

Antigen processing and presentation in cancer immunotherapy

Maxwell Y Lee, Jun W Jeon, Cem Sievers, Clint T Allen

To cite: Lee MY, Jeon JW, Sievers C, *et al.* Antigen processing and presentation in cancer immunotherapy. *Journal for ImmunoTherapy of Cancer* 2020;**8**:e001111. doi:10.1136/jitc-2020-001111

Accepted 19 July 2020

ABSTRACT

Background Knowledge about and identification of T cell tumor antigens may inform the development of T cell receptor-engineered adoptive cell transfer or personalized cancer vaccine immunotherapy. Here, we review antigen processing and presentation and discuss limitations in tumor antigen prediction approaches.

Methods Original articles covering antigen processing and presentation, epitope discovery, and in silico T cell epitope prediction were reviewed.

Results Natural processing and presentation of antigens is a complex process that involves proteasomal proteolysis of parental proteins, transportation of digested peptides into the endoplasmic reticulum, loading of peptides onto major histocompatibility complex (MHC) class I molecules, and shuttling of peptide:MHC complexes to the cell surface. A number of T cell tumor antigens have been experimentally validated in patients with cancer. Assessment of predicted MHC class I binding and total score for these validated T cell antigens demonstrated a wide range of values, with nearly one-third of validated antigens carrying an IC₅₀ of greater than 500 nM.

Conclusions Antigen processing and presentation is a complex, multistep process. In silico epitope prediction techniques can be a useful tool, but comprehensive experimental testing and validation on a patient-by-patient basis may be required to reliably identify T cell tumor antigens.

BACKGROUND

Immunotherapy has revolutionized cancer treatment. At present, various therapeutics for the enhancement or replacement of T cell antitumor immunity have Food and Drug Administration approval for a variety of cancers.^{1,2} The core principles driving current immunotherapies began to take shape in the late 20th century with the elucidation of the mechanism by which T cells recognize antigens presented by the major histocompatibility complexes (MHC).³ T cell antigens are specific peptide sequences that are recognized by CD8 or CD4 T cells when presented on MHC I or MHC II molecules, respectively. Neoantigens are peptides derived from tumor-specific mutations which have not been previously recognized by the body's immune system. Experimental evidence supports that tumor antigen recognition by T cells

is critical for antitumor immunity and that cancers can evade such immunity by immunodominance, display of immune checkpoints, or immunoediting for loss of specific tumor antigens.⁴ Some cancers are composed of subclonal tumor cell populations that harbor defects in antigen processing and presentation, suggesting that these cancers may not be curable despite maximal activation of T cell immunity.⁵ For those cancers without defects in antigen processing and presentation, knowledge of immunogenic T cell antigens may assist in the engineering of immunotherapeutics designed to control and eradicate cancer.⁶

For a peptide to serve as a natural T cell antigen, the parental protein must first be processed so that its peptides can be presented on an MHC I molecule. Natural processing and presentation of an antigen is a complex and multifactorial process, which is still subject to active research. The available peptide repertoire is constrained by how proteins are cleaved, trimmed, loaded onto MHC, and translocated to the cell surface. Specificity within this process is likely an explanation for why the vast majority of peptides predicted to be high-affinity MHC binders are unable to elicit T cell responses.⁷ Validated T cell antigens are often tumor-specific as they are derived from expressed tumor-specific mutated genes. An alternative theory for why many peptide candidates do not elicit T cell responses and, thus, are not antigens, is that these mutated peptides are very similar to their non-mutated counterparts, leading to central tolerance and elimination of potentially reactive T cells.

Despite challenges, there has been success in using T cells to target specific antigen–MHC combinations. The cancer testis antigen NY-ESO-1 is one of the best examples.⁸ As a validated, highly immunogenic antigen, the use of the NY-ESO-1 T cell experimental system has set the stage for an abundance of T cell and cancer biology research over the past two decades.⁹ Furthermore, its use in clinical trials has led to advances in



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

NIDCD, National Institutes of Health, Bethesda, Maryland, USA

Correspondence to

Dr Clint T Allen;
clint.allen@nih.gov

adoptive cell transfer immunotherapy, including the first successful treatment of a non-melanoma tumor using T cell receptor (TCR)-engineered T cells.⁹ The successful application of NY-ESO-1-specific T cells suggests that epitope-specific, T cell-based immunotherapy may allow personalized immunotherapy. In this review, we focus on the mechanisms behind natural processing and presentation of T cell antigens and summarize current methods and limitations of antigen prediction and validation.

BIOLOGY OF ANTIGEN PRESENTATION

MHC molecule

Class I MHC molecules are polypeptides composed of a polymorphic heavy chain that associates with a constant β 2-microglobulin (β 2m) subunit, whereas class II MHC molecules are composed of an α and β polypeptide chain. Class I MHC molecules are generally recognized by CD8 cytotoxic T cells, whereas class II MHC molecules are recognized by CD4 T cells. For the purposes of this review, we will only discuss class I MHC molecules and their interactions with CD8 T cells. The mechanisms underlying epitope prediction and discovery for class II MHC molecules, while important, are beyond the scope of this article. The peptide binding groove of an MHC I molecule preferentially binds 8–11mer peptides. The top surface of the groove, where both the MHC protein and bound peptide are exposed, is the portion of the peptide:MHC complex detected by a TCR.¹⁰

Humans possess three MHC I genes: human leucocyte antigen A (HLA-A), HLA-B, and HLA-C. These genes encode highly polymorphic proteins, particularly in the peptide interacting region, allowing different HLA molecules to bind different sets of peptides.¹¹ This presents a challenge to the prediction of MHC binding peptides as prediction algorithms must take into account HLA-specific binding preferences.

From proteins to peptides

For a peptide to be presented on an MHC I molecule, it must be processed by the cellular antigen processing machinery. This begins with proteolysis of parental protein precursors. Exogenous proteins that are internalized by the cell through processes such as phagocytosis and endocytosis are processed in the late endosome and primarily presented on MHC II. Endogenous proteins are processed by the proteasome and are presented mainly on MHC I.¹²

Proteasomes are proteases that break down misfolded, damaged, aberrant, or ubiquitinated proteins. The core particle of a proteasome is a barrel-shaped structure composed of two outer α -rings and two inner β -rings.¹³ Variants of the β -subunits are interferon gamma (IFN- γ) inducible and can replace their constitutive counterparts.¹⁴ When these are expressed and active, they form a complex known as the immunoproteasome. The constitutive proteasome is capped at each end by a 19S complex that helps proteins into the proteasome

lumen by deubiquitinating and unfolding them. The immunoproteasomes are capped by an alternative 11S complex that is also IFN- γ inducible.¹⁴ As a result of the different proteolytic subunits and capping structures, the cleavage site specificity differs between constitutive and immunoproteasomes.¹⁵

The characterization of proteasomal activity includes two basic approaches. In vitro study involves co-incubation of parental proteins with proteasomes. The resulting peptides are analyzed using mass spectrometry.¹⁶ Several proteolysis prediction techniques are trained on in vitro datasets.¹⁷ In vivo proteolysis study involves the measurement of peptides that have been eluted from an peptide:MHC complexes and subjected to mass spectrometry.¹⁸ A limitation of this strategy is that only a small proportion of cleavage sites can be identified in this manner as most cleaved peptides do not end up being presented on MHC complexes. Furthermore, the influence of other cellular proteases cannot be deconvoluted from pure proteasome function. Several prediction techniques are trained on in vivo datasets and are generally superior at predicting in vivo proteolysis as compared with prediction techniques trained on in vitro datasets.¹⁹

Peptide loading onto MHC I

The peptide loading complex (PLC) is a multi-subunit, endoplasmic reticulum (ER) membrane complex which coordinates peptide translocation into the ER, editing, and loading onto MHC I molecules. One of the core components is the transporter associated with antigen processing (TAP), which is made up of the TAP1 and TAP2 subunits. Each subunit contains a transmembrane domain and a nucleotide binding domain which interact to mediate a channel for peptide movement between the cytosol and ER lumen.¹² TAP preferentially translocates peptides with 9–16 residues, although longer peptides up to 25–30 amino acids in length can also be translocated with lower efficiency.²⁰

The specificity of TAP has been well studied. On the C-terminal end of the peptide, TAP selects for the presence of an aromatic, hydrophobic, or positively charged terminal amino acid such as phenylalanine, tyrosine, arginine, or leucine.²¹ On the N-terminal end, the first three residues have significant effects on peptide binding. Aromatic and hydrophobic side chains are favored and an N-terminal arginine is optimal.²² Proteasomal specificity at the C-terminal residue is relatively non-specific and aligns with the restraints of both TAP and MHC I molecules.²³ However, the N-terminal ends of TAP-translocated peptides require further processing within the ER to conform to MHC I binding requirements.²⁴ ER aminopeptidase 1 (ERAP1) and ERAP2 are ER luminal components tasked with the intraluminal processing of peptides.²⁵ ERAP1 trims N-terminal amino acids to create 8–11mer peptides capable of fitting within the MHC I binding groove.²⁶ Peptides undergo structural changes when reaching this size, preventing ERAP1 from further trimming.²⁷

Within the PLC, TAP is associated with a number of other proteins. Tapasin interacts with TAP and serves to recruit MHC I- β 2m complexes to the PLC.²⁸ Tapasin is further linked to ERp57 and calreticulin, which mediate the folding of newly synthesized MHC molecules.¹² Following translocation of peptides by TAP, peptides are released into a molecular basket, confined by two editing complexes composed of tapasin, CRT, ERp57, and MHC I. Peptides are then further edited by ERAP1 and loaded onto MHC I molecules.²⁹ Once loaded onto MHC I, peptides undergo additional processing and quality control mediated by tapasin and TAP binding protein-related protein (TAPBPR).³⁰ TAPBPR associates with empty MHC I molecules, stabilizing their structure. The choice of peptide is based on the stability of the final MHC I-peptide complex, and TAPBPR is released once a high-affinity MHC I-peptide complex is stably achieved.³¹ A similar mechanism is thought to underlie the function of tapasin in the PLC.^{29, 31} If the affinity for the bound peptide is high and the resulting complex is stable, the MHC I-peptide complex will be transported through the Golgi apparatus and to the cell surface.¹² The peptide specificity associated with each step of this process beginning from proteasome proteolysis to MHC I binding is what defines natural processing and presentation of epitopes. For a peptide to be clinically relevant, experimental validation should strive to mimic this process.

