REVIEW

Bioinformatics for cancer immunotherapy target discovery

**Lars Rønn Olsen · Benito Campos ·
Mike Stein Barnkob · Ole Winther · Vladimir Brusic ·
Mads Hald Andersen**

Abstract The mechanisms of immune response to cancer have been studied extensively and great effort has been invested into harnessing the therapeutic potential of the immune system. Immunotherapies have seen significant advances in the past 20 years, but the full potential of protective and therapeutic cancer immunotherapies has yet to be fulfilled. The insufficient efficacy of existing treatments can be attributed to a number of biological and technical issues. In this review, we detail the current limitations of immunotherapy target selection and design, and review computational methods to streamline therapy target discovery in a bioinformatics analysis pipeline. We describe specialized bioinformatics tools and databases for three main bottlenecks in immunotherapy target discovery: the cataloging of potentially antigenic proteins, the identification of potential HLA binders, and the selection epitopes and co-targets for single-epitope and multi-epitope strategies. We provide examples of application to the well-known tumor antigen HER2 and suggest bioinformatics methods to ameliorate therapy resistance and ensure efficient and lasting control of tumors.

Abbreviations

ANN	Artificial neural network
CAR	Chimeric antigen receptor
CTLs	Cytotoxic T lymphocytes
ER	Estrogen receptor
KEGG	Kyoto Encyclopedia of Genes and Genomes
pHLA	Peptide-HLA
RMA	Robust multichip average
TAAs	Tumor-associated antigens
TAP	Transporter associated with antigen processing
TCGA	The Cancer Genome Atlas
TCR	T cell receptor
TSAs	Tumor-specific antigens

Lars Rønn Olsen, Benito Campos, and Mike Stein Barnkob have contributed equally to this work.

L. R. Olsen (✉) · O. Winther
Department of Biology, Bioinformatics Centre, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark
e-mail: lro@binf.ku.dk

L. R. Olsen
Biotech Research and Innovation Center (BRIC), University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark

L. R. Olsen · V. Brusic
Cancer Vaccine Center, Dana-Farber Cancer Institute, 77 Avenue Louis Pasteur, Boston, MA 02115, USA

B. Campos
Division of Experimental Neurosurgery, Department of Neurosurgery, Heidelberg University Hospital, INF 400, 69120 Heidelberg, Germany

M. S. Barnkob
Department of Clinical Immunology, Odense University Hospital, University of Southern Denmark, Sdr. Boulevard 29, 5000 Odense C, Denmark

O. Winther
Cognitive Systems, DTU Compute, Technical University of Denmark, Matematiktorvet, Building 303B, 2800 Kgs. Lyngby, Denmark

V. Brusic
School of Science and Technology, Nazarbayev University, Astana, Kazakhstan

M. H. Andersen
Department of Hematology, Center for Cancer Immune Therapy, Herlev Hospital, Herlev Ringvej 75, 2730 Herlev, Denmark

Introduction

Cancer immunotherapy is a major modality of cancer treatment. It is often used in combination with other cancer treatments such as surgery, radiotherapy, and chemotherapy. Immunotherapy is an intervention that modulates immune responses for effective targeting and elimination of tumor cells. The term immunotherapy is quite broad and encompasses a variety of treatments; some enhance immune response in a very general way, while others direct the immune system to specifically target cancer cells. Although immunotherapy has been well documented for more than a century [1], the mechanisms of immune responses and tumor evasion remain poorly understood. Understanding the genetic background of the patient as well as the molecular characteristics of individual cancers at the time of diagnosis, as well as during therapy, is essential for stratification of patients, assessing their responses, and selection of optimal therapies [2]. Immunotherapies focus on tumor-specific antigens as well as on regulatory mechanisms that drive effective immune responses. Inducing immune responses to cancer is a delicate balancing act where effective immunity against the cancer cells should be enhanced, while the autoimmune responses against normal cells should be minimized [3]. Even when successful in this aspect, immunotherapies often lose their efficacy over time due to immune evasion by tumor cells. Immune evasion by tumor cells involves several mechanisms such as downregulation of HLA expression [4], infiltration of immune suppressive cells (e.g., Tregs [5], MDSC [6]), expression of immune suppressive molecules, e.g., IDO [7] and Arg-1 [8], lack of chemokine-mediated trafficking [9], poor innate immune cell activation [10], or immune checkpoint ligands like PD-L1 and PD-L2 [11]. Passive immunotherapies involve administration of antibodies and other immune system products that provide immunity but without activation of the host's immune responses, while active immunotherapies stimulate host immune responses against tumor cells by activating them to respond against tumor antigens. In this review, we will focus our attention to active immunotherapies using tumor antigens to target cancer cells.

Identification and detailed characterization of vaccine targets is an essential step in tumor vaccine development. Technical advances in instrumentation, sample processing, immunological assays, and bioinformatics techniques have generated large amounts of immunological data, including experimentally identified tumor antigens and T cell epitopes, novel tumor biomarkers, and differentially expressed genes or proteins identified through genomics, proteomics, or other high-throughput methods. Collection, analysis and management of these data require extensive use of bioinformatics applications. Here we describe bioinformatics tools and data resources used for the study of cancer immunotherapies, focusing primarily on T cell-based therapies.

Tumor antigens for T cell-based therapies

T cell epitopes are short peptide fragments of 8–12 and 13–25 amino acids in length for HLA class I and II, respectively [12, 13]. In the classic mode of CD8+ T cell activation, intracellular proteins are processed in the cytoplasm and cleaved by the proteasome into small peptides [14], which are then delivered into the endoplasmic reticulum by transporter associated with antigen processing (TAP) proteins [15], where they bind the HLA class I molecules and subsequently present on the cell surface as T cell epitopes. The T cell epitopes on the surface of target cells are screened and recognized by CD8+ cytotoxic T lymphocytes (CTLs). Those target cells that are recognized as foreign due to malignant transformations are killed by the cognizant CD8+ cells [16]. Initially, the CD8+ T cells are in their naïve state. After recognition of a peptide-HLA complex, they are activated to their effector CTL state and proliferate to clear cells presenting the peptide-HLA complex on their surface.

The cellular immune system responds to tumor cells by recognition of either tumor-associated antigens (TAAs, antigens that are overexpressed in cancer cells) [17, 18] or tumor-specific antigens (TSAs, antigens that are not expressed in most normal tissues) [19]. The therapeutic potential of a tumor antigen depends on an array of factors. In an effort to define the characteristics of the ideal cancer antigen, researchers at the American National Cancer Institute proposed to rank tumor antigens by therapeutic function, immunogenicity, oncogenicity, specificity, expression level and percentage of positive cells, stem cell expression, the number of patients with antigen-positive cancers, the number of validated epitopes in antigen, and cellular location of expression [20].

TAAs are not exclusive to tumor cells. They can, however, in certain instances, still elicit a tumor-specific response. TAAs can be divided into two subgroups: differentiation antigens and overexpressed antigens. Common for both types is the inherent risk that eliciting a sufficiently strong T cell response may induce systemic autoimmunity against healthy cells carrying the antigen. A number of overexpressed antigens have been characterized, as an array of genes involved in regulating tumor growth, replication, as well as apoptosis, and can severely affect the health of the cell if dysregulated [21].

Mutations in tumor cell genes can lead to changes in the primary or secondary protein structure that may affect immunogenicity of antigens [22]. Specifically, sequence changes in short peptides can change peptide binding affinity for HLA, and thus change subsequent responses by T

cells [23]. Similarly, mutations are likely to induce changes in secondary structure that can affect, to some extent T cell recognition, but primarily change the affinity and avidity of circulating antibodies for the target [24]. The main drawback of TSAs is that they are mostly patient specific, and therefore, not applicable for broadly neutralizing therapies. However, they are likely to be truly tumor specific and are often found in driver genes, thus making them less susceptible to immune escape by immunoediting [25].

Bioinformatics for T cell-based cancer immunotherapy

Identification and selection of antigens is a multifaceted task that depends both on the type and on the application of antigens. In 2000, Rino Rappuoli formalized the role of computational analyses in vaccinology in a conceptual framework termed "Reverse Vaccinology" [26]. Originally formulated to facilitate vaccine target discovery in pathogens, the concepts of reverse vaccinology can be expanded for applications in cancer immunology. Reverse vaccinology revolves around sequence analysis, whereby the genomic sequence is used to catalog all potential molecular antigens. In simple viral and prokaryotic pathogens, essentially all protein products are potential antigens, whereas the majority of tumor tissue proteins are not aberrant, thereby rendering them poor therapy targets. Therefore, cataloging antigens in tumor cells require additional pre-screening.

Once the catalog of potential antigens has been established, the reverse vaccinology pipeline calls for in silico prediction of vaccine targets—a process that is characterized as being completely naïve, and in that targets predicted from sequence, such as predicted HLA binders, are not characterized in terms of intracellular preprocessing, conservation, or in vivo expression. Additional computational and experimental pre-screening of epitope candidates must, therefore, be performed before they can be included as therapy targets. Potential T cell epitopes, must be examined in terms of preprocessing by the proteasome, transport by the TAP, HLA binding, stability of peptide-HLA complex [27], peptide-HLA binding to the T cell receptor (TCR) [28], and in vivo expression.

