table 5, Alkbh6.2, Slit3, and Atxn10.1 elicited modest CD8 T cell response that

**Table 5.** CMS5 and Meth A epitopes with highest DAI scores

Gene	Mut/WT sequence	Mut/WT Score	DAI	Measured IC <sub>50</sub> for K <sup>d</sup> (Mut/WT)	ICS	ELISpot	Tumor rejection
CMS5 Epitopes							
Dhx8.1	P(y/D)LTQYAIIML	9.3/-2.3	11.6	2192/1653	-	+++++	-
Alkbh6.1	D(y/D)VPMEQP	4.3/-7.3	11.6	60858/-	-	-	-
Dhx8.2	P(y/D)LTQYAI	9.9/-1.6	11.5	- <sup>a</sup> /-	-	+++++	-
Dhx8.3	P(y/D)LTQYAI	9.3/-2.2	11.5	244/1418	+	+++++	-
Alkbh6.2	D(y/D)VPMEQPR	6.8/-4.7	11.5	-/51229	-	++	+
Dhx8.4	P(y/D)LTQYAIIM	6.6/-4.9	11.5	6571/6256	-	+++	-
Alkbh6.3	D(y/D)VPMEQPRP	4.8/-6.7	11.5	-/23570	-	-	-
Alkbh6.4	D(y/D)VPMEQPRPP	4.7/-6.8	11.5	47053/2957	-	-	-
Rangap1	SEDKIKAI(I/P)	1.4/-5.4	6.8	-/5108	-	+++++	-
Stau1.1	LKSEEKT(I/P)	0.6/-6.2	6.8	-/-	-	-	-
Stau1.2	KPALKSEEKT(I/P)	0.1/-6.7	6.8	-/-	-	-	-
Stau1.3	PALKSEEKT(I/P)	1.3/-5.4	6.7	69546/-	-	-	-
Stau1.4	ALKSEEKT(I/P)	1.2/-5.5	6.7	-/-	-	-	+
9430016	SWSSRRSLLG(I/R)	5.9/-0.6	6.5	-/-	-	+++++	-
H08Rik							
Slit3	GFHGCIHEV(I/R)	4.7/-1.8	6.5	51640/-	+	+++	+
Atxn10.1	QVFPGLME(I/R)	3.4/-3.1	6.5	-/7054	+	+++	+
Sipa1l3	TTTPGGRPPY(I/R)	2.7/-3.8	6.5	-/-	-	++++	-
Atxn10.2	VFPGLME(I/R)	2.5/-4	6.5	-/1107	-	-	+
Ccdc136	ELQGLLEDE(I/R)	2.4/-4.1	6.5	-/4537	-	-	+
Mast2	KLQRQYRSR(I/R)	2.2/-4.3	6.5	10107/8511	-	-	-
Meth A Epitopes							
Tnp03.1	(sy/LD)MLQALCI	8.2/-5.2	13.4	82/146	+	+++++	+
Tnp03.2	(sy/LD)MLQALCIPT	7.1/-6.3	13.4	9964/85	+	+++	-
Tnp03.3	(sy/LD)MLQALCIP	3/-10.4	13.4	67/111	+	+++++	-
Tnp03.4	(sy/LD)MLQALC	5.9/-7.4	13.3	14927/89	-	-	-
Trim26.1	A(y/D)ILAALTKL	12.8/1.2	11.6	622/1.1	-	+++++	-
Nus1	P(y/D)LVLKFGPV	10.5/-1.1	11.6	2359/1.9	-	-	-
Tpst2.1	L(y/D)EAGVTDEV	10.3/-1.3	11.6	60/-	-	+++	-
Fiz1	H(y/D)LQGSNA	10.3/-1.3	11.6	2473/-	-	-	-
Kdm4b	L(y/D)HTRPTAL	10/-1.6	11.6	264/-	-	-	-
Dis3l2.1	I(y/D)GVVARNRAL	9.3/-2.3	11.6	143/-	+	+	-
Ube4a.1	A(y/D)AKQFAAI	9.3/-2.3	11.6	12/-	+	+	-
Ncdn	S(y/D)CEPALNQA	8.9/-2.7	11.6	664/-	+	-	-
Gapdh	V(y/D)LTCRLEKPA	8.9/-2.7	11.6	1150/-	-	-	-
Ckap5	K(y/D)MSMLEER	8.1/-3.5	11.6	58/-	-	+++	-
Prrc2a	P(y/D)KRLKAEP	7.9/-3.7	11.6	1450/261	-	-	-
Tmx3	D(y/D)HIEFAHRV	7.3/-4.3	11.6	7941/351	-	-	-
Nfkb1	G(y/D)SVLHLAI	6.9/-4.6	11.5	0.26/1615	+	-	+
Dis3l2.2	I(y/D)GVVARNRA	6.9/-4.7	11.6	1342/262	-	-	-
Ugdh	L(y/D)YERIHKKML	6.4/-5.2	11.6	600/12	-	-	-
Mil2	S(y/D)RPLPSSRKK	5.9/-5.7	11.6	6673/17	-	+	-
Galnt1	L(y/D)VSKLNGP	5.6/-6	11.6	27086/-	-	-	-
Tpst2.2	L(y/D)EAGVTDE	5.5/-6.1	11.6	-/-	-	+++	-
Cpsf2	L(y/D)DVDAAF	5.3/-6.3	11.6	1410/-	-	+++	+
Zfp236.1	E(y/D)LDLQQTQ	5.3/-6.3	11.6	1641/-	-	+++	+

<sup>a</sup>IC<sub>50</sub> > 70,000 nM

Note for ELISpot results: 1-9 spots/10<sup>6</sup> CD8 = +; 10-20 spots/10<sup>6</sup> CD8 = ++; 21-50 spots/10<sup>6</sup> CD8 = +++; 51-100 spots/10<sup>6</sup> CD8 = ++++; >100 spots/10<sup>6</sup> CD8 = +++++.

Table 5. (Continued)

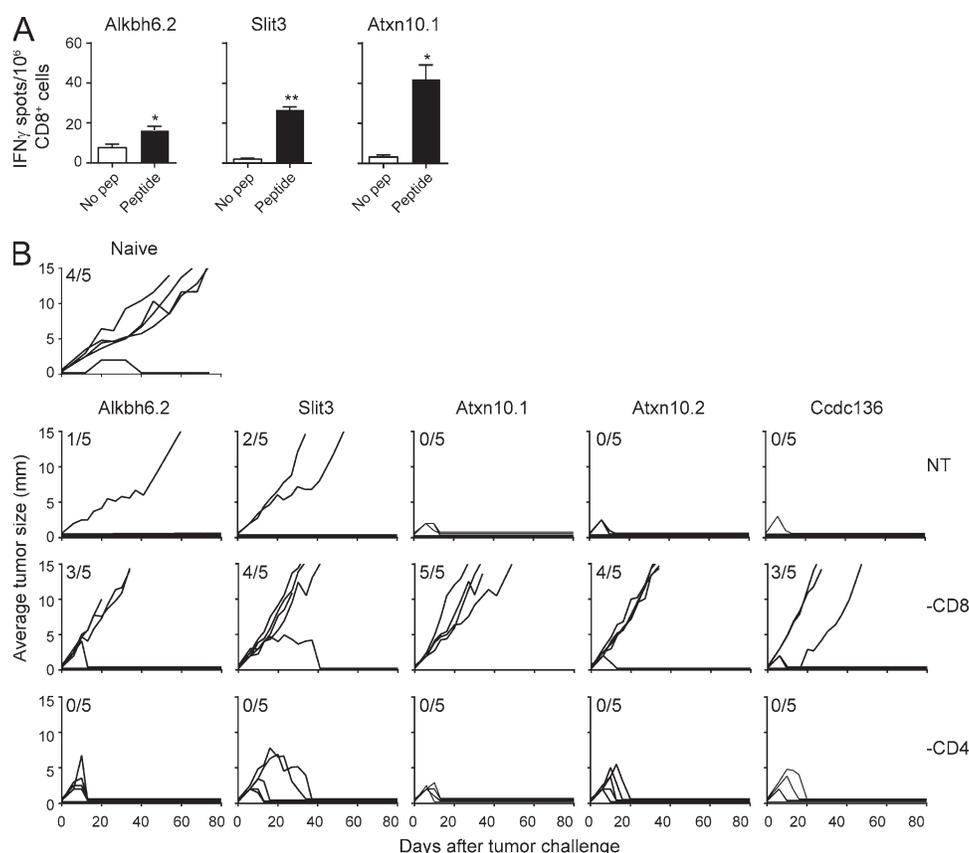
Gene	Mut/WT sequence	Mut/WT Score	DAI	Measured IC <sub>50</sub> for K <sup>d</sup> (Mut/WT)	ICS	ELISpot	Tumor rejection
Trim26.2	A(y/D)ILAALTKLQ	4.9/-6.7	11.6	17599/-	-	-	-
Zfp236.2	E(y/D)LDLQTQG	4.9/-6.7	11.6	10028/-	-	++++	-
Ube4a.2	A(y/D)AKQFAA	4.7/-6.9	11.6	28583/-	-	-	-
Dcaf6	A(y/D)RLEGDRS	3.7/-7.9	11.6	-/-	-	-	-

<sup>a</sup>IC<sub>50</sub> > 70,000 nM

Note for ELISpot results: 1–9 spots/10<sup>6</sup> CD8 = +; 10–20 spots/10<sup>6</sup> CD8 = ++; 21–50 spots/10<sup>6</sup> CD8 = +++; 51–100 spots/10<sup>6</sup> CD8 = ++++; >100 spots/10<sup>6</sup> CD8 = +++++.

was undetectable by the intracellular cytokine assay, and detectable only by the ELISpot assay (Fig. 3 A), whereas Atxn10.2 and Ccdc136 did not elicit a detectable response at all. Immunized mice were depleted of CD8 or CD4 cells in the effector phase only (i.e., after immunization but before tumor challenge) and were challenged with CMS5 cells as in Fig. 2. The results (Fig. 3 B) show that, compared with naive mice, each mutant peptide elicited potent tumor rejection.

