



Published in final edited form as:

Cell. 2017 February 09; 168(4): 600–612. doi:10.1016/j.cell.2017.01.014.

## Applications of Immunogenomics to Cancer

X. Shirley Liu\* and

Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard T.H. Chan School of Public Health, 450 Brookline Ave, Boston MA 02215, USA

Elaine R. Mardis\*

Institute for Genomic Medicine, Nationwide Children's Hospital, and The Ohio State University College of Medicine, 575 Children's Crossroad, Columbus OH 43205, USA

### Brief history of tumor-specific mutant antigens and immunogenomics

The underpinnings of modern immunogenomics resulted from hypotheses generated and tested by visionaries in cancer immunology during the late 1980s through the 1990's. Their central hypothesis was that cancer cells presented novel, tumor-specific (i.e. mutated) peptides on the cancer cell surface bound by the patient's HLA molecules. By virtue of this cell surface presentation, specific T cell immunity might be elicited to these "neoantigens". Supporting evidence for this hypothesis was demonstrated in cancers of non-viral origin by Old (Old and Boyse, 1964), Gross (Gross, L. Cancer Res. 3(5): 326–333 (1943), Foley (Foley, 1953), and Prehn (Prehn and Main, 1957). This foundational work led to the identification and characterization of the role of MHC proteins in antigen presentation (Babbitt et al., 1985); (Bjorkman et al., 1987). Concomitantly, methods to grow antigen-specific cytolytic T lymphocytes (CTLs) in culture also were developed (Cerottini et al., 1974; Gillis and Smith, 1977), as were the molecular biology procedures to clone and express gene products. Thierry Boon's laboratory combined these new methods to identify the first TSA, a point mutation in a protein called P91A (De Plaen et al., 1988). Subsequently, Hans Schreiber's laboratory demonstrated that TSAs also functioned as neoantigens using primary UV-induced mouse tumors (Monach et al., 1995). Similarly, groups studying human melanomas showed they could identify T cells in the peripheral circulation that bound melanoma cells preferentially over normal cells from the same patient (Dubey et al., 1997; Knuth et al., 1984; Robbins et al., 1996; Van den Eynde et al., 1989). Shortly thereafter, Boon's laboratory cloned the first human tumor antigen, called MAGEA1 (van der Bruggen et al., 1991) and Sahin's group demonstrated an autologous antibody-based method to clone and identify different human tumor antigens (Sahin et al., 1995). While these foundational studies established supporting evidence for the existence of tumor-specific peptide neoantigens, the lengthy and painstaking nature of these processes was unlikely to scale to clinical application for cancer patients.

More recently, these limitations have been alleviated by the application of new sequencing technologies and associated computational data analysis approaches. These methods,

Correspondence: xsliu@jimmy.harvard.edu (X.S.L.), elaine.mardis@nationwidechildrens.org (E.R.M.).

\*Joint Authorship:

collectively referred to as “immunogenomics”, have improved the facility with which individual cancers can be studied to predict their neoantigens for prognostic purposes or to inform immunotherapeutic interventions. Complementary methods have been developed to study the changes in the T-cell repertoire, to characterize the gene expression signatures of the immune cell types present in the tumor mass, and to design personalized vaccines or adoptive cell transfer (ACT) therapies. The now scalable nature of immunogenomic methods should permit their widespread clinical application, although there remain issues and challenges to be resolved. This primer will highlight the specific methods and describe the known strengths and weaknesses in modern immunogenomics.

## Somatic mutations generate neoantigens

It has long been known that cancer is caused by alterations to genomic DNA that impact protein functions, ultimately disrupting cellular control of pathways and resulting in the outgrowth of a tumor mass. Methods using next generation sequencing platforms generate data from tumor and normal DNA isolates that, once aligned to the Human Reference Genome sequence, can be interpreted to identify somatic alterations (Ley et al., 2008). In practice, such analyses aim to identify DNA alterations in known cancer genes, both oncogenes and tumor suppressors, that combine to transform the founder cell. For certain oncogenes, identified mutations indicate therapeutic interventions that may successfully halt the tumor cell growth. By contrast, immunogenomic approaches aim to identify tumor-specific DNA alterations that predict amino acid sequence changes in all encoded proteins, and then evaluate their potential as neoantigens. In practice, most tumor-specific antigens identified to-date are highly unique to each patient and generally do not involve known cancer genes.

Hence, the widespread use of next-generation sequencing (NGS) instrumentation has enabled immunogenomics, providing a facile way to generate data to predict tumor-specific neoantigens in a rapid, inexpensive and comprehensive manner (Gubin et al., 2015). NGS technologies have rapidly evolved over the past 10 years, resulting in dramatically increased amounts of sequencing data produced per instrument run at ever-decreasing costs (Mardis, 2017). In immunogenomics, since the focus is protein-coding genes, solution hybridization-based methods are used to select these sequences (“exome”) prior to sequencing (Bainbridge et al., 2010; Gnirke et al., 2009; Hodges et al., 2009). Importantly, the concomitant development of advanced variant detection algorithms that identify different classes of mutations from NGS data has enabled the identification of all classes of somatic variation. Accurate detection of variants in this setting is influenced by multiple factors, which are presented here in detail.

One important consideration for somatic variant detection is depth of coverage by NGS sequencing reads from the tumor. In principle, since tumor samples include variable percentages of normal cells, adjustments to the depth of NGS data generated must be flexible to ensure that a sufficient representation of tumor-derived sequence reads are obtained. Isolating DNA from selected, tumor-rich areas of a biopsy or resection sample is ideal, but not always possible, so average read depths of 300–500 fold exome coverage are typically attempted to compensate for the normal cell DNA-derived reads. A second reason

for high coverage of the tumor-derived DNA is to enable the evaluation of founder clone versus subclonal mutations in the resulting data. Here, we define founder mutations as the original set of mutations present in the cell that transformed from normal to neoplastic, whereas subclonal mutations are acquired as the daughter cells of this founder acquired additional mutations during growth of the tumor mass. Based on this definition, founder clone mutations in diploid regions of the exome have a proportional fraction of variant-containing sequencing reads (Variant Allege Fraction or VAF) that is around 50% (adjusted for normal DNA contribution), since most somatic mutations are heterozygous. In theory, neoantigens that result from founder clone mutations should elicit a T-cell response that targets all cancer cells rather than the subset of tumor cells that would be targeted by T-cell response to subclonal neoantigens in the vaccine.

Equally important to appropriate coverage depth for accurate prediction of variants is the algorithm or set of algorithms used to identify variants from the NGS exome data. The factors to consider here include the types of variants one wishes to consider in neoantigen discovery. For example, single nucleotide variants (point mutations) are easiest to predict with high accuracy because reads containing a single variant are readily aligned to their reference genome “match”, and because there are a variety of different algorithms that also can detect low VAF variants. Variant detection from NGS reads has been an area of rapid development and there are many algorithms to choose from, with variable performance, as has been evaluated (Cornish and Guda, 2015; Ghoneim et al., 2014; Krøigård et al., 2016). By contrast, variants resulting from insertion or deletion of one or a few nucleotides (“indels”) are significantly more difficult to identify due to issues of read alignment by standard alignment algorithms, that lead often to lower coverage in these regions for the variant-containing sequencing reads (Jiang et al., 2012, 2015; Ratan et al., 2015). However, indels may be important to immunogenomics efforts because they can introduce frameshift mutations that result in highly divergent amino acid sequences in the resulting protein and hence may produce strong predicted neoantigens. Increased read lengths on NGS platforms have improved indel detection, as has the use of gapped alignment or split-read algorithms that are computationally intensive but better able to align the indel-containing reads to the reference genome. Assembly-based realignment approaches also have been developed to improve the precision of indel variant detection (Mose et al., 2014; Narzisi et al., 2014).

Another type of somatic variation that can lead to highly altered amino acid sequences and as a result, create a neoantigenic peptide, is a structural variant which fuses two protein-coding sequences. These can result from inversion or deletion of a chromosomal segment, or from chromosomal translocations. Detecting these alterations from exome sequencing data is quite challenging and error-prone, but RNA-based analysis can identify the resulting fusion transcript (Li et al. 2011; Scolnick et al. 2015; Zhang et al. 2016; Kumar et al. 2016), and compare the predicted fusion sequence to NGS data from DNA (whole genome or exome sequencing) to identify supporting evidence of the genomic event causing the fusion. Recently, we adapted this approach for neoantigen prediction with a process called IntegrateNEO, using the TMPRSS2-ERG fusions common in prostate cancer to evaluate its ability to identify fusion peptide neoantigens (Zhang et al., 2016b). RNAseq data brings added value to immunogenomics efforts beyond the detection of fusion peptides, as will be described later.

