Epitope Landscape in Breast and Colorectal Cancer

Neil H. Segal,[1,2] D. Williams Parsons,[4] Karl S. Peggs,[2,3] Victor Velculescu,[4] Ken W. Kinzler,[4] Bert Vogelstein,[4] and James P. Allison[2,3]

[1]Department of Medicine, [2]Ludwig Center for Cancer Immunotherapy, and [3]Immunology Program, Memorial Sloan-Kettering Cancer Center, New York, New York; and [4]Ludwig Center for Cancer Genetics and Therapeutics at The Johns Hopkins Kimmel Cancer Center, Baltimore, Maryland

Abstract

The finding that individual cancers contain many mutant genes not present in normal tissues has prompted considerable interest in the cancer epitope landscape. To further understand such effects, we applied *in silico*–based epitope prediction algorithms and high throughput post hoc analysis to identify candidate tumor antigens. Analysis of 1,152 peptides containing missense mutations previously identified in breast and colorectal cancer revealed that individual cancers accumulate on average ~10 and ~7 novel and unique HLA-A*0201 epitopes, respectively, including genes implicated in the neoplastic process. These data suggest that, with appropriate manipulation of the immune system, tumor cell destruction *in situ* may provide a polyvalent tumor vaccine without a requirement for knowledge of the targeted antigens. [Cancer Res 2008;68(3):889–92]

Introduction

Several classes of tumor antigens have been described and named according to their distribution in normal and neoplastic tissues. These include the shared differentiation antigens, such as melan-A/MART1 (1, 2) in melanoma; cancer testes or germ cell antigens, such as MAGE-1 (3) and NY-ESO-1 (4) in adult testes and diverse tumor types; and unique tumor antigens, which generally carry mutations, such as CDK4 in melanoma (5) and CASP-8 in head and neck cancer (6).

The immunogenicity of the unique tumor antigens was recognized in several seminal studies including animal transplant models (7) and chemically or UV light–induced tumors (8, 9). They are of particular interest because they result from somatic mutations in individual tumors and are absent from normal tissues (10, 11), providing antitumor specificity without anticipated deleterious autoimmunity. Somatic mutations can be classified as either "drivers" or "passengers". Passenger mutations provide no positive or negative selective advantage to the tumor but are retained by chance during cell division and clonal expansion. In contrast, driver mutations provide a selective advantage that promotes the tumorigenic process. The generation of mutations is continuous due to the imperfect nature of DNA replication and repair. Thus, the generation of additional antigens during tumor progression, whether driver or passenger (12), provides a continuously renewable source of antigen.

Recent analyses of breast and colorectal cancers showed a remarkable number of somatic mutations in human cancer (13). Among >13,000 genes analyzed, a total of 1,307 somatic mutations were identified in 11 breast and 11 colorectal cancers. Approximately 83% were missense mutations, 6% nonsense, and the remainder were insertions, deletions, duplications, and changes in noncoding regions. When extrapolated to the whole genome, it was calculated that individual tumors harbored an average of ~90 amino acid–altering (i.e., nonsynonymous) mutations. The kind of information available from such large-scale sequencing studies of individual tumors has not heretofore been available but clearly has implications for tumor immunity.

In the current study, we designed an *in silico* approach to examine whether the mutations identified in Sjoblom et al. (13) have the potential to generate novel epitopes that might serve as targets for an immune response. Using epitope prediction algorithms and high throughput post hoc analysis, we found evidence to support the notion that the human tumorigenic process results in the generation of multiple immune targets. Individual breast and colorectal cancers accumulated an average of ~10 and ~7 novel and unique HLA-A*0201 epitopes, respectively; several within genes that may be drivers. These results provide insights into the unique immune profiles of individual tumors with potential clinical relevance.

Materials and Methods

Epitope prediction. Peptide sequences corresponded to missense mutations identified during the discovery phase by Sjoblom et al. (13), flanked by up to 10 amino acids on either side. Concatamers of these peptides were analyzed with several epitope prediction algorithms for HLA-A*0201 binders. Major histocompatibility complex (MHC)-I antigenic peptide processing prediction (MAPPP; ref 14), developed at the Max-Plank Institute, facilitates the prediction of epitopes that can bind to MHC class I molecules based on a score calculated for each subsequence. Each amino acid at a specific position within a subsequence is given a value that has been precalculated and stored in static matrices. The precalculation was done either by BIMAS (15) or SYFPEITHI (16). Depending on the algorithm selected, the values were then multiplied (BIMAS) or added (SYFPEITHI) to determine the score for the subsequence. Peptides qualified as positive if they scored $\geq$100 and $\geq$24, respectively (17, 18). RANKPEP (19) uses specific scoring matrices from sets of peptides known to bind to MHC molecules as the predictor of MHC-peptide binding. Peptides qualified as positive if the percentage optimum was $\geq$50% or higher. NetMHC (20, 21) predicts peptide-MHC binding using artificial neural networks (ANN) and weight matrices. For ANN, used for HLA-A*0201 prediction, peptides scored positive if IC_{50} is $\leq$500.

