Human Immunology xx (2010) xxx

Contents lists available at ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

A census of predicted mutational epitopes suitable for immunologic cancer control

René L. Warren [a] and Robert A. Holt [a,b,*]

[a] British Columbia Cancer Agency, Genome Sciences Centre, Vancouver, British Columbia, Canada
[b] Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada

ARTICLE INFO

Article history:
Received 2 October 2009
Accepted 17 December 2009
Available online 22 December 2009

Keywords:
Epitope prediction
T-cell epitope
Immunosurveillance
Cancer vaccine
Cancer immunity

ABSTRACT

The adaptive immune system can protect against spontaneously arising tumors, and the potential exists to reduce cancer incidence by priming adaptive immune responses with vaccines. Immunologic cancer control has been implemented for cancers caused by infectious agents, but not for spontaneous cancers caused by mutation. This is largely due to the high cost of preventative clinical trials and the lack of validated tumor epitopes. Here we evaluate, computationally, all known somatic mutations in human tumors for their antigenic potential. All possible human leukocyte antigen (HLA) class I presented peptides containing recurrent somatic cancer mutations with frequency > 5% were screened by three independent epitope prediction algorithms (SYFPEITHI, BIMAS, and IEDB). Using stringent filters, a total of 20 genes, 35 mutations, and 159 candidate epitopes were identified, each presented by up to four distinct HLA class I alleles. The top-ranking gene from our survey was KRAS, which figures prominently because there are frequent hotspot mutations in numerous, prevalent cancers, and mutant peptides are predicted to be presented by several common HLA alleles. From our data, we estimate that prophylactic vaccination could provide meaningful levels of prevention of tumors associated with common recurrent mutations.

1. Introduction

More than a century ago it was recognized independently by Ehrlich as well as Bashford and colleagues [1,2] that cells isolated from rodent tumors could be lethal when injected into naive mice but were unable to proliferate when injected at a new site in the same animal from which they were obtained, even as the original tumor continued to grow. This phenomenon was termed concomitant immunity and clearly illustrated that the mammalian immune system is effective in eliminating cancer if the burden of malignant cells is low, but is less effective for established tumors. Half a century later, a series of definitive experiments by Prehn and Main [3] showed that inoculation with tumor, but not normal tissue, was protective and, in time, the theory of cancer immunosurveillance was formally established [4]. Recently, it has been shown that Rag2$^{-/-}$ mice, which have no V(D)J recombination and therefore no repertoire of mature lymphocytes, have dramatically increased incidence of spontaneous tumors [5]. Further, it has been shown that tumors formed in immunologically permissive Rag2$^{-/-}$ hosts are more immunogenic when transferred to wild-type mice, and that host-immune pressures can maintain tumors in an equilibrium state [6]. Based on these observations, the early theory of immunosurveillance has been revised to that of immunoediting [7], which holds that spontaneously arising tumor cells are frequently eliminated by the immune system. Those that begin to grow are held in a state equilibrium with host immunity, from which they may eventually escape by various mechanisms such as loss of antigen, loss of antigen presentation pathways, or interference by regulatory T cells. The potential to use vaccination to mobilize adaptive immunity against cancer has been illustrated in mice engineered to express, for example, the SV40 T antigen [8], or activated rat Erbb2 [9–12]. In these animal models, pre-exposure to antigen can reproducibly provide complete protection against tumor development.

In humans, it is already well established that cancers of viral origin can be prevented by vaccination, and Gardasil® (Merck & Co., Inc., Whitehouse Station, NJ) now makes this a clinical reality for cervical cancer [13]. For tumors of mutational origin, the principal lines of evidence for antitumor adaptive immune responses are as follows. First, there are numerous case reports of donated organs in which occult cancers, initially held in check by donor immunity, undergo rapid and progressive outgrowth after transplantation because recipients are naive to the tumor antigens, and immunosuppressed [14–16]. Second, the only controlled study that has evaluated cancer rates in immunosuppressed individuals (Scandinavian kidney transplant recipients) reports a clear increase in the incidence of a wide variety of noninfectious primary cancers [17]. Third, it is well established in solid tumors that patients with detectable tumor infiltrating lymphocytes have better outcomes,

* Corresponding author.
E-mail address: rholt@bcgsc.ca (R.A. Holt).

0198-8859/10/$32.00 - see front matter
doi:10.1016/j.humimm.2009.12.007

both in terms of progression free interval and overall survival [18–20]. Finally, natural immune responses to tumor-specific antigens are detectable in patients with cancer. Humoral responses have been extensively characterized in patient serum and antibodies against common tumor antigens have been found, in some studies, in more than half of cancer patients but in very few healthy control individuals (reviewed in [21]). Natural cellular immune responses are more difficult to assess, and efforts have focused on characterization of tumor-infiltrating rather than circulating lymphocytes. $CD4^+$ and $CD8^+$ tumor-reactive T cells have been identified that recognize point mutations in numerous genes such as fibronectin [22], HSP70 [23], major histocompatibility complex class I [24], N-RAS [25], β-catenin [26], MART-2 [27], and p14ARF [24].

It is conspicuous that after more than a century of studying cancer immunology we still do not have immunologic cancer control (*i.e.*, vaccines) [28] for the most common and deadly cancers. The reasons for this include the fact that it is difficult to test preventative strategies against cancer. The likelihood of any individual developing a specific type of tumor is relatively low, and when they do appear, they occur over a wide age range such that preventative clinical trials require the involvement of a large number of subjects over many years. These trials are, therefore, expensive and require extraordinary justification. Further, for most cancers we do not have validated antigens. In cancers of viral origin, the human immune system responds to non-self antigens from the virus. In contrast, vaccination against tumors of mutational origin involves important and unique considerations. Antigens must be sufficiently common in the target cancer to have a meaningful public health impact and they must be reliably immunogenic (*i.e.*, nontolerated). They should also be causally involved in tumorigenicity and confer a selective advantage necessary for tumor survival, such that the risk of immune escape by antigen loss is minimized.

Most oncogenic mutations are incurred in intracellular signaling proteins. In nucleated cells, intracellular proteins are cleaved and a subset of the cleavage products are presented at the cell surface by human leukocyte antigen (HLA) class I, where they are subjected to surveillance by the repertoire of cytotoxic T lymphocytes. Because HLA class I restricted peptides are short (~8–11 amino acids), even single amino acid changes such as the hotspot mutations incurred by common tumor genes can be sufficient for cytotoxic T lymphocyte to single out and selectively kill cells presenting the mutant peptide [29,30]. For these reasons mutational epitopes (rather than expression or differentiation epitopes) will likely prove to be best suited for preventative vaccines against spontaneous cancers. As of this writing, only 39 HLA class I restricted mutational epitopes have been reported and compiled by the cancer immunity database (http://www.cancerimmunity.org). In this set there is over-representation of melanoma epitopes (17/39) and epitopes that are HLA-A2 restricted (11/39). HLA loci, particularly HLA antigen-binding regions, are the most highly polymorphic sites in the genome with 3,477 HLA class I allelic sequences now recognized (http://hla.alleles.org/alleles/index.html). Different HLA alleles have different antigen presentation characteristics, such that presentation of any given peptide is often restricted to, or strongly favored by, a specific HLA allele [31]. HLA restriction is a key consideration when assessing the immunogenicity of oncogenic mutations.