Potential T cell epitopes should then be examined for stability of their expression in tumor cells to ensure lasting immunological control or clearance of targeted tumor cells. However, the heterogenic nature of cancer means that no given cancer type has a uniform molecular profile and several cancers are subclassified into a number of characterized or uncharacterized classes with varying prognostic and therapeutic outcome [29–32]. Adding further to this property, evidence of intra-tumor heterogeneity is beginning to surface for certain tumors [33, 34]. Additionally,

given that immunoediting of tumor antigens is based on somatic mutations, it is extremely difficult to predict the antigenic phenotype after clonal selection. It is, therefore, often observed that tumors develop tolerance to immunotherapy after a limited period of time of successful treatment [35, 36].

Cataloging, predicting, and selecting immunotherapy targets can be extensively addressed using existing bioinformatics tools and biological databases. Figure 1 presents a schematic overview of the general process. In the following sections, we present examples of bioinformatics analyses for antigen cataloging and immunotherapy target discovery with examples of application for each task. A number of tools and databases were used for the example applications and many more exist. It is beyond the scope of this review to assess and catalog all existing tools, but in the following section we present an overview of some of the most commonly used tools and data resources for each of the tasks.

Cataloging potential antigens

Identification of potential antigens de novo from genomic sequence using bioinformatics tools is highly challenging, as expression of proteins is regulated by an array of complex regulatory mechanisms, many of which are poorly understood. Traditionally, tumor antigens are identified in vitro from serum by screening cDNA phage libraries using immunoassays [37] or proteomics-based screening [38], but bioinformatics tools are perfectly suited to aid this process, either by actively identifying novel tumor antigens or by organizing and accessing information about known tumor antigens in accessible databases.

In silico screening for novel tumor antigens

Large-scale screening of mRNA from public databases is a proposed method for active identification of novel tumor antigens [39, 40]. Specifically, comparing expression profiles of tumors and healthy tissue can elucidate genes that are overexpressed or expressed exclusively in malignant tumors. High-throughput genomics methods have enabled large-scale screening of gene expression. These include nucleotide microarray technologies [41] and next-generation RNA sequencing [42]. Preprocessing of raw outputs from expression data depends entirely on the platforms used for the experiments. Tools for most, if not all, platforms can be preprocessed and analyzed in the statistical environment R, with packages from the Bioconductor software project [43]. Once preprocessed, the most widely used computational tools for analysis of gene expression data are the packages RMA (Robust Multichip Average) [44] and Limma [45] for R or one of several
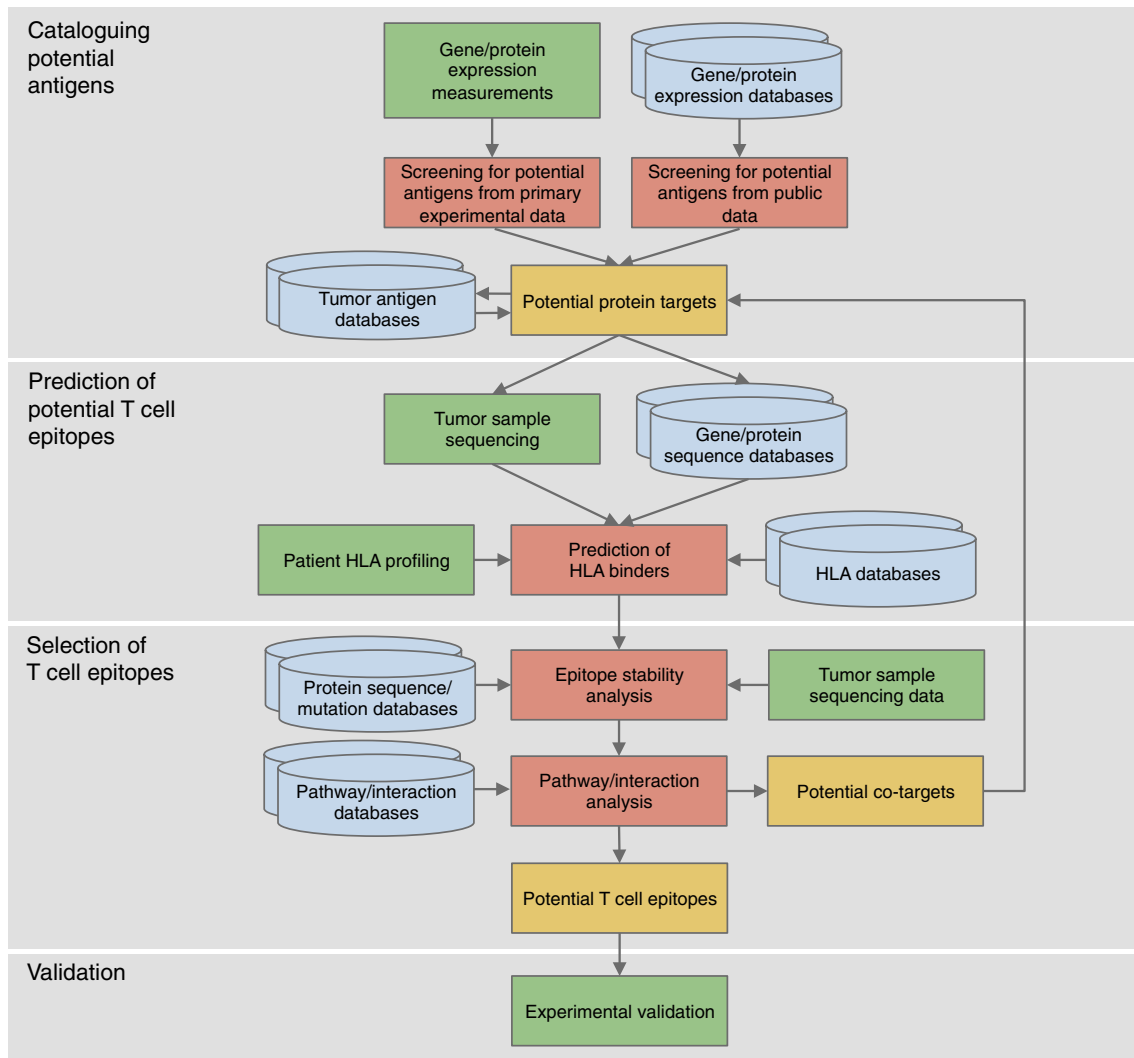
Fig. 1 General overview of bioinformatics supported discovery of potential T cell-based immunotherapy targets in cancer tumors. *Green operations* are preliminary and final laboratory analyses; *red operations* are bioinformatics analyses performed using computational tools (described in further detail in the following sections); *blue operations* are cross-references with biological databases (likewise described in further detail in the following sections), and *yellow operations* denote intermediary outputs from the analyses and cross-referencing

graphical user interface tools, most notably the Gitools software [46]. However, for an antigen to be suitable for immunotherapy, it must be expressed at protein level as well. Large-scale proteome studies include technologies of protein microarrays [47], antibody microarrays [48] mass spectrometry-based proteomics [49], and mass cytometry screening [50]. Recent studies of the proteome of breast cancer have revealed molecular features of tumorigenesis [51], and proteomics studies are gradually approaching a scale where whole proteome screening is feasible [52]. Databases such as the Gene Expression Omnibus [53] contain MIAME standard compliant expression data [54] and a number of specialized cancer genomics data repositories such as The Cancer Genome Atlas (TCGA) consortium

data portal (https://tcga-data.nci.nih.gov/tcga/), which contains standardized data (including protein expression) for a large number of cancer types, Oncomine [55] (mRNA expression data for multiple cancer types), and the Hema-Explorer [56, 57] (mRNA expression data for healthy and malignant hematopoietic cells). Although tumor antigens must be expressed at protein level to be useful for antigen-based cancer immunotherapies, mRNA expression profiles can serve as useful pre-screenings, subject to additional experimental validation, including proteomics verification, as correlation between RNA expression and protein expression vary between different protein families [58]. Lastly, an expression profile database such as UniGene (http://www.ncbi.nlm.nih.gov/unigene) and The Human Protein Atlas

[59] can be consulted to ensure that potential target tumor antigens are not similarly expressed in healthy tissues on genetic and protein levels. Nucleic Acids' Research offers a comprehensive catalog of maintained biological databases [60].

Cross-referencing expression data with known tumor antigens

A large number of studies presenting potential tumor antigens are published each year. Cross-referencing tumor gene expression and protein expression profiles with previous experimental efforts enable fast cataloging of potential targets. Data resources for tumor antigens include the Cancer Immunity peptide database of T cell-defined tumor antigens [61], a static listing of tumor T cell antigens provided by Parmiani and colleagues [21], CTdatabase of cancer-testis antigens [62], and the TANTIGEN database of T cell tumor antigens (http://cvc.dfci.harvard.edu/tadb/index.html). Genes or proteins previously identified as TSAs (and expressed in a target sample), or identified as TAAs (and overexpressed compared to normal tissue from the same patient), are subject to further investigations as a potential immunotherapy targets if they are expressed at appropriate levels in a given tumor sample.