Cross presentation

Endogenous proteins are processed through the proteasome pathway and loaded onto MHC I molecules for detection by CD8 T cells, whereas exogenous proteins are processed through the endosomal pathway and presented on MHC II molecules for detection by CD4 T cells. Tumor cells alone are unable to prime an effective, antigen-specific, CD8 cytotoxic T cell response.³² How do CD8 T cells targeting MHC I-restricted antigens become primed and activated against antigens that originated as exogenous proteins?

Cross presentation refers to the presentation of exogenous proteins on MHC I.³³ In 1990, Rock and colleagues found that soluble antigens in the extracellular fluid could be internalized, processed, and presented on MHC I molecules in specialized antigen presenting cells (APC).³⁴ One crucial APC is the dendritic cell (DC). Mechanistic studies in mice have demonstrated that the development of antigen-specific T cell responses depends on functional APC.³²

DCs can acquire antigens from multiple sources. Dying cells release antigens and they are an important source of antigenic materials.³⁵ Antigens ingested by DC through micropinocytosis, endocytosis, or phagocytosis are efficiently cross-presented.³⁶ While the exact mechanisms are unclear, it is hypothesized that the corresponding downstream intracellular compartments, such as the phagosome or endosome, contain cellular proteins, which are part of the MHC I presentation pathway.³³ Additionally, DCs can internalize cytosolic and membrane materials

from live cells to prime T cell responses.³⁷ Possible mechanisms for this transfer of antigens includes ingestion of secreted antigens and trogocytosis, where one cell 'nibbles' bits off another.^{38, 39}

Once a DC has acquired an antigen, processing and presentation occur through either a proteasome-independent vacuolar pathway or a proteasome-dependent cytosolic pathway. To establish a diverse repertoire of T cells that can recognize tumor-presented antigens, an ideal cross-presentation pathway would mimic the natural processing and presentation machinery of a tumor cell. In the vacuolar pathway, proteins are lysed into peptides within the endocytic compartments and then loaded onto MHC I molecules without ever encountering the cytosol.^{33, 40} Therefore, the proteasome and TAP are never encountered, meaning that the repertoire of selected epitopes may differ from those presented on a tumor.⁴¹ In the proteasome-dependent pathway, proteins progress through the cytoplasm and ER.⁴² Several studies have confirmed that exogenous antigens appear in the cytosol.⁴³ Once in the cytosol, the proteins undergo the same process described earlier. Studies have verified that peptides generated by the immunoproteasome from conventional endogenous antigens are identical to immunoproteasome-generated peptides from exogenously delivered proteins.⁴⁴ Presentation of tumor antigens by APC to naive T cells in lymphoid tissues then primes T cells for clonal expansion and cytotoxic effector function.

Antigen identification in cancer therapy

Many sources of cancer antigens exist. Epitopes derived from single nucleotide variations (SNVs), insertions/deletions (INDELs), transcript splice variants, gene fusions, and endogenous retroviral elements can all serve as tumor-specific antigens.^{45, 46} Some immunotherapies do not rely on the identification of tumor antigens for use, while other forms of immunotherapy do require knowledge of antigen and possibly HLA type. Immune checkpoint blockade immunotherapy is widely used but induces durable tumor control in only one subset of patients and does not require a priori knowledge of antigens.⁴⁷ Adoptive transfer of bulk cultured tumor infiltrating lymphocytes (TIL) as an immunotherapeutic approach has produced durable regression of malignancy in a limited subset of patients, and this approach does not rely on the identification of specific tumor antigens.⁴⁸ Conversely, adoptive cell transfer of TCR-engineered T cells does rely on the identification of a specific antigen target, the HLA restriction, and the TCR specific for this antigen:MHC complex. Autologous T cells are engineered to express TCRs that target antigen:MHC complexes presenting tumor-specific neoantigens.⁴⁹ Personalized cancer vaccine therapies similarly rely on the identification of tumor-specific neoantigens.^{50, 51} However, the utility of these potential therapies is constrained by the difficulties underlying the discovery of tumor-specific epitopes that elicit T cell responses.⁵² Here, we review the commonly

utilized epitope discovery techniques and key considerations underlying these methods.

Identifying tumor-specific epitopes and epitope-reactive T cells

A general approach to identify neoepitopes and accompany reactive T cells from humans is a reverse immunology approach based on tumor sequencing. A section of surgically resected tumor or metastasis is cultured in media containing cytokines selective for lymphocyte growth, notably IL-2, to expand TIL in vitro. In addition, another portion of the tumor is sent for genome sequencing to identify genetic mutations and the corresponding repertoire of potential neoepitopes. The resulting neoepitopes can be further analyzed in silico to predict the best candidates for MHC I cell surface presentation. Next, T cells are cocultured with APC expressing the antigens of interest and analyzed for activation signals. T cells reactive against antigens can be assessed further via TCR sequencing or expanded as primary effector cells.

Tran and colleagues have successfully applied this strategy to identify T cells reactive against mutant KRAS G12D in a patient with metastatic colorectal cancer.⁴⁸ Three metastatic lung lesions were resected and used to generate TIL cultures and whole exome sequencing (WES) data. In total, 61 mutations were identified. Cultures of TIL were screened against all identified mutant neoepitopes expressed by the tumor. For each mutation, a minigene encoding the mutation flanked by 12 amino acids on either side was generated and synthesized in tandem to create tandem minigene (TMG) constructs. Five total TMG constructs were made, in vitro transcribed into RNA, and electroporated into autologous DCs to undergo natural processing and presentation of epitopes. TIL cultures were reactive against TMG-1, so the corresponding peptides were synthesized and pulsed onto DCs, followed by coincubation with TIL. Adoptive transfer of four T cell clonotypes that demonstrated antigen-specific reactivity resulted in objective regression of all lung metastases.⁴⁸ The TMG approach to identify tumor-specific T cell antigens is highly comprehensive as it interrogates responses from all possible neoantigens.

Methodological considerations

Sequencing and mutation calling

The identification of expressed genetic mutations is critical to neoepitope discovery. After tumor and normal tissue samples are sequenced, typically using next-generation sequencing (NGS), the presence of somatic genetic variants is assessed through the application of variant calling pipelines. Comparisons of different pipelines have been summarized.⁵³

In the context of neoepitope prediction/analysis, somatic variants can be classified into SNVs, INDELS, and structural variants. Somatic SNVs are capable of generating immunogenic neoepitopes through amino acid substitutions, and reactive T cells to SNV-derived neoepitopes are detected with higher frequency in cancers with

higher mutational burdens.^{54–55} Less well studied are INDELS, which can lead to frameshift mutations that create neo open reading frames (neo-ORF). Frameshift INDELS lead to higher proportions of neoantigens per mutation compared with SNVs. Additionally, INDELS more commonly result in mutation-specific predicted high-affinity binding neoantigens compared with that of SNV mutations.⁵⁶ Several studies have discovered T cell clones reactive against frameshift INDELS.^{57–58} While tumors can harbor hundreds of neoantigens, the number of mutant peptides that can be tested in parallel is limited by time and cost. Therefore, prioritization of candidate antigens for experimental validation is important when conducting large-scale experiments.

As the advent of NGS has made the study of genetics-based precision medicine more clinically feasible, investigators must also be aware of its limitations. Study of putative neoantigens to date has primarily focused on the coding region of the genome, given that antigens must be transcribed and translated into protein that can be processed into peptides. The combination of WES and RNA-seq is commonly employed for consistent variant analysis and provides consistent results across different sequencing centers.⁵⁹ It is worth mentioning that many tumor-rejection antigens have been identified from non-canonical protein sequences derived from retained introns, non-coding regions, alternative ORF, or proteasome-mediated splice variants.^{60–62} To capture these possibilities, studies may use a combination of WGS, WES, RNA-seq, and mass spectrometry.

Lastly, there are considerations related to the method of tissue sample preservation. Formalin-fixed, paraffin-embedded (FFPE) tissues are an invaluable resource, but the process of sample fixation may modify nucleic acids to cause inaccurate or incomplete capture of information. Studies have shown that FFPE processing of tissue can lead to reduced transcript detection in RNAseq data compared with fresh frozen (FF) samples.⁶³ Currently, the use of FF tumor samples is the gold standard.