Post hoc analysis. First, we searched for unique epitopes within concatamers of wild-type and mutant peptides. Epitopes identified in the "wild-type concatamer" included both true wild-type epitopes and artifacts across the concatenation sites and were removed from further analysis. The "mutant concatamer" was then used to search for remaining epitopes. To ensure that potential mutant epitopes did not span concatenation sites, the "mutant concatamer" used in this confirmatory phase included additional redundant characters spaced between peptides, thereby permitting confirmation of epitopes contained entirely within a mutant

Requests for reprints: James P. Allison, Ludwig Center of Cancer Immunotherapy, Memorial Sloan-Kettering Cancer Center, 415 East 68th Street, Z-1560, New York, NY 10065. E-mail: allisonj@mskcc.org.

©2008 American Association for Cancer Research.

doi:10.1158/0008-5472.CAN-07-3095

Downloaded from http://aacrjournals.org/cancerres/article-pdf/68/3/889/2599635/889.pdf by guest on 08 August 2024

Table 1. HLA-A*0201 epitopes (SYFPEITHI)

Total epitopes	241			
CRC: ≥1 epitope	11			
BC: ≥1 epitope	11			
	Ave	SD	Min	Max
Per sample: CRC	8.2	4.7	2	17
: BC	13.6	6.7	8	30
: CRC + BC	10.9	6.3	2	30
Per peptide: CRC	0.11	0.36	0	2
: BC	0.15	0.42	0	3
: CRC+BC	0.13	0.40	0	3

Abbreviations: BC, breast cancers; CRC, colorectal cancers.

peptide only. These were then annotated for specimen and tumor type as described in supplementary information in Sjoblom et al. (13).

Estimates of epitope frequency. The total number of epitopes corresponding to each peptide, original specimen, and tumor type were calculated for the 11,721 genes that were successfully sequenced of the 13,023 CCDS genes. As per Sjoblom et al. (13), we extrapolated this number to the total number of genes in the human genome (conservatively estimated at 18,203) by dividing the number of identified epitopes by 0.64.

Results and Discussion

A total of 1,152 peptides containing missense mutations, previously identified in breast and colorectal cancer (13) were concatenated into a single string of 23,924 amino acids and analyzed for potential MHC class I binders. Epitopes were predicted using several algorithms then applied to post hoc analysis. This analysis entailed a series of manual steps, predefined calculations, and macro-based algorithms to identify epitopes that are absent from corresponding wild-type sequences but present within the mutant peptide. We restricted our analysis to HLA-A*0201 9-mer epitopes because this haplotype has been extensively studied and is represented in up to 27% of the population (22). All epitopes were confirmed to be unique 9-mers after BLAST analysis, searching for "short, nearly exact matches" in the "nr" database. Two hundred and forty-one epitopes were identified using MAPPP (14)/SYPEITHI (16). On average, 8.2 and 13.6 epitopes were found per specimen in colorectal and breast cancers, respectively (Table 1). Each tumor contained a minimum of two and as many as 30 epitopes in a case of breast cancer (Fig. 1).

Next, we focused our attention on the 191 candidate cancer (*CAN*) genes. *CAN* genes were identified as containing mutations in at least two independent tumors and were mutated at greater frequency than non-*CAN* genes when adjusted for size, nucleotide composition, and mutational spectra (13). We identified 47 potential epitopes in *CAN* genes (Table 2). These epitopes were identified in 6 of 11 colorectal cancers and 7 of 11 breast cancers.

Additional algorithms including NetMHC (20, 21), MAPPP (14)/BIMAS (15), and RANKPEP (19), were then used, identifying an average of 182 epitopes. The majority of epitopes predicted by each of the different algorithms were identified by at least two algorithms. Epitopes predicted by SYFPEITHI, BIMAS, RANKPEP, and NetMHC overlapped by 59%, 73%, 64%, and 75%, respectively.