Once variant detection is completed, each variant is annotated to predict the resulting amino acid change(s) that result from the altered DNA sequence (if any). There are widely utilized computational tools such as Annovar and VEP available to produce the translated peptides from the DNA data. The translated peptides constitute one type of input data for the neoantigen prediction software to calculate the class I or class II predicted binding affinities. The second type of data needed for neoantigen prediction are the HLA haplotypes of the patient, also derived from exome data since these reagents capture the HLA gene loci. Heretofore, HLA typing was performed using a PCR-based and Sanger sequencing-based clinical assay. The repetitive nature of the HLA genes requires a high stringency assembly of these genes, which can be achieved using the >500 bp read lengths from Sanger data. Sequence analysis of these regions based on hybrid capture-derived NGS reads, which are relatively short (~100bp), requires a stringent alignment of the read data to the IMGT/HLA database (Robinson et al., 2001) using a haplotype-resolved algorithm to interpret the HLA class I and II haplotypes. There now exist multiple algorithms for accomplishing these data interpretations, including Polysolver (Shukla et al., 2015), HLAMiner (Warren et al., 2012), and OptiType (Szolek et al., 2014). Typically, one interprets the normal tissue-derived exome data to obtain the HLA haplotypes. Clinical analysis of these genes also should include repeating the alignment of the tumor-derived exome data and identification of mutations in order to identify HLA alleles that are impacted by nonsense mutations, deletions, or other similarly deleterious types of somatic alterations that may influence the presence of that allele (Shukla et al., 2015). Some algorithms also can use RNA-derived data to interpret the HLA haplotypes (Warren et al., 2012).

Another critical component of identifying neoantigens is the *in silico* prediction of HLA class I and II binding affinities for specific peptides. These predictions are quite computationally complex and require machine learning-based approaches to establish models for the different types of binding site interactions. In particular, each peptide interacts with the binding pocket residues of the many different HLA proteins through the amino acid side chains of specific residues. Therefore, the binding affinity of any peptide is sequence-specific relative to that patient's HLA proteins, some of which may be common and some rare. There also are differences in the binding of peptides by class I or class II HLA that impact the precision of neoantigen prediction, as described later. Finally, there is considerable debate about an appropriate cutoff value for binding affinity in terms of what does or does not constitute a strong neoantigen candidate (Duan et al., 2014)(Bassani-Sternberg et al., 2016)

The initial approach to computational HLA binding predictions utilized a neural network-based learning method developed from a training set of experimentally derived binding affinities for class I HLA proteins and different peptides. This effort resulted in an HLA class I binding prediction software known as netMHC, devised by researchers in the Center for Biological Sequence Analysis at the Technical University of Denmark (Lundegaard et al., 2008a, 2008b; Nielsen et al., 2003). The predictor has improved over time with the availability of training data sets for HLA proteins that are more rare in the population, although calculated binding affinities for the most rare HLA alleles in humans remain less certain (Wang et al., 2010). An interim approach to address rare HLA class I binding calculations was PickPocket, which extrapolated from variants with known binding

specificity to those without existing experimental data (Zhang et al., 2009). The most recent version is netMHCstabpan (Rasmussen et al., 2016), which uses a neural network approach based on a dataset of stability values calculated for different peptide-MHC-1 complexes, rather than their binding affinity values, since the stability of their interaction has experimentally been shown to be more strongly correlated to T-cell immunogenicity. Another early method developed to generate class I binding predictions was based on a stabilized matrix method (SMM) algorithm developed by Peters and Sette (Peters and Sette, 2005). This approach models the sequence specificity of binding processes as a means of predicting outcomes for untested sequences. SMM not only predicted HLA binding but also evaluated peptide transport as a function of antigen presentation and proteasomal cleavage with the TAP algorithm. Subsequent efforts to develop new class I binding affinity prediction softwares have included the use of combined support vector machine-based (SVM) and random forest machine-learning approaches (Srivastava et al., 2013), or combining the information obtained from amino acid pairwise contact potentials and quantum topology molecular similarity descriptors (Saethang et al., 2013) to better model HLA class I peptide interactions.

With the requisite information generated by NGS to call somatic variants and interpret their impact on protein sequences, and to identify the HLA haplotypes specific to the patient, neoantigen prediction softwares can be used to predict both the class I and class II HLA binding affinities for each tumor-unique set of peptides. Considerations and specifics for these prediction approaches are described in detail below. There are a number of binding prediction softwares and associated immunogenomics algorithms available at the Immune Epitope DataBase (IEDB) analysis resource (Robinson et al., 2013), which can be found at this URL: <http://tools.immuneepitope.org/main/>. The IEDB web interface permits the input of peptide sequences for sequential evaluation by user-configured steps using the software of choice to predict neoantigens. Publicly available software pipelines also are available for local download and computing of neoantigen predictions by end-users, including pVAC-seq (<https://github.com/griffithlab/pVAC-Seq>) and epidisco (<https://github.com/hammerlab>).

## Class I predictions

Approaches to predict HLA class I neoantigens typically begin by parsing the tumor-specific peptides predicted from variant calling as 21mer peptides that encompass the variant amino acid(s) placed as near to the center of the 21mer as possible. This is easiest to envisage for simple non-synonymous amino acid substitutions, shown in Figure 1a, which then are tiled across the variant-containing peptides to define a set of 8mer to 11mers to input for binding calculations, based on HLA class I binding characteristics (Figure 1b). These peptide sets are parsed along with their corresponding wild type peptide sequences as input data for consideration by neoantigen prediction software, along with information about the HLA class I haplotypes determined for the patient. The resulting list of neoantigens can be quite extensive, depending upon the numbers and types of input peptide sequences and the diversity of the HLA haplotypes. The numbers of neoantigens can be winnowed by applying several criteria to narrow the potential set of peptides that are considered further in vaccine design, if desired. One conventional approach is to only consider variant peptides with a strong- to intermediate-binding affinity (typically lower than 500 nM) but this arbitrary cut-

off is controversial because strong neoantigens can have lower calculated affinities than actual. This sometimes is due to the presence of a rare HLA haplotype, for which the neural net software provides an inaccurate binding affinity prediction. Thus, for each altered locus, one can select the candidate peptide with the single best binding affinity to each corresponding HLA allele across all peptide lengths considered, or proceed with all candidates for all HLA alleles to additional filtering steps, as follows.

Three important additional filters should be applied to remove false positives, 1) RNA-based filtering to remove genes with no evidence of expression, 2) filtering based on exome data coverage depth at the variant loci, and 3) filtering based on variant allele fraction (VAF)-based metrics. The RNA expression filter ensures that each peptide is supported by evidence of RNA expression, wherein evidence of RNA expression is considered a reasonable, but not absolute, proxy that the gene is expressed in the tumor cell proteome. For the NGS coverage filter, a minimum level of normal read coverage depth is required to ensure there is sufficient sequencing data coverage from the normal tissue (i.e. supports a true positive somatic variant call). Finally, both DNA and RNA data should be evaluated to ascertain the percentage of variant-containing reads or variant allele fraction (VAF). As described earlier, this criterion helps to inform the final list of neoantigen candidates by providing information on whether a specific alteration is shared across all tumor cells (i.e. in the founder clone) or is subclonal, based on DNA sequencing data, and ensures that a variant is expressed in the tumor RNA. The latter is especially important in tumor types with a high mutation load such as those with chemical or UV damage to DNA, since upwards of 50% of mutations are typically not expressed in RNA (or protein by inference) for these tumors. With these filtering steps completed, a list of high confidence, predicted neoantigenic peptides and the HLA class I proteins predicted to bind them, their calculated binding affinity value(s), and the binding affinity of the cognate wild-type peptide values can be parsed for further consideration in vaccine design or other immunological evaluations such as neoantigen burden. In the former case, neoantigen predictions have been tested in clinical trials of personalized vaccines, with demonstrated ability to elicit specific T-cell responses (Schumacher et al. 2014; Carreno et al. 2015; Tran et al. 2014). In the latter approach, there are demonstrated correlations between neoantigen burden and the likelihood of response to checkpoint blockade inhibition therapies (Le et al., 2015; Rizvi et al., 2015; Snyder et al., 2014; Van Allen et al., 2015), and a demonstration that predicted neoantigens also are the epitopes targeted by checkpoint blockade immunotherapies (Gubin et al., 2014).

## Class II predictions

HLA class II predictions are significantly more difficult to generate with precision due to the nature of the HLA class II proteins. Firstly, class II HLA proteins are heterodimers of alpha and beta peptides encoded by four different loci in the human genome. Only one of these four loci is not highly polymorphic (Robinson, 2003), meaning there is extensive HLA class II polymorphism in the general population. This becomes somewhat less complex if neoantigen predictions focus on the most frequently expressed class II molecules (McKinney et al., 2013). Secondly, certain peptides bind to multiple different HLA class II molecules and are responsible for the majority of antigen-specific T-cell responses. These so-called 'promiscuous peptides' are difficult to predict using computational approaches.