Antigen testing

The question of natural processing and presentation is of great concern as most experimental systems permit T cell responses against non-naturally processed and presented epitopes.⁶⁴ In the study by Tran *et al*, natural processing and presentation of epitopes was ensured by using TMG constructs electroporated into APCs, meaning that the in vitro transcribed TMG mRNA directly entered the cytosol and was translated into protein, and the protein was processed in a proteasome-dependent manner. Coincubation of APCs with synthetic long (15–30mer) peptides is an alternative strategy for APCs to acquire and present exogenous antigens.⁶⁵ When designing peptide pulsing experiments, it is important to consider the peptide length. In a comparison of overlapping 24-mer long peptide versus whole protein, Zhang and colleagues found that 24-mer peptides trafficked to both endosomes and cytosol to activate CD8 and CD4 T cells, whereas whole protein (HIV

nef, 206 amino acids) was found only in the endosomal compartment and not in the cytosol, activating CD4 but not CD8 T cells.⁶⁶ Pulsing of minimal epitope or short peptides of length 8–16 may also be misleading as this does not guarantee natural processing by the immunoproteasome, transport by TAP, or trimming by ERAP1. In a study of the TEL-AML1 translocation, an immunogenic minimal epitope did not illicit T cell responses when pulsed into APCs as part of a long peptide, indicating that the minimal epitope was not naturally processed and presented.⁶⁷ Electroporation of mRNA encoding the full-length protein of interest (mimicking endogenous peptide) into APCs or pulsing APCs with long peptide (mimicking exogenous peptide) that require processing may be the most effective means of ensuring that identified antigens are naturally processed and presented from the parental protein. Small 8–11mer epitopes can then be used directly after narrowing down candidates to confirm the identity of the minimal epitope.

While TCR signaling strength is generally thought to be proportional to peptide–MHC affinity, TCR binding of peptide:MHC complexes does not always result in T cell activation.⁶⁸ Greater than 15% of T cell clones do not produce antigen-specific responses even after peptide–MHC recognition due to structural mechanisms.⁶⁸ The existence of naturally processed and presented epitope:MHC complexes that do not elicit T cell activation introduces a wrinkle into methods that attempt to predict antigenic targets based solely on MHC binding and tumor cell antigen processing. Several methods have been developed to directly characterize T cell specificities. Kula and colleagues developed T-scan as a high-throughput, unbiased method for the discovery of antigens on a genome scale. Using this system, T cell pools are used to recognize cells virally transduced with antigen libraries. Functionally recognized cells are then isolated using a reporter for granzyme B activity and sequenced, allowing identification of cognate antigens.⁶⁹ Studies by Baltimore and colleagues use a similar schema of querying antigen libraries, but use alternative reporter mechanisms of trogocytosis or activation of synthetic signaling molecules to perform high-throughput, unbiased epitope discovery.^{70–71} The application of these methods to screening of autologous TIL against patient-specific antigen libraries or the screening of candidate TCRs for off-target effects holds promise. Of note, although many T cell clones overlap between tumor and blood, tumors display increased clonality and harbor-enriched clones which are absent or rare in the blood.⁷² Therefore, TIL should be prioritized over PBMC as the preferred T cell source for experimental validations.

In silico prediction algorithms

The advent of in silico epitope prediction algorithms began with the hypothesis that certain peptide sequence motifs were more likely to bind to an MHC molecule compared with others.⁷³ These motifs were generally based on primary anchor residues that highly influenced

binding and secondary anchor positions that were less important for binding.⁷⁴ For example, positions 2 and 9 serve as anchor residues in HLA-A*02:01 and are most often occupied by small hydrophobic amino acids such as valine, leucine, and isoleucine.⁷⁵ From these motifs, matrices were developed which assigned a numerical score corresponding to the contribution of each amino acid at each position; values for each amino acid within a peptide were then combined to generate a score for each peptide:MHC combination.⁷⁶ By design, these models assumed that the contribution of an amino acid at a particular position was independent of its neighboring positions. This may not be true.⁷⁷

To handle these interactions and to improve predictions over motif-based models, researchers turned to machine learning algorithms. The machine learning algorithms used for T cell epitope prediction are trained on large-scale datasets, resulting in a model that deciphers patterns from a known set of input data and applies them to new data to generate predictions. Datasets used for T cell epitope prediction can be classified as datasets of peptides with known binding affinity to MHC molecules and datasets containing peptides that either bind or do not bind to MHC molecules.⁷⁸ Models trained on the combination of MHC binding and ligand elution data show superior performance compared with models trained on binding datasets alone.⁷⁹ An additional benefit of machine learning is that methods can be developed to predict binding for rare HLA alleles using data from well-known HLA alleles (pan-specific predictors).⁷⁹ This is important because rare HLA alleles often do not have sufficient data to train their own prediction algorithms. The quality of the available datasets determines the performance of the machine learning algorithms. Several databases compile epitope data in a consistent manner to enable proper training of machine learning algorithms.⁸⁰

Most in silico epitope prediction algorithms aim to predict binding affinities of peptides to MHC I molecules. This is based on the assumption that peptide binding to MHC is the single most selective step in the antigen presentation pathway.^{77–79} Comparison of a combined proteasome/TAP/MHC affinity predictor with a predictor based on MHC affinity alone determined that the MHC affinity predictor offers higher sensitivity whereas combined predictor offers higher specificity. For a hypothetical protein of 300 amino acids, one would need to functionally test 11 peptides to identify a new T cell-reactive epitope with 90% likelihood when using a combined predictor versus 13 peptides when using a MHC affinity predictor.⁸¹ Furthermore, widely used prediction algorithms do not incorporate peptide:MHC complex stability, which is a more important determinant of immunogenicity than binding affinity.⁸² Incorporation of peptide dissociation data into current algorithms may improve prediction performance.⁸³ Current immune epitope database (IEDB) guidelines recommend using MHC binding predictions alone by default and using prediction scores that

incorporate antigen processing as an additional filter if resources require limiting.⁸⁴

The ways in which researchers use MHC epitope prediction data are varied. Studies of MHC I binding affinity and T cell epitope recognition have shown that predicted MHC binding IC_{50} of 500 nM or less results in approximately 90% sensitivity for identification of true T cell reactive epitopes.⁸⁵ Genetic differences between HLA alleles create disparities; some alleles are better able to bind oncogenic mutations while others are predicted to have very few peptide binders with $IC_{50} < 500$ nM.⁸⁶ Therefore, the rank of an epitope's MHC binding affinity in comparison with all other possible epitopes (percentile rank) has been suggested as an alternative. When generating predictions across multiple HLA alleles, percentile ranks are preferred to avoid bias towards certain HLA alleles.⁸⁷

In the development of a personalized melanoma vaccine, Ott *et al* utilized NetMHCpan to identify up to 30 patient-specific MHC I-restricted high-affinity binders derived from mutant peptides to include in each vaccine.⁸⁸ The final list of epitopes was chosen based on a rank list that prioritized neo-ORFs over SNVs, mutations in anchor residues over secondary residues, and low MHC IC_{50} . For each patient, up to 20 putative neoantigens were ultimately incorporated into a personalized vaccine and delivered. All neoantigens had predicted MHC $IC_{50} < 500$ nM or rank in the top one percentile. Ex vivo analyses demonstrated that 15 of 91 (16%) predicted immunizing peptides-induced CD8 antigen-specific T cell responses across six patients. The total number of predicted epitopes that can be included in each vaccine is limited, but the use of MHC I binding prediction algorithms with astute prioritization criteria generated sufficient specificity to generate T cell reactive epitopes for each patient. Of six vaccinated patients, four were recurrence-free 25 months after vaccination and the remaining two experienced complete tumor regression after treatment with anti-programmed death receptor-1 (PD-1) therapy.⁸⁸

Assessment of in silico prediction

We utilized the IEDB proteasomal cleavage/TAP transport/MHC class I combined predictor tool to retrospectively generate total prediction scores and MHC binding affinities of a list of known and validated epitopes derived from the Cancer Antigenic Peptide Database (<https://caped.icp.ucl.ac.be/>).⁸⁹ The total score is a combined score incorporating proteasomal cleavage, TAP transport, and MHC binding predictions where higher scores represent higher efficiency for MHC presentation.⁹⁰

Validated tumor-specific neoantigens are summarized in [table 1](#) and validated tumor-associated antigens derived from germ-line or differentiation antigens are summarized in [table 2](#). Tumor-specific and tumor-associated antigens displayed similar MHC IC_{50} and total score profiles and were combined for further analyses. A significant proportion of epitopes displayed

weak predicted binding ($IC_{50} > 500$ nM, 30.3%) to their respective MHC alleles, with several epitopes exhibiting very weak predicted binding ($IC_{50} > 10,000$ nM, 9.7%) ([figure 1A](#)). One possible explanation for why low-affinity peptides elicit strong T cell responses is that these weakly immunogenic antigens have not been subjected to immunoediting and therefore represent the immunodominant antigen following elimination of more immunogenic targets. When considering discovery of tumor-specific putative neoantigens, nearly all studies utilized a comprehensive querying technique that did not use, or minimally used, epitope prediction algorithms. Methods that heavily incorporated epitope prediction methods were also successful, but rarer in this database. Robbins *et al* tested the top 55 predicted 9mer or 10mer MHC I binders from a melanoma cell line and demonstrated that the 5th, 18th, 19th, and 39th ranked peptides elicited T cell responses.⁷ The hit rate of 4/55 and seemingly random rank of peptides validated as antigens raises the question of whether peptide prediction algorithms perform better than random testing. Unfortunately, it is impossible to know how many similar studies failed to recognize successful epitopes. The only solution to achieve maximum sensitivity for detecting antigens is comprehensive screening of all possible neoantigens, which should be used when resources are available.