In sum, we found that individual colorectal and breast cancers accumulated an average of ~7 and ~10 novel and unique HLA-A*0201 epitopes, respectively. Approximately one new epitope was generated for every 10 mutations, and 45% of predicted epitopes were shown to be cleaved at their COOH terminal according to the PAProC proteasome prediction algorithm (refs. 23, 24; Table 3). Note that these numbers are underestimates because other MHC molecules, not studied here, can present additional mutant peptides depending on the haplotype of individual patients. Because individual tumors potentially contain six distinct MHC class I molecules, including two loci each for HLA-A, HLA-B, and HLA-C, the estimated frequency of novel epitopes may be multiplied up to 6-fold. Thereby estimating that individual colorectal and breast cancers potentially accumulate up to ~40 and ~60 novel MHC class I restricted epitopes, respectively.

On a more cautionary note, we have not yet shown that these epitopes are actually expressed in tumor cells; we have studied only

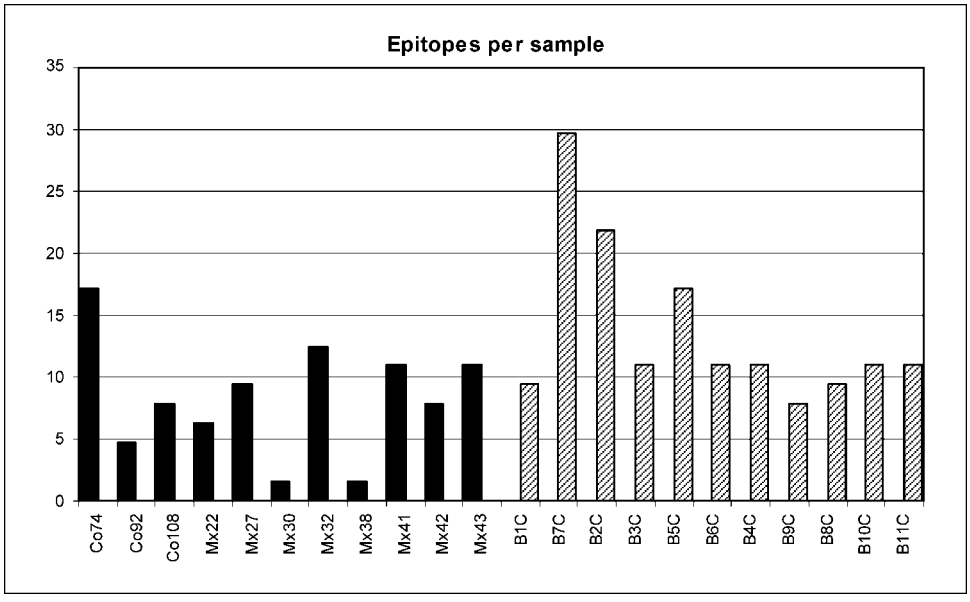

Figure 1. Distribution of HLA-A*0201 epitopes by sample. *Solid black bars,* colorectal cancer samples; *lined bars,* breast cancer samples.

Downloaded from http://aacrjournals.org/cancerres/article-pdf/68/3/889/2599635/889.pdf by guest on 08 August 2024

Table 2. HLA-A*0201 epitopes in CAN genes (SYFPEITHI)

Total epitopes	47			
CRC: ≥1 epitope	6			
BC: ≥1 epitope	7			
	Ave	SD	Min	Max
Per sample: CRC	1.4	1.5	0	3
: BC	2.8	3.6	0	11
: CRC+BC	2.1	2.8	0	11

genomic DNA, not RNA or protein. Nor have we shown that, if expressed, the protein is processed and presented by the antigen presentation machinery. It remains possible that the potential antigen may be destroyed by the proteasome, or the tumor cells may have lost other functions required for epitope presentation. It is also possible that the epitopes are recognized by the immune system as foreign and tumor cells that express them are eliminated, yielding a population of cells in which the genes containing the epitopes are silenced either randomly or after selection for evasion of immune responses (25). These issues will have to be tested experimentally.

However, even if the results of the algorithm are incorrect 90% of the time and only 1 of every 10 neoepitopes is appropriately presented to the immune system, each colorectal and breast cancer would still accumulate ~4 and ~6 immune targets, respectively. Given the sheer number of predicted epitopes, we believe that *in silico* epitope prediction provides meaningful insight into the unique antigen repertoire that accumulates by random mutation coupled with rounds of clonal expansion. Because the predicted and potential immune response is patient-specific and directed toward nonself, these mutations are not expected to have triggered tolerance and, in theory, have the capacity to result in potent tumor rejection antigens.