Depletion of CD8 cells completely abrogated the tumor-rejecting ability of each peptide. Depletion of CD4 cells had no such effect. These data show that the CMS5 neopeptides that elicit tumor rejection do so through elicitation of specific CD8<sup>+</sup> T cells, and not through nonimmunological means. However, the data do not imply that CD4<sup>+</sup> T cells do not have a role in tumor rejection; an examination of the kinetics of tumor rejection shows that in the absence of CD4<sup>+</sup> cells,



**Figure 3. Protective tumor immunity elicited by tumor-specific peptides is CD8-dependent.** (A) Mice were immunized with mutant peptides Alkbh6.2, Slit3, and Atxn10.1, Atxn10.2, and Ccdc136. Draining LNs harvested 7 d after immunization were not stimulated or stimulated in vitro with the cognate peptides for 20 h, and were analyzed by ELISpot. Data for mice immunized with Alkbh6.2, Slit3, and Atxn10.1 peptides are shown. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ . (B) Naive mice or mice immunized twice with indicated peptides (Alkbh6.2, Slit3, Atxn10.1, Atxn10.2, and Ccdc136) were challenged intradermally with 300,000 live CMS5 tumor cells, and tumor growth was monitored. The numbers in the top left corner of each set of tumor growth curves denote the number of mice with growing tumors/total number of mice. Immunized mice were not depleted (NT), depleted of CD8 or CD4 cells, as indicated. Each line shows the kinetics of tumor growth in a single mouse. The experiments were performed three times.

tumors actually do begin to grow in almost all the immunized mice, but begin to regress by days 7–10.

### Conformational stability as an indicator of immunogenicity

The CD8-dependence of neoepitopes that have low predicted and measured affinity for their restricting allele led us to seek other determinants of immunogenicity. Although the DAI algorithm selects for mutations at primary anchor positions that improve peptide–MHC binding affinity, unless the mutations alter the structural properties of the peptide in the binding groove as discussed above, potentially responding T cells may be centrally or peripherally tolerized. Studies have shown that anchor modification can have a range of effects on MHC-bound peptides, in some cases having no apparent influence (Borbulevych et al., 2005, 2007), and in others leading to alterations in structural and motional properties (Sharma et al., 2001; Borbulevych et al., 2007; Insaïdo et al., 2011).

To examine the consequences of anchor modification, we used computational modeling to examine the structural properties of pairs of mutant and WT peptides bound to K<sup>d</sup>. We adapted an approach recently used to model the structures of peptides bound to HLA-A\*0201 (Park et al., 2013), using the crystallographic structure of a viral peptide–H-2K<sup>d</sup> complex as a template (Mitaksov and Fremont, 2006). As described in Materials and methods, we used a workflow that included homology modeling, simulated annealing, and molecular dynamics simulations to predict structural properties. Because the conformations of peptide backbones in MHC I proteins vary considerably with peptide length (Borbulevych et al., 2007), we restricted the modeling to pairs of nonamers in Table 5, matching the length of the peptide in the template structure. This necessary restriction leads to noninclusion in our modeling analyses of some of the more immunogenic and protective neoepitopes in Table 5, simply because they are 8- or 10-mers, and not 9-mers. To ascertain how well the modeling procedure was transferable from HLA-A\*0201 to H-2K<sup>d</sup>, we applied the modeling procedure to the complex of the immunodominant and highly immunogenic HBV core peptide with K<sup>d</sup>, a complex for which the crystallographic structure is known (Zhou et al., 2004).

We began by looking for the presence of structural differences between the mutant and WT peptides. In each instance, there were differences between the predicted conformations of the mutant and WT peptides bound to H-2K<sup>d</sup>. In all cases, these differences propagated away from the site of the mutations (Fig. 4 A). These observations were not surprising, as the mutations were all quite drastic, and even conservative mutations at class I MHC-presented peptide anchor residues can influence downstream conformation (Sharma et al., 2001; Borbulevych et al., 2007).

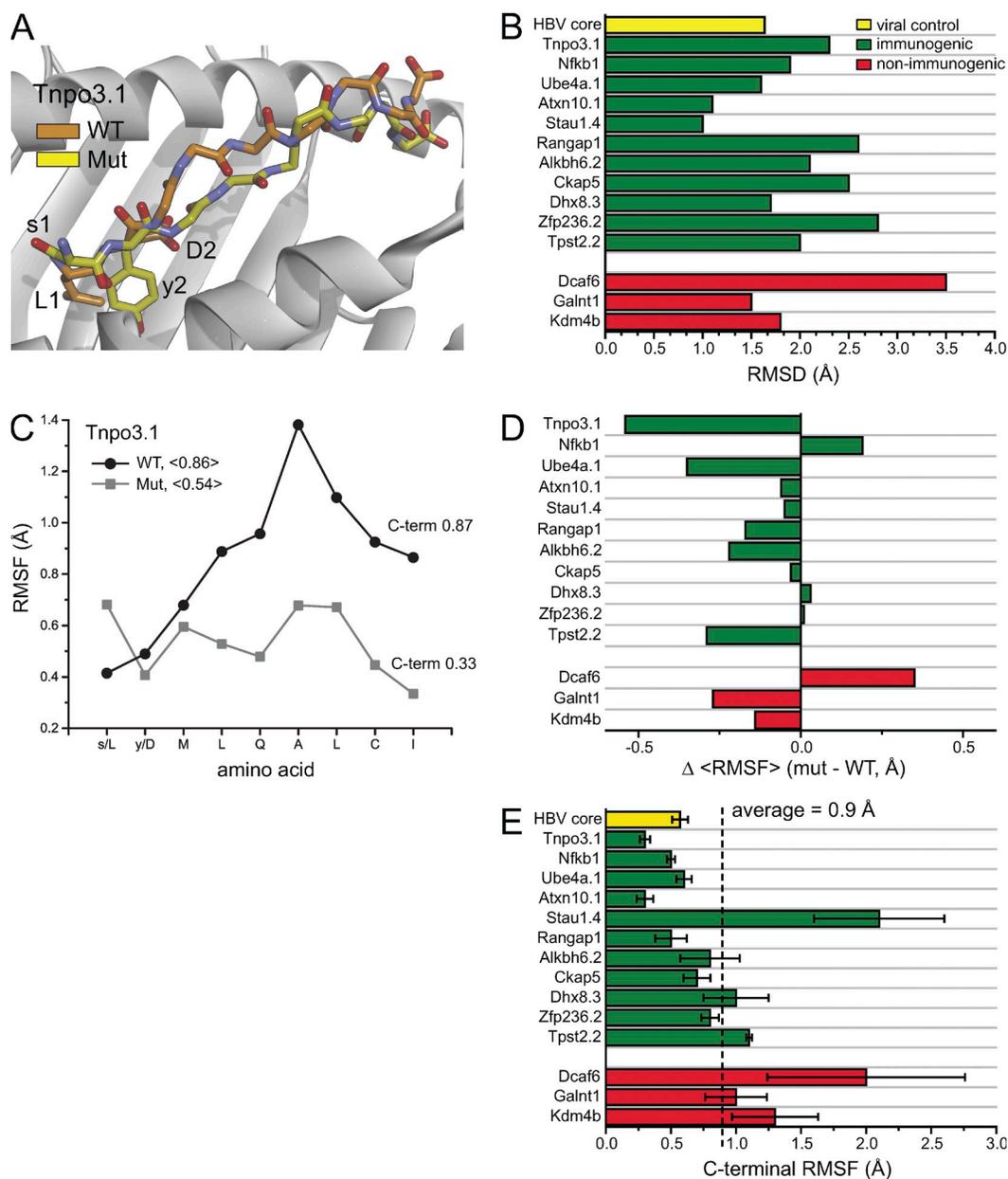
However, although T cell receptors can be exquisitely sensitive to changes in peptide conformation (Ding et al., 1999), peptide immunogenicity (defined as having positive ELISpot or ICS results) did not correlate with the magnitude of structural differences, expressed as root mean square

deviations (RMSD) in Ångstroms when all common atoms of each peptide pair were superimposed (Fig. 4 B). Examining the modeling data more closely, however, revealed a correlation, albeit imperfect, with conformational stability. In 10 out of 14 cases, the mutant peptides were more rigid, sampling fewer conformations during the molecular dynamics phase of the modeling, expressed as the average root mean square fluctuations (RMSF) of all nine  $\alpha$  carbons of each peptide (Fig. 4, C–D; and Fig. S1). Of the 10 instances in which the mutant peptides were more rigid, eight of these had positive ICS or ELISpot results. Of the four instances in which the WT peptides were more rigid, one did not yield any positive immunological outcomes. Altogether, greater structural stability was a predictor of immunogenicity in 9 out of 14, or 65% of the nonameric, high DAI-ranking neoepitopes.