There was high correlation between MHC IC_{50} and total prediction score in each group ([figure 1B](#)), indicating that use of MHC binding predictions alone compared with total prediction scores is likely to generate similar results. When comparing validated neoantigens with their wild-type (WT) counterparts, many neoantigens had predicted MHC I binding affinities and total prediction scores that were similar to that of the WT sequence ([figure 1C](#)). Since WT sequences are widely presented throughout the body, central T cell tolerance should have protected the body against auto-immunity. Shaping of T cell repertoires through thymic selection is a critical step in the determination of tolerance on exposure to various antigen:HLA complexes.⁹¹ The simplest experimental models of thymic negative selection have demonstrated that negative selection is specific enough that single amino acid substitutions in self-peptides can effectively be recognized as foreign by T cells and serve as neoantigens.⁹²

Why predictions are inconsistent

The stability of the peptide:HLA class I complex may determine presentation to and development of an activated T cell clone to a greater degree than simple peptide binding affinity.⁹³ The algorithms detailed earlier do not incorporate stability as a variable, illustrating one deficiency of current, widely utilized modeling approaches. Weighted inclusion of stability in predicting immunogenic epitopes outperformed predicting based on binding affinity alone, and such parameters may be incorporated into future prediction pipelines.⁸³

Table 1 Table of mutated tumor-specific antigens from the Cancer Antigenic Peptide Database

Gene/protein	Tumor type	T cell origin	HLA restriction element	Mutant peptide sequence	WT peptide sequence	Reference	Total score mutant	Total score WT	MHC IC ₅₀ (nM) mutant peptide	MHC IC ₅₀ (nM) WT peptide	T cells reactive to WT peptide?
alpha-actinin-4	LC	Tumor	A2	FIASNGV/KLV	FIASGV/KLV	Echchakir, 2001	-1.14	-1.41	280.8	379.8	No
BCR-ABL fusion protein (b3a2)	CML	BM	B8	GFKOSSKAL	-	Yotnda, 1998	-1.5		2679		
BCR-ABL fusion protein (b3a2)	CML	BM	A2	SSKALQRPV	-	Yotnda, 1998	-3.48		23712.2		
beta-catenin	M	Tumor	A24	SYLDSGIHF	SYLDSGIHS	Robbins, 1996	0.92	-4.11	74.7	25856.4	Yes at high conc
CASP-5	Various	Blood	A2	FLIIVQNTM	-	Schwitalle, 2004	-0.39		64		
CASP-8	HNC	Blood	B35	FPSDSWCYF	-	Mandruzzato, 1997	1.8		2.6		
CDK12	M	Tumor	A11	CILGKLFTK	CILGELFTK	Robbins, 2013 ⁷	0.18	0.05	15.4	21.3	No
CDK4	M	Blood	A2	ACDPHSGHFV	ARDPHSGHFV	Wolfe, 1995	-2.98	-3.18	18383.1	27546.2	No
CDKN2A	M	Tumor	A11	AVCPWTWLR	-	Huang, 2004 ⁵⁷	0.4		37.2		
CLPP	M	LN	A2	ILDKVLVHL	ILDKVLVHP	Corbière, 2011	0.67	-2.91	16.4	5424	Yes at high conc
CSNK1A1	M	Tumor	A2	GLFGDIYLA	GSGFDIYLA	Robbins, 2013 ⁷	-0.05	-1.65	6.1	625.8	No
EFTUD2	M	Blood	A3	KILDAVAQK	KILDAVAQE	Lennerz, 2005	-0.31	-3.62	41.8	9290.9	No
Elongation factor 2	LC	Blood	A68	ETVSEQSNV	ETVSEESNV	Hogan, 1998	-2.7	-2.99	5755.2	9457.9	No
ETV6-AML1 fusion protein	ALL	BM	A2	RIAECILGM	-	Yotnda, 1998	-1.02		188.2		
FLT3-ITD	AML	Blood	A1	YVDFREYEEY	YVDFREYED	Graf, 2007	1.22	-2.42	18.6	361.2	No
FNDC3B	CLL	Blood	A2	VYMSWAPPV	VL-SWAPPV	Rajasagi, 2004	0.2	-0.12	5.9	15.4	No
GAS7	M	Tumor	A2	SLADEAEVYL	SLADEAEVHL	Robbins, 2013 ⁷	0.67	0.02	25	111	No
GNMB	M	Blood	A3	TLDWLLQTPK	TLGWLLQTPK	Lennerz, 2005	-1.27	-0.84	86	57.7	No
HAUS3	M	Tumor	A2	ILNAMIKI	ILNAMITKI	Robbins, 2013 ⁷	-0.34	-0.3	44.4	50.5	No
HSDL1	OC	Tumor	Cw14	CYMEAVL	CYMEAVLAL	Wick, 2013	-0.18	1.37	147.9	4.5	No
hsp70-2	RCC	Tumor	A2	SLFEGIDIYT	SLFEGIDFYT	Gaudin, 1999	-0.66	-0.32	24.5	7.6	Yes at high conc
K-ras	CRC	Tumor	Cw8	GADGVGKSA	GAGGVGKSA	Tran, 2016 ⁴⁸	-3.36	-3.76	14448.1	37042.9	No
K-ras	CRC	Tumor	Cw8	GADGVGKSA	GAGGVGKSA	Tran, 2016 ⁴⁸	-1.93	-2.81	3976.7	30765	No
K-ras	PC	Blood	B35	VVVGAVGVG	VVVGAGGVG	Gjertsen, 1997	-3.96	-3.93	19846	20374.7	No
KIAA0205	BC	Blood	B44	AEPIDIQTV	AEPINIQTV	Guéguen, 1998	0.09	-0.11	96.7	84.2	No
MART2	M	Tumor	A1	FLEGNEVGKTY	FLGGNEVGKTY	Kawakami, 2001	-0.89	-1.44	4010.9	10257.4	No
MATN	M	Tumor	A11	KTLSVFQK	ETLSVFQK	Robbins, 2013 ⁷	0.4	-0.52	8.4	38	No
ME1	LC	Tumor	A2	FLDEFMEGV	FLDEFMEAV	Karanikas, 2001	0.66	0.87	2.3	2	No
MUM-1	M	Tumor	B44	EEKLIWLF	EEKLSVLF	Coullie, 1995	0.01	0.23	166.3	154.2	No

Continued

Table 1 Continued

Gene/protein	Tumor type	T cell origin	HLA restriction element	Mutant peptide sequence	WT peptide sequence	Reference	Total score mutant	Total score WT	MHC IC ₅₀ (nM) mutant peptide	MHC IC ₅₀ (nM) WT peptide	T cells reactive to WT peptide?
MUM-2	M	Blood	B44	SELFRRSGLDSY	SELFRRRLDSY	Chiari, 1999	-0.28	-0.17	713.2	501	No
MUM-2	M	Blood	Cw6	FRSGLDSYV	FRRLDSYV	Chiari, 1999	-1.34	-0.93	157.1	92.3	No
MUM-3	M	Blood	A68	EAFIQPITR	EAFSIQPITR	Baurain, 2000	0.75	1.08	13.1	7	No
Myosin class I	M	Tumor	A3	KINKPKYK	EINKPKYK	Zorn, 1999	0.48	-2.63	63.3	5884.6	No
N-ras	M	Tumor	A1	ILDTAGREEY	ILDTAGQEEY	Linard, 2002	0.03	0.51	255.8	140.1	No
NFYC	LC	LN	B52	QQITKTEV	QQITQTEV	Takenoyama, 2006	-3.34	-3.15	28088.8	24411.9	No
OGT	CRC	Blood	A2	SLYKFSFPPL	-	Ripberger, 2003	0.47	11	-	-	-
OS-9	M	Blood	B44	KELEGILL	KELEGILLP	Vigneron, 2002	-0.9	-2.9	522.6	3473.2	No
p53	HNC	Blood	A2	VWPCPEPEV	VWPEPEPEV	Ito, 2007	-1.85	-2.21	1268.1	3017.2	No
PPP1R3B	M	Tumor	A1	YTDHFCQYV	YTDFFCQYV	Robbins, 2013 ⁷	-0.96	-1.9	107.4	194.4	No
PRDX5	M	Blood	A2	LLLDLLVSI	LLDDSLVSI	Sensi, 2005	0.45	0.47	14.2	16.2	No
RBAF600	M	Blood	B7	RPHVPESAF	GPHVPESAF	Lennerz, 2005	1.27	0.24	9.4	41.6	No
SIRT2	M	Blood	A3	KIFSEVTLK	KIFSEVTPK	Lennerz, 2005	0.2	-0.3	11.8	15.1	No
SNRPD1	M	Blood	B38	SHETVIEL	SHETVTIEL	Lennerz, 2005	-0.19	-0.26	184	168.5	No
SYT-SSX1 or SYT-SSX2 fusion protein	Sarcoma	Blood	B7	QRPYGYDQIM	-	Worley, 2001	-1.75	1033.6	-	-	-
TGF-betaRIII	CRC	Blood	A2	RLSSCVPA	-	Linnebacher, 2001	-0.86	82.7	-	-	-
TP53	Various	Tumor	A2	VWPCPEPEV	VWPEPEPEV	Malekzadeh, 2019	-1.85	-2.21	1268.1	3017.2	No

These result from unique mutations in genes that are expressed ubiquitously. Total score and MHC IC₅₀ were predicted using the IEDB proteasomal cleavage/TAP transport/MHC class I combined predictor tool for mutant and wild-type peptides. Additional information about T cell reactivity for native peptide, methods used to find epitopes, and T cell origin were collected from the references listed in the table. Each line corresponds to a peptide that is considered to be a tumor antigen that is recognized by T cells. For each antigenic peptide, evidence of natural processing and presentation and isolation of stable human T cell clones that recognize the peptide were required for inclusion in the table. The MHC I binding predictions were made on March 27, 2020 using the IEDB Analysis Resource Consensus Tool⁸⁴ that combines predictions from artificial neural networks (ANN) a.k.a. NetMHC, ⁸⁵⁻⁸⁷ stabilized matrix method, ⁸⁸ and CombiBl. ⁸⁹

HLA, human leucocyte antigen; IEDB, immune epitope database; MHC, major histocompatibility complex; TAP, transporter associated with antigen processing; WT, wild type.