Approximately half the cancers contain novel epitopes in a subset of mutant genes, which may contribute to the neoplastic process, designated *CAN* genes (13). Most of the mutations occurring in cancers are likely to be passengers, providing no selective advantage or disadvantage. One of the critical challenges in cancer genomics is the distinction between such passengers and drivers. The *CAN* genes are thought to be the best candidates for drivers among those studied. Mutations in drivers may be advantageous in that the tumor might not be able to "down-regulate" expression of the epitope if it is required for continued neoplastic growth (addicted). However, the exploitation of mutations in cancer immunotherapeutics is not dependent on the mutation being a driver. As long as the epitope is presented to the immune system appropriately, it can, in theory, stimulate an antitumor response. All of the multiple nonsynonymous mutations in cancers therefore potentially contribute to multiple targets for immune attack, providing the basis for the concept of combinatorial targeted immunotherapy.

Several methods for exploiting these findings can be devised. Initially, they depend on sequencing each patient's cancer to some extent. Although this is too difficult and expensive to be done routinely at present, improvements in technology will soon make such sequencing feasible, at least for a limited number of genes most likely to harbor missense mutations. The simplest approach to exploit the resultant sequencing information would be to administer vaccines to individual patients made from their own tumor. Although this strategy has been attempted before in humans, it has often been difficult to judge the magnitude of the immune response elicited because the antigens, if any, were not known. Once potential antigens are known through sequencing, patients' responses can be quantitatively assessed, and vaccines that are more rational than the present ones will be possibly designed.

A second way to exploit the epitopes defined by cancer genome sequencing would be through the administration of particular epitopes in the context of immune stimulants. One can imagine a cocktail of synthesized peptide epitopes presented to patients together with adjuvants or the patients' own antigen-presenting cells. As with whole cell vaccines, the ability to evaluate the resultant immune response would provide pivotal data for optimization of such strategies.

A third, and arguably more accessible, way to exploit this may be emphasized during the course of standard antitumor therapy, including chemotherapy, radiation, hormonal therapy, or treatment with the newer class of targeted biological agents, such as Imatinib or Erlotinib. In this manner, tumor cell destruction *in situ* can potentially provide a polyvalent tumor vaccine to the host immune system, without an absolute requirement for knowledge of the targeted antigens. Amplification of these responses by interference with immune regulatory circuits, such

Downloaded from http://aacrjournals.org/cancerres/article-pdf/68/3/889/2599635/889.pdf by guest on 08 August 2024

Table 3. HLA-A*0201 epitopes identified by algorithm

	SYFPEITHI	NetMHC	BIMAS	RANKPEP	Ave
Total epitopes	241	169	128	189	182
Epitopes in *CAN* genes	47	30	20	33	32
CRC: ≥1 epitope	11	11	11	10	11
BC: ≥1 epitope	11	11	11	11	11
Per sample: CRC	8.2	6.7	5.0	6.5	6.6
: BC	13.6	8.7	6.7	10.7	9.9
: CRC+BC	10.9	7.7	5.8	8.6	8.3
Per peptide: CRC	0.11	0.09	0.07	0.09	0.09
: BC	0.15	0.10	0.07	0.12	0.11
: CRC+BC	0.13	0.09	0.07	0.11	0.10
Cleaved at COOH terminal	44%	38%	48%	50%	45%

as CTLA-4 blockade, may prove to be an obligate element of such strategies (26).

In conclusion, we have shown that genomic complexity, although often frustrating targeted therapies, is actually a gold mine for the immune system given the large number of potential antigenic targets for T cells. The liberation of these antigens during conventional therapy coupled with blockade of the checkpoints that normally limit immune responses could provide a powerful approach to cancer treatment. These findings encourage both the continued development of augmented immunotherapy in cancer without an absolute requirement for knowledge of the targeted antigens, and also future strategies that incorporate high throughput sequence analysis toward individualized multivalent cancer vaccines.

Acknowledgments

Received 8/13/2007; revised 10/24/2007; accepted 11/20/2007.