Here, we present a meta-analysis of predicted HLA class I restricted epitopes from recurrent tumor mutations using established databases and epitope prediction tools. We have evaluated, computationally, all mutant peptides derived from all known tumor mutations. Epitopes have been ranked according to (1) the estimated frequency of mutation in a given tumor type, (2) the incidence of that tumor, (3) the likelihood of the mutation producing an HLA presented T-cell epitope, and (4) the estimated population frequency of the presenting HLA allele. Because of the stringency of our analysis our predictions are very conservative, but offer the first quantitative estimate of the minimum benefit that might be expected from preventative cancer vaccination, and provide a short list of the best predicted epitopes for further study.

2. Methods

2.1. Selection of candidate tumor mutations

Researchers at the Sanger Institute [32–34] have undertaken the daunting task of curating and cataloging over 50,000 published somatic mutations from almost 4,800 genes in 250,000 tumors scattered across the vast cancer literature. Their database, COSMIC (Catalogue of Somatic Mutations in Cancer), tracks each gene and somatic change observed, its frequency, and the primary tissue where it occurs. Using COSMIC v43 Release (http://www.sanger.ac.uk/genetics/CGP/cosmic/), every combination of primary tissue, gene, and somatic mutation was considered for our study. Only somatic mutations seen in 5% or more of tissue samples and where at least five or more positives identified from 100 or more samples were retained for further analysis. A further condition for tumor mutation selection was that the cancer incidence be known for each tissue type where the mutation occurred. Cancer incidence data from the Surveillance, Epidemiology and End Results (SEER) program (http://seer.cancer.gov), which is the definitive source for cancer statistics in the United States, was used for this purpose.

First, we defined a Mutation Impact Score (MIS) as follows:

$$\text{MIS} = \text{Frequency of mutation in given tumor} \times \text{tumor incidence}$$

Each mutation was considered independently, regardless of whether there were other mutations in the same gene, or even at the same codon (*e.g.*, KRAS G12V and KRAS G12D were treated as independent mutations). For comparing mutations we also calculated a cumulative MIS, which was the sum of the MIS for all tumor sites in which that mutation was detected (Figure 1).

2.2. Identification of putative HLA ligands

We used computational HLA-binding prediction tools to identify mutations likely to be presented by specific HLA alleles as peptide ligands and, by extension, T-cell epitopes. Using web-based SYFPEITHI [35] (http://www.syfpeithi.de), BIMAS [36] (http://www-bimas.cit.nih.gov/molbio/hla_bind/), and IEDB [37,38] (Stabilized Matrix Method; http://tools.immuneepitope.org/analyze/html/mhc_binding.html), we queried each candidate mutation in a systematic and high-throughput fashion, testing all combinations of peptide lengths and HLA class I alleles available to query on the hosting server. These prediction algorithms use empiric peptide-HLA binding data to build models of peptide sequence specificity, and give a predictive outcome for yet untested sequences. In these models, the amino acids found at each position of short peptides eluted from real HLA molecules are compiled in a database. For SYFPEITHI and BIMAS a two-dimensional matrix is built using the observed frequency of each amino acid at each position within the HLA peptide binding pocket. Stabilized matrix method (IEDB) considers, in addition, pairwise interactions between amino acid positions and is thought to yield more accurate predictions [38]. These three tools are the most heavily used and highly cited tools for *in silico* HLA-binding predictions [39].

A script using PERL LWP and HTML::Form modules was developed to handle high-throughput form submission and post-prediction data extraction. This script was run in a manner that minimized impact on the hosting service, with a 5-second buffer between each submission. All overlapping 8-, 9-, 10-, and 11-mer mutational peptides were tested against all HLA class I combinations supported by the host servers. For each peptide containing the mutation of interest, the output in terms of HLA-binding prediction method, rank, score,

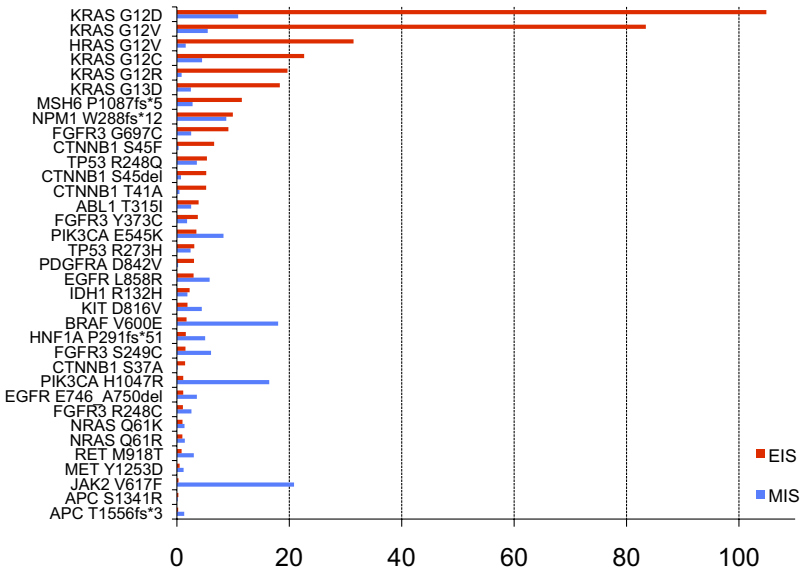

Fig. 1. Sum of mutation impact score (MIS) and sum of epitope impact score (EIS) for each candidate tumor-associated antigen identified by our study. The EIS, which takes into account the average HLA frequency in the population, and the HLA-binding prediction ranking is high for all KRAS mutants, which suggests that their predicted epitopes would be expected to have the greatest utility for immunologic cancer control.

start coordinate, length, peptide sequence, and HLA allele were stored in a local custom MySQL database for future querying. To be considered further, each peptide-HLA allele combination had to be found by at least two of the three epitope prediction algorithms and the predicted HLA ligand had to be within the top scoring 5% of all possible peptides from that protein. Only HLA class I coding variants were considered (*i.e.*, 4-digit resolution). Candidate epitopes that passed this filter were given an epitope score (ES).

$$\mathrm{ES} = 100/(\mathrm{AVG\ RANK} \times \mathrm{EXP}(\mathrm{STDEV}/100))$$

The ES is based on the mean prediction rank and standard deviation from the mean for two or more prediction algorithms and ranges from 0 to 100. A peptide predicted to bind to a given HLA and ranked first by any two or all three prediction programs will score 100 (*e.g.*, KRAS G12R VVVGARGVGK). We used the rank as a metric because the epitope prediction scores are determined differently by the three programs and are not directly comparable.

Next, we calculated Epitope Impact Score (EIS).

$$\mathrm{EIS} = \mathrm{ES} \times \text{average population frequency of presenting HLA allele}$$

The EIS considers the US population frequencies for each HLA class I allele (http://www.allelefrequencies.net). Only allelic frequencies with a sample size equal to or greater than 500 were used. The EIS will decrease the ES by a factor proportional to the allele frequency itself. Thus, even high-ranking peptides will have a low EIS score if they are predicted to be presented by rare HLA alleles. For example, top-ranking KRAS G12R peptide VVVGARGVGK (ES = 100), predicted to be presented by HLA class I coding variant A*1101 (average population frequency = 7.07%) has a low EIS of 7.07.