Table 2 Table of germ line and differentiation antigens from the Cancer Antigenic Peptide Database

Gene/protein	Tumor type	HLA restriction element	Peptide sequence	Reference	Total score mutant	MHC IC ₅₀ (nM) mutant
Germ-line epitopes						
BAGE-1	M	Cw16	AARAVFLAL	Boel, 1995	-0.53	419.6
CT37/FMR1NB	LC	A2	YLCSGSSYFV	Flores-Villanueva, 2018	0.3	8.5
Cyclin-A1	AML	A2	SLIAAAAFCLA	Ochsenreither, 2012	-2.08	894.4
Cyclin-A1	AML	A2	FLDRFLSCM	Ochsenreither, 2012	-0.46	29
GAGE-1,2,8	M	Cw6	YRPRPRRY	Van den Eynde, 1995	-0.26	1140.7
GAGE-3,4,5,6,7	M	A29	YYWPRPRRY	De Backer, 1999	1.65	13.1
GnTV	M	A2	VLDPVFIRC(V)	Guilloux, 1996	-1.41	438.5
HERV-E	RCC	A11	ATFLGSLTWK	Takahashi, 2008	0.21	7.8
HERV-K-MEL	M	A2	MLAVISCAV	Schiavetti, 2002	0.43	6.9
KK-LC-1	LC	B15	RQKRILVNL	Fukuyama, 2006	-0.6	517.9
KM-HN-1	EC	A24	NYNNFYRFL	Monji, 2004	-0.42	223.7
KM-HN-1	EC	A24	EYSKECLKEF	Monji, 2004	-0.35	774.8
KM-HN-1	EC	A24	EYLSLSDKI	Monji, 2004	-1.34	502.4
LAGE-1	M	A2	MLMAQEALAFI	Aarnoudse, 1999	1.01	11.7
LAGE-1	M	A2	SLLMWITQC	Rimoldi, 2000	-1.33	390.1
LAGE-1	Various	A31	LAAQERRVPR	Wang, 1998	-0.63	113.5
LAGE-1	M	B7	APRGVRMAV	Slager, 2004	0.61	3.3
LAGE-1	BC	A68	ELVRRILSR	Sun, 2006	-0.31	93
LAGE-1	M	B7	APRGVRMAV	Slager, 2004	0.61	3.3
LRPAP1	Various	A2	FLGPWAAS	Marijt, 2019	-3.82	5299.3
LY6K	Various	A24	RYCNLEGPPI	Suda, 2007	-1.37	519.5
MAGE-A1	M	A1	EADPTGHSY	Traversari, 1992	0.47	107.3
MAGE-A1	CC	A2	KVLEYVIKV	Pascolo, 2001	0.64	6.2
MAGE-A1	M	A3	SLFRAVITK	Chaux, 1999	0.14	16.2
MAGE-A1	BC	A2	KVLEYVIKV	Ottaviani, 2005	0.64	6.2
MAGE-A1	M	A68	EVYDGREHSA	Chaux, 1999	-2.92	7161.9
MAGE-A1	M	B7	RVRFFPSL	Luiten, 2000	0.14	108.4
MAGE-A1	M	B35	EADPTGHSY	Luiten, 2000	1.24	16.1
MAGE-A1	M	B37	REPVTKAEML	Tanzarella, 1999	-1.49	3338.6
MAGE-A1	M	B44	KEADPTGHSY	Stroobant, 2012	0.59	144.9
MAGE-A1	M	B53	DPARYEFLW	Chaux, 1999	0.1	24.7
MAGE-A1	M	B57	ITKKVADLVGF	Corbière, 2004	-0.72	1775.4
MAGE-A1	M	Cw2	SAFPTTINF	Chaux, 1999	0.02	382.7
MAGE-A1	M	Cw3	SAYGEPRKL	Chaux, 1999	-0.43	219.7
MAGE-A1	MM	Cw7	RVRFFPSL	Goodyear, 2011	-1.97	6638.8
MAGE-A1	M	Cw16	SAYGEPRKL	van der Bruggen, 1994	-0.75	525.8
MAGE-A10	M	A2	GLYDGM EHL	Huang, 1999	1.38	5.1
MAGE-A10	M	B53	DPARYEFLW	Chaux, 1999	0.1	24.7
MAGE-A12 m	M	A2	FLWGPRLV	van der Bruggen, 1994	0.14	11.1
MAGE-A12 m	BC	Cw7	VRIGHLYIL	Heidecker, 2000	-0.2	212.5
MAGE-A12 m	M	Cw7	VRIGHLYIL	Panelli, 2000	-0.2	212.5
MAGE-A12 m	M	Cw7	EGDCAPEEK	Breckpot, 2004	-3.8	43,575.9
MAGE-A2	M	A2	YLQLVFGIEV	Kawashima, 1998	-0.41	43.2
MAGE-A2	Various	A24	EYLQLVFGI	Tahara, 1999	-0.92	147.8

Continued

Table 2 Continued

Gene/protein	Tumor type	HLA restriction element	Peptide sequence	Reference	Total score mutant	MHC IC ₅₀ (nM) mutant
MAGE-A2	M	Cw7	EGDCAPEEK	Breckpot, 2004	-3.8	43,575.9
MAGE-A2	M	B37	REPVTKAEML	Tanzarella, 1999	-1.49	3338.6
MAGE-A3	M	A2	KVAELVHFL	Kawashima, 1998	0.98	11.2
MAGE-A3	M	A1	EVDPIGHLY	Gaugler, 1994	1.18	17.9
MAGE-A3	HNC	A24	TFPDLESEF	Oiso, 1999	-0.54	981.9
MAGE-A3	M	B18	MEVDPIGHLY	Bilsborough, 2002	0.94	46.1
MAGE-A3	M	B35	EVDPIGHLY	Schultz, 2001	0.51	74.6
MAGE-A3	M	B37	REPVTKAEML	Tanzarella, 1999	-1.49	3338.6
MAGE-A3	M	B44	MEVDPIGHLY	Herman, 1996	0.87	45.1
MAGE-A3	M	B40	AELVHFLLL	Schultz, 2002	0.6	19.9
MAGE-A3	M	B52	WQYFFPVIF	Russo, 2000	-0.52	1441.5
MAGE-A3	M	Cw7	EGDCAPEEK	Breckpot, 2004	-3.8	43,575.9
MAGE-A3	M	A2	FLWGPRLV	van der Bruggen, 1994	0.14	11.1
MAGE-A3	CRC	A24	VAELVHFL	Miyagawa, 2006	-2.19	9428.9
MAGE-A4	Various	A2	GVYDGREHTV	Duffour, 1999	-1.47	971.6
MAGE-A4	M	A1	EVDPASNTY	Kobayashi, 2003	0.68	93.6
MAGE-A4	RCC	A24	NYKRCFPVI	Ottaviani, 2006	0	60.2
MAGE-A4	Various	A24	NYKRCFPVI	Miyahara, 2005	0	60.2
MAGE-A4	M	B37	SESLKMIF	Zhang, 2002	-1.68	11,466.1
MAGE-A6	M	B35	EVDPIGHVY	Benlalam, 2003	1.24	28.4
MAGE-A6	M	Cw7	EGDCAPEEK	Breckpot, 2004	-3.8	43,575.9
MAGE-A6	M	Cw16	ISGGPRISY	Vantomme, 2003	-0.11	628.4
#REF!	M	B37	REPVTKAEML	Tanzarella, 1999	-1.49	3338.6
MAGE-A9	RCC	A2	ALSVMGVYV	Oehlich, 2005	-0.32	31.4
MAGE-C1	MM	A2	ILFGISLREV	Anderson, 2011	0.2	11.2
MAGE-C1	MM	A2	KVVEFLAML	Anderson, 2011	0.84	19.7
MAGE-C2	M	A2	ALKDVEERV	Ma, 2004	-1.26	342.2
MAGE-C2	M	B44	SESIKKKVL	Godelaide, 2007	-1.48	3608.4
MAGE-C2	M	A2	LLFGLALIEV	Ma, 2004	0.52	9.9
MAGE-C2	M	B57	ASSTLYLVF	Ma, 2011	0.31	226.5
NA88-A	M	B13	QGQHFLQKV	Moreau-Aubry, 2000	-3.07	16,389.9
NY-ESO-1/LAGE-2	M	A2	SLLMWITQC	Jäger, 1998	-1.33	390.1
NY-ESO-1/LAGE-2	M	A2	SLLMWITQC	Chen, 2000	-1.33	390.1
NY-ESO-1/LAGE-2	M	A2	SLLMWITQC	Valmori, 2000	-1.33	390.1
NY-ESO-1/LAGE-2	M	A2	MLMAQEALAF	Aarnoudse, 1999	1.01	11.7
NY-ESO-1/LAGE-2	Various	A24	YLAMPFATPME	Eikawa, 2013	-4.36	32,521.3
NY-ESO-1/LAGE-2	Various	A31	ASGPGGGAPR	Wang, 1998	-1.4	498.5
NY-ESO-1/LAGE-2	Various	A31	LAAQERRVPR	Wang, 1998	-0.63	113.5
NY-ESO-1/LAGE-2	OC	A68	TVSGNILTIR	Matsuzaki, 2008	0.55	16.2
NY-ESO-1/LAGE-2	M	B7	APRPHGGAASGL	Ebert, 2009	0.11	46.5
NY-ESO-1/LAGE-2	Various	B35	MPFATPMEAE	Eikawa, 2013	-0.62	253.7
NY-ESO-1/LAGE-2	M	B49	KEFTVSGNILT	Knights, 2009	-0.66	200.9
NY-ESO-1/LAGE-2	Various	B52	FATPMEAE	Eikawa, 2013	-2.28	11,772.9
NY-ESO-1/LAGE-2	M	Cw3	LAMPFATPM	Gnjatic, 2000	0.48	2.6
NY-ESO-1/LAGE-2	M	Cw6	ARGPESRLL	Gnjatic, 2000	-1.38	2001