In some instances though, the differences in the fluctuations between the mutant and WT peptides were very small (for example, Ckap5, Dhx8.3, and Zfp236.2 have differences  $\leq 0.03$  Å). This led us to question the fidelity of the mean RMSF as a predictive tool. Examining the data for all peptides more closely, we observed a tendency for high structural instability in the peptide C terminus, particularly in the WT peptides (Fig. 4 C and Fig. S1). Previous studies have shown that weak binding peptides can detach and dissociate from MHC I proteins at the C terminus (Winkler et al., 2007; Narzi et al., 2012), suggesting that in these cases the high structural instability reflects at least partial peptide dissociation. However, some of the mutant peptides also had high instability at the C-terminus. Most notably, all three mutant peptides which failed to elicit an immunological response had high C-terminal instability. To quantitate this, we calculated the mean C-terminal RMSF for all mutant peptides. The C-terminal RMSF of the nonimmunogenic peptides were all above this mean value (Fig. 4 E). Only three of the 14 immunogenic peptides had C-terminal RMSF values above the mean (Stau1.4, Dhx8.3, and Tpst2.2). Thus, the presence or absence of C-terminal instability was a predictor of immunological outcome in 11 out of 14, or 79% of the nonameric high DAI ranking neoepitopes. (We note that the HBV control peptide was relatively stable in the K<sup>d</sup> binding groove, with a mean C $\alpha$  RMSD from the starting coordinates of 1.7 Å and an average C-terminal RMSF of 0.57 Å, as would be predicted from an immunogenic viral peptide).

### Natural presentation of a neoepitope

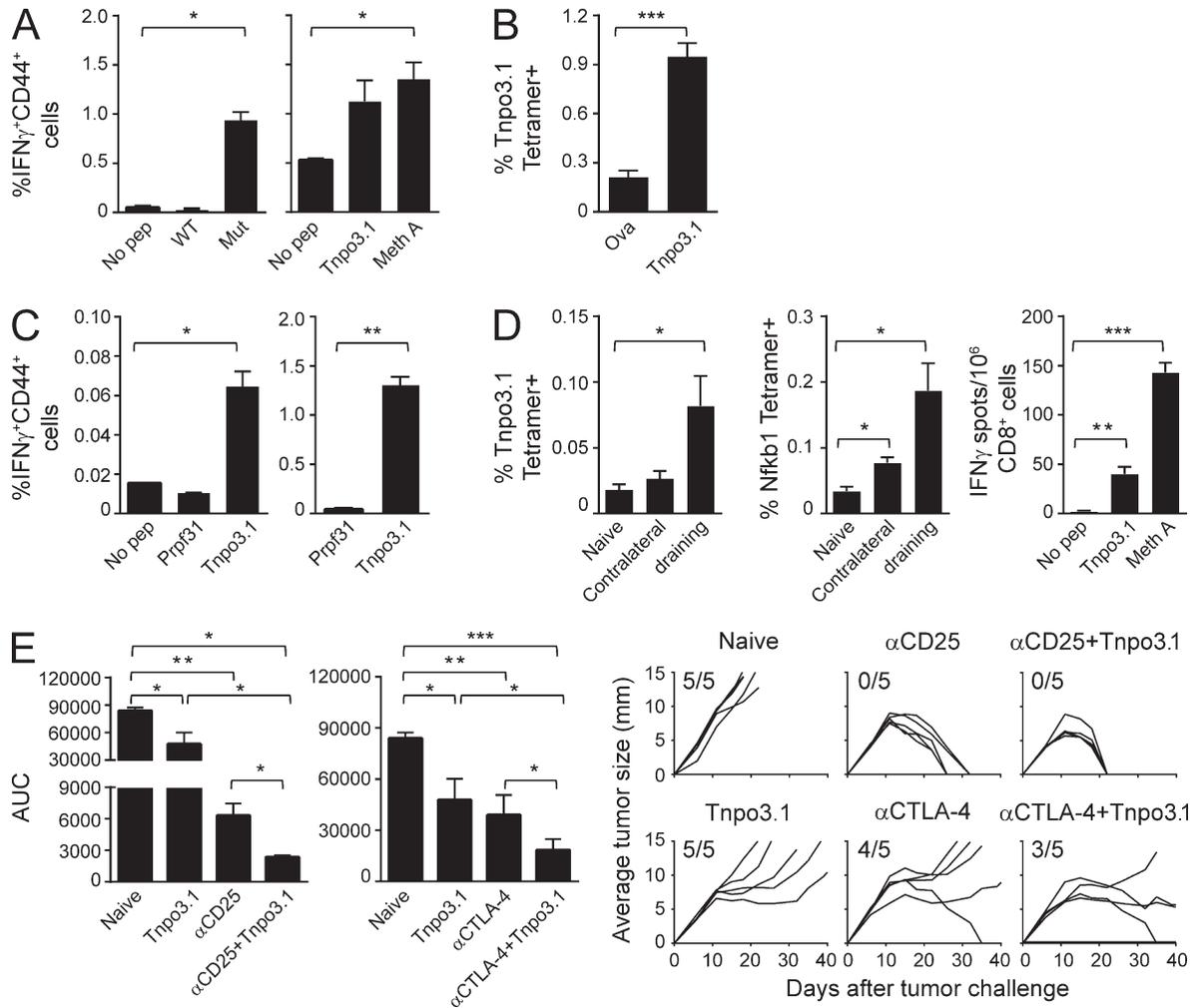
The tumor-protective immunogenicity of the epitope syMLQALCI (WT LDMLQALCI), the mutated Transportin 3 (Tnpo3)-derived epitope, the highest ranking (by DAI) epitope of Meth A (Table 5), was investigated in more detail. Tnpo3 is a nuclear import receptor and is not a driver protein for any tumor type reported thus far (Brass et al., 2008). The mutant Tnpo3 epitope was shown to elicit strictly tumor-specific CD8<sup>+</sup> immune response, as seen by the ability of mutant Tnpo3-immunized mice to show strong tumor-specific CD8<sup>+</sup>, CD44<sup>+</sup>, IFN- $\gamma$ <sup>+</sup> response to the mutant but not the WT peptide ex vivo or after stimulation in vitro (Fig. 5 A).



**Figure 4. Structural stability as a correlate with immunogenicity.** (A) Mutations within neoepitopes lead to structural alterations across the peptide backbone, as illustrated with structural snapshots from the simulations of the mutant and WT Tnp3.1 epitope bound to H-2K<sup>d</sup>. (B) Summary of structural differences for highly DAI ranked nonamers. Differences were quantified by superimposing average peptide conformations from the molecular dynamics simulations and computing RMSDs for all common atoms. Green and red bars indicate epitopes that led to either positive or negative immunological responses, respectively. The yellow bar shows the results for control calculations for an immunogenic HBV core epitope. (alue for the HBV epitope is the average C $\alpha$  RMSD relative to the starting coordinates). (C) The numbers in the legend give the mean RMSF for the all amino acids of each peptide; those at the right give the value for only the C-terminal  $\alpha$  carbon. Mutated amino acids are indicated by lower case in the x-axis. Results for all nonamers are in Fig. S1. (D) Effects of mutations on the conformational stability of all nonamers, calculated as the difference between the mean RMSF of the mutant and the WT peptide. (E) Fluctuations at the peptide C-terminal ends and immunogenicity. The dashed vertical line shows the mean value for all mutant nonamers. The yellow bar shows the C-terminal stability of the HBV core epitope control. Error bars represent SEM.