B. Vogelstein and J.P. Allison are investigators of the Howard Hughes Medical Institute.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

References

1. Coulie PG, Brichard V, Van Pel A, et al. A new gene coding for a differentiation antigen recognized by autologous cytolytic T lymphocytes on HLA-A2 melanomas. J Exp Med 1994;180:35–42.
2. Kawakami Y, Eliyahu S, Delgado CH, et al. Cloning of the gene coding for a shared human melanoma antigen recognized by autologous T cells infiltrating into tumor. Proc Natl Acad Sci U S A 1994;91:3515–9.
3. Traversari C, van der Bruggen P, Luescher IF, et al. A nonapeptide encoded by human gene MAGE-1 is recognized on HLA-A1 by cytolytic T lymphocytes directed against tumor antigen MZ2-E. J Exp Med 1992;176:1453–7.
4. Chen YT, Scanlan MJ, Sahin U, et al. A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening. Proc Natl Acad Sci U S A 1997;94:1914–8.
5. Wolfel T, Hauer M, Schneider J, et al. A p16INK4a-insensitive CDK4 mutant targeted by cytolytic T lymphocytes in a human melanoma. Science 1995;269:1281–4.
6. Mandruzzato S, Brasseur F, Andry G, Boon T, van der Bruggen P. A CASP-8 mutation recognized by cytolytic T lymphocytes on a human head and neck carcinoma. J Exp Med 1997;186:785–93.
7. Prehn RT, Main JM. Immunity to methylcholanthrene-induced sarcomas. J Natl Cancer Inst 1957;18:769–78.
8. Klein G, Sjogren HO, Klein E, Hellstrom KE. Demonstration of resistance against methylcholanthrene-induced sarcomas in the primary autochthonous host. Cancer Res 1960;20:1561–72.
9. Kripke ML. Antigenicity of murine skin tumors induced by ultraviolet light. J Natl Cancer Inst 1974;53:1333–6.
10. Wortzel RD, Philipps C, Schreiber H. Multiple tumour-specific antigens expressed on a single tumour cell. Nature 1983;304:165–7.
11. Gilboa E. The makings of a tumor rejection antigen. Immunity 1999;11:263–70.
12. Lennerz V, Fatho M, Gentilini C, et al. The response of autologous T cells to a human melanoma is dominated by mutated neoantigens. Proc Natl Acad Sci U S A 2005;102:16013–8.
13. Sjoblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. Science 2006;314:268–74.
14. Hakenberg J, Nussbaum AK, Schild H, et al. MAPPP: MHC class I antigenic peptide processing prediction. Appl Bioinformatics 2003;2:155–8.
15. 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.
16. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 1999;50:213–9.
17. Gomez-Nunez M, Pinilla-Ibarz J, Dao T, et al. Peptide binding motif predictive algorithms correspond with experimental binding of leukemia vaccine candidate peptides to HLA-A*0201 molecules. Leuk Res 2006;30:1293–8.
18. Elkington R, Walker S, Crough T, et al. *Ex vivo* profiling of CD8+-T-cell responses to human cytomegalovirus reveals broad and multispecific reactivities in healthy virus carriers. J Virol 2003;77:5226–40.
19. Reche PA, Glutting JP, Reinherz EL. Prediction of MHC class I binding peptides using profile motifs. Hum Immunol 2002;63:701–9.
20. Buus S, Lauemoller SL, Worning P, et al. Sensitive quantitative predictions of peptide-MHC binding by a "Query by Committee" artificial neural network approach. Tissue Antigens 2003;62:378–84.
21. Nielsen M, Lundegaard C, Worning P, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci 2003;12:1007–17.
22. Cao K, Hollenbach J, Shi X, Shi W, Chopek M, Fernandez-Vina MA. 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 2001;62:1009–30.
23. Nussbaum AK, Kuttler C, Hadeler KP, Rammensee HG, Schild H. PAProC: a prediction algorithm for proteasomal cleavages available on the WWW. Immunogenetics 2001;53:87–94.
24. Kuttler C, Nussbaum AK, Dick TP, Rammensee HG, Schild H, Hadeler KP. An algorithm for the prediction of proteasomal cleavages. J Mol Biol 2000;298:417–29.
25. Dunn GP, Old LJ, Schreiber RD. The three Es of cancer immunoediting. Annu Rev Immunol 2004;22:329–60.
26. Korman AJ, Peggs KS, Allison JP. Checkpoint blockade in cancer immunotherapy. Adv Immunol 2006;90:297–339.

Downloaded from http://aacrjournals.org/cancerres/article-pdf/68/3/889/2599635/889.pdf by guest on 08 August 2024