Continued

Table 2 Continued

Gene/protein	Tumor type	HLA restriction element	Peptide sequence	Reference	Total score mutant	MHC IC ₅₀ (nM) mutant
NY-ESO-1/LAGE-2	Various	C12	FATPMEAELAR	Eikawa, 2013	-2.63	16,931.9
NY-ESO-1/LAGE-2	M	B51	MPFATPMEA	Jäger, 2002	-2.09	704.3
SAGE	Various	A24	LYATVIHDI	Miyahara, 2005	-0.41	43.1
Sp17	MM	A1	ILDSSEEDK	Chiriva-Internati, 2003	-3.5	18,212.4
SSX-2	M	A2	KASEKIFYV	Ayyoub, 2002	0.01	16.6
TAG-1	M	A2	SLGWLFLLL	Adair, 2008	0.17	52
TAG-2	M	B8	LSRLSNRLL	Adair, 2008	-1.68	4019.9
TRP2-INT2	M	A68	EVISCKLIKR	Lupetti, 1998	0.52	19
XAGE-1b/GAGED2a	M	A2	RQKKIRIQL	Ohue, 2012	-2.19	19,497
Differentiation epitopes						
CEA	GIC	A2	YLSGANLNL	Tsang, 1995	0.71	18.8
CEA	GIC	A24	TYACFVSNL	Nukaya, 1999	-0.05	60.1
CEA	GIC	A2	GVLVGVALI	Alves, 2007	-1.64	637.3
CEA	GIC	A3	HLFGYSWYK	Kawashima, 1999	0.17	7.9
CEA	GIC	A24	QYSWFVNGTF	Nukaya, 1999	1.26	17
CEA	GIC	A2	IMIGVLVGV	Kawashima, 1998	0.78	4.6
gp100/Pmel17	M	A2	KTWGQYWQV	Bakker, 1995	0.46	11.8
gp100/Pmel17	M	A2	MLGHTTMEV	Tsai, 1997	0.33	7.5
gp100/Pmel17	M	A2	KTWGQYWQV	Kawakami, 1995	0.46	11.8
gp100/Pmel17	M	A2	ITDQVPFSV	Kawakami, 1995	-1.2	211.5
gp100/Pmel17	M	A2	YLEPGPVTA	Cox, 1994	-1.09	172.7
gp100/Pmel17	M	A2	VLYRYGSFSV	Kawakami, 1995	0.12	12.3
gp100/Pmel17	M	A2	LLDGTATLRL	Kawakami, 1994	-0.51	347.7
gp100/Pmel17	M	A2	RLMKQDFSV	Kawakami, 1998	0.66	4.9
gp100/Pmel17	M	A2	SLADTNSLAV	Tsai, 1997	-0.32	47.7
gp100/Pmel17	M	A3	LIYRRRLMK	Kawakami, 1998	0.53	6.1
gp100/Pmel17	M	A2	RLPRIFCSC	Kawakami, 1998	-2.06	1491.6
gp100/Pmel17	M	A3	IALNFPGSQK	Kawashima, 1998	-0.19	31
gp100/Pmel17	M	A3	ALLAVGATK	Skipper, 1996	-0.35	45.2
gp100/Pmel17	M	A3	RSYVPLAHR	Michaux, 2014	0.3	68
gp100/Pmel17	M	A3	ALNFPGSQK	Kawashima, 1998	-0.03	19
gp100/Pmel17	M	A24	VYFFLPDHL	Robbins, 1997	-0.05	77.6
gp100/Pmel17	M	A11	ALNFPGSQK	Kawashima, 1998	-0.67	46.7
gp100/Pmel17	M	B7	SSPGCQPPA	Lennerz, 2005	-3.78	23,794.9
gp100/Pmel17	M	A68	HTMEVTYVYHR	Sensi, 2002	1.34	2.6
gp100/Pmel17	M	A32	RTKQLYPEW	Vigneron, 2004	-0.02	69
gp100/Pmel17	M	B35	VPLDCVLYRY	Benlalam, 2003	0.23	154.7
gp100/Pmel17	M	B35	LPHSSSHWL	Vigneron, 2005	-0.73	240.8
gp100/Pmel17	M	Cw8	SNDGPTLI	Castelli, 1999	-3.07	31,048.6
mammaglobin-A	BC	A3	PLLENVISK	Jaramillo, 2002	-2.42	2471.7
Melan-A/MART-1	M	A2	ILTVILGVL	Castelli, 1995	-0.6	529.6
Melan-A/MART-1	M	A2	AAGIGILTV	Kawakami, 1994	-2.13	3674
Melan-A/MART-1	M	Cw7	RNGYRALMDKS	Larrieu, 2008	-4.73	42,544.4
Melan-A/MART-1	M	B35	EAAGIGILTV	Benlalam, 2003	-3.08	15,289.2
Melan-A/MART-1	M	A2	SLSKILDTV	Wang, 2006	0.04	15.1

Continued

Table 2 Continued

Gene/protein	Tumor type	HLA restriction element	Peptide sequence	Reference	Total score mutant	MHC IC ₅₀ (nM) mutant
NY-BR-1	BC	A24	LYSACFWWL	Touloukian, 2003	0.04	91.7
OA1	M	A2	TLMSAMTNL	Olson, 2010	0.89	15.3
PAP	PC	A2	ALDVYNGLL	Olson, 2010	-0.66	240.5
PAP	PC	A2	FLFLLFFWL	Olson, 2010	0.22	42.5
PAP	PC	A2	FLTPKKLQCV	Correale, 1997	-0.84	126.4
PSA	PC	A2	VISNDVCAQV	Correale, 1997	-1.21	359
PSA	PC	A2	VLHWDPETV	Walton, 2006	-0.58	116.7
RAB38/NY-MEL-1	M	A31	MSLQRQFLR	Wang, 1996	0.77	5.4
TRP-1/gp75	M	A2	SVYDFFVWL	Parkhurst, 1998	0.87	21.6
TRP-2	M	A2	TLDSQVMSL	Noppen, 2000	0.26	52.5
TRP-2	M	A31	LLGPRPYR	Wang, 1996	0.27	24.1
TRP-2	M	Cw8	ANDPIFVVL	Castelli, 1999	-1.2	2060.1
TRP-2	M	A33	LLGPRPYR	Wang, 1998	-0.47	115.1
TRP-2	M	A1	KCDICTDEY	Kittlesen, 1998	-0.62	1397.5
tyrosinase	M	A1	SSDYVIPIGTY	Kawakami, 1998	0.43	147.3
tyrosinase	M	A2	MLLAVLYCL	Wölfel, 1994	1.21	8.3
tyrosinase	M	A2	CLLWSFQ TSA	Riley, 2001	-1.08	99.6
tyrosinase	M	A2	YMDGTMSQV	Wölfel, 1994	0.47	5.7
tyrosinase	M	A2	YMDGTMSQV	Skipper, 1996	0.47	5.7
tyrosinase	M	A24	IYMDGTADFSF	Dalet, 2011	1.09	23.4
tyrosinase	M	A24	AFLPWHRLF	Kang, 1995	0.85	44.6
tyrosinase	M	A26	QCSGNFMGF	Lennerz, 2005	-1.83	14,811.2
tyrosinase	M	B35	LPSSADVEF	Morel, 1999	1.82	4.1
tyrosinase	M	B35	TPRLPSSADVEF	Benlalam, 2003	-1.07	2534.5
tyrosinase	M	B38	LHHAFVDSIF	Lennerz, 2005	-0.48	620.8
tyrosinase	M	B44	SEIWRDIDF	Brichard, 1996	-0.16	250.7

Germ-line antigens are expressed in many tumors but not in normal tissues. Differentiation antigens are also expressed in the normal tissue of origin of the malignancy. Total score and MHC IC₅₀ were predicted using the IEDB proteasomal cleavage/TAP transport/MHC class I combined predictor tool for mutant and wild-type peptides. Each line corresponds to a peptide that is considered to be a tumor antigen that is recognized by T cells. For each antigenic peptide, evidence of natural processing and presentation and isolation of stable human T cell clones that recognize the peptide were required for inclusion in the table. The MHC I binding predictions were made on March 27, 2020 using the IEDB Analysis Resource Consensus Tool⁹⁴ that combines predictions from ANN a.k.a. NetMHC,⁹⁵⁻⁹⁷ SMM,⁹⁸ and Comblib.⁹⁹

HLA, human leucocyte antigen; IEDB, immune epitope database; MHC, major histocompatibility complex; TAP, transporter associated with antigen processing.