Assessing potential tumor antigens

Once one or several potential tumor antigens have been identified, further information can aid in the pre-experimental assessment of the antigens. This process is best illustrated with an example of information collection and antigen assessment of the well-characterized tumor antigen, HER2.

Information about HER2, relevant to assessing its role as a tumor antigen is located and extracted from a number of different biological databases. Table 1 lists information relevant to assessing the suitability of HER2 as an antigen in a number of different cancers. HER2 is an epidermal growth factor that is amplified in about 20–40 % of invasive breast cancers [63]. Whereas normal tissue generally has low expression of HER2, breast cancer cells can have up to 50 copies of the encoding ERBB2 gene and up to 100-fold increased protein expression [64], with heavy correlation to a poor clinical outcome. These properties make HER2 a good marker for tumor tissue. HER2 has four isoforms produced by alternative splicing and alternative initiation. The isoforms overlap in identity by slightly less than half of the protein sequences, and a number of somatic mutations detectable on protein level are characterized in HER2. Since HER2 is present on the cell surface in large numbers, it is suitable for targeting by both cellular and humoral immunities, and a number of both T cell and B cell HER2

epitopes have been identified [65, 66]. A selection of tools and databases useful for cataloging potential tumor antigens can be found in Table 2.

Prediction of potential T cell epitopes

Each of the cellular processes responsible for T cell epitope preprocessing is rate limiting in the classical T cell-mediated immunity. The prediction of immunogenicity is, therefore, a non-trivial task, which is divided into predictions of each epitope processing step.

HLA binding prediction

Prediction of peptide processing events, such as proteasomal cleavage [74–76] and TAP transport [77–79], has been explored, but evaluations suggest that these methods are still not optimal [80]. Algorithms for predicting HLA binding affinity are superior in accuracy and highly accurate for a number of HLA alleles. Most currently maintained algorithms have been thoroughly reviewed and benchmarked in [81].

Prediction of peptide binding affinity to HLA class I and class II can be performed with a host of prediction algorithms (extensively reviewed in [82, 83]). The overall best performing predictor for HLA class I binding is the artificial neural network (ANN) and weight matrix-based prediction tool, netMHC 3.2 [84], and the best for class II is the ANN predictor netMHCII 2.2 [85]. Other highly accurate classification algorithms include BIMAS [86], SYFPEITHI [87], novel ensemble methods PM and AvgTanh [88], and various averaging methods [89]. Pan-specific prediction algorithms such as NetMHCpan [90] include HLA allele sequence in the binding prediction, which has been shown to increase accuracy for certain alleles [81]. Additionally, these methods enable prediction to a large number specific HLA alleles as well as HLA supertypes. A number of prediction algorithms combines prediction of HLA binding with prediction of proteasomal cleavage and TAP transport, for example, NetCTL [91] (for an extensive review of CTL prediction algorithms, please refer to [91]), but as peptide preprocessing predictions are not nearly as accurate as HLA binding predictions, including these additional algorithms do not enrich the prediction.

A number of known HLA binders have been shown to be unable to elicit immune response, a phenomenon referred to as holes in the T cell repertoire [92]. Recently, it has been shown that stability of the peptide-HLA (pHLA) complex is likely a better predictor for immunogenicity of a peptide than the affinity of the binding, as immunogenic peptides are generally more stably bound to HLA [27]. Similarly, assessment of pHLA complex binding to the TCR has been explored as a predictor of

Table 1 Annotation of HER2 for assessing potential as tumor antigen

Information type	Information	Source (ID/accession)
Protein name	HER2	GeneCards (GC17P037844)
Gene name	ERBB2	GeneCards (GC17P037844)
Full name	Human epidermal growth factor receptor 2	GeneCards (GC17P037844)
Synonyms	CD340, HER-2, HER2, NEU, NGL	GeneCards (GC17P037844)
Function	HER2 is a member of the epidermal growth factor (EGF) receptor family of receptor tyrosine kinases. HER2 cannot bind growth factors itself, since it lacks a ligand binding domain. Rather, it binds to other EGF receptor/ligand complexes, with which it forms a heterodimer, thus stabilizing binding and enhancing activation of downstream signaling pathways.	OMIM (164870)
Role in disease	The ERBB2 gene is amplified and HER2 is overexpressed in 25–30 % of breast cancers, correlating with increased aggressiveness of the tumor.	OMIM (164870)
Localization	Cell surface, plasma membrane bound.	OMIM (164870)
Isoforms	Isoform 1: canonical sequence Isoform 2: 1–610: missing Isoform 3: 1–686: missing Isoform 4: 1–23: MELAALCRWGLLLALLPPGAAST → MPRGSWKP	UniProt (P04626)
Mutations	Pos: 20, A → T Pos: 49, L → H Pos: 49, L → P Pos: 92, R → G Pos: 101, I → S	COSMIC (COSG165)
Gene expression, normal tissue	Low expression in most healthy tissue, moderate expression in intestine and mammary glands.	UniGene (241389)
Gene expression, tumor tissue	Overexpressed in 25–30 % mammary gland tumors (strong correlation with clinical outcome) and in approximately 15 % colorectal tumors (no correlation with clinical outcome).	UniGene (241389)
Protein expression, normal tissue	No-to-low expression in hematopoietic, digestive, and respiratory tissue. No-to-medium expression in female tissues, placenta, male tissues, and urinary tract tissue.	The human protein atlas
Protein expression, tumor tissue	Strong expression observed in some breast cancer, colorectal cancer, glioma, head and neck cancer, ovarian cancer, and urothelial cancer	The human protein atlas
T cell epitopes	50 reported CD8+ T cell epitopes 13 reported HLA class I ligands	TANTIGEN (Ag000001)

immunogenicity [93]. However, at present, only 21 crystal structures of pHLA-TCR complex are completed, which is not sufficient basis to train a generally applicable classifier. Other approaches to evaluating immunogenicity include prediction of T cell reactivity based on an array of physiochemical properties [94, 95].

To predict HLA binders of tumor cell samples, the tumor tissue exome should ideally be sequenced and potential epitopes should be predicted from the translated sequence. If tumor tissue sequence is not available, canonical sequences can be extracted from protein sequence databases such as UniProt [72] or from primary repositories for cancer sequences such as GChub (https://cghub.ucsc.edu/). Predicted binders can be cross-referenced with a tumor antigen database such as TANTIGEN to check whether experimental validation has been previously performed.

Predicting HLA binders for 9-meric peptides in HER2 using netMHC 3.4 yields potential binders to a number

of HLA alleles. A closer look at HLA A*02:01 reveals 52 predicted binders, of which one is an experimentally validated binder, found by cross-referencing with TANTIGEN. Some candidate binders are conserved across all isoforms and mutated forms of HER2, while others are found only in some isoforms. Table 3 shows peptides binding HLA A*02:01 that are either conserved in all isoforms, or positions where all variant peptides bind HLA A*02:01, discovered by block conservation analysis [96]).

Prior to preclinical testing of the predicted epitopes, experimental validation of the candidates is important. Since peptide preprocessing is still unknown for predicted HLA binders, appropriate peptide processing and in vivo binding should be confirmed experimentally before epitopes are included in vaccine constructs. Large-scale T cell epitope validation is enabled by mass spectrometry [98] and flow cytometry-based methods [99].

Table 2 Sample of analytical tools for cataloging tumor antigens

Resource	Purpose	URL	Ref.
Genomics and proteomics expression analysis tools			
Gitools	Expression data analysis (GUI)	http://www.gitools.org/	[46]
Limma	Expression data analysis (R package)	http://www.bioconductor.org/packages/release/bioc/html/limma.html	[44]
RMA	Expression data analysis (R package, integrated in "affy" package)	http://www.bioconductor.org/packages/release/bioc/html/affy.html	[45]
Databases for gene and protein expression			
GEO	mRNA expression data	http://www.ncbi.nlm.nih.gov/geo/	[53]
ArrayExpress	mRNA expression data (among other data)	http://www.ebi.ac.uk/arrayexpress/	[67]
TCGA data portal	mRNA expression data, cancer specific (various types)	https://tcga-data.nci.nih.gov/tcga/	–
Oncomine	mRNA expression data, cancer specific (various types)	https://www.oncomine.org/	[55]
HemaExplorer	mRNA expression profiles in normal and malignant haematopoiesis	http://servers.binf.ku.dk/hemaexplorer/	[56]
UniGene	mRNA expression data by tissue	http://www.ncbi.nlm.nih.gov/unigene	–
Human protein atlas	Protein expression by tissue	http://www.proteinatlas.org/	[59]
Databases for tumor antigen information			
TANTIGEN	Tumor T cell antigens	http://cvc.dfci.harvard.edu/tadb/	–
CTdatabase	Cancer-testis antigens	http://www.cta.lncc.br/	[62]
Cancer immunity peptide database	Tumor T cell antigens	http://cancerimmunity.org/peptide/	[61]
Databases for gene and protein information			
OMIM	Information about human genes and genetic phenotypes	http://omim.org/	[68]
Ensembl	Annotated gene sequences	http://www.ensembl.org	[69]
Genecards	Functional information about human genes	http://www.genecards.org/	[70]
Databases for gene and protein sequences			
GenBank	Gene sequences and information	http://www.ncbi.nlm.nih.gov/genbank	[71]
GC hub	Cancer genome sequences	https://cghub.ucsc.edu/	–
UniProt	Proteomic sequences and functional annotations	http://www.uniprot.org/	[72]
COSMIC	Somatic mutations in cancer	http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/	[73]