For each candidate peptide-HLA pair we calculated a global score (GS) that takes into account the MIS as

$$\mathrm{GS} = \mathrm{MIS} \times \mathrm{EIS}$$

And for each gene, we calculated an overall score (OS) which is the sum of the GS for all predicted epitopes for that gene across all tumor sites where the mutation is found.

$$\mathrm{OS} = \sum \mathrm{GS}$$

To estimate the proportion of the population that could benefit from vaccination with a combination of predicted epitopes, it is necessary to determine the proportion of the population that would carry at least one presenting HLA allele. The proportion (P) of the population that contains at least one of some larger number of alleles (K) is determined as described in Gulukota and DeLisi [40]. Assuming that two or more alleles occur with correlated probabilities, the overall coverage is the sum of individual allele coverage corrected for the overlaps.

$$P_{tot} = \sum_{i=1}^{i=k} P_i - \sum_{pairs} P_{ij} + \sum_{triplets} P_{ijk} - \ldots,$$

3. Results

We used the COSMIC database as a starting point to identify the most promising tumor mutations for epitope prediction. COSMIC compiles data on somatic tumor mutations from the public scientific literature, and it is the most comprehensive repository of curated somatic mutations in cancer [32–34]. Using our selection criteria for somatic mutations (minimum frequency 5%, minimum of five positives from at least 100 samples) we identified 36 distinct somatic mutations in 20 genes and 19 distinct tumor sites (Table 1). These consisted of point mutations (29 cases), indels causing frameshifts (five cases) and codon deletions (two cases). These results demonstrate that based on mutational screens undertaken to date, there appear to be relatively few recurrent mutations in cancer. The highest scoring mutation was JAK2 V617F in hematopoietic/lymphoid tumors (MIS = 20.8), followed by PIK3CA H1047R in breast tumors (MIS = 15.1), and the NPM1 W288-frameshift in, again, hematopoietic/lymphoid tumors (MIS = 8.8). Interestingly, JAK2 is prominent because it is mutated at high frequency (45.1%) in a single category of common tumors (hematopoietic and lymphoid). NPM1, though also mutated only in this same category of tumors, ranks lower because of lower mutation frequency (19%).

Our screen prioritizes genes where mutations occur at specific locations and in multiple cancers. These are the candidates where the smallest number of epitopes could have the greatest potential impact when delivered as a vaccine. As a result, there are some well-recognized cancer genes that do not figure prominently in our results. For example, TP53 is one of the most frequently mutated cancer genes, but it does not rank highly here because mutations occur at many sites throughout the gene rather than at one or a small number of specific hotspots. For the present study a low TP53

Table 1
Predicted mutational tumor epitopes from cancer genes sorted by overall score for each gene. The mutated amino acid is underlined

Gene	Mutation[a]	Tissue	Cancer incidence[b] (per 100k)	Mutation frequency[c]	MIS[d]	Peptide	Presenting allele	HLA frequency[e]	EIS[f]	GS[g]	OS[h]
KRAS	G13D	Large intestine	45.94	5.42%	2.49	VVVGAGDVGK	A*1101	7.07%	5.272	13.127	506.52
							A*6801	3.50%	0.671	1.670	
						KLVVVGAGDV	A*0201	13.66%	5.349	13.319	
						GAGDVGKSAL	B*3801	1.03%	0.221	0.549	
						VVGAGDVGK	A*1101	7.07%	2.089	5.201	
						AGDVGKSAL	B*0702	6.71%	3.307	8.235	
							B*3801	1.03%	0.227	0.565	
							B*3901	0.87%	0.120	0.299	
						DVGKSALTI	B*5101	4.29%	1.057	2.631	
	G12V	Biliary tract	1.14	5.89%	0.07	VVGAVGVGK	A*1101	7.07%	3.013	0.202	
						GAVGVGKSAL	B*0702	6.71%	1.072	0.072	
							B*3801	1.03%	0.221	0.015	
						KLVVVGAVGV	A*0201	13.66%	8.102	0.544	
						AVGVGKSAL	B*0702	6.71%	4.441	0.298	
						VVVGAVGVGK	A*1101	7.07%	7.070	0.475	
							A*6801	3.50%	0.751	0.050	
						YKLVVVGAV	A*0203	0.90%	0.256	0.017	
							B*3902	0.11%	0.017	0.001	
						LVVVGAVGV	A*0201	13.66%	2.732	0.183	
							A*0203	0.90%	0.138	0.009	
		Large intestine	45.94	7.29%	3.35	VVGAVGVGK	A*1101	7.07%	3.013	10.089	
						GAVGVGKSAL	B*0702	6.71%	1.072	3.589	
							B*3801	1.03%	0.221	0.738	
						KLVVVGAVGV	A*0201	13.66%	8.102	27.133	
						AVGVGKSAL	B*0702	6.71%	4.441	14.873	
						VVVGAVGVGK	A*1101	7.07%	7.070	23.678	
							A*6801	3.50%	0.751	2.514	
						YKLVVVGAV	A*0203	0.90%	0.256	0.859	
							B*3902	0.11%	0.017	0.056	
						LVVVGAVGV	A*0201	13.66%	2.732	9.149	
							A*0203	0.90%	0.138	0.462	
		Pancreas	11.70	17.67%	2.07	VVGAVGVGK	A*1101	7.07%	3.013	6.228	
						GAVGVGKSAL	B*0702	6.71%	1.072	2.216	
							B*3801	1.03%	0.221	0.456	
						KLVVVGAVGV	A*0201	13.66%	8.102	16.749	
						AVGVGKSAL	B*0702	6.71%	4.441	9.182	
						VVVGAVGVGK	A*1101	7.07%	7.070	14.616	
							A*6801	3.50%	0.751	1.552	
						YKLVVVGAV	A*0203	0.90%	0.256	0.530	
							B*3902	0.11%	0.017	0.035	
						LVVVGAVGV	A*0201	13.66%	2.732	5.648	
							A*0203	0.90%	0.138	0.285	
	G12R	Pancreas	11.70	7.17%	0.84	ARGVGKSAL	B*0702	6.71%	1.608	1.349	
							B*2705	1.38%	0.331	0.278	
							B*3901	0.87%	0.239	0.200	
						GARGVGKSAL	B*0702	6.71%	1.286	1.079	
						EYKLVVVGAR	A*3101	2.40%	0.600	0.503	
						KLVVVGARGV	A*0201	13.66%	3.633	3.048	
						VVGARGVGK	A*1101	7.07%	3.013	2.527	
						RGVGKSALTI	B*0702	6.71%	1.087	0.912	
						VVVGARGVGK	A*1101	7.07%	7.070	5.931	
							A*6801	3.50%	0.671	0.563	
						LVVVGARGV	A*0203	0.90%	0.138	0.116	
	G12D	Biliary tract	1.14	15.43%	0.18	VVGADGVGK	A*1101	7.07%	2.333	0.410	
						LVVVGADGV	A*0201	13.66%	1.683	0.296	
							A*0203	0.90%	0.127	0.022	
						KLVVVGADGV	A*0201	13.66%	6.713	1.181	
						GADGVGKSAL	B*3801	1.03%	0.680	0.120	
						VVVGADGVGK	A*1101	7.07%	5.272	0.927	
							A*6801	3.50%	0.671	0.118	
		Endometrium	23.49	5.67%	1.33	VVGADGVGK	A*1101	7.07%	2.333	3.108	
						LVVVGADGV	A*0201	13.66%	1.683	2.242	
							A*0203	0.90%	0.127	0.170	
						KLVVVGADGV	A*0201	13.66%	6.713	8.940	
						GADGVGKSAL	B*3801	1.03%	0.680	0.906	
						VVVGADGVGK	A*1101	7.07%	5.272	7.022	
							A*6801	3.50%	0.671	0.893	
		Large intestine	45.94	11.32%	5.20	VVGADGVGK	A*1101	7.07%	2.333	12.134	
						LVVVGADGV	A*0201	13.66%	1.683	8.755	
							A*0203	0.90%	0.127	0.662	
						KLVVVGADGV	A*0201	13.66%	6.713	34.908	
						GADGVGKSAL	B*3801	1.03%	0.680	3.539	
						VVVGADGVGK	A*1101	7.07%	5.272	27.416	
							A*6801	3.50%	0.671	3.488	