A Tnp3-specific immune response was also detectable ex vivo upon staining of cells with a Tnp3-specific tetramer (Fig. 5 B). Conversely, CD8<sup>+</sup> CD44<sup>+</sup> IFN- $\gamma$ <sup>+</sup> cells isolated from mice immunized with Meth A cells recognize mutant Tnp3-pulsed cells but not cells pulsed with an irrelevant K<sup>d</sup>-binding peptide

Prpf31 ex vivo, as well as after stimulation in vitro (Fig. 5 C). Although <0.1% Tnp3-specific CD8<sup>+</sup> T cells seen ex vivo is a truly small response, we consider it real because ex vivo responses are bound to be weak, the response is statistically significant, and the response amplifies very significantly (>1.2%),



**Figure 5. Antigen presentation of neopeptide Tnp3 and immune response and tumor protection elicited by it.** (A) Mice were immunized with mutant Tnp3 peptide. Draining LNs, harvested 1 wk after immunization, were briefly stimulated *ex vivo* without (No pep) or with WT or mutant Tnp3 peptides (left), or with a weekly *in vitro* stimulation with 1  $\mu$ M mutant Tnp3 peptide (right). After 5 d, cells were tested for the responsiveness to mutant Tnp3-pulsed cells (Tnp3) or Meth A cells (Meth A) by ELISpot. IFN- $\gamma$ <sup>+</sup> CD44<sup>+</sup> CD8<sup>+</sup> T cells were counted. (B) Mice were immunized twice with ovalbumin peptide (SIINFEKL) or Tnp3 mutant peptide. 6 d after the second immunization, splenocytes from both groups were stained with K<sup>d</sup>/SYMLQALCI tetramer. Tetramer positive cells were counted in CD8<sup>+</sup> gate. (C) Mice were immunized with irradiated Meth A cells. (left) 6 d later, inguinal LN cells were stimulated overnight without peptide, irrelevant Prpf31 peptide or Tnp3 peptide. % activated effector CD8<sup>+</sup> cells is shown, as assessed by flow cytometry. (right) Splenocytes were stimulated *in vitro* in multiple rounds with 1  $\mu$ M of indicated peptides for a total of 19 d. Irrelevant peptide from Prpf31 was used as a control. 5 d after stimulation, cells were tested for the responsiveness to indicated peptides by flow cytometry. Typically, for each sample, 150,000 lymphocytes, or at least 19,000 CD8<sup>+</sup>CD4<sup>-</sup> cells, were acquired. See in Fig. S3 for FACS gating strategy and representative primary data. (D) Mice were injected with 200,000 Meth A cells on the right flank. 21 d later, tumor-draining LNs and contralateral LNs were harvested and stained with anti-CD8 antibody and Tnp3 and NfkB1 tetramers K<sup>d</sup>/SYMLQALCI and K<sup>d</sup>/GYSVLHLAI, respectively (left and middle). Splenocytes were used to purify CD8<sup>+</sup> cells to assess the responsiveness to mutant Tnp3-pulsed cells (Tnp3) or Meth A cells (Meth A) by ELISpot assay with no peptide (No pep) stimulation as negative control (right). (E) Naive mice or Tnp3 mutant peptide-immunized mice were challenged with Meth A cells and treated with anti-CD25 antibody or anti-CTLA-4 antibody as indicated. AUC for each group is plotted (Duan et al., 2012), and complete tumor growth curves for all the mice in all groups are shown. Between four and six mice per group were used in each experiment, and each experiment was repeated between three and five times. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ .

while maintaining specificity, after weekly *in vitro* stimulation with Tnp3 peptide for 19 d (Fig. 5 C, right). Interestingly, Meth A tumor-bearing mice (day 21 after inoculation) harbor a low frequency of T cells recognizing two K<sup>d</sup>-binding peptides (Tnp3 and NfkB1, measured by tetramer staining) in the tumor-draining LNs (Fig. 5 D). These observations

confirm that the mutant Tnp3 peptide is naturally presented by Meth A cells and also that immune response to it is elicited upon immunization by whole tumor cells, as well as in tumor-bearing mice. Attempts to identify this mutant peptide by mass spectroscopic analysis of MHC I-eluted peptides from Meth A were unsuccessful, presumably because of

the higher sensitivity of the T cell assays as compared with mass spectroscopy. The structural modeling predicts that the mutant Tnp3 peptide is substantially more stable than the WT peptide across the center of the peptide (Fig. 4, A and C).

### Enhancement of tumor-protectivity of neopeptides by immune modulators

Combination of immunization with mutant neopeptides was tested using the Meth A neopeptide Tnp3. This neopeptide is only modestly tumor protective in monotherapy, thus allowing more dynamic range for testing of an enhanced effect by combination therapy. Combination of immunization with mutant Tnp3 with antagonistic antibodies to CD25, which has been shown to target regulatory T cells, showed synergy; the anti-CD25 alone showed complete regression in all mice ( $P = 0.008$ ) and Tnp3 alone also elicited significant protection ( $P = 0.03$ ). The combination showed more significant protection than either agent alone ( $P = 0.016$ ; Fig. 5 E, left); although tumors eventually regressed in all mice in both groups, the kinetics of tumor regression was significantly steeper in the combination group. A similar result was obtained with anti-CTLA4 antibody, which releases T cells from checkpoint blockade (Egen et al., 2002; Callahan et al., 2010). Each agent alone elicited statistically significant protection and the combination was significantly more effective than Tnp3 alone ( $P = 0.04$ ) or anti-CTLA4 antibody alone ( $P = 0.04$ ; Fig. 5 E, right). Only a single tumor regressed in the anti-CTLA4 antibody group, and no tumors regressed in the Tnp3 alone group (although the tumor growth was very significantly retarded); the combination group showed complete regression of two tumors and a sustained trend toward regression in two additional tumors.

### DISCUSSION

Our results uncover a trove of truly tumor-specific antigenic epitopes. Using novel tools reported in this study, we are able to identify with high accuracy the small number of neopeptides (among the vast numbers of potential neopeptides) that truly elicit immunological protection against tumor growth. The application of our method is described here for two independent tumors. In actuality, the pipeline, including the DAI algorithm, was first derived empirically on the data from the Meth A tumor, and was then tested on CMS5. Clearly, the antitumor activity predicted from the DAI algorithm is significantly stronger in CMS5 than in Meth A; this variation is most likely a reflection of the immuno-suppressive mechanisms unique to the Meth A tumor (Levey et al., 2001), and thus unrelated to the merits of the DAI algorithm per se. The DAI algorithm has since been tested in yet another mouse tumor, the B16 melanoma, and data on T cell responses in this line as well, are consistent with significant superiority of DAI over NetMHC alone (unpublished data). Although the present study is focused on identification of MHC I-restricted epitopes of CD8 T cells, the analysis can also be extended to MHC II-restricted epitopes of CD4 T cells.

Although T cells play an unambiguously central role in cancer immunity, they have been poor probes for identification of immunoprotective epitopes thus far. Extensive and laborious analyses of T cell-defined tumor-specific antigens of Meth A and CMS5 sarcomas over the years managed to yield a total of five epitopes (Buckwalter and Srivastava, 2008), none of which elicit particularly robust tumor rejection; in contrast, this single study has uncovered nearly a dozen potent tumor-protective epitopes of these two tumors. It is instructive to ponder the reason for this discrepancy. The use of T cells as probes inherently requires generation of T cell lines or clones, which is itself a highly selective process. It would appear that the diversity of effector T cells in vivo is not readily captured by the T cell lines or clones generated in vitro, leading to a distorted, and sparse, view of the T cell immunomes of tumors. The genomics-driven analysis of the immunome described in this study cuts through the bias in selection of T cells, and thus illuminates the entire field of neopeptides.

We introduce here the DAI score (the numerical difference between the NetMHC scores of the mutated epitope and its unmutated counterpart), which allows significant enrichment for the extremely small number of truly immunoprotective neopeptides from among the hundreds of putative neopeptides identified by the NetMHC algorithm. The demonstrated utility of the DAI score underscores the validity of its premise: a tumor-protective immune response requires neopeptides that differ from their WT counterparts, and the DAI score is a means to quantify and rank such differences. Understandably, because our existing ideas about immunogenicity are derived entirely from the study of viral and model antigens, which have no self-counterparts, there was no necessity to devise a DAI for their studies. As follows from the design of the NetMHC algorithm, amino acid substitutions at primary anchor residues make for the biggest contributions to the DAI. Indeed, every neopeptide with a high DAI ranking replaces aspartic acid with tyrosine at position 2 or proline/arginine at the C terminus with leucine. From structural considerations, these substitutions would be expected to significantly impact peptide binding, as tyrosine at P2 and leucine at the peptide C terminus are the most optimal  $K^d$  anchor residues (indeed, aspartic acid at P2 or arginine at the C terminus would be expected to be considerably unfavorable due to substantial charge repulsion; Mitaksov and Fremont, 2006). Peptide conformational stability, expressed as the fluctuations observed during molecular dynamics simulations, is another tool that suggests a novel correlate with immunogenicity. The majority of the neopeptides with high DAI rankings are predicted to interact with  $K^d$  in a more stable fashion than their wild-type counterparts; in these cases, alteration of the anchor residues yields a more rigidly bound peptide. For mutations that replace charges at the P2 or C-terminal positions, greater conformational stability may follow from elimination of charge repulsion as noted above. Effect of anchor modification on peptide conformational stability has been noted previously, and notably, increased peptide flexibility correlates with a loss of immunogenicity. This may occur by reducing the opportunities

for productive interactions with T cell receptors (Sharma et al., 2001; Borbulevych et al., 2007; Insaiddo et al., 2011) or increasing the lifetime of the MHC-bound peptide.