Human immune system diversity

HLAs are among the most polymorphic molecules in the human genome and represent the most variable factor of human immune recognition. Comprised of more than 200 genes located on chromosome 6, three different HLA classes are defined [100]. Only class I and class II are involved in adaptive immunity and thus are a main focus in this review. For each class, several major and minor proteins are defined, which are in turn classified into supertypes [101, 102] and 9,310 individual alleles are reported in Release 3.12.0, (April 17, 2013) of the IMGT/HLA database [103]. Additional data resources for HLA allele sequences, clinical data, and population frequencies can be found in dbMHC [104] and The Allele Frequency Net Database [105] among others. Further bioinformatics resources for HLA research have been thoroughly reviewed in [106].

Specificity of HLA molecules is instrumental in determining resistance and susceptibility to invading pathogens and cancers. Owing to hereditability of HLA loci, specific alleles are often geographically clustered, meaning that some populations are more susceptible to, for example, EBV-related cancers [107]. T cell-mediated immunosurveillance of cancerous cells involves HLA restriction, which further complicates formulation of T cell-based therapies. Even if we ignore the variability of tumor antigens, the diversity of human immune response to T cell epitopes renders the identification of broadly applicable T cell-based immunotherapy targets highly challenging, and increases the search space of useful T cell targets in personalized therapies immensely.

Table 3 Peptides from HER2 predicted to bind HLA A*02:01

Position	Number of variants in given position	Peptide	Predicted HLA A*0201 binding affinity (nM)	Experimental status	Reference
689	1	RLLQETELV	34	T cell epitope	[97]
767	4	ILDEAYVMA	34	N/A	–
		ILDEAYAMA	33	N/A	–
		ILHEAYVMA	73	N/A	–
		MLDEAYVMA	15	N/A	–
823	2	LLNWCMQIA	359	N/A	–
		LLNWCMQTA	89	N/A	–
949	1	TIDVYMIMV	86	N/A	–
953	1	YMIMVKCWM	101	N/A	–
954	1	MIMVKCWMI	22	N/A	–

All peptides in this table are either found in all isoforms and mutated types of HER2, or are peptides from a position on which all peptides are predicted to bind HLA A*02:01. Peptides with a predicted binding affinity of <50 nM are strong binders and <500 nM are weak binders. All predictions were done using netMHC 3.4

Host HLA profile must be identified before predicting personalized targets, a task that can be done by DNA sequencing [108], RNA sequencing [109], or microarray-based approaches [110] (thoroughly reviewed in [100]). Analysis of sequencing data can be performed by a large number of tools reviewed in [111]. If the aim is to produce more general therapies, population coverage tools such as the PopCover algorithm [112], the Block Conservation analysis web server (http://met-hilab.bu.edu/blockcons), or the IEDB Population Coverage Calculation tool [113] can be applied to ensure broad coverage by a combination of immune targets. A selection of useful bioinformatics resources for prediction of prediction of potential T cell epitopes is listed in Table 4.

Selection of potential epitopes for immunotherapy

Conservation and variability analysis for immunotherapy target selection is a multidimensional problem. If one aims to define targets for general immunotherapies applicable to a broad cohort of patients, antigen diversity must be studied across the patient population. Due to high variability, even personalized vaccine targets are likely to be unstable over time, and the somatic process driving the selection is difficult to predict. There are, however, strategies for inducing lasting immune response against tumors, including epitope selection with increased emphasis on sequence stability, the combination of multiple therapy targets, and multi-epitope vaccination strategies [122].

Selection of stable epitopes

Comparative analysis of gene and protein sequences can reveal de novo SNPs and other somatic mutations. Peptides found in mutated protein regions unique to the tumor are candidates for epitope prediction. However, peptides found in highly variable regions are likely subject to frequent mutations and potentially lead to loss of antigen immunogenicity. Owing to the heterogenic nature of cancer and complex processes such as immunoediting, the landscape of tumor mutation is far from fully understood. However, potential epitopes can be analyzed for stability in the context of known mutations cataloged in databases such as COSMIC [123]. Similarly, splice variations of proteins and structural variations in the genome can influence the stability of a given epitope. This can be examined using databases such as DECIPHER [124] for chromosomal variations and UniProt [125] for protein isoforms. Identifying regions of limited variability and high stability, and choosing potential epitopes in these regions may increase the likelihood for sustained immune response. Variability also depends on the selection pressure exerted by the immune system during therapy, but regions of known high variability can be excluded.

Multi-epitope strategies

Epitopes in mutating regions of target proteins are not necessarily excluded as a valuable target for immunotherapy. Treatments can be composed of multiple epitopes from one or more proteins [122], such that if immunogenicity of one epitope is lost, the remaining set of epitopes may continue to confer protection. In a multi-epitope setting, mutating regions of target proteins can be of value if they contain epitopes—even if just for a limited time or a limited fraction of the tumor cells. Analysis of metabolic pathways of tumor cells may reveal potential epitopes in multiple proteins complementary to each other, which collectively

Table 4 Sample of analytical tools for discovery of T cell epitopes for cancer immunotherapy

Tool	Purpose	URL	Ref.
HLA allele prediction from sequence data			
HLAminer	Prediction of HLA haplotypes from sequence data	http://www.bcgsc.ca/platform/bioinfo/software/hlaminer	[114]
HLAforest	Prediction of HLA haplotypes from sequence data	http://code.google.com/p/hlaforest	[115]
HLA data resources			
IMGT/HLA	HLA sequences and nomenclature	http://www.ebi.ac.uk/ipd/imgt/hla/	[103]
dbMHC	DNA and clinical data-related HLA	http://www.ncbi.nlm.nih.gov/projects/gv/mhc/	[104]
The allele frequency net database	Allele frequencies from different polymorphic areas in the Human genome	http://www.allelefrequencies.net/	[105]
Peptide-HLA binding affinity predictions			
netMHC	Prediction of HLA class I binders	http://www.cbs.dtu.dk/services/NetMHC/	[116]
netMHCII	Prediction of HLA class II binders	http://www.cbs.dtu.dk/services/NetMHCII/	[85]
Epitope conservation analysis			
ClustalW	Multiple sequence alignment	http://www.genome.jp/tools/clustalw/	[117]
MAFFT	Multiple sequence alignment	http://www.ebi.ac.uk/Tools/msa/mafft/	[118]
MUSCLE	Multiple sequence alignment	http://www.ebi.ac.uk/Tools/msa/muscle/	[119]
IEDB analysis resource	Epitope conservancy analysis	http://tools.immuneepitope.org/tools/conservancy/	[120]
IEDB analysis resource	Epitope cluster analysis	http://tools.immuneepitope.org/tools/cluster/	–
IEDB analysis resource	Population coverage calculation	http://tools.immuneepitope.org/tools/population/	[113]
PopCover	Epitope selection for population coverage	–	[112]
BlockLogo	Visualization of immunological motifs	http://research4.dfci.harvard.edu/cvc/blocklogo/	[121]
Block conservation	Conservation analysis for multi-epitope strategies	http://met-hilab.bu.edu/blockcons/	–

Databases of gene and protein sequences useful for HLA binding predictions are listed in Table 2. For a comprehensive review and performance benchmarking of HLA binding prediction algorithms, please refer to [81–83]

can provide protection. Theoretically, co-targeting multiple epitopes in multiple proteins in pathways essential for tumor fitness should increase probability of sustained response [126]. The network analysis of signaling pathways and the perturbations by oncogenes was recently shown to successfully identify oncogenic targets. A sequential application of anticancer drugs increased the collective efficiency of the drugs targeting oncogenic signaling pathways [127]. The sequential administration of immunotherapies targeting different epitopes may also be advantageous. Multi-epitope approaches carry the inherent risk of raising a dominant response against one, or a few, of the administered epitopes. However, this can, in theory, be avoided by multiple site vaccinations [128]. Additionally, it has been shown that immunotherapies sometimes facilitate immune responses against additional antigens, not included as targets in the therapy, by a process referred to as "epitope spreading" [129] or "provoked immunity" [130].