Table 1
(*continued*)

Gene	Mutation[a]	Tissue	Cancer incidence[b] (per 100k)	Mutation frequency[c]	MIS[d]	Peptide	Presenting allele	HLA frequency[e]	EIS[f]	GS[g]	OS[h]
		Ovary	12.62	5.47%	0.69	VVGADGVGK	A*1101	7.07%	2.333	1.611	
						LVVVGADGV	A*0201	13.66%	1.683	1.162	
							A*0203	0.90%	0.127	0.088	
						KLVVVGADGV	A*0201	13.66%	6.713	4.634	
						GADGVGKSAL	B*3801	1.03%	0.680	0.470	
						VVVGADGVGK	A*1101	7.07%	5.272	3.639	
							A*6801	3.50%	0.671	0.463	
		Pancreas	11.70	28.57%	3.34	VVGADGVGK	A*1101	7.07%	2.333	7.799	
						LVVVGADGV	A*0201	13.66%	1.683	5.627	
							A*0203	0.90%	0.127	0.426	
						KLVVVGADGV	A*0201	13.66%	6.713	22.438	
						GADGVGKSAL	B*3801	1.03%	0.680	2.275	
						VVVGADGVGK	A*1101	7.07%	5.272	17.622	
							A*6801	3.50%	0.671	2.242	
		Small intestine	1.92	8.33%	0.16	VVGADGVGK	A*1101	7.07%	2.333	0.373	
						LVVVGADGV	A*0201	13.66%	1.683	0.269	
							A*0203	0.90%	0.127	0.020	
						KLVVVGADGV	A*0201	13.66%	6.713	1.074	
						GADGVGKSAL	B*3801	1.03%	0.680	0.109	
						VVVGADGVGK	A*1101	7.07%	5.272	0.843	
							A*6801	3.50%	0.671	0.107	
	G12C	Lung	60.66	7.38%	4.48	ACGVGKSAL	B*0702	6.71%	1.850	8.283	
						VVVGACGVGK	A*1101	7.07%	7.070	31.650	
							A*6801	3.50%	0.671	3.003	
						LVVVGACGV	A*0201	13.66%	2.245	10.049	
							A*0203	0.90%	0.151	0.674	
						VVGACGVGK	A*1101	7.07%	2.333	10.445	
						GACGVGKSAL	B*3801	1.03%	0.221	0.987	
						KLVVVGACGV	A*0201	13.66%	8.102	36.269	
NPM1	W288fs*12	Hematopoietic and lymphoid tissue	46.19	19.03%	8.79	DLCLAVEEV	A*0201	13.66%	1.635	14.368	87.49
						DQEAIQDLCL	B*3801	1.03%	0.118	1.041	
						EAIQDLCLA	A*6801	3.50%	0.250	2.197	
							B*4501	2.35%	0.261	2.295	
						AIQDLCLAV	A*0201	13.66%	4.051	35.607	
						QEAIQDLCL	B*4001	3.26%	0.924	8.121	
							B*4403	5.12%	0.711	6.247	
						QEAIQDLCLA	B*4403	5.12%	2.004	17.615	
HRAS	G12V	Urinary tract	20.27	7.73%	1.57	VVGAVGVGK	A*1101	7.07%	4.218	6.608	49.24
						GAVGVGKSAL	B*0702	6.71%	1.102	1.727	
							B*3801	1.03%	0.250	0.391	
						KLVVVGAVGV	A*0201	13.66%	8.102	12.694	
						AVGVGKSAL	B*0702	6.71%	6.709	10.512	
						VVVGAVGVGK	A*1101	7.07%	7.070	11.078	
							A*6801	3.50%	0.838	1.314	
						YKLVVVGAV	A*0203	0.90%	0.256	0.402	
						LVVVGAVGV	A*0201	13.66%	2.732	4.281	
							A*0203	0.90%	0.151	0.236	
FGFR3	Y373C	Urinary tract	20.27	8.95%	1.81	DEAGSVCAG	B*4403	5.12%	0.188	0.341	42.06
						DEAGSVCAGI	B*4402	2.85%	0.214	0.388	
							B*4403	5.12%	0.557	1.010	
						SVCAGILSY	A*1101	7.07%	0.307	0.557	
							B*1501	2.64%	1.749	3.173	
						CAGILSYGV	B*5101	4.29%	0.144	0.261	
						GSVCAGILSY	A*1101	7.07%	0.264	0.479	
						EAGSVCAGI	B*5101	4.29%	0.300	0.544	
	S249C	Urinary tract	20.27	29.97%	6.07	ERCPHRPIL	B*2705	1.38%	0.064	0.389	
							B*3901	0.87%	0.054	0.326	
						LERCPHRPI	B*4403	5.12%	0.136	0.828	
						TYTLDVLERC	A*2402	9.29%	0.247	1.502	
						DVLERCPHR	A*1101	7.07%	0.224	1.359	
							A*3101	2.40%	0.115	0.702	
							A*6801	3.50%	0.376	2.281	
						CPHRPILQA	B*0702	6.71%	0.311	1.889	
	R248C	Skin	19.79	13.07%	2.59	TYTLDVLECS	A*2402	9.29%	0.247	0.639	
						DVLECSPHR	A*1101	7.07%	0.224	0.579	
							A*3101	2.40%	0.095	0.245	
							A*6801	3.50%	0.312	0.808	
						LECSPHRPI	B*4403	5.12%	0.190	0.492	
	G697C	Upper aerodigestive tract	13.51	18.80%	2.54	CIPVEELFK	A*1101	7.07%	0.408	1.035	
						LGGSPYPCI	B*5101	4.29%	0.163	0.413	
						YPCIPVEELF	A*2402	9.29%	0.266	0.674	
							B*3501	5.82%	0.916	2.327	
						PCIPVEELF	A*2402	9.29%	0.492	1.250	
						TLGGSPYPCI	A*0201	13.66%	0.455	1.156	

Table 1
(*continued*)