CONCLUSIONS

The future of immunotherapy will hinge on its ability to achieve durable responses in a large patient population. Given low rates of durable tumor regression observed with immune checkpoint blockade in patients with relapsed malignancy, other T cell-based immunotherapeutic approaches such as personalized cancer vaccines and adoptive transfer of TCR-engineered T cells may be able to improve outcomes. These approaches require the determination of T cell antigens. The discovery of such antigens is largely constrained to the study of few tumor types (mostly melanoma) in the most common HLA alleles. While it seems like bona fide, immunogenic epitopes are the exception and not the rule, new methods

of epitope discovery may allow for high-throughput, comprehensive searches of proteins and mutations to the point where personalized cancer immunotherapy is feasible for patients whose tumors harbor intact antigen and processing machinery. However, numerous limitations based on current technologies exist which make reliable identification of a naturally processed and presented antigen difficult. Recent studies have demonstrated proof of concept that epitope discovery can lead to remarkable tumor regression in highly metastatic patients.^{48 49 88} The development of strategies that allow for efficient and accurate antigen discovery, such as incorporation of peptide:HLA complex stability, is poised to make huge contributions to cancer treatment in the near

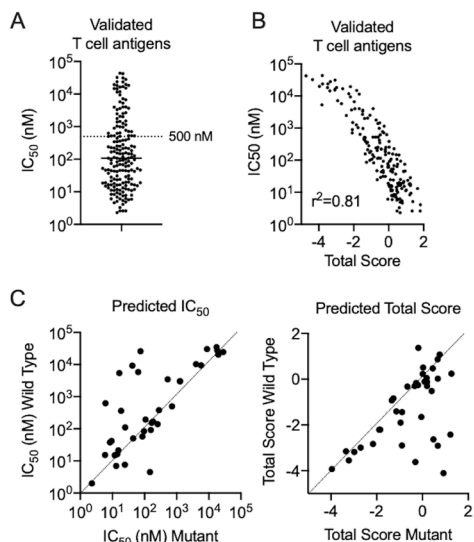


Figure 1 Performance of in silico predictions in validated T cell antigens. (A) Dot plots of predicted MHC IC₅₀ for a panel of validated T cell antigens. A reference level of IC₅₀ 500 nM is provided to delineate a common cut-off score used by researchers. (B) Scatterplot of MHC IC₅₀ and total prediction score showing a linear correlation (Pearson correlation, $p < 0.001$). (C) Scatterplots of predicted MHC IC₅₀ and total score between wild-type and mutant peptides for validated neopeptides. MHC, major histocompatibility complex.

future as fundamental building blocks of TCR discovery for adoptive T cell transfer and personalized cancer vaccine therapy.

Contributors Concept: MYL and CA. Literature reviews: MYL, JWJ, CS, and CA. Computational work: MYL and JWJ. Authorship of manuscript: MYL, JWJ, CS, and CA. Final approval: MYL, JWJ, CS, and CA.

Funding This work was supported by the Intramural Research Program of the NIH, National Institute on Deafness and Other Communication Disorders, project number ZIA-DC00008. This research was also made possible through the NIH Medical Research Scholars Program, a public-private partnership supported jointly by the NIH and contributions to the Foundation for the NIH from the Doris Duke Charitable Foundation (DDCF Grant #2014194), the American Association for Dental Research, the Colgate-Palmolive Company, Genentech, Elsevier, and other private donors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- Horn L, Mansfield AS, Szczyńska A, *et al*. First-Line Atezolizumab plus chemotherapy in extensive-stage small-cell lung cancer. *N Engl J Med* 2018;379:2220–9.
- Burtneš B, Harrington KJ, Greil R, *et al*. Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study. *Lancet* 2019;394:1915–28.
- Doherty PC. The keys to cell-mediated immunity. *JAMA* 1995;274:1067–8.

- Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science* 2015;348:69–74.
- McGranahan N, Swanton C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* 2015;27:15–26.
- van der Burg SH, Arens R, Ossendorp F, *et al*. Vaccines for established cancer: overcoming the challenges posed by immune evasion. *Nat Rev Cancer* 2016;16:219–33.
- Robbins PF, Lu Y-C, El-Gamil M, *et al*. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat Med* 2013;19:747–52.
- Chen Y-T, 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.
- Matsuzaki J, Grnjatic S, Mhawech-Fauceglia P, *et al*. Tumor-infiltrating NY-ESO-1-specific CD8⁺ T cells are negatively regulated by LAG-3 and PD-1 in human ovarian cancer. *Proc Natl Acad Sci U S A* 2010;107:7875–80.
- Madden DR, Gorga JC, Strominger JL, *et al*. The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight peptide binding to MHC. *Cell* 1992;70:1035–48.
- Robbins PF, Morgan RA, Feldman SA, *et al*. Tumor regression in patients with metastatic synovial cell sarcoma and melanoma using genetically engineered lymphocytes reactive with NY-ESO-1. *J Clin Oncol* 2011;29:917–24.
- Blum JS, Wearsch PA, Cresswell P. Pathways of antigen processing. *Annu Rev Immunol* 2013;31:443–73.
- Groll M, Ditzel L, Löwe J, *et al*. Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 1997;386:463–71.
- Griffin TA, Nandi D, Cruz M, *et al*. Immunoproteasome Assembly: Cooperative Incorporation of Interferon γ (IFN- γ)-inducible Subunits. *J Exp Med* 1998;187:97–104.
- Baumeister W, Walz J, Zühl F, *et al*. The proteasome: paradigm of a self-compartmentalizing protease. *Cell* 1998;92:367–80.
- Tenzen S, Peters B, Bulik S, *et al*. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *CMLS, Cell. Mol. Life Sci.* 2005;62:1025–37.
- Tenzen S, Stoltze L, Schönfisch B, *et al*. Quantitative analysis of prion-protein degradation by constitutive and immuno-20S proteasomes indicates differences correlated with disease susceptibility. *J Immunol* 2004;172:1083–91.
- Schirle M, Keilholz W, Weber B, *et al*. Identification of tumor-associated MHC class I ligands by a novel T cell-independent approach. *Eur J Immunol* 2000;30:2216–25.
- Nielsen M, Lundegaard C, Lund O, *et al*. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* 2005;57:33–41.
- van Enderd PM, Tampé R, Meyer TH, *et al*. A sequential model for peptide binding and transport by the transporters associated with antigen processing. *Immunity* 1994;1:491–500.
- Schumacher TN, Kantesaria DV, Heemels MT, *et al*. Peptide length and sequence specificity of the mouse TAP1/TAP2 translocator. *J Exp Med* 1994;179:533–40.
- Uebel S, Meyer TH, Kraas W, *et al*. Requirements for peptide binding to the human transporter associated with antigen processing revealed by peptide scans and complex peptide libraries. *J Biol. Chem.* 1995;270:18512–6.
- Uebel S, Kraas W, Kienle S, *et al*. Recognition principle of the TAP transporter disclosed by combinatorial peptide libraries. *Proc Natl Acad Sci U S A* 1997;94:8976–81.
- Kochan G, Krojer T, Harvey D, *et al*. Crystal structures of the endoplasmic reticulum aminopeptidase-1 (ERAP1) reveal the molecular basis for N-terminal peptide trimming. *Proc Natl Acad Sci U S A* 2011;108:7745–50.
- Serwold T, Gonzalez F, Kim J, *et al*. ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature* 2002;419:480–3.
- Hammer GE, Gonzalez F, Champsaur M, *et al*. The aminopeptidase ERAAP shapes the peptide repertoire displayed by major histocompatibility complex class I molecules. *Nat Immunol* 2006;7:103–12.
- Nguyen TT, Chang S-C, Evnouchidou I, *et al*. Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAP1. *Nat Struct Mol Biol* 2011;18:604–13.
- Sadasivan B, Lehner PJ, Ortmann B, *et al*. Roles for calreticulin and a novel glycoprotein, tapasin, in the interaction of MHC class I molecules with TAP. *Immunity* 1996;5:103–14.
- Blees A, Janulienė D, Hofmann T, *et al*. Structure of the human MHC-I peptide-loading complex. *Nature* 2017;551:525–8.