Thirdly, the HLA class II binding groove is open on both ends, and although the core binding motif is a 9mer amino acid, variable length peptides are allowed to bind. Many of the HLA-II polymorphic sites comprise other regions of the binding groove outside the core motif binding region, which allows the flanking amino acid sequences on either side of the motif sequence to influence its binding affinity. As a result, binding affinities are difficult to predict with a high degree of precision. Input data for MHC class II binding predictions consist of 15mer representatives of each somatic neoantigen candidate peptide, along with the patient's HLA class II haplotypes. A cutoff binding of <1000 nM may be utilized to distinguish strong binders but given the vagaries of binding affinity predictions described above, this cutoff may not be appropriate. RNA expression level has been identified as a critical filtering parameter for predicted class II neoantigen candidates, whereby those peptides corresponding to genes with higher relative expression values from RNAseq data analysis are considered to be the strongest candidates (Kreiter et al., 2015).

Computational predictions, based on the aforementioned caveats for both class I and II, therefore only offer putative neoantigen candidates that may be subject to a variety of errors or sources of inaccuracy. In addition to what we already have described, there are other challenges to accurate neoantigen prediction. First, even though RNA evidence supports a variant as being expressed, the most accurate evidence of a peptide's presence in the cell is identifying that peptide from mass spectrometry-based proteomic data derived from the specific tumor under study. Second, binding affinity calculations are more accurate for the common class I HLA haplotypes, less so for rarer haplotypes. Third, a significant biological confounder of neoantigen discovery is our inability to predict precisely which of the putative neoantigen peptides will be processed in the tumor cell degradasome, then bound to and properly presented by HLA molecules on the cell surface. This critical component of T-cell activation must occur for the neoantigen to stimulate a specific immune response, yet it is presently not possible to computationally predict the processing and presentation of peptides by HLA. One way to inform neoantigen prediction methods is using experimental measurements of T-cell based immune responses to the predicted peptide epitopes. There are conventional methods such as EliSpot (IFN-gamma release) assays (Cole), flow cytometry-based dextramer assays (Carreno et al., 2015), and mass spectrometry-based evaluation of HLA-bound peptides (Gubin et al., 2014). However, scalable, high-throughput methods are in development at present and will require time and testing.

## Immune repertoire profiling

Cellular immune responses from T cells and humoral immune responses from B cells are stimulated by exposures to antigens, including pathogens, allergens and neoantigens. V(D)J recombination in the primary lymphoid organs creates the incredibly diverse and unique repertoire of the hypervariable regions of B cell receptors (BCR) and T cell receptors (TCR), and somatic hypermutations contribute to additional BCR diversity. During B and T cell development, self-antigens are presented to B and T cells to select out self-reacting types, and to ensure only B and T cells that recognize and attack foreign antigens are in the circulation. T cells only recognize foreign proteins presented on MHC, while B cells can also target foreign DNA, lipids, or carbohydrates. Upon recognition of foreign antigens and with the presence of co-stimulatory molecules, B and T cells express cell surface activation

markers, attack foreign antigens, secrete cytokines, stimulate each other, and proliferate (Pasternack, 1994). One goal of immunogenomic studies is to characterize the repertoire of B and T cells in patients with cancer, especially before and after immunotherapy-based interventions.

DNA sequencing approaches have enabled the characterization of immune repertoires (Pasternack 1994; Robins 2013). After a pioneering study introduced the technique (Freeman et al., 2009), a plethora of immune repertoire methods have been published and commercial solutions are also available. Several studies (Calis and Rosenberg, 2014; Hou et al., 2016; Yaari and Kleinstein, 2015) have evaluated the experimental techniques and practical advice needed for immune repertoire profiling. Basically, multiplex PCR can amplify the recombined V(D)J regions from either mRNA or DNA in the B or T cells. The V(D)J, and most importantly, the variable complementarity-determining region CDR3 sequences, and their respective abundance can be resolved by high throughput sequencing. Paired-end sequencing with additional PCR primers in the middle of the fragment permits full length TCR repertoire sequencing with short read NGS technology to resolve the V / J pairing (Cole et al., 2016). One caveat to this approach is that PCR biases and sequencing errors can falsely increase the total repertoire with deeper sequencing coverage, so unique molecular identifier barcodes should be used to eliminate such artifacts (Cole et al., 2016), although such an approach is presently only available for RNA based repertoire profiling.

Computational methods, as summarized in (Greiff et al., 2015a, 2015b) are important components for the analysis, annotation, and visualization of immune repertoires. To this end, IMGT (Giudicelli et al., 1997) is the most widely cited immunogenetics database and provides many useful tools such as V-QUEST and HighV-QUEST (Alamyar et al., 2012) as well as statistical metrics (Aouinti et al., 2015) for the analysis and annotation of immune repertoire data. VDJtools (Shugay et al., 2015) is a comprehensive analysis framework for T-cell and B-cell repertoire sequencing data. It includes MIXCR for fast alignment and clonal type assembly (Bolotin et al., 2015), MIGEC for removing duplicates and combining barcodes (Shugay et al., 2014), and VDJviz for visualization (Bagaev et al., 2016), and provides basic statistical analyses for characterizing and comparing different immune repertoires.

The initial output from a repertoire profiling analysis is a list of BCR/TCR CDR3 sequences, sometimes including the adjoining V and J sequences, each followed by an abundance estimate. This output allows samples to be compared and clustered, if desired. For example, common CDR3 sequences that are shared among individuals indicate BCR / TCR clones that recognize common antigens such as herpes or common cold viruses. In comparison, CDR3 clones that are rare among patients but are abundant within a tumor, and more importantly for BCR lineage-related CDR3s with small amount of mutations, indicate T/B cell recognition of the patient tumor-specific antigens (Saul et al., 2016). Repertoire profiles from individuals with similar ethnic backgrounds, lifestyles or environmental exposures are often clustered. Two independent metrics, diversity (often measured by the Shannon entropy), and evenness (indicative of the degree of clonal expansion), have been proposed as important characteristics of immune repertoires (Greiff et al., 2015b). Since V(D)J recombination in TCR only happens in children, TCR diversity once established, generally



declines with age. In contrast, V(D)J recombination in BCR occurs throughout life although at reduced levels in adults, and activated BCR undergoes somatic hypermutation to improve the antibody affinity to the recognized antigen, so the BCR diversity distributions assume more complex patterns. Although immune-stimulating events such as allergy or vaccination could shift the abundance of some clones, the immune repertoire has been suggested as a means to monitor an individual's immune health (Johnson et al., 2014). The utility of this metric depends on the accurate measure of clonal abundance, which requires linear amplification from multiplex PCR products and additional normalization of TCR/BCR expression levels for RNA-based profiles. Furthermore, the method and time-span of sample storage can also influence sample quality for repertoire profiling.

While immune repertoires are informative, profiling them over large sample cohorts can be expensive. Computational methods have been developed to directly infer immune repertoires from unselected bulk tumor RNAseq data, such as TRUST for TCR (Li et al., 2016a) and V'DJer for BCR (Mose et al., 2016). The hypervariability of the CDR3 regions of TCR and BCR renders the RNAseq reads from these regions unmappable to the human reference genome sequence, and somatic hypermutation adds additional challenges to BCR mapping and alignment. Both of the aforementioned methods select unmappable RNAseq reads, align these unmapped reads to each other with de Bruijn graphing methods, *de novo* assemble these alignments into contigs, and use IMGT (Giudicelli et al., 1997) to annotate those containing CDR3 motifs as potential BCR or TCR. Although these approaches only recover the most abundant of the immune repertoires, they were used to analyze RNA-seq data across tumor samples profiled by The Cancer Genome Atlas (TCGA) and resulted in novel findings. For example, TRUST revealed increased T cell clonal diversity in tumor types with higher mutational loads and potential neoantigens based on their co-occurrence with CDR3-containing sequences in the tumors (Li et al., 2016a), while V'DJer reported higher somatic hypermutation in IgG and IgA than in IgM (Mose et al., 2016).

Published studies have made fascinating observations on how immune repertoires can reflect an individual patient's immune health and predict their response to therapy. The ability to reconstruct a more diverse TCR repertoire after autologous hematopoietic stem cell transplantation has been observed to predict better transplant outcomes in multiple sclerosis patients (Johnson et al., 2014; Muraro et al., 2014). Another study used TCR repertoire sequencing to compare each patient's TCR before and after dendritic cell-based neoantigen vaccine dosing, illustrating expanded TCRs for the vaccine peptides that elicited a T-cell response (Carreno et al., 2015). For metastatic melanoma patients, the anti-CTLA4 antibody Ipilimumab has been shown to increase peripheral blood TCR diversity (Robert et al., 2014), and those patients with higher peripheral TCR diversity before treatment were reported to respond better to ipilimumab (Postow et al., 2015). In contrast, the anti-PD-1 antibody pembrolizumab showed better efficacy in melanoma patients whose pre-treatment tumor-infiltrating T-cells were less diverse and more clonal (Tumeh et al., 2014). This study also demonstrated that more tumor-infiltrating T-cell clones expanded after treatment in the therapy responsive group than in the (non-responding) disease progression group. Although these pioneering studies were conducted on a limited number of patients, they do suggest TCR repertoire as a universal cancer immunotherapy biomarker (McNeel 2016)(McNeel, 2016). Potentially overall patient immune health from the peripheral TCR and signs of

neoantigen recognition and clonal expansion from the tumor TCR before treatment could predict better patient response to cancer immunotherapies. As an example, one bioinformatics study using a Potential Support Vector Machine-based approach reported the ability to predict an individual's age, health, transplantation status, and development of lymphoid cancer based on repertoire profiles (Greiff et al., 2015b).