Gene	Mutation[a]	Tissue	Cancer incidence[b] (per 100k)	Mutation frequency[c]	MIS[d]	Peptide	Presenting allele	HLA frequency[e]	EIS[f]	GS[g]	OS[h]
						CIPVEELFKL	A*0201	13.66%	0.963	2.446	
							B*3801	1.03%	0.030	0.075	
						PYPCIPVEE	A*2402	9.29%	0.430	1.092	
						PYPCIPVEEL	A*2402	9.29%	3.637	9.238	
						YPCIPVEEL	B*0702	6.71%	0.793	2.013	
							B*3501	5.82%	0.404	1.026	
							B*5101	4.29%	0.206	0.523	
MSH6	P1087fs*5	Large intestine	45.94	6.07%	2.79	LPEDTPPLL	B*0702	6.71%	0.876	2.442	32.21
							B*3501	5.82%	0.260	0.724	
							B*3801	1.03%	0.053	0.148	
							B*5101	4.29%	0.443	1.236	
						ILLPEDTPPL	A*0201	13.66%	4.475	12.479	
						LLPEDTPPLL	A*0201	13.66%	1.631	4.548	
							A*2402	9.29%	0.878	2.449	
							B*3801	1.03%	0.026	0.071	
						LLPEDTPPL	A*0201	13.66%	2.910	8.115	
TP53	R273H	Large intestine	45.94	5.31%	2.44	GRNSFEVHV	B*2705	1.38%	0.108	0.264	26.52
						HVCACPGRDR	A*3101	2.40%	0.209	0.510	
							A*6801	3.50%	0.322	0.785	
						EVHVCACPGR	A*1101	7.07%	0.610	1.489	
							A*6801	3.50%	1.725	4.208	
						GRNSFEVHVC	B*2705	1.38%	0.119	0.291	
	R248Q	Hematopoietic and lymphoid tissue	46.19	7.69%	3.55	QRPILTIITL	A*2402	9.29%	0.587	2.084	
						MGGMNQRPIL	B*5101	4.29%	0.352	1.251	
						SCMGGMNQR	A*3101	2.40%	0.269	0.954	
						MGGMNQRPI	B*5101	4.29%	0.778	2.765	
						SSCMGGMNQR	A*1101	7.07%	0.859	3.052	
							A*6801	3.50%	0.301	1.068	
						GMNQRPILTI	A*0201	13.66%	2.196	7.800	
EGFR	L858R	Lung	60.66	9.60%	5.82	ITDFGRAKLL	B*3801	1.03%	0.054	0.312	21.28
						HVKITDFGR	A*1101	7.07%	0.252	1.470	
							A*3101	2.40%	1.200	6.988	
							A*6801	3.50%	0.084	0.491	
						KITDFGRAK	A*1101	7.07%	0.249	1.449	
						RAKLLGAEEK	A*1101	7.07%	0.447	2.600	
						KITDFGRAKL	A*0201	13.66%	0.594	3.461	
						GRAKLLGAE	B*2705	1.38%	0.026	0.153	
						ITDFGRAKL	B*3801	1.03%	0.062	0.359	
	E746_A750del	Lung	60.66	5.86%	3.55	IPVAIKTSPK	A*1101	7.07%	0.240	0.854	
						PVAIKTSPK	A*1101	7.07%	0.287	1.020	
						AIKTSPKANK	A*1101	7.07%	0.418	1.485	
						TSPKANKEI	A*2402	9.29%	0.180	0.641	
PIK3CA	H1047R	Breast	120.81	12.51%	15.11	FMKQMNDAR	A*3101	2.40%	0.085	1.279	18.84
						YFMKQMNDAR	A*3101	2.40%	0.106	1.603	
						ARHGGWTTK	B*2705	1.38%	0.336	5.073	
						ARHGGWTTKM	B*2705	1.38%	0.037	0.557	
		Endometrium	23.49	5.59%	1.31	FMKQMNDAR	A*3101	2.40%	0.085	0.111	
						YFMKQMNDAR	A*3101	2.40%	0.106	0.139	
						ARHGGWTTK	B*2705	1.38%	0.336	0.441	
						ARHGGWTTKM	B*2705	1.38%	0.037	0.048	
	E545K	Breast	120.81	5.39%	6.51	SEITKQEKD	B*4403	5.12%	0.112	0.731	
						SEITKQEKDF	B*4402	2.85%	0.374	2.436	
							B*4403	5.12%	0.332	2.163	
						LSEITKQEK	A*1101	7.07%	0.191	1.241	
						PLSEITKQEK	A*1101	7.07%	0.148	0.961	
		Cervix	7.98	5.04%	0.40	SEITKQEKD	B*4403	5.12%	0.112	0.045	
						SEITKQEKDF	B*4402	2.85%	0.374	0.150	
							B*4403	5.12%	0.332	0.134	
						LSEITKQEK	A*1101	7.07%	0.191	0.077	
						PLSEITKQEK	A*1101	7.07%	0.148	0.059	
		Urinary tract	20.27	6.79%	1.38	SEITKQEKD	B*4403	5.12%	0.112	0.155	
						SEITKQEKDF	B*4402	2.85%	0.374	0.515	
							B*4403	5.12%	0.332	0.457	
						LSEITKQEK	A*1101	7.07%	0.191	0.262	
						PLSEITKQEK	A*1101	7.07%	0.148	0.203	
ABL1	T315I	Hematopoietic and lymphoid tissue	46.19	5.49%	2.54	FYIIIEFMTY	A*2402	9.29%	0.221	0.561	9.80
						IIIEFMTYG	A*0201	13.66%	0.329	0.835	
						YIIIEFMTY	B*1501	2.64%	0.257	0.652	
							B*3501	5.82%	0.148	0.374	
						IEFMTYGNL	B*4001	3.26%	0.780	1.979	
							B*4403	5.12%	0.121	0.307	
						REPPFYIII	A*2402	9.29%	0.195	0.495	
							B*4001	3.26%	0.202	0.512	
							B*4403	5.12%	0.193	0.490	

Table 1
(*continued*)