The predicted HLA binders of HER2 shown in Table 3 are filtered by conservation in known isoforms and mutated forms. As can be seen, four of the predicted binders are found in all known forms of HER2 (positions 689, 949, 953, and 954), whereas six are found only in some

(positions 767 and 823). The latter six peptides are located in potentially unstable regions, but as observed on position 767, each of the four variants are predicted to bind HLA A*02:01, making them potentially useful in a multi-epitope setting. Note, however, that only 244 uniquely mutated samples of HER2 have been identified (UniProt, May 28, 2013), giving an estimate of the variability of HER2. Additionally, the frequency of each mutation is unknown, so some mutations may impact a peptide's suitability as an immunotherapy target more than others.

Targeting multiple antigenic protein

Targeting multiple epitopes within the same protein can be valuable to avoid loss of immunogenicity caused by mutations or splice variation. However, targeting a single protein does not address loss of immunogenicity by downregulation or complete loss of protein expression. It can, therefore, be valuable to target multiple epitopes in different proteins. Combined targeting of several antigens increases therapy flexibility and may increase the magnitude of the response. This approach is especially valuable when targeting proteins of similar or compensatory function in redundant

pathways [131]. An examination of HER2 interactions recorded in the STRING (a database of direct (physical) and indirect (functional) associations [132]) with known tumor antigens (from TANTIGEN) reveals a large number of proteins homologous to, co-expressed with, or interacting with HER2, based on recorded co-expression, data mining primary literature, or recorded interactions in specialized databases. The ten tumor antigens with highest scoring confidence relationship to HER2 are shown in Fig. 2. One of these is EGFR, which has previously been examined as a co-target with HER2 [131, 133]. Cross-referencing TANTIGEN shows that EGFR harbors T cell epitopes for potential immunotherapy targeting. In a similar fashion, functional homologues can be examined as novel targets.

Another strategy is to target multiple epitopes in proteins from different interacting pathways. The HER2 and the estrogen receptor (ER) signaling pathways are the dominant drivers of cell proliferation in 85 % of breast cancers, which make antigens of these pathways desirable therapy targets [134]. Another multi-epitope strategy could, therefore, involve targeting several antigens in both pathways to avoid therapy resistance if a single epitope is lost. Examining the HER2 and ER pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG) the molecular signatures database MSigDB [135] reveals multiple potential antigens in each pathway. Lastly, exploring the ClinicalTrials.gov database (http://www.clinicaltrials.gov) may help direct the selection of useful targets. Table 5 lists a sample of data resources useful for co-target discovery.

Future perspectives

At present, accurate epitope predictions are limited to cellular responses although prediction of antibody response is highly studied [139]. Computational methods for identification of T cell epitopes also have limitations, in that peptide preprocessing predictions are not yet as accurate as peptide binding prediction algorithms. Additionally, availability of tumor sequences represents a bottleneck in conservation and variability analyses, but this is likely to be remedied in a near future, as high-throughput sequencing becomes cheaper and more efficient. Another issue currently being addressed is that intra-tumor diversity may not be adequately captured by current methods, and may, therefore, impact the efficacy of immunotherapies and other therapies alike. Lastly, immunotherapy is largely a field of research that must be addressed using proteomics analyses rather than genomics analyses, the latter having been far more prolific in the past decade. Common to all these limitations is that they are currently being addressed in the wet laboratory to different extents, and all aspects of this progress will increase the need for bioinformatics tools and experts.
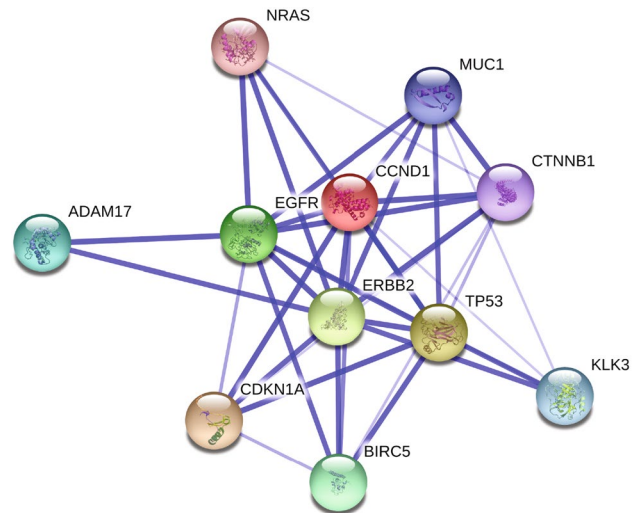
Fig. 2 Protein–protein interaction (PPI) network for HER2 and interacting tumor antigens (from TANTIGEN). PPI networks may indicate possible compensatory activity such as that for HER2 and EGFR, making interaction networks useful for elucidating additional potential targets. *Nodes* represent proteins and *edges* correspond to functional interactions. *Thicker edges* signify higher confidence in the interaction. Image was generated using the STRING database

Bioinformatics for antibody-based therapies

Immunotherapies utilizing cellular surface proteins, such as antibody-based therapies [140] or the adoptive transfer of chimeric antigen receptor (CAR)-modified T cells [141] also stands to benefit from bioinformatics supported target selection. There are, however, limitations to the current computational tools for B cell epitope prediction as well as a lack of a dedicated data resource for cellular expression of surface proteins. Prediction of B cell epitopes have long been explored using a variety of approaches, but the accuracy of these algorithms—particularly for discontinuous epitopes—remains suboptimal, and their practical utility is limited [139, 142]. Computational prediction of membrane spanning protein regions has been explored for several decades [143], but before prediction of surface epitopes or selection of novel CAR targets can take place, a comprehensive data resource of validated surface protein expression should be assembled to provide the foundation for such exploratory analyses.

In silico assessment of susceptibility to immunotherapy

Successfully induced T cells, raised against a number of tumor T cell antigens, have been observed in peripheral blood, and yet clinical responses to immunotherapies have been limited. This indicates that barriers to immune response exist within the tumor environment, and as such, play a significant role in planning appropriate treatment

Table 5 Sample of data resources for interaction and pathway analysis useful for co-target discovery

Database	Content	URL	Ref.
Interaction/pathway data			
STRING	Known and predicted protein interactions	http://string-db.org/	[132]
BioGRID	Biological interaction data	http://thebiogrid.org/	[136]
KEGG	Information about genomes, enzymatic pathways, and biological chemicals	http://www.genome.jp/kegg/	[137]
Reactome	Information about biological pathways	http://www.reactome.org/	[138]
MSigDB	Molecular signatures annotated using gene set enrichment analysis methods	http://www.broadinstitute.org/gsea/msigdb	[135]
Clinical trials			
ClinicalTrials.gov	Database of active and completed clinical trials	http://www.clinicaltrials.gov	–

modalities [144]. Therefore, personalized immunotherapy treatments are likely to benefit from thorough analysis of genetic and proteomic host factors related to tumor immune escape mechanisms and thereby a patient's susceptibility to a given immunotherapy.

In response to the immune system's role in curbing cancer cell growth and immunoediting it has recently been proposed that the immune system should be included in the traditional histopathological classification of tumors. The classical TNM staging system describes the extent of tumorigenesis based on tumor burden (T), the presence of cancer cells in draining lymph nodes (N), and status of metastases (M). In addition to these scores, a so-called immunoscore (I) can be determined on the basis of two leukocyte populations, namely cytotoxic CD8+ T cells and memory CD8+ T cells [145, 146]. Comparison of the infiltration-rate of these two cell types in the center of the tumor and in the invasive margin of the tumor, determines the "I" score for the tumor. In two independent cohorts, patients with a high I score had significantly less relapse and overall improved survival compared with patients with low I scores [147].

The TNM-I classification scheme in conjunction with predictive biomarkers for immune response could provide a reasonable estimate of the suitability of immunotherapy as part of a treatment modality in any given patient. However, the applicability of this approach is lessened by lack of functional studies on the topic. Firstly, not all cancers are resected, and even fewer have significant material left after normal histological assessment has been completed. Markers and leukocyte profiles that can classify patients on the basis of immunological markers in peripheral blood are, therefore, desirable although no such markers have been successfully correlated with clinical response to antigen-based immunotherapies [144]. Secondly, the complex interplay between genetic and proteomic elements makes it hard to elucidate single predictive biomarkers for accurate predictions. Therefore, no tools or data resources for immunotherapy susceptibility biomarkers exist as of yet.