## Distribution of tumor infiltrating lymphocytes

Large-scale molecular tumor profiling often selects samples with high tumor purity to best characterize the molecular signatures of the tumor. While most cancer genomics studies are focused on the cancerous cells in the tumor tissue, the impurities, such as stromal cells, endothelial cells, and immune cells, could have major impact on the development and progression of cancer. With genomic profiling, tumor purity could be estimated from DNA copy number (Carter et al., 2012), SNP allele frequency (Li and Li, 2014), RNA-seq (Yoshihara et al., 2013), or DNA methylation (Zhang et al., 2015; Zheng et al., 2014) data. Interestingly, these methods using orthogonal tumor profiling modalities yield very consistent tumor purity estimates, in distinct contrast to the estimates provided by pathologists, suggesting that molecular and morphological changes in the tumor do not appear simultaneously.

Pertinent to immunogenomic studies of cancer is the evaluation of tumor-infiltrating lymphocytes (TILs), which can involve traditional approaches such as flow cytometry and multiplex immunohistochemistry. Flow cytometry uses antibodies against proteins uniquely expressed on different subpopulations of immune cells to isolate specific subsets of these cells from blood or tissues. The resulting cell counts characterize the relative abundance of different subpopulations in individual cancer samples and can reveal changes following treatment. Flow-cytometry requires relatively large fresh tissue samples for study, but the resulting isolated cells, once sorted, can be cultured and profiled. Multiplex immunohistochemistry (IHC) can simultaneously capture the expression levels of multiple proteins in formalin-fixed paraffin-embedded (FFPE) tissue, with the advantage of capturing their spatial organization and co-expression patterns, although the number of proteins that can be differentially stained on each tissue slide is limited.

In addition to these conventional approaches, recent computational methods have also advanced our understanding of TILs. In a seminal study (Rooney et al., 2015), Rooney and colleagues used Granzyme A and perforin expression levels to model the immune cytolytic activities in tumors studied in TCGA, observing increased cytolytic activities in tumors with higher mutation load, copy number aberration, viral infection, and lower tumor stage. This signature-gene based approach has been employed by two recent studies (Angelova et al., 2015; enbabao lu et al., 2016) to estimate immune subset abundance based on a collection of pre-selected markers. CIBERSORT (Newman et al., 2015) used an expert-selected signature of about 500 genes to infer the abundance of 22 different tumor infiltrating immune components. In contrast, TIMER (Li et al., 2016b) selected cancer-specific signature genes to eliminate the bias from highly expressed genes in cancer cells and deconvolved only six immune components to ensure that colinear expression between closely related immune cells did not affect the deconvolution accuracy. These studies

confirmed previous observations (Bindea et al., 2013; Rooney et al., 2015) and reported that CD8+ T cells are associated with better overall survival and fewer relapses, whereas macrophages are associated with worse clinical outcome in many cancer types (Li et al., 2016b)

There have been inconsistent observations on whether the abundance of B cells is associated with improved cancer survival (DiLillo et al., 2010; Perricone et al., 2004; Qin et al., 1998; Schultz et al., 1990). One potential cause is that B cells at different activation statuses may either inhibit or promote T cell functions (Nelson, 2010). Another possible reason is that B cells are sometimes enriched at the margin of tumor capsules instead of evenly distributed throughout the tumor tissue (Kroeger et al., 2016; Lao et al., 2016; Nelson, 2010; Shi et al., 2013). Therefore, abundance estimates of B cells may be variable due to the specific tumor section under assay. By contrast, TCR-seq of different sections of a large ovarian tumor (Emerson et al., 2013) revealed that T cells are spatially homogeneous within the tumor, similar to peripheral blood. Therefore, it is possible that the correlation of TIL abundance with patient outcome will depend on the homogeneity of TIL distribution for different cancer types.

## Applications

The culmination of our renewed understanding of the immune system and its interaction potential with cancer cells has been a decades-long effort to develop therapeutic approaches that boost existing immune responses against neoplastic cells. These efforts span widely variable approaches, and a comprehensive review has been recently published that explores the broad landscape of cancer immunotherapies (Galluzzi et al., 2014).

Certain types of cancer immunotherapies act to re-invigorate existing immunity that has been suppressed in the tumor microenvironment. These so-called “checkpoint blockade” therapies were devised to address our fundamental understanding of immunosuppression and T-cell exhaustion, and provide a relatively tumor-specific immune response. However, there often are attendant side effects of variable severity, because their action targets native immune molecules such as CTLA-4, PD-1 and PD-L1. Potentially, more specific targeting could result from using putative neoantigens predicted by NGS-based analysis, described above, delivered as patient-specific vaccines meant to stimulate an immune response that is highly specific for the tumor cells. In this paradigm, several different vaccine types (or ‘platforms’) have emerged and are actively being tested in pre-clinical and clinical settings, as follows (Hirayama and Nishimura, 2016; Overwijk et al., 2013; Vormehr et al., 2015; Zhang et al., 2016c).

*DNA minicassette vaccines:* One vaccine platform is based on piecing together the individual coding sequences for each predicted neoantigen peptide into a DNA construct that may also contain a specific human promoter element to drive peptide production, once introduced into the patient. The vaccine construct, once sequence-verified, can be electroporated into patient-derived dendritic cells and the DCs then re-infused into the patient. Synthetic DNA is relatively cheaply and quickly obtained, even with the attendant GMP requirements for sequence verification prior

to use in a human vaccine. Hence, concerns about cost and scalability of this approach are minimal. One design consideration is that no self-antigens are potentially encoded by the junctions between each neoantigen sequence, but this is relatively easy to confirm computationally once the proposed vaccine design is conceptualized.

*Peptide vaccines:* Synthetic peptides representing computationally identified neoantigens can be combined and solubilized in the presence of one or more immune-stimulatory adjuvants to create patient-specific peptide vaccines. These can be directly injected intramuscularly, intradermally or subcutaneously as a means of presenting the neoantigenic peptides during maturation of native dendritic cells, which then can prime a robust and specific immune response. Short neoantigen peptides of 8–12 amino acids can directly bind to HLAs expressed on the surface of antigen-presenting dendritic cells, thereby priming a T-cell specific response. Peptide vaccines also can be comprised of synthetic long peptides (25–30 amino acids) which require uptake, processing and presentation by antigen-presenting cells in order to elicit an immune response. While GMP-grade peptides are expensive to manufacture, this is a scalable enterprise and when coupled with the simplicity of the peptide vaccine design, is being applied in clinical trials of patient-specific vaccines (W. Gillanders, personal communication).

*RNA vaccines:* Conceptually similar to DNA and peptide vaccines are RNA-based neoantigen vaccines, wherein the RNA encodes the various predicted neoantigens that are unique to each patient's tumor. As with all RNA-based therapeutics, the lability of RNA invokes a need to stabilize the RNA molecules and to provide for appropriate uptake by antigen presenting cells so the encoded peptides can be processed and presented. Cost and scalability of RNA synthesis are similarly straightforward as for DNA, so the packaging and stabilization are the challenging puzzles for this platform, which is being actively pursued in the research setting.

*Autologous dendritic cell vaccines:* Dendritic cells (DCs) occupy a unique position at the interface of innate and adaptive immunity, and have been shown to effect a robust, therapeutically relevant anti-neoplastic immune response. In particular, autologous dendritic cells can be isolated from patients and conditioned *ex-vivo* to mature, thereby providing immune-stimulatory functions. When coupled with neoantigenic peptides from patient-specific analyses, the resulting dendritic cell vaccine can be re-infused and has been shown to elicit neoantigen-specific T-cell immunity and an attendant expansion of the neoantigen-specific TCR (Carreno et al., 2015; Galluzzi et al., 2014). Emphasizing their specificity for tumor cells, no severe adverse events were recorded in this initial trial of patient-specific DC vaccines. However, not all of the predicted neoantigens elicited a T-cell response, indicating that our ability to predict even class I neoantigens will require additional precision, as discussed herein. While these early first-in-human results are exciting, the preparation of dendritic cell vaccines requires significant amounts of peripheral blood mononuclear cells for dendritic cell isolation, as well as time- and effort-intensive laboratory work to culture and mature the DCs *ex-vivo*. Hence, their scalability may be in question for broad-based clinical use.