In a recent study, Fritsch et al. (2014) have performed a retrospective of 40 neoepitopes of human cancers, “that induced immune responses, most of which were associated with regression or long-term disease stability.” Interestingly, in all of these neoepitopes, the mutations were located in the TCR-facing residues of the neoepitopes rather than the anchor residues. At first look, these data appear to contradict our finding that the vast majority of bonafide neoepitopes arise due to changes in anchor residues. However, a careful look at the data shows that the comparison cannot really be made: the criterion “associated with aggression or long-term disease stability” as determined subjectively by a dozen different groups of clinical investigators, is a very different criterion from actually testing tumor regression, (as we have done experimentally with mice). The true test of this question in a human study would require a randomized trial where patients would be immunized based upon one set of criteria or the other. Fritsch et al. (2014) have just begun a study using their criteria in human melanoma, and we are starting a clinical study in epithelial ovarian cancer based on our criteria. These studies will allow us to compare the two approaches.

A most surprising observation that emerges from our study is that 10/10 neoepitopes that elicit protective immunity are classified as nonbinders of  $K^d$  by NetMHC (cut-off value of 8.72). Correspondingly, the affinity of 8/10 neoepitopes for  $K^d$  is well over 500 nM, the traditional threshold for fruitful interaction of viral epitopes with MHC I molecules. In five of five instances tested, these presumed “non-binders” elicit classical CD8 T cell-dependent tumor immunity. This observation challenges some of our basic assumptions about MHC I-peptide-T cell receptor interactions, and exposes a far wider universe of potential neoantigens than assumed thus far.

The observed dissociation between detectable CD8 responses and immunoprotection (Table 5) from tumor growth merits comment. The neoepitopes that elicit immunoprotection and a CD8 response are straightforward and require no comment. The neoepitopes that elicit immunoprotection but not a detectable CD8 response (Fig. 3) may also be understood with the explanation that the CD8 response elicited is too weak to be detected by the ELISpot assay, thus highlighting the need for developing more sensitive assays for CD8 cells and their activities. It is, however, the epitopes that elicit potent CD8 responses but not immunoprotection that are difficult to understand. However, data in Table 5 may provide guidance in thinking about this dissociation. Note the neoepitopes Tnp03.1, 3.2, 3.3, and 3.4 for Meth A in Table 5. They share the same N-terminal mutations (sy/LD), but differ in the extent of their extension on the C termini. Whereas the first three elicit strong CD8 responses, Tnp03.4 does not. More interestingly, of the three neoepitopes that are immunogenic, only one, Tnp03.1, elicits protection from tumor growth. It is conceivable that Tnp03.1 is the only neoepitope that is naturally presented, while the others are not. This entirely

testable hypothesis provides a framework for testing the dissociation between T cell responses and immunoprotection.

We note with satisfaction, but also with some surprise that not a single WT epitope among the more than 100 tested (66 epitopes listed in Tables 4 and 5, and >35 additional epitopes) elicited a measurable, amplifiable CD8 immune response. The immune responses, when detected after a first immunization, were abrogated rather than enhanced after a second immunization, consistent with them being peripherally tolerized responses. This study represents perhaps the largest in which the immune responses to such a large number of self-epitopes have been systematically tested, and testifies strongly to the powerful scope of mechanisms of negative selection and peripheral tolerance.

These results present clear opportunities for clinical translation in human cancer immunotherapy. With the advent of high-throughput and inexpensive DNA sequencing, it is now possible to routinely sequence the exomes of cancers and normal tissues of each cancer patient, and compare the two to identify cancer-specific mis-sense mutations. The NetMHC or other such commonly available algorithms can then be used to identify the potential neoepitopes generated by the mis-sense mutations, for each of the three to six HLA I alleles of each patient. Peptides corresponding to the neoepitopes can then be chemically synthesized and used to immunize patients. However, the numbers of potential neoepitopes can be vast, and it is impractical to immunize patients with such vast numbers of peptides. The pioneering study of van Rooij et al. (2013) is an excellent case in point. Using exome comparison, indexed to the tumor transcriptome, these authors identified 448 potential neoepitopes in a melanoma patient; of these, only two were truly immunogenic. The combination of the NetMHC algorithm with the DAI and the C-terminal stability algorithms, as identified here, now makes it possible to reduce the large numbers of potential neoepitopes to a much smaller number of truly immunogenic epitopes, which can now be used to immunize patients in a realistic manner.

Our observations also present new opportunities to address some long-standing questions in immunology. Antigenic heterogeneity of cancers has been the subject of much discussion, but due to the lack of bona fide tumor-specific antigens in significant numbers, the debate has been largely theoretical. Uncovering of a large repertoire of true tumor-specific neoepitopes now allows the questions regarding antigenic heterogeneity (and antigen escape) to be asked in an unprecedentedly robust manner. Inherently linked to the issue of antigenic heterogeneity is the role of immunoediting of growing cancers. Since the classical studies on immune surveillance against tumors (Burnet, 1970), Dunn et al. (2004) have suggested that tumors go through elimination, equilibrium, and escape phases of immunoediting and have demonstrated evidence supporting the idea (Matsushita et al., 2012). The availability of a large repertoire of tumor-specific neoepitopes of any given tumor allows immuno-editing to be addressed in far more granularity. Epitope spreading is one of the more exciting and underexplored ideas in cancer immunity

(Markiewicz et al., 2001; Corbière et al., 2011). The major reason that it is underexplored is the paucity of a suitable number of antigenic neopeptides. Identification of a large number of neopeptides for practically any mouse or human tumor now permits a vigorous exploration of epitope spreading, along with its clinical utilization.

## MATERIALS AND METHODS

**Mice and tumors.** The BALB/c mice (6–8-wk-old female) were purchased from The Jackson Laboratory. Mice were maintained in the virus-free mouse facilities at the University of Connecticut Health Center and their use was approved and monitored by the Institutional Animal Care and Use Committee.

**Sample preparation.** Samples were prepared using the Illumina protocol outlined in “Preparing Samples for Sequencing of mRNA” (Part# 1004898 Rev. A September 2008). The protocol consists of two parts: cDNA synthesis and paired-end library preparation. First, mRNA was purified from total RNA using magnetic oligo(dT) beads, and then fragmented using divalent cations under elevated temperature. cDNA was synthesized from the fragmented mRNA using Superscript II (Invitrogen), followed by second strand synthesis. cDNA fragment ends were repaired and phosphorylated using Klenow, T4 DNA Polymerase, and T4 Polynucleotide Kinase. Next, an ‘A’ base was added to the 3’ end of the blunted fragments, followed by ligation of Illumina Paired-End adapters via T–A-mediated ligation. The ligated products were size selected by gel purification and then PCR amplified using Illumina Paired-End primers. The library size and concentration were determined using an Agilent Bioanalyzer.

**GAI run conditions.** The RNA-seq library was seeded onto the flowcell at 8 pM, yielding ~282–384 K clusters per tile. The library was sequenced using 61 cycles of chemistry and imaging.

**Analysis of sequencing data.** Initial data processing and base calling, including extraction of cluster intensities, was done using RTA (SCS version 2.6 and SCS version 2.61). Sequence quality filtering script was executed in the Illumina CASAVA software (ver 1.6.0, Illumina).