- 30 Morozov GI, Zhao H, Mage MG, *et al.* Interaction of TAPBPR, a tapasin homolog, with MHC-I molecules promotes peptide editing. *Proc Natl Acad Sci U S A* 2016;113:E1006–15.
- 31 McShan AC, Natarajan K, Kumirov VK, *et al.* Peptide exchange on MHC-I by TAPBPR is driven by a negative allosteric release cycle. *Nat Chem Biol* 2018;14:811–20.
- 32 Fuertes MB, Kacha AK, Kline J, *et al.* Host type I IFN signals are required for antitumor CD8⁺ T cell responses through CD8 α ⁺ dendritic cells. *J Exp Med* 2011;208:2005–16.
- 33 Cruz FM, Colbert JD, Merino E, *et al.* The biology and underlying mechanisms of cross-presentation of exogenous antigens on MHC-I molecules. *Annu Rev Immunol* 2017;35:149–76.
- 34 Rock K, Gamble S, Rothstein L. Presentation of exogenous antigen with class I major histocompatibility complex molecules. *Science* 1990;249:918–21.
- 35 Iborra S, Izquierdo HM, Martínez-López M, *et al.* The DC receptor DNGR-1 mediates cross-priming of CTLs during vaccinia virus infection in mice. *J Clin Invest* 2012;122:1628–43.
- 36 Norbury CC, Hewlett LJ, Prescott AR, *et al.* Class I MHC presentation of exogenous soluble antigen via macropinocytosis in bone marrow macrophages. *Immunity* 1995;3:783–91.
- 37 Matheoud D, Perié L, Hoeffel G, *et al.* Cross-Presentation by dendritic cells from live cells induces protective immune responses in vivo. *Blood* 2010;115:4412–20.
- 38 Wolfers J, Lozier A, Raposo G, *et al.* Tumor-Derived exosomes are a source of shared tumor rejection antigens for CTL cross-priming. *Nat Med* 2001;7:297–303.
- 39 Dance A. Core concept: cells nibble one another via the under-appreciated process of trogocytosis. *Proc Natl Acad Sci U S A* 2019;116:17608–10.
- 40 Shen L, Sigal LJ, Boes M, *et al.* Important role of cathepsin S in generating peptides for TAP-independent MHC class I cross-presentation in vivo. *Immunity* 2004;21:155–65.
- 41 Ma W, Zhang Y, Vigneron N, *et al.* Long-Peptide cross-presentation by human dendritic cells occurs in vacuoles by peptide exchange on nascent MHC class I molecules. *J Immunol* 2016;196:1711–20.
- 42 Kovacsics-Bankowski M, Rock K. A phagosome-to-cytosol pathway for exogenous antigens presented on MHC class I molecules. *Science* 1995;267:243–6.
- 43 Burgdorf S, Schölz C, Kautz A, *et al.* Spatial and mechanistic separation of cross-presentation and endogenous antigen presentation. *Nat Immunol* 2008;9:558–66.
- 44 Palmowski MJ, Gileadi U, Salio M, *et al.* Role of immunoproteasomes in cross-presentation. *J Immunol* 2006;177:983–90.
- 45 Smith CC, Selitsky SR, Chai S, *et al.* Alternative tumour-specific antigens. *Nat Rev Cancer* 2019;19:465–78.
- 46 Ebstein F, Textoris-Taube K, Keller C, *et al.* Proteasomes generate spliced epitopes by two different mechanisms and as efficiently as non-spliced epitopes. *Sci Rep* 2016;6:24032.
- 47 Robert C, Schachter J, Long GV, *et al.* Pembrolizumab versus ipilimumab in advanced melanoma. *N Engl J Med* 2015;372:2521–32.
- 48 Tran E, Robbins PF, Lu Y-C, *et al.* T-Cell transfer therapy targeting mutant KRAS in cancer. *N Engl J Med Overseas Ed* 2016;375:2255–62.
- 49 Doran SL, Stevanović S, Adhikari S, *et al.* T-Cell receptor gene therapy for human Papillomavirus-Associated epithelial cancers: a first-in-human, phase I/II study. *JCO* 2019;37:2759–68.
- 50 Tanyi JL, Bobisse S, Ophir E, *et al.* Personalized cancer vaccine effectively mobilizes antitumor T cell immunity in ovarian cancer. *Sci Transl Med* 2018;10:eaa05931.
- 51 Hilf N, Kuttruff-Coqui S, Frenzel K, *et al.* Actively personalized vaccination trial for newly diagnosed glioblastoma. *Nature* 2019;565:240–5.
- 52 Strønen E, Toebes M, Kelderman S, *et al.* Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science* 2016;352:1337–41.
- 53 Xu C. A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. *Comput Struct Biotechnol J* 2018;16:15–24.
- 54 McGranahan N, Furness AJS, Rosenthal R, *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* 2016;351:1463–9.
- 55 Cohen CJ, Gartner JJ, Horovitz-Fried M, *et al.* Isolation of neoantigen-specific T cells from tumor and peripheral lymphocytes. *J Clin Invest* 2015;125:3981–91.
- 56 Turajlic S, Litchfield K, Xu H, *et al.* Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol* 2017;18:1009–21.
- 57 Huang J, El-Gamil M, Dudley ME, *et al.* T Cells Associated with Tumor Regression Recognize Frameshifted Products of the *CDKN2A* Tumor Suppressor Gene Locus and a Mutated HLA Class I Gene Product. *J Immunol* 2004;172:6057–64.
- 58 Maby P, Galon J, Latouche J-B. Frameshift mutations, neoantigens and tumor-specific CD8⁺ T cells in microsatellite unstable colorectal cancers. *Oncotarget* 2016;5:e1115943.
- 59 Van Allen EM, Robinson D, Morrissey C, *et al.* A comparative assessment of clinical whole exome and transcriptome profiling across sequencing centers: implications for precision cancer medicine. *Oncotarget* 2016;7:52888–99.
- 60 Laumont CM, Vincent K, Hesnard L, *et al.* Noncoding regions are the main source of targetable tumor-specific antigens. *Sci Transl Med* 2018;10:eaa05516.
- 61 Smart AC, Margolis CA, Pimentel H, *et al.* Intron retention is a source of neoepitopes in cancer. *Nat Biotechnol* 2018;36:1056–8.
- 62 Hanada K-ichi, Yewdell JW, Yang JC. Immune recognition of a human renal cancer antigen through post-translational protein splicing. *Nature* 2004;427:252–6.
- 63 Hedegaard J, Thorsen K, Lund MK, *et al.* Next-Generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. *PLoS One* 2014;9:e98187.
- 64 Sharma G, Rive CM, Holt RA. Rapid selection and identification of functional CD8⁺ T cell epitopes from large peptide-coding libraries. *Nat Commun* 2019;10:1–13.
- 65 Townsend Aet al. The epitopes of influenza nucleoprotein recognized by cytotoxic T lymphocytes can be defined with short synthetic peptides. *Cell* 1986;44:959–68.
- 66 Zhang H, Hong H, Li D, *et al.* Comparing Pooled Peptides with Intact Protein for Accessing Cross-presentation Pathways for Protective CD8⁺ and CD4⁺ T Cells. *J. Biol. Chem.* 2009;284:9184–91.
- 67 Popović J, Li L-P, Kloetzel PM, *et al.* The only proposed T-cell epitope derived from the TEL-AML1 translocation is not naturally processed. *Blood* 2011;118:946–54.
- 68 Sibener LV, Fernandes RA, Kolawole EM, *et al.* Isolation of a structural mechanism for uncoupling T cell receptor signaling from peptide-MHC binding. *Cell* 2018;174:672–87.
- 69 Kula T, Dezfulian MH, Wang CI, *et al.* T-Scan: a genome-wide method for the systematic discovery of T cell epitopes. *Cell* 2019;178:1016–28.
- 70 Joglekar AV, Leonard MT, Jeppson JD, *et al.* T cell antigen discovery via signaling and antigen-presenting bifunctional receptors. *Nat Methods* 2019;16:191–8.
- 71 Li G, Bethune MT, Wong S, *et al.* T cell antigen discovery via trogocytosis. *Nat Methods* 2019;16:183–90.
- 72 Beausang JF, Wheeler AJ, Chan NH, *et al.* T cell receptor sequencing of early-stage breast cancer tumors identifies altered clonal structure of the T cell repertoire. *Proc Natl Acad Sci U S A* 2017;114:E10409–17.
- 73 Sette A, Buus S, Appella E, *et al.* Prediction of major histocompatibility complex binding regions of protein antigens by sequence pattern analysis. *Proc Natl Acad Sci U S A* 1989;86:3296–300.
- 74 Ruppert J, Sidney J, Celis E, *et al.* Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. *Cell* 1993;74:929–37.
- 75 Falk K, Rötzschke O, Stevanović S, *et al.* Allele-Specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* 1991;351:290–6.
- 76 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.
- 77 Peters B, Nielsen M, Sette A. T cell epitope predictions. *Annu Rev Immunol* 2020;38:123–45.
- 78 Abelin JG, Keskin DB, Sarkizova S, *et al.* Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity* 2017;46:315–26.
- 79 Jurtz V, Paul S, Andreatta M, *et al.* NetMHCpan-4.0: improved Peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J.I.* 2017;199:3360–8.
- 80 Backert L, Kohlbacher O. Immunoinformatics and epitope prediction in the age of genomic medicine. *Genome Med* 2015;7:119.
- 81 Stranzl T, Larsen MV, Lundegaard C, *et al.* NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics* 2010;62:357–68.
- 82 van der Burg SH, Visseren MJ, Brandt RM, *et al.* Immunogenicity of peptides bound to MHC class I molecules depends on the MHC-peptide complex stability. *J Immunol* 1996;156:3308–14.
- 83 Rasmussen M, Fenoy E, Harndahl M, *et al.* Pan-Specific prediction of peptide-MHC class I complex stability, a correlate of T cell immunogenicity. *J Immunol* 2016;197:1517–24.

- 84 Peters B. IEDB workshop: the immune epitope database analysis resource: MHC class I peptide processing and immunogenicity predictions; 2015.
- 85 Paul S, Weiskopf D, Angelo MA, *et al.* HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity. *J.i.* 2013;191:5831–9.
- 86 Marty R, Kaabinejadian S, Rossell D, *et al.* MHC-I genotype restricts the oncogenic mutational landscape. *Cell* 2017;171:1272–83.
- 87 Koşaloğlu-Yalçın Z, Lanka M, Frentzen A, *et al.* Predicting T cell recognition of MHC class I restricted neopeptides. *Oncoimmunology* 2018;7:e1492508.
- 88 Ott PA, Hu Z, Keskin DB, *et al.* An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* 2017;547:217–21.
- 89 Vita R, Mahajan S, Overton JA, *et al.* The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res* 2019;47:D339–43.
- 90 Vita R, Overton JA, Greenbaum JA, *et al.* The immune epitope database (IEDB) 3.0. *Nucleic Acids Res* 2015;43:D405–12.
- 91 Correia-Neves M, Waltzinger C, Mathis D, *et al.* The shaping of the T cell repertoire. *Immunity* 2001;14:21–32.
- 92 George JT, Kessler DA, Levine H. Effects of thymic selection on T cell recognition of foreign and tumor antigenic peptides. *Proc Natl Acad Sci U S A* 2017;114:E7875–81.
- 93 Micheletti F, Bazzaro M, Canella A, *et al.* The lifespan of major histocompatibility complex class I/peptide complexes determines the efficiency of cytotoxic T-lymphocyte responses. *Immunology* 1999;96:411–5.
- 94 Kim Y, Ponomarenko J, Zhu Z, *et al.* Immune epitope database analysis resource. *Nucleic Acids Res* 2012;40:W525–30.
- 95 Nielsen M, Lundegaard C, Worning P, *et al.* Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Science* 2003;12:1007–17.
- 96 Lundegaard C, Lamberth K, Harndahl M, *et al.* 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* 2008;36:W509–12.
- 97 Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* 2016;32:511–7.
- 98 Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics* 2005;6:132.
- 99 Sidney J, Assarsson E, Moore C, *et al.* Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. *Immunome Res* 2008;4:2.