Conclusions

Traditionally, mass experimental screening has been the primary tool to elucidate cancer immunotherapy targets, a process which could be streamlined by systematic application of bioinformatics on patient antigen expression and sequence profiles in conjunction with publically available biological data. The conceptual framework put forth in this review was applied to HER2 as an example of the proposed computational analyses. These methods will increase in accuracy as the genomic and proteomic landscape of tumor antigens is uncovered in greater breadth and depth in the wet laboratory. As tumor cells' response to immunotherapies is gradually uncovered and the body of biological tumor data grows, so will the need for bioinformatics to organize, store, and analyze these data.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Kirkwood JM, Butterfield LH, Tarhini AA et al (2012) Immunotherapy of cancer in 2012. CA Cancer J Clin 62:309–335. doi:10.3322/caac.20132

2. Rees R, Laversin S, Murray C, Ball G (2012) Current approaches to identify and evaluate cancer biomarkers for patient stratification. In: Morrow WJW, Sheikh NA, Schmidt CS, Davies DH (eds) Vaccinology: principles and Practice. Wiley, Oxford, pp 452–463

3. Wei W-Z, Morris GP, Kong Y-CM (2004) Anti-tumor immunity and autoimmunity: a balancing act of regulatory T

cells. Cancer Immunol Immunother 53:73–78. doi:10.1007/s00262-003-0444-1

4. Blades RA, Keating PJ, McWilliam LJ et al (1995) Loss of HLA class I expression in prostate cancer: implications for immunotherapy. Urology 46:681–686. doi:10.1016/S0090-4295(99)80301-X (discussion 686–7)

5. Curiel TJ, Coukos G, Zou L et al (2004) Specific recruitment of regulatory T cells in ovarian carcinoma fosters immune privilege and predicts reduced survival. Nat Med 10:942–949. doi:10.1038/nm1093

6. Gabrilovich DI, Nagaraj S (2009) Myeloid-derived suppressor cells as regulators of the immune system. Nat Rev Immunol 9:162–174. doi:10.1038/nri2506

7. Prendergast GC (2008) Immune escape as a fundamental trait of cancer: focus on IDO. Oncogene 27:3889–3900. doi:10.1038/onc.2008.35

8. Bronte V, Serafini P, De Santo C et al (2003) IL-4-induced arginase 1 suppresses alloreactive T cells in tumor-bearing mice. J Immunol 170:270–278

9. Ben-Baruch A (2006) Inflammation-associated immune suppression in cancer: the roles played by cytokines, chemokines and additional mediators. Semin Cancer Biol 16:38–52

10. Wang T, Niu G, Kortylewski M et al (2004) Regulation of the innate and adaptive immune responses by Stat-3 signaling in tumor cells. Nat Med 10:48–54. doi:10.1038/nm976

11. Latchman Y, Wood CR, Chernova T et al (2001) PD-L2 is a second ligand for PD-1 and inhibits T cell activation. Nat Immunol 2:261–268. doi:10.1038/85330

12. Falk K, Rötzschke O, Stevanović S et al (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature 351:290–296. doi:10.1038/351290a0

13. Reinherz EL, Tan K, Tang L et al (1999) The crystal structure of a T cell receptor in complex with peptide and MHC class II. Science 286:1913–1921

14. Rock KL, Gramm C, Rothstein L et al (1994) Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. Cell 78:761–771

15. Androlewicz MJ, Anderson KS, Cresswell P (1993) Evidence that transporters associated with antigen processing translocate a major histocompatibility complex class I-binding peptide into the endoplasmic reticulum in an ATP-dependent manner. Proc Natl Acad Sci USA 90:9130–9134

16. Henney CS, Gaffney J, Bloom BR (1974) On the relation of products of activated lymphocytes to cell-mediated cytolysis. J Exp Med 140:837–852

17. Van der Bruggen P, Traversari C, Chomez P et al (1991) A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. Science 254:1643–1647. doi:10.1126/science.1840703

18. Boon T, van der Bruggen P (1996) Human tumor antigens recognized by T lymphocytes. J Exp Med 183:725–729

19. Van den Eynde BJ, van der Bruggen P (1997) T cell defined tumor antigens. Curr Opin Immunol 9:684–693

20. Cheever MA, Allison JP, Ferris AS et al (2009) The prioritization of cancer antigens: a national cancer institute pilot project for the acceleration of translational research. Clin Cancer Res 15:5323–5337. doi:10.1158/1078-0432.CCR-09-0737

21. Novellino L, Castelli C, Parmiani G (2005) A listing of human tumor antigens recognized by T cells: March 2004 update. Cancer Immunol Immunother 54:187–207. doi:10.1007/s00262-004-0560-6

22. Seremet T, Brasseur F, Coulie PG (2011) Tumor-specific antigens and immunologic adjuvants in cancer immunotherapy. Cancer J 17:325–330. doi:10.1097/PPO.0b013e3182326004

23. Lurquin C, Van Pel A, Mariamé B et al (1989) Structure of the gene of tum- transplantation antigen P91A: the mutated exon encodes a peptide recognized with Ld by cytolytic T cells. Cell 58:293–303

24. Mariuzza RA, Phillips SE, Poljak RJ (1987) The structural basis of antigen-antibody recognition. Annu Rev Biophys Biophys Chem 16:139–159. doi:10.1146/annurev.bb.16.060187.001035

25. Dunn GP, Bruce AT, Ikeda H et al (2002) Cancer immunoediting: from immunosurveillance to tumor escape. Nat Immunol 3:991–998. doi:10.1038/ni1102-991

26. Rappuoli R (2000) Reverse vaccinology. Curr Opin Microbiol 3:445–450

27. Harndahl M, Rasmussen M, Roder G et al (2012) Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity. Eur J Immunol 42:1405–1416. doi:10.1002/eji.201141774

28. Rudolph MG, Stanfield RL, Wilson IA (2006) How TCRs bind MHCs, peptides, and coreceptors. Annu Rev Immunol 24:419–466. doi:10.1146/annurev.immunol.23.021704.115658

29. Dhanasekaran SM, Barrette TR, Ghosh D et al (2001) Delineation of prognostic biomarkers in prostate cancer. Nature 412:822–826. doi:10.1038/35090585

30. Spentzos D, Levine DA, Ramoni MF et al (2004) Gene expression signature with independent prognostic significance in epithelial ovarian cancer. J Clin Oncol 22:4700–4710. doi:10.1200/JCO.2004.04.070

31. Buyse M, Loi S, van't Veer L et al (2006) Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. J Natl Cancer Inst 98:1183–1192. doi:10.1093/jnci/djj329

32. Verhaak RGW, Hoadley KA, Purdom E et al (2010) Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 17:98–110. doi:10.1016/j.ccr.2009.12.020

33. Snuderl M, Fazlollahi L, Le LP et al (2011) Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. Cancer Cell 20:810–817. doi:10.1016/j.ccr.2011.11.005

34. Sottoriva A, Spiteri I, Piccirillo SGM et al (2013) Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proc Natl Acad Sci USA 110:4009–4014. doi:10.1073/pnas.1219747110

35. Campoli M, Chang C-C, Ferrone S (2002) HLA class I antigen loss, tumor immune escape and immune selection. Vaccine 20(Suppl 4):A40–A45

36. Chang C-C, Campoli M, Restifo NP et al (2005) Immune selection of hot-spot beta 2-microglobulin gene mutations, HLA-A2 allospecificity loss, and antigen-processing machinery component down-regulation in melanoma cells derived from recurrent metastases following immunotherapy. J Immunol 174:1462–1471

37. Sioud M, Hansen M, Dybwad A (2000) Profiling the immune responses in patient sera with peptide and cDNA display libraries. Int J Mol Med 6:123–128

38. Hanash S (2003) Disease proteomics. Nature 422:226–232. doi:10.1038/nature01514

39. Loging WT, Lal A, Siu IM et al (2000) Identifying potential tumor markers and antigens by database mining and rapid expression screening. Genome Res 10:1393–1402

40. Scanlan MJ, Gordon CM, Williamson B et al (2002) Identification of cancer/testis genes by database mining and mRNA expression analysis. Int J Cancer 98:485–492

41. Ramsay G (1998) DNA chips: state-of-the-art. Nat Biotechnol 16:40–44. doi:10.1038/nbt0198-40

42. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63. doi:10.1038/nrg2484

43. Gentleman RC, Carey VJ, Bates DM et al (2004) Bioconductor: open software development for computational biology and

bioinformatics. Genome Biol 5:R80. doi:10.1186/gb-2004-5-10-r80

44. Irizarry RA, Hobbs B, Collin F et al (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249–264. doi:10.1093/biostatistics/4.2.249

45. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3:Article3. doi:10.2202/1544-6115.1027

46. Perez-Llamas C, Lopez-Bigas N (2011) Gitools: analysis and visualisation of genomic data using interactive heat-maps. PLoS ONE 6:e19541. doi:10.1371/journal.pone.0019541

47. Chatterjee M, Draghici S, Tainsky MA (2006) Immunotheranostics: breaking tolerance in immunotherapy using tumor autoantigens identified on protein microarrays. Curr Opin Drug Discov Dev 9:380–385

48. Miller JC, Zhou H, Kwekel J et al (2003) Antibody microarray profiling of human prostate cancer sera: antibody screening and identification of potential biomarkers. Proteomics 3:56–63. doi:10.1002/pmic.200390009

49. Hanash SM, Pitteri SJ, Faca VM (2008) Mining the plasma proteome for cancer biomarkers. Nature 452:571–579. doi:10.1038/nature06916