**Epi-Seq bioinformatics pipeline.** The pipeline starts by mapping RNA-Seq reads against the strain-specific genome sequences downloaded from the Sanger Mouse Genomes Project (Keane et al., 2011) and a strain-specific haploid transcript library derived from CCDS annotations (Li et al., 2009). We used BALB/c genome/transcriptome sequences for CMS5 and Meth A cell lines. DatabaseSNP polymorphisms were removed. Instead of comparing to the mm9 reference genome, and then excluding SNPs in dbSNP, we created a strain-specific genome by applying strain SNVs to the mm9 reference genome. The SNVs were downloaded from the mouse genomes. The created genome was used to map the reads and call the mutations. Reads were mapped using Bowtie (Langmead et al., 2009) with the default seed length of 28, maximum of 2 mismatches in the seed, and maximum sum of phred quality scores at mismatch positions of 125. After an initial round of mapping, we calculated mismatch statistics for each read position and each sample. Based on this analysis, 2 bases from the 5’ end and 10 bases from the 3’ end were clipped from all aligned reads. The resulting read alignments were merged using the HardMerge algorithm (Duitama et al., 2012). HardMerge discards reads that align to multiple locations in the genome and/or transcriptome, as well as reads that align uniquely to both, but at discordant locations. To reduce the effect of bias introduced by PCR amplification during library preparation (Aird et al., 2011), we replaced multiple reads with alignments starting at the same genomic location with their consensus. The SNVQ algorithm (Aird et al., 2011) was then used to call single nucleotide variants (SNVs) from the filtered set of aligned reads. SNVQ uses Bayes’ rule to call the genotype with the highest probability while taking base quality scores into account. High confidence SNVs were selected by requiring a minimum phred

quality score of 50 for each called genotype, a minimum of 3 reads supporting the alternative allele, with at least one read mapping on each strand. Haplotype inference over called SNV genotypes was performed using the RefHap Single Individual Haplotyping algorithm in (Duitama et al., 2012) that uses read evidence to phase blocks of proximal SNVs. Because residual heterozygosity in the inbred mice used in our experiments is predicted to be low (Bailey 1978 [specifically, Table 1 therein]), unique heterozygous SNVs were considered to be novel somatic mutations. Homozygous SNVs as well as heterozygous SNVs shared by more than one tumor with the same genome background were assumed to be germ-line mutations and were not used for epitope prediction unless located near a unique heterozygous SNV. For each unique heterozygous SNV, reference and alternative peptide sequences were generated based on the two inferred haplotypes for each CCDS transcript. Generated amino acid sequences were then run through the NetMHC 3.0 epitope prediction program (Lundegaard et al., 2008) and scored using the Profile Weight Matrix (PWM) algorithm with default detection thresholds.

**Binding assays.** Binding of peptides to H-2 K<sup>d</sup> was determined using quantitative assays based on the inhibition of binding of a radiolabeled standard peptide to purified MHC molecules essentially as described previously (van der Most et al., 1996; Sidney et al., 2013). Peptides were typically tested at six different concentrations covering a 100,000-fold dose range, and in three or more independent assays. Under the conditions used, where [label] < [MHC] and IC<sub>50</sub> ≥ [MHC], the measured IC<sub>50</sub> values are reasonable approximations of the dissociation constant values (Gulukota et al., 1997).

**Intracellular IFN-γ assay by FACS and ELISpot.** Lymphocytes were incubated either with or without 1–10 µg/ml peptide. GolgiPlug (BD) was added 1 h later. After incubation of 12 to 16 h, cells were stained for CD44 (clone IM7), CD4 (clone GK1.5) and CD8 (clone 53–6.7; BD), fixed and permeabilized using the Cytotfix/Cytoperm kit (BD), and stained for intracellular IFN-γ using Phycoerythrin-conjugated anti-mouse IFN-γ (clone XMG1.2, BD). Cells were stained with 1 µl antibody/million cells in 50 µl staining buffer (PBS with 1% bovine serum albumin) and incubated for 20 min at 4°C in the dark, or according to the manufacturer’s instructions. Cells without peptide stimulation were used as a negative control, and the values for these controls was very close to the values seen with the negative control peptide. Typically, 95,000–129,000 lymphocytes (14,500–17,000 CD8<sup>+</sup>CD4<sup>+</sup> cells) were acquired. Our background is consistently very low (10% of the signal).

For the ELISpot assays, our negative controls are CD8<sup>+</sup> cells from immunized mice without peptide stimulation. We consider the peptides to be positive or immunogenic when spots from peptide-stimulated wells are significant higher by Mann-Whitney test, compared with wells without cognate peptide stimulation. We rate the magnitude of responses by mean spot numbers per million CD8<sup>+</sup> cells: 5–10(+); 11–20(++); 21–50(+++); 51–100(++++) and >100(++++)+.

**Tumor challenge and representation of tumor growth.** AUC as a tool to measure tumor growth has been described previously (Duan et al., 2012). In brief, AUC was calculated by selecting “Curves & Regression” and then “Area under curve” from the “analyze” tool, using the Prism 5.0 (GraphPad). Grubbs’ test was used to remove up to one outlier from each group.

**Depletion of T cell subsets.** Immunized mice were depleted of CD8 cells using anti-CD8 rat IgG2b monoclonal antibody 2.43, or depleted of CD4 cells using anti-CD4 rat IgG2b monoclonal antibody GK1.5. Depleting antibodies were given in PBS i.p. 2 d before tumor challenge and every 7 d for the duration of the experiment. The first 3 injections of depleting antibodies were 250 µg per mouse; later injections were 500 µg per mouse. For treatment with antagonistic antibodies, mice were treated with anti-CD25 antibody (clone PC61, 250 µg, 2 d before tumor challenge) or anti-CTLA-4 antibody (clone 9D9; 100 µg; 7 d before and every 3 d after tumor challenge). The appropriate T cell subsets were depleted by >95%.

**Modeling of peptide/H-2K<sup>d</sup> complexes.** Models of peptide/H-2K<sup>d</sup> complexes were built by adapting a previous method by Park et al. (2013) used to identify immunogenic epitopes. Although developed on HLA-A\*0201, the approach is generally applicable to class I MHC proteins in general, taking advantage of common class I MHC structural features.

An initial model of each complex was generated by MODELLER and the ‘build mutants’ functionality implemented in Accelrys Discovery Studio (Feyfant et al., 2007; Webb and Sali 2014). The protocol uses a heavy-atom representation of the protein and includes homology-derived restraints combined with energy minimization and molecular dynamics/simulated annealing. The structure of the Flu peptide bound to H-2K<sup>d</sup> was used as a template (Mitaksov and Fremont 2006). Atoms within 4.5 Å of each altered residue were allowed to repack during the modeling. For each pMHC, one hundred initial models were generated, and the lowest energy model from this first set was subjected to a more exhaustive, second phase of fully atomistic simulated annealing and molecular dynamics. For the second phase, after adding hydrogens, the structure was heated to 1500 K over 200 ps of simulation followed by cooling to 300 K over 800 ps. The annealed structure was then subjected to five independent 10 ns molecular dynamics runs at 300 K, each time beginning with the structure that resulted from the second phase annealing step.

All second phase dynamics calculations were performed using AMBER 12 running on Nvidia GPU accelerators (Case et al., 2005). The ff99SB force field was used with a 2 fs time step. Solvent was treated implicitly using the generalized Born model to accelerate sampling (Götz et al., 2012). The SHAKE algorithm was used to constrain all bonds to hydrogens. A 20-kcal/mol harmonic restraint was applied to the  $\alpha 1$  and  $\alpha 2$  helices (residues 56–85 and 138–175). Two additional 20 kcal/mol distance restraints were applied to hydrogen bonds at the N- and C-terminal ends of the peptide. The first was between the P1 backbone oxygen and the hydroxyl of Tyr 159 of H-2K<sup>d</sup>. The second was between the P8 backbone oxygen and the ring nitrogen of Trp 147 of H-2K<sup>d</sup>. As a positive control, we performed the simulated annealing and molecular dynamics steps of the procedure on the structure of an HBV peptide presented by K<sup>d</sup> (Zhou et al., 2004). As shown in Figs. 4 B and 4D, the viral peptide was predicted to be relatively rigid in the K<sup>d</sup> binding groove.

For the data in Fig. 4 B, average peptide structures were calculated from the 50 ns of simulation data for each pMHC, and all common atoms of pairs of mutant and WT peptides superimposed to generate the RMSD. For the data in Figs. 4 (C–E) and Fig. S1, RMSFs were computed for the  $\alpha$  carbons of the peptides from the 50 ns of simulations.

**Statistical analysis.** P-values for group comparisons were calculated using a two-tailed nonparametric Mann–Whitney test, using GraphPad Prism 5.0 (GraphPad). For tumor rejection assays, Grubb’ test was used to remove up to one outlier from each group. Fisher’s exact test was used to test association between pairs of categorical parameters. Statistical significance of a Pearson correlation coefficient was computed using two-sided Student’s *t* test as described in (Cohen et al., 2003).

**Online supplemental material.** Table S1 contains the entire output of the EpiSeq pipeline for Meth A and CMS5 tumors. Fig. S1, which accompanies Fig. 4, shows the complete data for root mean square fluctuations for the  $\alpha$  carbons of all top DAI ranked nonamers from the structural modeling. Figs. S2 and S3 show the FACS gating strategies and representative primary data for Fig. 1 and Fig. 5 C, respectively. Online supplemental material is available at <http://www.jem.org/cgi/content/full/jem.20141308/DC1>.

We acknowledge Drs. Anthony T. Vella and Adam Adler of UConn School of Medicine for helpful discussions.