Gene	Mutation[a]	Tissue	Cancer incidence[b] (per 100k)	Mutation frequency[c]	MIS[d]	Peptide	Presenting allele	HLA frequency[e]	EIS[f]	GS[g]	OS[h]
						IEFMTYGNLL	B*4402	2.85%	0.412	1.044	
							B*4403	5.12%	0.120	0.305	
						IIEFMTYGNL	B*3801	1.03%	0.030	0.077	
						TREPPFYIII	B*2705	1.38%	0.042	0.107	
						EPPFYIIIEF	A*2402	9.29%	0.371	0.941	
							B*3501	5.82%	0.339	0.859	
							B*3801	1.03%	0.028	0.071	
							B*4402	2.85%	0.073	0.185	
CTNNB1	T41A	Soft tissue	3.03	15.02%	0.46	IHSGATATA	A*0203	0.90%	0.039	0.018	8.36
							B*3801	1.03%	0.084	0.038	
						TATAPSLSGK	A*1101	7.07%	0.326	0.148	
							A*6801	3.50%	0.236	0.108	
						ATAPSLSGK	A*1101	7.07%	3.535	1.609	
							A*3101	2.40%	0.069	0.031	
							A*6801	3.50%	0.493	0.224	
						GATATAPSL	B*5101	4.29%	0.143	0.065	
						GIHSGATATA	A*0203	0.90%	0.066	0.030	
						ATATAPSLSG	A*1101	7.07%	0.213	0.097	
	S45del	Kidney	14.12	5.15%	0.73	SGATTTAPL	B*0702	6.71%	0.257	0.187	
						ATTTAPLSGK	A*1101	7.07%	0.935	0.680	
							A*6801	3.50%	0.200	0.146	
						TTTAPLSGK	A*1101	7.07%	2.621	1.906	
							A*6801	3.50%	0.690	0.502	
						APLSGKGNPE	B*0702	6.71%	0.239	0.174	
						HSGATTTAPL	B*0702	6.71%	0.276	0.201	
	S45F	Soft tissue	3.03	9.86%	0.30	GATTTAPFL	B*5101	4.29%	0.143	0.043	
						TTAPFLSGK	A*1101	7.07%	3.535	1.056	
							A*3101	2.40%	0.100	0.030	
							A*6801	3.50%	0.965	0.288	
						TTTAPFLSGK	A*1101	7.07%	0.713	0.213	
							A*6801	3.50%	0.466	0.139	
						TTTAPFLSG	A*1101	7.07%	0.183	0.055	
						HSGATTTAPF	B*3501	5.82%	0.520	0.155	
	S37A	Small intestine	1.92	7.77%	0.15	SYLDSGIHAG	A*2402	9.29%	0.398	0.059	
						SYLDSGIHA	A*2402	9.29%	0.422	0.063	
						YLDSGIHAGA	A*0203	0.90%	0.115	0.017	
						GIHAGATTTA	A*0203	0.90%	0.038	0.006	
						AGATTTAPSL	B*0702	6.71%	0.362	0.054	
						IHAGATTTA	A*0203	0.90%	0.039	0.006	
							B*3801	1.03%	0.084	0.013	
KIT	D816V	Hematopoietic and lymphoid tissue	46.19	9.59%	4.43	RVIKNDSNY	B*1501	2.64%	0.372	1.648	8.35
						KICDFGLARV	A*0201	13.66%	1.392	6.165	
							A*3101	2.40%	0.058	0.257	
						ARVIKNDSN	B*2705	1.38%	0.062	0.275	
HNF1A	P291fs*51	Large intestine	45.94	10.94%	5.03	PPQGQARDL	B*0702	6.71%	1.072	5.386	7.97
						GPPQGQARDL	B*5101	4.29%	0.514	2.581	
JAK2	V617F	Hematopoietic and lymphoid tissue	46.19	45.12%	20.84	CFCGDENIL	A*2402	9.29%	0.189	3.936	5.96
						LVLNYGVCF	B*1501	2.64%	0.097	2.022	
BRAF	V600E	Biliary tract	1.14	10.06%	0.11	GDFGLATEK	A*1101	7.07%	0.286	0.033	5.16
		Eye	0.77	11.35%	0.09	GDFGLATEK	A*1101	7.07%	0.286	0.025	
		Large intestine	45.94	12.78%	5.87	GDFGLATEK	A*1101	7.07%	0.286	1.681	
		Ovary	12.62	8.70%	1.10	GDFGLATEK	A*1101	7.07%	0.286	0.314	
		Skin	19.79	33.51%	6.63	GDFGLATEK	A*1101	7.07%	0.286	1.899	
		Thyroid	10.60	39.71%	4.21	GDFGLATEK	A*1101	7.07%	0.286	1.205	
IDH1	R132H	Central nervous system	6.09	30.77%	1.87	KPIIIGHHA	B*0702	6.71%	1.025	1.921	4.24
						KPIIIGHHAY	B*3501	5.82%	0.903	1.693	
						PIIIGHHAY	B*1501	2.64%	0.335	0.628	
NRAS	Q61R	Skin	19.79	7.12%	1.41	LLDILDTAGR	A*6801	3.50%	0.527	0.743	2.72
						GREEYSAMR	B*2705	1.38%	0.448	0.631	
NRAS	Q61K	Skin	19.79	6.84%	1.35	LLDILDTAGK	A*1101	7.07%	0.996	1.348	
RET	M918T	Thyroid	10.60	28.38%	3.01	VKWTAIESL	B*3902	0.11%	0.042	0.126	2.51
						VKWTAIESLF	A*2402	9.29%	0.403	1.213	
						GRIPVKWTA	B*2705	1.38%	0.036	0.108	
						GRIPVKWTAI	B*2705	1.38%	0.050	0.151	
						KWTAIESLF	A*2402	9.29%	0.304	0.915	
PDGFRA	D842V	Soft tissue	3.03	7.00%	0.21	ARVIMHDSN	B*2705	1.38%	0.060	0.013	0.65
						RVIMHDSNY	A*1101	7.07%	0.144	0.031	
							B*1501	2.64%	0.575	0.122	
						KICDFGLARV	A*0201	13.66%	1.212	0.257	
						VIMHDSNYV	A*0201	13.66%	0.500	0.106	
						RVIMHDSNYV	A*0201	13.66%	0.553	0.117	
MET	Y1253D	Upper aerodigestive tract	13.51	8.81%	1.19	DMYDKEYDSV	A*0201	13.66%	0.297	0.353	0.58
						KEYDSVHNK	A*1101	7.07%	0.156	0.186	
							B*2705	1.38%	0.032	0.039	

Table 1
(*continued*)

Gene	Mutation[a]	Tissue	Cancer incidence[b] (per 100k)	Mutation frequency[c]	MIS[d]	Peptide	Presenting allele	HLA frequency[e]	EIS[f]	GS[g]	OS[h]
APC	T1556fs*3	Pancreas	11.70	6.54%	0.77	QEKEAEKNY	B*4403	5.12%	0.075	0.058	0.14
		Small intestine	1.92	5.16%	0.10	QEKEAEKNY	B*4403	5.12%	0.075	0.007	
		Stomach	7.48	5.84%	0.44	QEKEAEKNY	B*4403	5.12%	0.075	0.033	
	S1341R	Soft tissue	3.03	5.25%	0.16	SRLSSESAR	B*2705	1.38%	0.030	0.005	
						SRLQGSRLS	B*2705	1.38%	0.012	0.002	
						SSRLQGSRL	B*0702	6.71%	0.045	0.007	
						SRLSSESARH	B*2705	1.38%	0.014	0.002	
						RLSSESARHK	A*1101	7.07%	0.049	0.008	
						KSSRLQGSR	A*1101	7.07%	0.051	0.008	
						SRLQGSRLSS	B*2705	1.38%	0.009	0.001	
						KSSRLQGSRL	B*0702	6.71%	0.053	0.008	

[a]HGVS notations for describing mutations is a system produced by the Human Genome Variation Society (http://www.hgvs.org/).
[b]Cancer incidence for US population from http://seer.cancer.gov.
[c]Mutation frequency calculated from the Catalogue of Somatic Mutations in Cancer (COSMIC; Forbes *et al.* 2008).
[d]MIS (Mutation Impact Score) = Frequency of mutation in given cancer × cancer incidence.
[e]Average HLA allele frequency calculated from www.allelefrequencies.net for all US ethnic groups with $n >= 500$.
[f]EIS (Epitope Impact Score) = Epitope prediction score × average frequency of presenting HLA allele.
[g]GS (Global Score) = MIS × EIS.
[h]OS (Overall Score) = Σ GS.

score is beneficial, because epitopes from distributed mutations are likely to be inferior to those derived from hotspot mutations. This is because unlike hotspot mutations, which are typically early occurring, gain of function mutations essential for tumor viability, distributed mutations are usually loss of function mutations that arise later in tumorigenesis. Loss of function mutations are undesirable immunologic targets because genes mutated in this manner become nonessential to the cell, and immune evasion can result from downregulation of mutant epitope presentation. Mutations incurred late in the tumorigenic process are generally undesirable immunologic targets for the added reason that well-developed tumors are immunoresistant, and the opportunity for prevention has passed.

To be useful as a vaccine epitope, a mutation must be immunogenic. That is, it must be contained within a peptide that is presented at the cell surface by HLA class I molecules and, ultimately, be sufficiently distinct from the wild-type version of the same peptide to cause a T-cell response. Here, we used computational epitope prediction methods to identify common tumor mutations that may also be immunogenic. It is important to note that this approach predicts only peptide-HLA binding, which is taken as a correlate of immunogenicity. Whether a given peptide is truly immunogenic also depends on it being appropriately processed and loaded on HLA molecules, and being sufficiently distinct from the presumably tolerated, nonmutated version. In the present study, we estimate the proportion of tumor mutations that could be immunogenic, and should be prioritized for further investigation. At present, the cost and throughput of laboratory immunoassays are prohibitive for screening large numbers of candidate mutations and a computational approach that shortlists candidates is warranted. For *in silico* epitope prediction tools, false-positive predictions are known to occur [39,41]. To curb this problem, only peptides predicted by two or more tools and scoring within the top ranking 5% are considered here.