Besides cancer vaccines, another type of genomics-driven patient-specific cancer immunotherapy is adoptive cell transfer (ACT), as pioneered by Rosenberg and colleagues (Rosenberg and Restifo, 2015). Basically, T-cells extracted from a cancer patient, either from peripheral blood or the resected tumor, can be activated and expanded *ex vivo* by IL2 treatment, before infusing them back to the same patient to kill the cancer cells. Preparatory lymphodepletion either by chemotherapy or radiation of the patient is an important step done prior to infusion, to improve the engraftment and persistence of the adoptively transferred T-cells, thus increasing durability of tumor regression (Dudley et al., 2002). ACT cells not only persist months after infusion, but also expand in the patient. Two additional genomic approaches below have been shown to further enhance tumor-specific killing and broaden the applicable cancer types suitable for ACT. Despite the cost and the technical and logistical challenges of ACT, this personalized immunotherapy has demonstrated promising rates and duration of response.

*Genetically engineering T-cells:* T-cells extracted from patients can be genetically engineered to express TCRs that specifically recognize proteins expressed only in the patient's cancer cells, such as the melanoma / melanocyte specific MART-1 antigen (Morgan et al., 2006) or the cancer-testis antigen NY-ESO-1 (Robbins et al., 2011). T-cells also can be engineered by viral transduction to express a chimeric antigen receptor (CAR) that uniquely recognizes the B-cell specific CD19 (Kalos et al., 2011; Kochenderfer et al., 2010) on the cell surface. Linking the CAR with a co-stimulatory domain such as CD137 (Imai et al., 2004; Milone et al., 2009) or engineering the cells to express another chimeric costimulatory receptor recognizing a second antigen (Kloss et al., 2013) have both improved T-cell antitumor activity. Recently a new clinical trial has been proposed, where CRISPR technology is applied to further engineer the NY-ESO-1-targeting CAR T-cells. Using a small number of CRISPR guide RNAs to knock out the PD-1 gene and the cells' intrinsic TCR, this approach aims to eliminate immune suppression and improve the NY-ESO-1 receptor response. If proven effective, genome engineering technology could provide new opportunities to manipulate other genes in immune cells *ex vivo* using the CRISPR technology to achieve desired cancer killing phenotypes.

*Expanding tumor-specific T-cells:* Instead of engineering the autologous T-cells *ex vivo*, this approach separately culture tumor-infiltrating T-cell clones or subpopulations, then select those reacting against tumor cells for massively expansion before patient infusion. With the emergence of exome-sequencing, scientists can call somatic mutations from the tumor and computationally predict immunogenic neoantigens. Testing the immunogenicity of these mutations in parallel uses minigenes constructs (described above) encoding the mutated peptides into expression vectors and the *in vitro* transcribed RNA from the vectors can be electroporated into antigen presenting cells (APC). Culturing the tumor-infiltrating T-cells for reactivity against these APCs selects the tumor-specific T-cells and identify the immunogenic mutant minigenes (Robbins et al., 2013). Compared to genetically engineered T-cells, the final T-cells infused into patients using this approach are comprised of populations of different T-cells recognizing different

neoantigens. While most of these neoantigens are hypothesized to be passenger mutations, recently the Rosenberg group identified four T-cell clones that specifically react to the KRAS G12D mutation in colon cancer (Tran et al., 2016), thereby drugging the undruggable.

## Future perspectives

As high throughput technologies improve and our immunology knowledge grows, the future of immunogenomics-based application to cancer appears quite promising and likely will continue to broaden. Technological and computational innovations will be instrumental to overcome existing challenges and move the field forward. First, despite the advances offered by algorithms such as NetMHC-pan, both the accuracy of MHC presentation prediction, especially for rarer alleles, and of MHC class II presentation await improvements. In addition, most studies use MHC presentation of somatic mutations as a proxy to predict immunogenicity, although it is unclear which presented somatic mutations will elicit immune responses. Experimental assays such as EliSpot are currently used to validate the predicted neoantigens, although such assays are still conducted in a low throughput fashion.

Second, TIL deconvolution methods such as CIBERSORT and TIMER use reference expression profiling data on sorted immune components from peripheral blood. These methods could be combined with Nanostring-based measures of immune marker genes in addition to bulk tissue RNA-seq data for inexpensive profiling of large archival tumor cohorts. However, expression of immune cells in tumors might differ significantly from that in peripheral blood, which could influence the accuracy of these inference methods. Recent developments in single cell analyses techniques, such as CyTOF (Newell et al., 2012) and single-cell RNA-seq (Klein et al., 2015; Macosko et al., 2015; Tirosh et al., 2016), might offer more quantitative alternatives. However, for very detailed TIL deconvolution on large sample cohorts, the required starting tumor material and cost of single cell experiments need to decrease significantly for widespread use.

Third, monitoring an individual patient's immune repertoire in peripheral blood or tumors provides insights into their immune health as well as their response to allergens, vaccines or therapies (Robins, 2013). However, there are still many challenges ahead, such as how to identify the specific TCR / BCR that recognize each specific somatic mutation and how accurate is immune repertoire at predicting patient response to immunotherapy. Other challenges include how to robustly estimate the total immune repertoire in different samples from the same individual, normalize bias from minor immune events, and distinguish immune repertoire signals from normal versus pathogenic immune events.

Last but not least, predicting response to immunotherapies, including tumor killing effects and autoimmune side effects, is still an open question. So far, higher T cell infiltration (Taube et al., 2012; Tumeh et al., 2014), higher PD-1 or PD-L1/L2 expression (Garon et al., 2015; Herbst et al., 2014; Taube et al., 2012), higher neoantigen load from BRCA or somatic mutations in DNA repair pathway genes (Hugo et al., 2016; Snyder et al., 2014; Van Allen et al., 2015), or microsatellite instability (Le et al., 2015), higher peripheral baseline TCR diversity (Postow et al., 2015), lower tumor infiltrating TCR diversity (Tumeh et al., 2014),

lack of mutations in interferon gamma (INFG) (Gao et al., 2016), beta-2-microglobulin (B2M) (Zaretsky et al., 2016), or JAK1/JAK2 (Zaretsky et al., 2016) have been associated with better response to immunotherapies in various cancer types. A comprehensive model that integrates all of these factors to accurately predict patient response to immunotherapy is still lacking, and likely will require much more data to train and refine. In addition, methods to predict the optimal combination of immunotherapies or with other targeted, chemo, or radiation therapies for individual patients still await development. Despite all the aforementioned challenges, the exciting results obtained to-date from cancer immunotherapies will continue to motivate the biomedical research community to overcome these challenges and explore new frontiers.

## References cited

- Alamyar E, Duroux P, Lefranc M-P, Giudicelli V. IMGT(®) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods Mol. Biol.* 2012; 882:569–604. [PubMed: 22665256]
- Angelova M, Charoentong P, Hackl H, Fischer ML, Snajder R, Krogsdam AM, Waldner MJ, Bindea G, Mlecnik B, Galon J, et al. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. *Genome Biol.* 2015; 16:64. [PubMed: 25853550]
- Aouinti S, Malouche D, Giudicelli V, Kossida S, Lefranc M-P. IMGT/HighV-QUEST Statistical Significance of IMGT Clonotype (AA) Diversity per Gene for Standardized Comparisons of Next Generation Sequencing Immunoprofiles of Immunoglobulins and T Cell Receptors. *PLoS One.* 2015; 10:e0142353. [PubMed: 26540440]
- Babbitt BP, Allen PM, Matsueda G, Haber E, Unanue ER. Binding of immunogenic peptides to Ia histocompatibility molecules. *Nature.* 1985; 317:359–361. [PubMed: 3876513]
- Bagaev DV, Zvyagin IV, Putintseva EV, Izraelson M, Britanova OV, Chudakov DM, Shugay M. VDJviz: a versatile browser for immunogenomics data. *BMC Genomics.* 2016; 17:453. [PubMed: 27297497]
- Bainbridge MN, Wang M, Burgess DL, Kovar C, Rodesch MJ, D'Ascenzo M, Kitzman J, Wu Y-Q, Newsham I, Richmond TA, et al. Whole exome capture in solution with 3 Gbp of data. *Genome Biol.* 2010; 11:R62. [PubMed: 20565776]
- Bassani-Sternberg M, Bräunlein E, Klar R, Engleitner T, Sinitcyn P, Audehm S, Straub M, Weber J, Slotta-Huspenina J, Specht K, et al. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat. Commun.* 2016; 7:13404. [PubMed: 27869121]
- Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, Angell H, Fredriksen T, Lafontaine L, Berger A, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity.* 2013; 39:782–795. [PubMed: 24138885]
- Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature.* 1987; 329:506–512. [PubMed: 3309677]
- Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, Chudakov DM. MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods.* 2015; 12:380–381. [PubMed: 25924071]
- van der Bruggen P, Traversari C, Chomez P, Lurquin C, De Plaen E, Van den Eynde B, Knuth A, Boon T. A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. *Science.* 1991; 254:1643–1647. [PubMed: 1840703]
- Calis JJA, Rosenberg BR. Characterizing immune repertoires by high throughput sequencing: strategies and applications. *Trends Immunol.* 2014; 35:581–590. [PubMed: 25306219]
- Carreno BM, Magrini V, Becker-Hapak M, Kaabinejadian S, Hundal J, Petti AA, Ly A, Lie W-R, Hildebrand WH, Mardis ER, et al. Cancer immunotherapy. A dendritic cell vaccine increases the