50. Bendall SC, Simonds EF, Qiu P et al (2011) Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science 332:687–696. doi:10.1126/science.1198704

51. Imielinski M, Cha S, Rejtar T et al (2012) Integrated proteomic, transcriptomic, and biological network analysis of breast carcinoma reveals molecular features of tumorigenesis and clinical relapse. Mol Cell Proteomics 11(M111):014910. doi:10.1074/mcp.M111.014910

52. Nilsson T, Mann M, Aebersold R et al (2010) Mass spectrometry in high-throughput proteomics: ready for the big time. Nat Methods 7:681–685. doi:10.1038/nmeth0910-681

53. Barrett T, Wilhite SE, Ledoux P et al (2013) NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res 41:D991–D995. doi:10.1093/nar/gks1193

54. Brazma A, Hingamp P, Quackenbush J et al (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 29:365–371. doi:10.1038/ng1201-365

55. Rhodes DR, Kalyana-Sundaram S, Mahavisno V et al (2007) Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia 9:166–180

56. Bagger FO, Rapin N, Theilgaard-Mönch K et al (2013) HemaExplorer: a database of mRNA expression profiles in normal and malignant haematopoiesis. Nucleic Acids Res 41:D1034–D1039. doi:10.1093/nar/gks1021

57. Bagger FO, Rapin N, Theilgaard-Mönch K et al (2012) HemaExplorer: a Web server for easy and fast visualization of gene expression in normal and malignant hematopoiesis. Blood 119:6394–6395. doi:10.1182/blood-2012-05-427310

58. Gry M, Rimini R, Strömberg S et al (2009) Correlations between RNA and protein expression profiles in 23 human cell lines. BMC Genom 10:365. doi:10.1186/1471-2164-10-365

59. Uhlen M, Oksvold P, Fagerberg L et al (2010) Towards a knowledge-based human protein atlas. Nat Biotechnol 28:1248–1250. doi:10.1038/nbt1210-1248

60. Fernández-Suárez XM, Galperin MY (2013) The 2013 nucleic acids research database issue and the online molecular biology database collection. Nucleic Acids Res 41:D1–D7. doi:10.1093/nar/gks1297

61. Vigneron N, Stroobant V, Van den Eynde BJ, van der Bruggen P (2013) Database of T cell-defined human tumor antigens: the 2013 update. Cancer Immun 13:15

62. Almeida LG, Sakabe NJ, deOliveira AR et al (2009) CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. Nucleic Acids Res 37:D816–D819. doi:10.1093/nar/gkn673

63. Slamon DJ, Clark GM, Wong SG et al (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 235:177–182

64. Allred DC, Clark GM, Molina R et al (1992) Overexpression of HER-2/neu and its relationship with other prognostic factors change during the progression of in situ to invasive breast cancer. Hum Pathol 23:974–979

65. Fisk B, Blevins TL, Wharton JT, Ioannides CG (1995) Identification of an immunodominant peptide of HER-2/neu protooncogene recognized by ovarian tumor-specific cytotoxic T lymphocyte lines. J Exp Med 181:2109–2117

66. Carter P, Presta L, Gorman CM et al (1992) Humanization of an anti-p185HER2 antibody for human cancer therapy. Proc Natl Acad Sci USA 89:4285–4289

67. Rustici G, Kolesnikov N, Brandizi M et al (2013) ArrayExpress update—trends in database growth and links to data analysis tools. Nucleic Acids Res 41:D987–D990. doi:10.1093/nar/gks1174

68. Amberger J, Bocchini CA, Scott AF, Hamosh A (2009) McKusick's online Mendelian inheritance in man (OMIM). Nucleic Acids Res 37:D793–D796. doi:10.1093/nar/gkn665

69. Flicek P, Ahmed I, Amode MR et al (2013) Ensembl 2013. Nucleic Acids Res 41:D48–D55. doi:10.1093/nar/gks1236

70. Safran M, Dalah I, Alexander J et al (2010) GeneCards version 3: the human gene integrator. database (Oxford) baq020. doi:10.1093/database/baq020

71. Benson DA, Cavanaugh M, Clark K et al (2013) GenBank. Nucleic Acids Res 41:D36–D42. doi:10.1093/nar/gks1195

72. The UniProt Consortium (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013. Nucleic Acids Res 41:D43–D47. doi:10.1093/nar/gks1068

73. Forbes SA, Bindal N, Bamford S et al (2011) COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer. Nucleic Acids Res 39:D945–D950. doi:10.1093/nar/gkq929

74. Holzhütter HG, Kloetzel PM (2000) A kinetic model of vertebrate 20S proteasome accounting for the generation of major proteolytic fragments from oligomeric peptide substrates. Biophys J 79:1196–1205. doi:10.1016/S0006-3495(00)76374-0

75. Kuttler C, Nussbaum AK, Dick TP et al (2000) An algorithm for the prediction of proteasomal cleavages. J Mol Biol 298:417–429. doi:10.1006/jmbi.2000.3683

76. Keşmir C, Nussbaum AK, Schild H et al (2002) Prediction of proteasome cleavage motifs by neural networks. Protein Eng 15:287–296

77. Peters B, Bulik S, Tampe R et al (2003) Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. J Immunol 171:1741–1749

78. Zhang GL, Petrovsky N, Kwoh CK et al (2006) PRED(TAP): a system for prediction of peptide binding to the human transporter associated with antigen processing. Immunome Res 2:3. doi:10.1186/1745-7580-2-3

79. Bhasin M, Lata S, Raghava GPS (2007) TAPPred prediction of TAP-binding peptides in antigens. Methods Mol Biol 409:381–386. doi:10.1007/978-1-60327-118-9_28

80. Saxová P, Buus S, Brunak S, Keşmir C (2003) Predicting proteasomal cleavage sites: a comparison of available methods. Int Immunol 15:781–787

81. Zhang GL, Ansari HR, Bradley P et al (2011) Machine learning competition in immunology—prediction of HLA class I binding peptides. J Immunol Methods 374:1–4. doi:10.1016/j.jim.2011.09.010

82. Lin HH, Ray S, Tongchusak S et al (2008) Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. BMC Immunol 9:8. doi:10.1186/1471-2172-9-8

83. Lin HH, Zhang GL, Tongchusak S et al (2008) Evaluation of MHC-II peptide binding prediction servers: applications for vaccine research. BMC Bioinform 9(Suppl 12):S22. doi:10.1186/1471-2105-9-S12-S22

84. Lundegaard C, Lund O, Nielsen M (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. doi:10.1093/bioinformatics/btn128

85. Nielsen M, Lund O (2009) NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinform 10:296. doi:10.1186/1471-2105-10-296

86. Parker KC, Bednarek MA, Coligan JE (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J Immunol 152:163–175

87. Schuler MM, Nastke M-D, Stevanovikć S (2007) SYFPEITHI: database for searching and T-cell epitope prediction. Methods Mol Biol 409:75–93

88. Hu X, Mamitsuka H, Zhu S (2011) Ensemble approaches for improving HLA class I-peptide binding prediction. J Immunol Methods 374:47–52. doi:10.1016/j.jim.2010.09.007

89. Huang JC, Jojic N (2011) Modeling major histocompatibility complex binding by nonparametric averaging of multiple predictors and sequence encodings. J Immunol Methods 374:35–42. doi:10.1016/j.jim.2010.10.001

90. Nielsen M, Lundegaard C, Blicher T et al (2007) NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. PLoS ONE 2:e796. doi:10.1371/journal.pone.0000796

91. Larsen MV, Lundegaard C, Lamberth K et al (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinform 8:424. doi:10.1186/1471-2105-8-424

92. Frankild S, de Boer RJ, Lund O et al (2008) Amino acid similarity accounts for T cell cross-reactivity and for "holes" in the T cell repertoire. PLoS ONE 3:e1831. doi:10.1371/journal.pone.0001831

93. Pierce BG, Weng Z (2013) A flexible docking approach for prediction of T cell receptor–peptide–MHC complexes. Protein Sci 22:35–46. doi:10.1002/pro.2181

94. Tung C-W, Ziehm M, Kämper A et al (2011) POPISK: T-cell reactivity prediction using support vector machines and string kernels. BMC Bioinform 12:446. doi:10.1186/1471-2105-12-446

95. Saethang T, Hirose O, Kimkong I et al (2013) PAAQD: predicting immunogenicity of MHC class I binding peptides using amino acid pairwise contact potentials and quantum topological molecular similarity descriptors. J Immunol Methods 387:293–302. doi:10.1016/j.jim.2012.09.016

96. Olsen LR, Zhang GL, Keskin DB et al (2011) Conservation analysis of dengue virus T-cell epitope-based vaccine candidates using peptide block entropy. Front Immunol 2:69. doi:10.3389/fimmu.2011.00069

97. Rongcun Y, Salazar-Onfray F, Charo J et al (1999) Identification of new HER2/neu-derived peptide epitopes that can elicit specific CTL against autologous and allogeneic carcinomas and melanomas. J Immunol 163:1037–1044