Taking advantage of three well-established *in silico* epitope prediction programs, we queried in bulk every mutation or frameshift-peptide from candidate cancer genes identified, as described above, using the COSMIC database. Again, only peptides ranking within the top 5% of all possible peptide degradation products from a given gene, and predicted independently by at least two of the three programs were retained for analysis. These predictions, in conjunction with HLA class I allele frequencies for the US population (http://www.allelefrequencies.org) were used to produce an EIS for each mutant peptide (see Methods).

We screened all possible tiled peptides derived from the 20 genes and 36 mutations identified against all HLA class I variants represented by the three *in silico* epitope prediction programs, for 54,432 individual queries. For each gene, and 35 of the 36 mutations, we obtained one or more epitopes that passed filtering. There were 229 total and 159 unique peptides predicted to be presented by up to four distinct HLA class I alleles. Interestingly, of the 10 peptides with the highest EIS, seven are predicted to arise from the KRAS oncogene. Six candidates are position G12 mutants (two G12V, two G12C, one G12R, and one G12D) and one from G13D. Most KRAS mutational peptides were predicted to be high-ranking HLA binders by some of the most frequently represented alleles in the US population, which is an important factor in evaluating their EIS.

Finally, we ranked the 20 cancer genes according to an OS, which took into account all of our criteria, including the frequency of mutations within each gene in specific tumors, the prevalence of those tumors, the likelihood of mutations to be presented as HLA class I–restricted epitopes, and the population frequency of HLA alleles. The top-ranking gene was KRAS (OS =506.52), followed by NPM1 (OS =87.49) and HRAS (OS =49.24). The lowest ranked of the 20 genes was APC, with an OS of just 0.14. KRAS figures prominently because there are frequent hotspot mutations in numerous, prevalent cancers, and mutant peptides are predicted to be presented by several common HLA alleles (Table 1). Relative to KRAS, APC mutations occur infrequently in rare cancers, and are predicted to be presented weakly by rare HLA alleles. Our data illustrate other interesting features of predicted tumor epitopes. For example, we observe that a single peptide, KICDFGLARV, is shared by PDGFRA D842V and KIT D816V and, therefore, constitutes an attractive candidate epitope for simultaneous targeting of hematopoietic/lymphoid (KIT) and soft-tissue tumors (PDGFRA).

4. Discussion

The predicted epitopes presented here are almost certainly an underestimate of those that actually occur in cancer. There are several reasons for this. First, we have applied stringent filters for mutation frequency and epitope prediction scores that may have excluded many real epitopes. Second, our results are limited by the fact that epitope prediction tools support binding models for only a fraction of several hundred known HLA class I coding variants. For instance, BIMAS, SYFPEITHI, and IEDB support 18, 22, and 41 HLA class I coding variants, respectively. Only nine HLA class I coding variants are shared by at least two of the three epitope prediction tools and only four variants are shared by all three. Considering the

19 alleles supported by at least two programs, and frequencies of these alleles in the US population, we calculate that approximately 55% of the total US population would have at least one of these alleles, and our study is essentially blind to the remaining 45%. This is an inherent limitation of existing epitope prediction tools, and affects all epitope prediction studies. Finally, it is important to consider that mutation detection strategies have, to date, relied heavily on capillary sequencing of polymerase chain reaction–amplified coding sequence. This is a relatively insensitive approach where mutations can be obscured by coamplification of wild-type sequence from stromal cells. Single-molecule sequencing methods produce a digital readout of allele sequences and promise greater sensitivity [42]. Thus, with time, the number of candidate tumor mutational epitopes will increase. Mutations now thought to be rare may be shown to be present at more significant frequencies and new mutations are likely to be discovered by ongoing cancer genome screening efforts.

While the concept of prophylactic cancer vaccination has been discussed for some time, there have been no previous attempts to estimate the potential utility of such vaccines, in terms of the proportion of individuals that may benefit. As discussed above, the key factors to consider are tumor incidence, mutation frequency, and HLA restriction. HLA restriction rarely receives adequate attention, but is a critical issue in vaccine development, given the extensive allelic variation at this locus and the fact that efficient presentation of a given epitope is limited to a subset of HLA alleles. Here, for demonstration, we have estimated the proportion of the US population that could benefit from hypothetical multivalent vaccines targeting the relatively common colorectal and hematological/lymphoid cancers. Specifically, we determined the proportion of the US population that carries a presenting HLA allele for at least one of the mutant peptides associated with the tumor in question (Table 1). By this approach we estimate that hypothetically, 16.8% of those who would otherwise develop a hematopoietic or lymphoid cancer could be protected by vaccination with mutational epitopes predicted here, and of those who would otherwise develop colorectal cancer, 11.5% could be protected. Interestingly, given the prevalence of KRAS mutations in both pancreatic and colorectal cancer, a multivalent vaccine targeting colorectal cancer could also protect 12% of individuals who would otherwise develop pancreatic cancer.

The results presented here are purely for demonstration, but suggest meaningful levels of cancer prophylaxis could be achievable by vaccination. The results are limited by the fact that that only about half of the US population is estimated to carry at least one of the alleles considered by the epitope prediction tools employed. This taken with our incomplete knowledge of recurrent tumor mutations means that we are likely underestimating significantly the population reach of hypothetical vaccine constructs. Our projections are also based on the premise that predicted epitopes are immunogenic, and that immunogenic responses would be protective, both of which require experimental validation. There is little certainty as to which predicted epitopes correspond to high-value early driver mutations, and some may be from later onset mutations that occur in established tumors, and therefore have limited utility for immunoprevention. Finally it is, of course, not possible to predict who will or will not develop a spontaneous tumor; thus, our estimates are based on vaccination of the entire population. In principle, one could restrict vaccination to individuals carrying HLA alleles known to present epitopes contained in a vaccine construct, but the extra efforts required for HLA typing would detract from the utility of population cancer control. A more feasible strategy for targeted vaccination is in association with a cancer screening program, for example, colonoscopy for detection of colorectal tumors. Individuals with benign adenomatous polyps, identified by colonoscopy, are at risk of progression to adenocarcinoma. Progression from adenoma to adenocarcinoma is typically mediated by mutations in KRAS, BRAF, and other oncogenes, as well as loss of function mutations in TP53 and other tumor suppressors [43]. Thus, a progression blocking vaccine based on KRAS and related epitopes could be an effective preventative strategy for this tumor site.

Acknowledgments

The mutation data was obtained from the Sanger Institute Catalogue of Somatic Mutations in Cancer ftp site, ftp://ftp.sanger.ac.uk/pub/CGP/cosmic/. This work was funded by Genome Canada, Genome British Columbia, and the Canadian Institutes of Health Research. R.A.H. is a Michael Smith Foundation for Health Research Scholar.