This research was funded in part by an award from the Cancer Research Institute, New York (P.K. Srivastava), the Northeastern Utilities Chair in Experimental Oncology (P.K. Srivastava), the Personalized Immunotherapy Core Interest Group of the Connecticut Institute for Clinical and Translational Science (P.K. Srivastava), a SPARK award (P.K. Srivastava, I.I. Mandoiu, F. Duan, S. Al Seesi), award IIS-0916948 (I.I. Mandoiu) from National Science Foundation, awards GM067079 and GM103773 from National Institutes of Health (B.M. Baker), Collaborative Research Grant AG110891 on “Software for Robust Transcript Discovery and Quantification” from

Life Technologies (I.I. Mandoiu), and the Agriculture and Food Research Initiative Competitive Grant no. 2011-67016-30331 from the USDA National Institute of Food and Agriculture (to I.I. Mandoiu).

This paper is dedicated to the cherished memory of Lloyd J. Old for his mentorship and friendship to P.K. Srivastava.

P.K. Srivastava has a significant equity interest in Accuragen Inc., which has obtained an option to license the intellectual property relating to the results described in this manuscript. The authors have no additional financial interests.

Submitted: 11 July 2014

Accepted: 5 September 2014

## REFERENCES

- Aird, D., M.G. Ross, W.S. Chen, M. Danielsson, T. Fennell, C. Russ, D.B. Jaffe, C. Nusbaum, and A. Gnirke. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 12:R18. <http://dx.doi.org/10.1186/gb-2011-12-2-r18>
- Assarsson, E., J. Sidney, C. Oseroff, V. Paschetto, H.H. Bui, N. Frahm, C. Brander, B. Peters, H. Grey, and A. Sette. 2007. A quantitative analysis of the variables affecting the repertoire of T cell specificities recognized after vaccinia virus infection. *J. Immunol.* 178:7890–7901. <http://dx.doi.org/10.4049/jimmunol.178.12.7890>
- Bailey, D.W. 1978. Sources of subline divergences and their relative importance for inbred mice of six major inbred strains of mice. In *Origins of inbred mice*. H.C. Morse III, editor. Elsevier <http://dx.doi.org/10.1016/B978-0-12-507850-4.50020-2>
- Borbulevych, O.Y., T.K. Baxter, Z. Yu, N.P. Restifo, and B.M. Baker. 2005. Increased immunogenicity of an anchor-modified tumor-associated antigen is due to the enhanced stability of the peptide/MHC complex: implications for vaccine design. *J. Immunol.* 174:4812–4820. <http://dx.doi.org/10.4049/jimmunol.174.8.4812>
- Borbulevych, O.Y., F.K. Insaïdo, T.K. Baxter, D.J. Powell Jr., L.A. Johnson, N.P. Restifo, and B.M. Baker. 2007. Structures of MART-126/27–35 Peptide/HLA-A2 complexes reveal a remarkable disconnect between antigen structural homology and T cell recognition. *J. Mol. Biol.* 372:1123–1136. <http://dx.doi.org/10.1016/j.jmb.2007.07.025>
- Brass, A.L., D.M. Dykxhoorn, Y. Benita, N. Yan, A. Engelman, R.J. Xavier, J. Lieberman, and S.J. Elledge. 2008. Identification of host proteins required for HIV infection through a functional genomic screen. *Science.* 319:921–926. <http://dx.doi.org/10.1126/science.1152725>
- Buckwalter, M.R., and P.K. Srivastava. 2008. “It is the antigen(s), stupid” and other lessons from over a decade of vaccitherapy of human cancer. *Semin. Immunol.* 20:296–300. <http://dx.doi.org/10.1016/j.smim.2008.07.003>
- Burnet, F.M. 1970. The concept of immunological surveillance. *Prog. Exp. Tumor Res.* 13:1–27.
- Callahan, M.K., J.D. Wolchok, and J.P. Allison. 2010. Anti-CTLA-4 antibody therapy: immune monitoring during clinical development of a novel immunotherapy. *Semin. Oncol.* 37:473–484. <http://dx.doi.org/10.1053/j.seminoncol.2010.09.001>
- Cao, K., J. Hollenbach, X. Shi, W. Shi, M. Chopek, and M.A. Fernández-Viña. 2001. Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. *Hum. Immunol.* 62:1009–1030. [http://dx.doi.org/10.1016/S0198-8859\(01\)00298-1](http://dx.doi.org/10.1016/S0198-8859(01)00298-1)
- Case, D.A., T.E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K.M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang, and R.J. Woods. 2005. The Amber biomolecular simulation programs. *J. Comput. Chem.* 26:1668–1688. <http://dx.doi.org/10.1002/jcc.20290>
- Castle, J.C., S. Kreiter, J. Diekmann, M. Löwer, N. van de Roemer, J. de Graaf, A. Selmi, M. Diken, S. Boegel, C. Paret, et al. 2012. Exploiting the mutanome for tumor vaccination. *Cancer Res.* 72:1081–1091. <http://dx.doi.org/10.1158/0008-5472.CAN-11-3722>
- Cohen, J., P. Cohen, S.G. West, and L.S. Aiken. 2003. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Third edition. Lawrence Erlbaum Associates, Mahwah, NJ.
- Corbière, V., J. Chapiro, V. Stroobant, W. Ma, C. Lurquin, B. Lethé, N. van Baren, B.J. Van den Eynde, T. Boon, and P.G. Coulié. 2011. Antigen spreading contributes to MAGE vaccination-induced regression