- breadth and diversity of melanoma neoantigen-specific T cells. *Science*. 2015; 348:803–808. [PubMed: 25837513]
- Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* 2012; 30:413–421. [PubMed: 22544022]
- Cerottini JC, Engers HD, Macdonald HR, Brunner T. Generation of cytotoxic T lymphocytes in vitro. I. Response of normal and immune mouse spleen cells in mixed leukocyte cultures. *J. Exp. Med.* 1974; 140:703–717. [PubMed: 4278108]
- Cole GA. Interferon- $\gamma$  ELISPOT Assay for the Quantitative Measurement of Antigen-Specific Murine CD8 T-Cells. In *Handbook of ELISPOT*. :191–204.
- Cole C, Volden R, Dharmadhikari S, Scelfo-Dalbey C, Vollmers C. Highly Accurate Sequencing of Full-Length Immune Repertoire Amplicons Using Tn5-Enabled and Molecular Identifier-Guided Amplicon Assembly. *J. Immunol.* 2016; 196:2902–2907. [PubMed: 26856699]
- Cornish A, Guda C. A Comparison of Variant Calling Pipelines Using Genome in a Bottle as a Reference. *Biomed Res. Int.* 2015; 2015:456479. [PubMed: 26539496]
- De Plaen E, Lurquin C, Van Pel A, Mariamé B, Szikora JP, Wölfel T, Sibille C, Chomez P, Boon T. Immunogenic (tum-) variants of mouse tumor P815: cloning of the gene of tum- antigen P91A and identification of the tum- mutation. *Proc. Natl. Acad. Sci. U. S. A.* 1988; 85:2274–2278. [PubMed: 3127830]
- DiLillo DJ, Yanaba K, Tedder TF. B cells are required for optimal CD4+ and CD8+ T cell tumor immunity: therapeutic B cell depletion enhances B16 melanoma growth in mice. *J. Immunol.* 2010; 184:4006–4016. [PubMed: 20194720]
- Duan F, Duitama J, Al Seesi S, Ayres CM, Corcelli SA, Pawashe AP, Blanchard T, McMahon D, Sidney J, Sette A, et al. Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J. Exp. Med.* 2014; 211:2231–2248. [PubMed: 25245761]
- Dubey P, Hendrickson RC, Meredith SC, Siegel CT, Shabanowitz J, Skipper JC, Engelhard VH, Hunt DF, Schreiber H. The immunodominant antigen of an ultraviolet-induced regressor tumor is generated by a somatic point mutation in the DEAD box helicase p68. *J. Exp. Med.* 1997; 185:695–705. [PubMed: 9034148]
- Dudley ME, Wunderlich JR, Robbins PF, Yang JC, Hwu P, Schwartzentruber DJ, Topalian SL, Sherry R, Restifo NP, Hubicki AM, et al. Cancer regression and autoimmunity in patients after clonal repopulation with antitumor lymphocytes. *Science*. 2002; 298:850–854. [PubMed: 12242449]
- Emerson RO, Sherwood AM, Rieder MJ, Guenthoer J, Williamson DW, Carlson CS, Drescher CW, Tewari M, Bielas JH, Robins HS. High-throughput sequencing of T-cell receptors reveals a homogeneous repertoire of tumour-infiltrating lymphocytes in ovarian cancer. *J. Pathol.* 2013; 231:433–440. [PubMed: 24027095]
- Foley EJ. Antigenic properties of methylcholanthrene-induced tumors in mice of the strain of origin. *Cancer Res.* 1953; 13:835–837. [PubMed: 13116120]
- Freeman JD, Warren RL, Webb JR, Nelson BH, Holt RA. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res.* 2009; 19:1817–1824. [PubMed: 19541912]
- Galluzzi L, Vacchelli E, Bravo-San Pedro J-M, Buqué A, Senovilla L, Baracco EE, Bloy N, Castoldi F, Abastado J-P, Agostinis P, et al. Classification of current anticancer immunotherapies. *Oncotarget*. 2014; 5:12472–12508. [PubMed: 25537519]
- Gao J, Shi LZ, Zhao H, Chen J, Xiong L, He Q, Chen T, Roszik J, Bernatchez C, Woodman SE, et al. Loss of IFN- $\gamma$  Pathway Genes in Tumor Cells as a Mechanism of Resistance to Anti-CTLA-4 Therapy. *Cell*. 2016; 167:397–404.e9. [PubMed: 27667683]
- Garon EB, Rizvi NA, Hui R, Leigh N, Balmanoukian AS, Eder JP, Patnaik A, Aggarwal C, Gubens M, Horn L, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. *N. Engl. J. Med.* 2015; 372:2018–2028. [PubMed: 25891174]
- Ghoneim DH, Myers JR, Tuttle E, Paciorkowski AR. Comparison of insertion/deletion calling algorithms on human next-generation sequencing data. *BMC Res. Notes*. 2014; 7:864. [PubMed: 25435282]



- Gillis S, Smith KA. Long term culture of tumour-specific cytotoxic T cells. *Nature*. 1977; 268:154–156. [PubMed: 145543]
- Giudicelli V, Chaume D, Bodmer J, Müller W, Busin C, Marsh S, Bontrop R, Marc L, Malik A, Lefranc MP. IMGT, the international ImmunoGeneTics database. *Nucleic Acids Res*. 1997; 25:206–211. [PubMed: 9016537]
- Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol*. 2009; 27:182–189. [PubMed: 19182786]
- Greiff V, Miho E, Menzel U, Reddy ST. Bioinformatic and Statistical Analysis of Adaptive Immune Repertoires. *Trends Immunol*. 2015a; 36:738–749. [PubMed: 26508293]
- Greiff V, Bhat P, Cook SC, Menzel U, Kang W, Reddy ST. A bioinformatic framework for immune repertoire diversity profiling enables detection of immunological status. *Genome Med*. 2015b; 7:49. [PubMed: 26140055]
- Gubin MM, Zhang X, Schuster H, Caron E, Ward JP, Noguchi T, Ivanova Y, Hundal J, Arthur CD, Krebber W-J, et al. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature*. 2014; 515:577–581. [PubMed: 25428507]
- Gubin MM, Artyomov MN, Mardis ER, Schreiber RD. Tumor neoantigens: building a framework for personalized cancer immunotherapy. *J. Clin. Invest*. 2015; 125:3413–3421. [PubMed: 26258412]
- Herbst RS, Soria J-C, Kowanetz M, Fine GD, Hamid O, Gordon MS, Sosman JA, McDermott DF, Powderly JD, Gettinger SN, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature*. 2014; 515:563–567. [PubMed: 25428504]
- Hirayama M, Nishimura Y. The present status and future prospects of peptide-based cancer vaccines. *Int. Immunol*. 2016; 28:319–328. [PubMed: 27235694]
- Hodges E, Rooks M, Xuan Z, Bhattacharjee A, Benjamin Gordon D, Brizuela L, Richard McCombie W, Hannon GJ. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nat. Protoc*. 2009; 4:960–974. [PubMed: 19478811]
- Hou X-L, Wang L, Ding Y-L, Xie Q, Diao H-Y. Current status and recent advances of next generation sequencing techniques in immunological repertoire. *Genes Immun*. 2016; 17:153–164. [PubMed: 26963138]
- Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, Berent-Maoz B, Pang J, Chmielowski B, Cherry G, et al. Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell*. 2016; 165:35–44. [PubMed: 26997480]
- Imai C, Mihara K, Andreansky M, Nicholson IC, Pui C-H, Geiger TL, Campana D. Chimeric receptors with 4-1BB signaling capacity provoke potent cytotoxicity against acute lymphoblastic leukemia. *Leukemia*. 2004; 18:676–684. [PubMed: 14961035]
- Jiang Y, Wang Y, Brudno M. PRISM: pair-read informed split-read mapping for base-pair level detection of insertion, deletion and structural variants. *Bioinformatics*. 2012; 28:2576–2583. [PubMed: 22851530]
- Jiang Y, Turinsky AL, Brudno M. The missing indels: an estimate of indel variation in a human genome and analysis of factors that impede detection. *Nucleic Acids Res*. 2015; 43:7217–7228. [PubMed: 26130710]
- Johnson PLF, Goronzy JJ, Antia R. A population biological approach to understanding the maintenance and loss of the T-cell repertoire during aging. *Immunology*. 2014; 142:167–175. [PubMed: 24405293]
- Kalos M, Levine BL, Porter DL, Katz S, Grupp SA, Bagg A, June CH. T cells with chimeric antigen receptors have potent antitumor effects and can establish memory in patients with advanced leukemia. *Sci. Transl. Med*. 2011; 3:95ra73.
- Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*. 2015; 161:1187–1201. [PubMed: 26000487]
- Kloss CC, Condomines M, Cartellieri M, Bachmann M, Sadelain M. Combinatorial antigen recognition with balanced signaling promotes selective tumor eradication by engineered T cells. *Nat. Biotechnol*. 2013; 31:71–75. [PubMed: 23242161]