References

[1] Ehrlich P. Experimentelle Karzinom-studien on Mausen. Arch Inst Exp Ther 1906;1:65.
[2] Bashford E, Murray J, Haaland M, Bowen W. General results of propagation of malignant new growths. In Bashford E, ed. Third Scientific Report on the Investigations of the Imperial Cancer Research Fund. London: Taylor and Francis; 1908.
[3] Prehn RT, Main JM. Immunity to methylcholanthrene-induced sarcomas. J Natl Cancer Inst 1957;18:769–78.
[4] Burnet FM. The concept of immunological surveillance. Prog Exp Tumor Res 1970;13:1–27.
[5] Shankaran V, Ikeda H, Bruce AT, White JM, Swanson PE, Old LJ, et al. IFNgamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity. Nature 2001;410:1107–11.
[6] Koebel CM, Vermi W, Swann JB, Zerafa N, Rodig SJ, Old LJ, et al. Adaptive immunity maintains occult cancer in an equilibrium state. Nature 2007;450:903–7.
[7] Dunn GP, Old LJ, Schreiber RD. The three Es of cancer immunoediting. Annu Rev Immunol 2004;22:329–60.
[8] Willimsky G, Blankenstein T. Sporadic immunogenic tumours avoid destruction by inducing T-cell tolerance. Nature 2005;437:141–46.
[9] Cefai D, Morrison BW, Sckell A, Favre L, Balli M, Leunig M, et al. Targeting HER-2/neu for active-specific immunotherapy in a mouse model of spontaneous breast cancer. Int J Cancer 1999;83:393–400.
[10] Dakappagari NK, Douglas DB, Triozzi PL, Stevens VC, Kaumaya PT. Prevention of mammary tumors with a chimeric HER-2 B-cell epitope peptide vaccine. Cancer Res 2000;60:3782–89.
[11] Nanni P, Nicoletti G, De Giovanni C, Landuzzi L, Di Carlo E, Cavallo F, et al. Combined allogeneic tumor cell vaccination and systemic interleukin 12 prevents mammary carcinogenesis in HER-2/neu transgenic mice. J Exp Med 2001;194:1195–205.
[12] Park JM, Terabe M, Sakai Y, Munasinghe J, Forni G, Morris JC, et al. Early role of $CD4^+$ Th1 cells and antibodies in HER-2 adenovirus vaccine protection against autochthonous mammary carcinomas. J Immunol 2005;174:4228–36.
[13] Villa LL, Costa RL, Petta CA, Andrade RP, Paavonen J, Iversen OE, et al. High sustained efficacy of a prophylactic quadrivalent human papillomavirus types 6/11/16/18 L1 virus-like particle vaccine through 5 years of follow-up. Br J Cancer 2006;95:1459–66.
[14] Penn I. Malignant melanoma in organ allograft recipients. Transplantation 1996;61:274–78.
[15] Suranyi MG, Hogan PG, Falk MC, Axelsen RA, Rigby R, Hawley C, et al. Advanced donor-origin melanoma in a renal transplant recipient: Immunotherapy, cure, and retransplantation. Transplantation 1998;66:655–61.
[16] MacKie RM, Reid R, Junor B. Fatal melanoma transferred in a donated kidney 16 years after melanoma surgery. N Engl J Med 2003;348:567–68.
[17] Birkeland SA, Storm HH, Lamm LU, Barlow L, Blohme I, Forsberg B, et al. Cancer risk after renal transplantation in the Nordic countries, 1964–1986. Int J Cancer 1995;60:183–89.
[18] Mihm MC, Clemente CG, Cascinelli N. Tumor infiltrating lymphocytes in lymph node melanoma metastases: A histopathologic prognostic indicator and an expression of local immune response. Lab Invest 1996;74:43–7.
[19] Clemente CG, Mihm MC, Bufalino R, Zurrida S, Collini P, Cascinelli N. Prognostic value of tumor infiltrating lymphocytes in the vertical growth phase of primary cutaneous melanoma. Cancer 1996;77:1303–10.
[20] Zhang L, Conejo-Garcia JR, Katsaros D, Gimotty PA, Massobrio M, Regnani G, et al. Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. N Engl J Med 2003;348:203–13.
[21] Reuschenbach M, von Knebel Doeberitz M, Wentzensen N. A systematic review of humoral immune responses against tumor antigens. Cancer Immunol Immunother 2009;58:1535–44.
[22] Wang HY, Zhou J, Zhu K, Riker AI, Marincola FM, Wang RF. Identification of a mutated fibronectin as a tumor antigen recognized by CD4+ T cells: Its role in extracellular matrix formation and tumor metastasis. J Exp Med 2002;195:1397–406.
[23] Gaudin C, Kremer F, Angevin E, Scott V, Triebel F. A hsp70-2 mutation recognized by CTL on a human renal cell carcinoma. J Immunol 1999;162:1730–8.

[24] Huang J, El-Gamil M, Dudley ME, Li YF, Rosenberg SA, Robbins PF. T cells associated with tumor regression recognize frameshifted products of the CDKN2A tumor suppressor gene locus and a mutated HLA class I gene product. J Immunol 2004;172:6057–64.
[25] Linard B, Bezieau S, Benlalam H, Labarriere N, Guilloux Y, Diez E, et al. A ras-mutated peptide targeted by CTL infiltrating a human melanoma lesion. J Immunol 2002;168:4802–8.
[26] Robbins PF, El-Gamil M, Li YF, Kawakami Y, Loftus D, Appella E, et al. A mutated beta-catenin gene encodes a melanoma-specific antigen recognized by tumor infiltrating lymphocytes. J Exp Med 1996;183:1185–92.
[27] Kawakami Y, Wang X, Shofuda T, Sumimoto H, Tupesis J, Fitzgerald E, et al. Isolation of a new melanoma antigen, MART-2, containing a mutated epitope recognized by autologous tumor-infiltrating T lymphocytes. J Immunol 2001; 166:2871–7.
[28] Lollini PL, Cavallo F, Nanni P, Forni G. Vaccines for tumour prevention. Nat Rev Cancer 2006;6:204–16.
[29] Gjertsen MK, Bjorheim J, Saeterdal I, Myklebust J, Gaudernack G. Cytotoxic CD4+ and CD8+ T lymphocytes, generated by mutant p21-ras (12Val) peptide vaccination of a patient, recognize 12Val-dependent nested epitopes present within the vaccine peptide and kill autologous tumour cells carrying this mutation. Int J Cancer 1997;72:784–90.
[30] Ito D, Visus C, Hoffmann TK, Balz V, Bier H, Appella E, et al. Immunological characterization of missense mutations occurring within cytotoxic T cell-defined p53 epitopes in HLA-A*0201+ squamous cell carcinomas of the head and neck. Int J Cancer 2007;120:2618–24.
[31] Messaoudi I, Guevara Patino JA, Dyall R, LeMaoult J, Nikolich-Zugich J. Direct link between *mhc* polymorphism, T cell avidity, and diversity in immune defense. Science 2002;298:1797–800.
[32] Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer 2004;91:355–58.
[33] Forbes S, Clements J, Dawson E, Bamford S, Webb T, Dogan A, et al. COSMIC 2005. Br J Cancer 2006;94:318–22.
[34] Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, et al. The Catalogue of Somatic Mutations in Cancer (COSMIC). Curr Protoc Hum Genet 2008;Chapter 10:Unit 10.11.
[35] Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI: Database for MHC ligands and peptide motifs. Immunogenetics 1999;50:213–19.
[36] Parker KC, Bednarek MA, Coligan JE. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J Immunol 1994;152:163–75.
[37] Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, et al. The design and implementation of the immune epitope database and analysis resource. Immunogenetics 2005;57:326–36.
[38] Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinformatics 2005;6:132.
[39] Kessler JH, Melief CJ. Identification of T-cell epitopes for cancer immunotherapy. Leukemia 2007;21:1859–74.
[40] Gulukota K, DeLisi C. HLA allele selection for designing peptide vaccines. Genet Anal 1996;13:81–6.
[41] Kessler JH, Mommaas B, Mutis T, Huijbers I, Vissers D, Benckhuijsen WE. Competition-based cellular peptide binding assays for 13 prevalent HLA class I alleles using fluorescein-labeled synthetic peptides. Hum Immunol 2003;64: 245–55.
[42] Pushkarev D, Neff NF, Quake SR. Single-molecule sequencing of an individual human genome. Nat Biotechnol 2009;27:847–52.
[43] Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic alterations during colorectal-tumor development. N Engl J Med 1988; 319:525–32.