- of melanoma metastases. *Cancer Res.* 71:1253–1262. <http://dx.doi.org/10.1158/0008-5472.CAN-10-2693>
- Ding, Y.H., B.M. Baker, D.N. Garboczi, W.E. Biddison, and D.C. Wiley. 1999. Four A6-TCR/peptide/HLA-A2 structures that generate very different T cell signals are nearly identical. *Immunity.* 11:45–56. [http://dx.doi.org/10.1016/S1074-7613\(00\)80080-1](http://dx.doi.org/10.1016/S1074-7613(00)80080-1)
- Duan, F., Y. Lin, C. Liu, M.E. Engelhorn, A.D. Cohen, M. Curran, S. Sakaguchi, T. Merghoub, S. Terzulli, J.D. Wolchok, and A.N. Houghton. 2009. Immune rejection of mouse tumors expressing mutated self. *Cancer Res.* 69:3545–3553. <http://dx.doi.org/10.1158/0008-5472.CAN-08-2779>
- Duan, F., S. Simeone, R. Wu, J. Grady, I. Mandoiu, and P.K. Srivastava. 2012. Area under the curve as a tool to measure kinetics of tumor growth in experimental animals. *J. Immunol. Methods.* 382:224–228. <http://dx.doi.org/10.1016/j.jim.2012.06.005>
- Duitama, J., P.K. Srivastava, and I.I. Mandoiu. 2012. Towards accurate detection and genotyping of expressed variants from whole transcriptome sequencing data. *BMC Genomics.* 13:S6. <http://dx.doi.org/10.1186/1471-2164-13-S2-S6>
- Dunn, G.P., L.J. Old, and R.D. Schreiber. 2004. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity.* 21:137–148. <http://dx.doi.org/10.1016/j.immuni.2004.07.017>
- Egen, J.G., M.S. Kuhns, and J.P. Allison. 2002. CTLA-4: new insights into its biological function and use in tumor immunotherapy. *Nat. Immunol.* 3:611–618. <http://dx.doi.org/10.1038/ni0702-611>
- Feyfant, E., A. Sali, and A. Fiser. 2007. Modeling mutations in protein structures. *Protein Sci.* 16:2030–2041. <http://dx.doi.org/10.1110/ps.072855507>
- Fritsch, E.F., M. Rajasagi, P.A. Ott, V. Brusic, N. Hacohen, and C.J. Wu. 2014. HLA-binding properties of tumor neoepitopes in humans. *Cancer Immunol Res.* 2:522–529. <http://dx.doi.org/10.1158/2326-6066.CIR-13-0227>
- Götz, A.W., M.J. Williamson, D. Xu, D. Poole, S. Le Grand, and R.C. Walker. 2012. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* 8:1542–1555. <http://dx.doi.org/10.1021/ct200909j>
- Gulukota, K., J. Sidney, A. Sette, and C. DeLisi. 1997. Two complementary methods for predicting peptides binding major histocompatibility complex molecules. *J. Mol. Biol.* 267:1258–1267. <http://dx.doi.org/10.1006/jmbi.1997.0937>
- Hanson, H.L., D.L. Donermeyer, H. Ikeda, J.M. White, V. Shankaran, L.J. Old, H. Shiku, R.D. Schreiber, and P.M. Allen. 2000. Eradication of established tumors by CD8+ T cell adoptive immunotherapy. *Immunity.* 13:265–276. [http://dx.doi.org/10.1016/S1074-7613\(00\)00026-1](http://dx.doi.org/10.1016/S1074-7613(00)00026-1)
- Ikeda, H., N. Ohta, K. Furukawa, H. Miyazaki, L. Wang, K. Kuribayashi, L.J. Old, and H. Shiku. 1997. Mutated mitogen-activated protein kinase: a tumor rejection antigen of mouse sarcoma. *Proc. Natl. Acad. Sci. USA.* 94:6375–6379. <http://dx.doi.org/10.1073/pnas.94.12.6375>
- Insaiddo, F.K., O.Y. Borbulevych, M. Hossain, S.M. Santhanagopalan, T.K. Baxter, and B.M. Baker. 2011. Loss of T cell antigen recognition arising from changes in peptide and major histocompatibility complex protein flexibility: implications for vaccine design. *J. Biol. Chem.* 286:40163–40173. <http://dx.doi.org/10.1074/jbc.M111.283564>
- Keane, T.M., L. Goodstadt, P. Danecek, M.A. White, K. Wong, B. Yalcin, A. Heger, A. Agam, G. Slater, M. Goodson, et al. 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature.* 477:289–294. <http://dx.doi.org/10.1038/nature10413>
- Langmead, B., C. Trapnell, M. Pop, and S.L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. <http://dx.doi.org/10.1186/gb-2009-10-3-r25>
- Lennerz, V., M. Fatho, C. Gentilini, R.A. Frye, A. Lifke, D. Ferel, C. Wölfel, C. Huber, and T. Wölfel. 2005. The response of autologous T cells to a human melanoma is dominated by mutated neoantigens. *Proc. Natl. Acad. Sci. USA.* 102:16013–16018. <http://dx.doi.org/10.1073/pnas.0500090102>
- Levey, D.L., H. Udono, M. Heike, and P.K. Srivastava. 2001. Identification of a tumor-associated contact-dependent activity which reversibly downregulates cytolytic function of CD8+ T cells. *Cancer Immun.* 1:5.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 25:2078–2079. <http://dx.doi.org/10.1093/bioinformatics/btp352>
- Lu, Y.C., X. Yao, Y.F. Li, M. El-Gamil, M.E. Dudley, J.C. Yang, J.R. Almeida, D.C. Douek, Y. Samuels, S.A. Rosenberg, and P.F. Robbins. 2013. Mutated PPP1R3B is recognized by T cells used to treat a melanoma patient who experienced a durable complete tumor regression. *J. Immunol.* 190:6034–6042. <http://dx.doi.org/10.4049/jimmunol.1202830>
- Lundegaard, C., O. Lund, and M. Nielsen. 2008. Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers. *Bioinformatics.* 24:1397–1398. <http://dx.doi.org/10.1093/bioinformatics/btn128>
- Markiewicz, M.A., F. Fallarino, A. Ashikari, and T.F. Gajewski. 2001. Epitope spreading upon P815 tumor rejection triggered by vaccination with the single class I MHC-restricted peptide P1A. *Int. Immunol.* 13:625–632. <http://dx.doi.org/10.1093/intimm/13.5.625>
- Matsushita, H., M.D. Vesely, D.C. Koboldt, C.G. Rickert, R. Uppaluri, V.J. Magrini, C.D. Arthur, J.M. White, Y.S. Chen, L.K. Shea, et al. 2012. Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. *Nature.* 482:400–404. <http://dx.doi.org/10.1038/nature10755>
- Mitaksov, V., and D.H. Fremont. 2006. Structural definition of the H-2Kd peptide-binding motif. *J. Biol. Chem.* 281:10618–10625. <http://dx.doi.org/10.1074/jbc.M51051200>
- Narzi, D., C.M. Becker, M.T. Fiorillo, B. Uchanska-Ziegler, A. Ziegler, and R.A. Böckmann. 2012. Dynamical characterization of two differentially disease associated MHC class I proteins in complex with viral and self-peptides. *J. Mol. Biol.* 415:429–442. <http://dx.doi.org/10.1016/j.jmb.2011.11.021>
- Park, M.S., S.Y. Park, K.R. Miller, E.J. Collins, and H.Y. Lee. 2013. Accurate structure prediction of peptide-MHC complexes for identifying highly immunogenic antigens. *Mol. Immunol.* 56:81–90. <http://dx.doi.org/10.1016/j.molimm.2013.04.011>
- Rajasagi, M., S.A. Shukla, E.F. Fritsch, D.B. Keskin, D. DeLuca, E. Carmona, W. Zhang, C. Sougnez, K. Cibulskis, J. Sidney, et al. 2014. Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood.* 124:453–462. <http://dx.doi.org/10.1182/blood-2014-04-567933>
- Ren, S., Z. Peng, J.H. Mao, Y. Yu, C. Yin, X. Gao, Z. Cui, J. Zhang, K. Yi, W. Xu, et al. 2012. RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicings. *Cell Res.* 22: 806–821. <http://dx.doi.org/10.1038/cr.2012.30>
- Segal, N.H., D.W. Parsons, K.S. Peggs, V. Velculescu, K.W. Kinzler, B. Vogelstein, and J.P. Allison. 2008. Epitope landscape in breast and colorectal cancer. *Cancer Res.* 68:889–892. <http://dx.doi.org/10.1158/0008-5472.CAN-07-3095>
- Sharma, A.K., J.J. Kuhns, S. Yan, R.H. Friedline, B. Long, R. Tisch, and E.J. Collins. 2001. Class I major histocompatibility complex anchor substitutions alter the conformation of T cell receptor contacts. *J. Biol. Chem.* 276:21443–21449. <http://dx.doi.org/10.1074/jbc.M010791200>
- Sidney, J., S. Southwood, C. Moore, C. Oseroff, C. Pinilla, H.M. Grey, and A. Sette. Measurement of MHC/Peptide Interactions by Gel Filtration or Monoclonal Antibody Capture. *Curr Protoc Immunol.* 2013; Chapter 18:Unit18.3.
- Srivastava, P.K. 1993. Peptide-binding heat shock proteins in the endoplasmic reticulum: role in immune response to cancer and in antigen presentation. *Adv. Cancer Res.* 62:153–177. [http://dx.doi.org/10.1016/S0065-230X\(08\)60318-8](http://dx.doi.org/10.1016/S0065-230X(08)60318-8)
- Srivastava, N., and P.K. Srivastava. 2009. Modeling the repertoire of true tumor-specific MHC I epitopes in a human tumor. *PLoS ONE.* 4:e6094. <http://dx.doi.org/10.1371/journal.pone.0006094>
- van der Bruggen, P., V. Stroobant, N. Vigneron, and B. Van den Eynde. Peptide database: T cell-defined tumor antigens. *Cancer Immun* 2013 <http://www.cancerimmunity.org/peptide/>
- van der Most, R.G., A. Sette, C. Oseroff, J. Alexander, K. Murali-Krishna, L.L. Lau, S. Southwood, J. Sidney, R.W. Chesnut, M. Matloubian, and R.A. Ahmed. 1996. Analysis of cytotoxic T cell responses to dominant and subdominant epitopes during acute and chronic lymphocytic choriomeningitis virus infection. *J. Immunol.* 157:5543–5554.
- van Rooij, N., M.M. van Buuren, D. Philips, A. Velds, M. Toebes, B. Heemskerk, L.J. van Dijk, S. Behjati, H. Hilkmann, D. El Atmioui,

- et al. 2013. Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma. *J. Clin. Oncol.* 31:e439–e442. <http://dx.doi.org/10.1200/JCO.2012.47.7521>
- Vogelstein, B., N. Papadopoulos, V.E. Velculescu, S. Zhou, L.A. Diaz Jr., K.W. Kinzler, L.J. van Dijk, B. Vogelstein, N. Papadopoulos, V.E. Velculescu, et al. 2013. Cancer genome landscapes. *Science*. 339:1546–1558. <http://dx.doi.org/10.1126/science.1235122>
- Webb, B., and A. Sali. 2014. Protein structure modeling with MODELLER. *Methods Mol. Biol.* 1137:1–15. [http://dx.doi.org/10.1007/978-1-4939-0366-5\\_1](http://dx.doi.org/10.1007/978-1-4939-0366-5_1)
- Wei, X., V. Walia, J.C. Lin, J.K. Teer, T.D. Prickett, J. Gartner, S. Davis, K. Stemke-Hale, M.A. Davies, J.E. Gershenwald, et al. NISC Comparative Sequencing Program. 2011. Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nat. Genet.* 43:442–446. <http://dx.doi.org/10.1038/ng.810>
- Winkler, K., A. Winter, C. Rueckert, B. Uchanska-Ziegler, and U. Alexiev. 2007. Natural MHC class I polymorphism controls the pathway of peptide dissociation from HLA-B27 complexes. *Biophys. J.* 93:2743–2755. <http://dx.doi.org/10.1529/biophysj.106.096602>
- Zhou, M., Y. Xu, Z. Lou, D.K. Cole, X. Li, Y. Liu, P. Tien, Z. Rao, and G.F. Gao. 2004. Complex assembly, crystallization and preliminary X-ray crystallographic studies of MHC H-2Kd complexed with an HBV-core nonapeptide. *Acta Crystallogr. D Biol. Crystallogr.* 60:1473–1475. <http://dx.doi.org/10.1107/S0907444904013587>