98. Reinhold B, Keskin DB, Reinherz EL (2010) Molecular detection of targeted major histocompatibility complex I-bound peptides using a probabilistic measure and nanospray MS(3) on a hybrid quadrupole-linear ion trap. Anal Chem 82:9090–9099. doi:10.1021/ac102387t

99. Andersen RS, Kvistborg P, Frøsig TM et al (2012) Parallel detection of antigen-specific T cell responses by combinatorial encoding of MHC multimers. Nat Protoc 7:891–902. doi:10.1038/nprot.2012.037

100. Erlich H (2012) HLA DNA typing: past, present, and future. Tissue Antigens 80:1–11. doi:10.1111/j.1399-0039.2012.01881.x

101. Sette A, Sidney J (1998) HLA supertypes and supermotifs: a functional perspective on HLA polymorphism. Curr Opin Immunol 10:478–482

102. Sette A, Sidney J (1999) Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immunogenetics 50:201–212

103. Robinson J, Halliwell JA, McWilliam H et al (2013) The IMGT/HLA database. Nucleic Acids Res 41:D1222–D1227. doi:10.1093/nar/gks949

104. Helmberg W, Dunivin R, Feolo M (2004) The sequencing-based typing tool of dbMHC: typing highly polymorphic gene sequences. Nucleic Acids Res 32:W173–W175. doi:10.1093/nar/gkh424

105. Gonzalez-Galarza FF, Christmas S, Middleton D, Jones AR (2011) Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. Nucleic Acids Res 39:D913–D919. doi:10.1093/nar/gkq1128

106. Helmberg W (2012) Bioinformatic databases and resources in the public domain to aid HLA research. Tissue Antigens 80:295–304. doi:10.1111/tan.12000

107. Hildesheim A, Apple RJ, Chen C-J et al (2002) Association of HLA class I and II alleles and extended haplotypes with nasopharyngeal carcinoma in Taiwan. J Natl Cancer Inst 94:1780–1789

108. Wang C, Krishnakumar S, Wilhelmy J et al (2012) High-throughput, high-fidelity HLA genotyping with deep sequencing. Proc Natl Acad Sci USA 109:8676–8681. doi:10.1073/pnas.1206614109

109. Boegel S, Löwer M, Schäfer M et al (2013) HLA typing from RNA-Seq sequence reads. Genome Med 4:102. doi:10.1186/gm403

110. Zhang GL, Keskin DB, Reinherz EL, Brusic V (2011) A cDNA microarray for rapid and economical identification of HLA profiles of individuals. In: IEEE international conference on bioinformatics and biomedical work, pp 677–679

111. Pabinger S, Dander A, Fischer M et al (2014) A survey of tools for variant analysis of next-generation genome sequencing data. Brief Bioinform 15:256–278. doi:10.1093/bib/bbs086

112. Buggert M, Norström MM, Czarnecki C et al (2012) Characterization of HIV-specific CD4+ T cell responses against peptides selected with broad population and pathogen coverage. PLoS ONE 7:e39874. doi:10.1371/journal.pone.0039874

113. Bui H-H, Sidney J, Dinh K et al (2006) Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinform 7:153. doi:10.1186/1471-2105-7-153

114. Warren RL, Choe G, Freeman DJ et al (2012) Derivation of HLA types from shotgun sequence datasets. Genome Med 4:95. doi:10.1186/gm396

115. Kim HJ, Pourmand N (2013) HLA haplotyping from RNA-seq data using hierarchical read weighting. PLoS ONE 8:e67885. doi:10.1371/journal.pone.0067885

116. Lundegaard C, Lamberth K, Harndahl M et al (2008) NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. Nucleic Acids Res 36:W509–W512. doi:10.1093/nar/gkn202

117. Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinform Chapter 2:Unit 2.3. doi:10.1002/0471250953.bi0203s00

118. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. doi:10.1093/molbev/mst010

119. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform 5:113. doi:10.1186/1471-2105-5-113

120. Bui H-H, Sidney J, Li W et al (2007) Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinform 8:361. doi:10.1186/1471-2105-8-361

121. Olsen LR, Kudahl UJ, Simon C et al (2013) BlockLogo: visualization of peptide and sequence motif conservation. J Immunol Methods 400–401:37–44. doi:10.1016/j.jim.2013.08.014

122. Kawashima I, Hudson SJ, Tsai V et al (1998) The multi-epitope approach for immunotherapy for cancer: identification of several CTL epitopes from various tumor-associated antigens expressed on solid epithelial tumors. Hum Immunol 59:1–14

123. Forbes SA, Bhamra G, Bamford S et al (2008) The catalogue of somatic mutations in cancer (COSMIC). Curr Protoc Hum Genet Chapter 10:Unit 10.11. doi:10.1002/0471142905.hg1011s57

124. Firth HV, Richards SM, Bevan AP et al (2009) DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. Am J Hum Genet 84:524–533. doi:10.1016/j.ajhg.2009.03.010

125. Magrane M (2011) UniProt knowledgebase: a hub of integrated protein data. Database (Oxford) bar009. doi:10.1093/database/bar009

126. Andersen MH, Becker JC, Thor Straten P (2005) Regulators of apoptosis: suitable targets for immune therapy of cancer. Nat Rev Drug Discov 4:399–409. doi:10.1038/nrd1717

127. Lee MJ, Ye AS, Gardino AK et al (2012) Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. Cell 149:780–794. doi:10.1016/j.cell.2012.03.031

128. Liu J, Ewald BA, Lynch DM et al (2006) Modulation of DNA vaccine-elicited CD8+ T-lymphocyte epitope immunodominance hierarchies. J Virol 80:11991–11997. doi:10.1128/JVI.01348-06

129. Ranieri E, Kierstead LS, Zarour H et al (2000) Dendritic cell/peptide cancer vaccines: clinical responsiveness and epitope spreading. Immunol Investig 29:121–125

130. Henderson RA, Finn OJ (1996) Human tumor antigens are ready to fly. Adv Immunol 62:217–256

131. Reid A, Vidal L, Shaw H, de Bono J (2007) Dual inhibition of ErbB1 (EGFR/HER1) and ErbB2 (HER2/neu). Eur J Cancer 43:481–489. doi:10.1016/j.ejca.2006.11.007

132. Franceschini A, Szklarczyk D, Frankild S et al (2013) STRING v9.1: protein–protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41:D808–D815. doi:10.1093/nar/gks1094

133. Kawaguchi Y, Kono K, Mimura K et al (2007) Targeting EGFR and HER-2 with cetuximab- and trastuzumab-mediated immunotherapy in oesophageal squamous cell carcinoma. Br J Cancer 97:494–501. doi:10.1038/sj.bjc.6603885

134. Gutierrez C, Schiff R (2011) HER2: biology, detection, and clinical implications. Arch Pathol Lab Med 135:55–62. doi:10.1043/2010-0454-RAR.1

135. Subramanian A, Tamayo P, Mootha VK et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102:15545–15550. doi:10.1073/pnas.0506580102

136. Chatr-Aryamontri A, Breitkreutz B-J, Heinicke S et al (2013) The BioGRID interaction database: 2013 update. Nucleic Acids Res 41:D816–D823. doi:10.1093/nar/gks1158

137. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30

138. Croft D, O'Kelly G, Wu G et al (2011) Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res 39:D691–D697. doi:10.1093/nar/gkq1018

139. Caoili SEC (2010) Benchmarking B-cell epitope prediction for the design of peptide-based vaccines: problems and prospects. J Biomed Biotechnol 2010:910524. doi:10.1155/2010/910524

140. Adams GP, Weiner LM (2005) Monoclonal antibody therapy of cancer. Nat Biotechnol 23:1147–1157. doi:10.1038/nbt1137

141. Porter DL, Levine BL, Kalos M et al (2011) Chimeric antigen receptor-modified T cells in chronic lymphoid leukemia. N Engl J Med 365:725–733. doi:10.1056/NEJMoa1103849

142. El-Manzalawy Y, Honavar V (2010) Recent advances in B-cell epitope prediction methods. Immunome Res 6(Suppl 2):S2. doi:10.1186/1745-7580-6-S2-S2

143. Möller S, Croning MD, Apweiler R (2001) Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics 17:646–653

144. Gajewski TF (2011) Molecular profiling of melanoma and the evolution of patient-specific therapy. Semin Oncol 38:236–242. doi:10.1053/j.seminoncol.2011.01.004

145. Galon J, Franck P, Marincola FM et al (2012) Cancer classification using the Immunoscore: a worldwide task force. J Transl Med 10:205. doi:10.1186/1479-5876-10-205

146. Galon J, Pagès F, Marincola FM et al (2012) The immune score as a new possible approach for the classification of cancer. J Transl Med 10:1. doi:10.1186/1479-5876-10-1

147. Pagès F, Kirilovsky A, Mlecnik B et al (2009) In situ cytotoxic and memory T cells predict outcome in patients with early-stage colorectal cancer. J Clin Oncol 27:5944–5951. doi:10.1200/JCO.2008.19.6147