- Knuth A, Danowski B, Oettgen HF, Old LJ. T-cell-mediated cytotoxicity against autologous malignant melanoma: analysis with interleukin 2-dependent T-cell cultures. *Proc. Natl. Acad. Sci. U. S. A.* 1984; 81:3511–3515. [PubMed: 6610177]
- Kochenderfer JN, Wilson WH, Janik JE, Dudley ME, Stetler-Stevenson M, Feldman SA, Maric I, Raffeld M, Nathan D-AN, Lanier BJ, et al. Eradication of B-lineage cells and regression of lymphoma in a patient treated with autologous T cells genetically engineered to recognize CD19. *Blood.* 2010; 116:4099–4102. [PubMed: 20668228]
- Kreiter S, Vormehr M, van de Roemer N, Diken M, Löwer M, Diekmann J, Boegel S, Schrörs B, Vascotto F, Castle JC, et al. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature.* 2015; 520:692–696. [PubMed: 25901682]
- Kroeger DR, Milne K, Nelson BH. Tumor-Infiltrating Plasma Cells Are Associated with Tertiary Lymphoid Structures, Cytolytic T-Cell Responses, and Superior Prognosis in Ovarian Cancer. *Clin. Cancer Res.* 2016; 22:3005–3015. [PubMed: 26763251]
- Krøigård AB, Thomassen M, Lænkholm A-V, Kruse TA, Larsen MJ. Evaluation of Nine Somatic Variant Callers for Detection of Somatic Mutations in Exome and Targeted Deep Sequencing Data. *PLoS One.* 2016; 11:e0151664. [PubMed: 27002637]
- Kumar S, Razzaq SK, Vo AD, Gautam M, Li H. Identifying fusion transcripts using next generation sequencing. *Wiley Interdiscip. Rev. RNA.* 2016; 7:811–823. [PubMed: 27485475]
- Lao X-M, Liang Y-J, Su Y-X, Zhang S-E, Zhou XI, Liao G-Q. Distribution and significance of interstitial fibrosis and stroma-infiltrating B cells in tongue squamous cell carcinoma. *Oncol. Lett.* 2016; 11:2027–2034. [PubMed: 26998116]
- Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Luber BS, Azad NS, Laheru D, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N. Engl. J. Med.* 2015; 372:2509–2520. [PubMed: 26028255]
- Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbotham M, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature.* 2008; 456:66–72. [PubMed: 18987736]
- Li B, Li JZ. A general framework for analyzing tumor subclonality using SNP array and DNA sequencing data. *Genome Biol.* 2014; 15:473. [PubMed: 25253082]
- Li B, Li T, Pignon J-C, Wang B, Wang J, Shukla SA, Dou R, Chen Q, Hodi FS, Choueiri TK, et al. Landscape of tumor-infiltrating T cell repertoire of human cancers. *Nat. Genet.* 2016a; 48:725–732. [PubMed: 27240091]
- Li B, Severson E, Pignon J-C, Zhao H, Li T, Novak J, Jiang P, Shen H, Aster JC, Rodig S, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol.* 2016b; 17:174. [PubMed: 27549193]
- Li Y, Chien J, Smith DI, Ma J. FusionHunter: identifying fusion transcripts in cancer using paired-end RNA-seq. *Bioinformatics.* 2011; 27:1708–1710. [PubMed: 21546395]
- Lundegaard C, Lund O, Nielsen M. 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.* 2008a; 24:1397–1398. [PubMed: 18413329]
- Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. 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.* 2008b; 36:W509–W512. [PubMed: 18463140]
- Macosko EZ, Basu A, Satija R, Nemes J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell.* 2015; 161:1202–1214. [PubMed: 26000488]
- Mardis ER. DNA sequencing technologies: 2006–2016. *Nat. Protoc.* 2017; 12:213–218. [PubMed: 28055035]
- McKinney DM, Southwood S, Hinz D, Oseroff C, Arlehamn CSL, Schulten V, Taplitz R, Broide D, Hanekom WA, Scriba TJ, et al. A strategy to determine HLA class II restriction broadly covering the DR, DP, and DQ allelic variants most commonly expressed in the general population. *Immunogenetics.* 2013; 65:357–370. [PubMed: 23392739]
- Milone MC, Fish JD, Carpenito C, Carroll RG, Binder GK, Teachey D, Samanta M, Lakhai M, Gloss B, Danet-Desnoyers G, et al. Chimeric receptors containing CD137 signal transduction domains

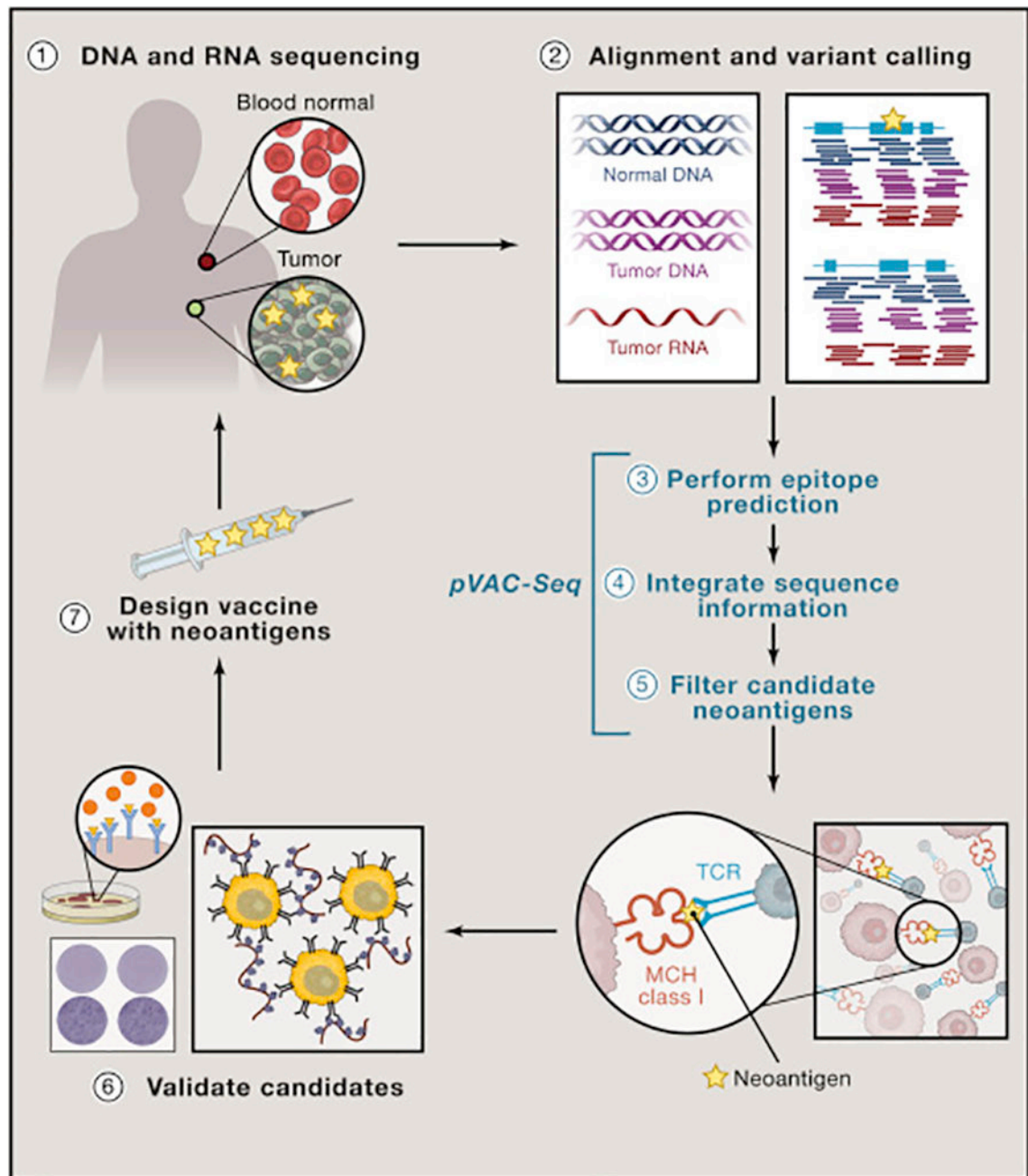
- mediate enhanced survival of T cells and increased antileukemic efficacy in vivo. *Mol. Ther.* 2009; 17:1453–1464. [PubMed: 19384291]
- Monach PA, Meredith SC, Siegel CT, Schreiber H. A unique tumor antigen produced by a single amino acid substitution. *Immunity.* 1995; 2:45–59. [PubMed: 7600302]
- Morgan RA, Dudley ME, Wunderlich JR, Hughes MS, Yang JC, Sherry RM, Royal RE, Topalian SL, Kammula US, Restifo NP, et al. Cancer regression in patients after transfer of genetically engineered lymphocytes. *Science.* 2006; 314:126–129. [PubMed: 16946036]
- Mose LE, Wilkerson MD, Hayes DN, Perou CM, Parker JS. ABRA: improved coding indel detection via assembly-based realignment. *Bioinformatics.* 2014; 30:2813–2815. [PubMed: 24907369]
- Mose LE, Selitsky SR, Bixby LM, Marron DL, Iglesia MD, Serody JS, Perou CM, Vincent BG, Parker JS. Assembly-based inference of B-cell receptor repertoires from short read RNA sequencing data with V<sup>2</sup>DJer. *Bioinformatics.* 2016
- Muraro PA, Robins H, Malhotra S, Howell M, Phippard D, Desmarais C, de Paula Alves Sousa A, Griffith LM, Lim N, Nash RA, et al. T cell repertoire following autologous stem cell transplantation for multiple sclerosis. *J. Clin. Invest.* 2014; 124:1168–1172. [PubMed: 24531550]
- Narzisi G, O'Rawe JA, Iossifov I, Fang H, Lee Y-H, Wang Z, Wu Y, Lyon GJ, Wigler M, Schatz MC. Accurate de novo and transmitted indel detection in exome-capture data using microassembly. *Nat. Methods.* 2014; 11:1033–1036. [PubMed: 25128977]
- Nelson BH. CD20+ B cells: the other tumor-infiltrating lymphocytes. *J. Immunol.* 2010; 185:4977–4982. [PubMed: 20962266]
- Newell EW, Sigal N, Bendall SC, Nolan GP, Davis MM. Cytometry by time-of-flight shows combinatorial cytokine expression and virus-specific cell niches within a continuum of CD8+ T cell phenotypes. *Immunity.* 2012; 36:142–152. [PubMed: 22265676]
- Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods.* 2015; 12:453–457. [PubMed: 25822800]
- Nielsen M, Lundegaard C, Wornig P, Lauemøller SL, Lamberth K, Buus S, Brunak S, Lund O. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci.* 2003; 12:1007–1017. [PubMed: 12717023]
- Old LJ, Boyse EA. IMMUNOLOGY OF EXPERIMENTAL TUMORS. *Annu. Rev. Med.* 1964; 15:167–186. [PubMed: 14139934]
- Overwijk WW, Wang E, Marincola FM, Rammensee H-G, Restifo NP. Mining the mutanome: developing highly personalized Immunotherapies based on mutational analysis of tumors. *J Immunother Cancer.* 2013; 1:11. [PubMed: 24829748]
- Pasternack, MS. *Fundamental Immunology*. 3. Paul, William E., editor. New York: Raven Press; 1994. 1993. p. 1,490illustrated. \$95. *Clin. Infect. Dis.* 19, 996–997
- Perricone MA, Smith KA, Claussen KA, Plog MS, Hempel DM, Roberts BL, St George JA, Kaplan JM. Enhanced efficacy of melanoma vaccines in the absence of B lymphocytes. *J. Immunother.* 2004; 27:273–281. [PubMed: 15235388]
- Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics.* 2005; 6:132. [PubMed: 15927070]
- Postow MA, Manuel M, Wong P, Yuan J, Dong Z, Liu C, Perez S, Tanneau I, Noel M, Courtier A, et al. Peripheral T cell receptor diversity is associated with clinical outcomes following ipilimumab treatment in metastatic melanoma. *J Immunother Cancer.* 2015; 3:23. [PubMed: 26085931]
- Prehn RT, Main JM. Immunity to methylcholanthrene-induced sarcomas. *J. Natl. Cancer Inst.* 1957; 18:769–778. [PubMed: 13502695]
- Qin Z, Richter G, Schüler T, Ibe S, Cao X, Blankenstein T. B cells inhibit induction of T cell-dependent tumor immunity. *Nat. Med.* 1998; 4:627–630. [PubMed: 9585241]
- Rasmussen M, Fenoy E, Harndahl M, Kristensen AB, Nielsen IK, Nielsen M, Buus S. Pan-Specific Prediction of Peptide-MHC Class I Complex Stability, a Correlate of T Cell Immunogenicity. *J. Immunol.* 2016; 197:1517–1524. [PubMed: 27402703]
- Ratan A, Olson TL, Loughran TP Jr, Miller W. Identification of indels in next-generation sequencing data. *BMC Bioinformatics.* 2015; 16:42. [PubMed: 25879703]

- Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, Lee W, Yuan J, Wong P, Ho TS, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*. 2015; 348:124–128. [PubMed: 25765070]
- Robbins PF, El-Gamil M, Li YF, Kawakami Y, Loftus D, Appella E, Rosenberg SA. A mutated beta-catenin gene encodes a melanoma-specific antigen recognized by tumor infiltrating lymphocytes. *J. Exp. Med.* 1996; 183:1185–1192. [PubMed: 8642260]
- Robbins PF, Morgan RA, Feldman SA, Yang JC, Sherry RM, Dudley ME, Wunderlich JR, Nahvi AV, Helman LJ, Mackall CL, 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; 29:917–924. [PubMed: 21282551]
- Robbins PF, Lu Y-C, El-Gamil M, Li YF, Gross C, Gartner J, Lin JC, Teer JK, Cliften P, Tycksen E, et al. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat. Med.* 2013; 19:747–752. [PubMed: 23644516]
- Robert L, Tsoi J, Wang X, Emerson R, Homet B, Chodon T, Mok S, Huang RR, Cochran AJ, Comin-Anduix B, et al. CTLA4 blockade broadens the peripheral T-cell receptor repertoire. *Clin. Cancer Res.* 2014; 20:2424–2432. [PubMed: 24583799]
- Robins H. Immunosequencing: applications of immune repertoire deep sequencing. *Curr. Opin. Immunol.* 2013; 25:646–652. [PubMed: 24140071]
- Robinson J. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res.* 2003; 31:311–314. [PubMed: 12520010]
- Robinson J, Waller MJ, Parham P, Bodmer JG, Marsh SG. IMGT/HLA Database—a sequence database for the human major histocompatibility complex. *Nucleic Acids Res.* 2001; 29:210–213. [PubMed: 11125094]
- Robinson J, Halliwell JA, McWilliam H, Lopez R, Parham P, Marsh SGE. The IMGT/HLA database. *Nucleic Acids Res.* 2013; 41:D1222–D1227. [PubMed: 23080122]
- Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell*. 2015; 160:48–61. [PubMed: 25594174]
- Rosenberg SA, Restifo NP. Adoptive cell transfer as personalized immunotherapy for human cancer. *Science*. 2015; 348:62–68. [PubMed: 25838374]
- Saethang T, Hirose O, Kimkong I, Tran VA, Dang XT, Nguyen LAT, Le TKT, Kubo M, Yamada Y, Satou K. PAAQD: Predicting immunogenicity of MHC class I binding peptides using amino acid pairwise contact potentials and quantum topological molecular similarity descriptors. *J. Immunol. Methods*. 2013; 387:293–302. [PubMed: 23058674]
- Sahin U, Türeci O, Schmitt H, Cochlovius B, Johannes T, Schmits R, Stenner F, Luo G, Schobert I, Pfreundschuh M. Human neoplasms elicit multiple specific immune responses in the autologous host. *Proc. Natl. Acad. Sci. U. S. A.* 1995; 92:11810–11813. [PubMed: 8524854]
- Saul L, Ilieva KM, Bax HJ, Karagiannis P, Correa I, Rodriguez-Hernandez I, Josephs DH, Tosi I, Egbuniwe IU, Lombardi S, et al. IgG subclass switching and clonal expansion in cutaneous melanoma and normal skin. *Sci. Rep.* 2016; 6:29736. [PubMed: 27411958]
- Schultz KR, Klarnet JP, Gienni RS, HayGlass KT, Greenberg PD. The role of B cells for in vivo T cell responses to a Friend virus-induced leukemia. *Science*. 1990; 249:921–923. [PubMed: 2118273]
- Schumacher T, Bunse L, Pusch S, Sahm F, Wiestler B, Quandt J, Menn O, Osswald M, Oezen I, Ott M, et al. A vaccine targeting mutant IDH1 induces antitumour immunity. *Nature*. 2014; 512:324–327. [PubMed: 25043048]
- Scolnick JA, Dimon M, Wang I-C, Huelga SC, Amorese DA. An Efficient Method for Identifying Gene Fusions by Targeted RNA Sequencing from Fresh Frozen and FFPE Samples. *PLoS One*. 2015; 10:e0128916. [PubMed: 26132974]
- enbabao lu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, Miao D, Ostrovskaya I, Drill E, Luna A, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biol.* 2016; 17:231. [PubMed: 27855702]
- Shi J-Y, Gao Q, Wang Z-C, Zhou J, Wang X-Y, Min Z-H, Shi Y-H, Shi G-M, Ding Z-B, Ke A-W, et al. Margin-infiltrating CD20(+) B cells display an atypical memory phenotype and correlate with

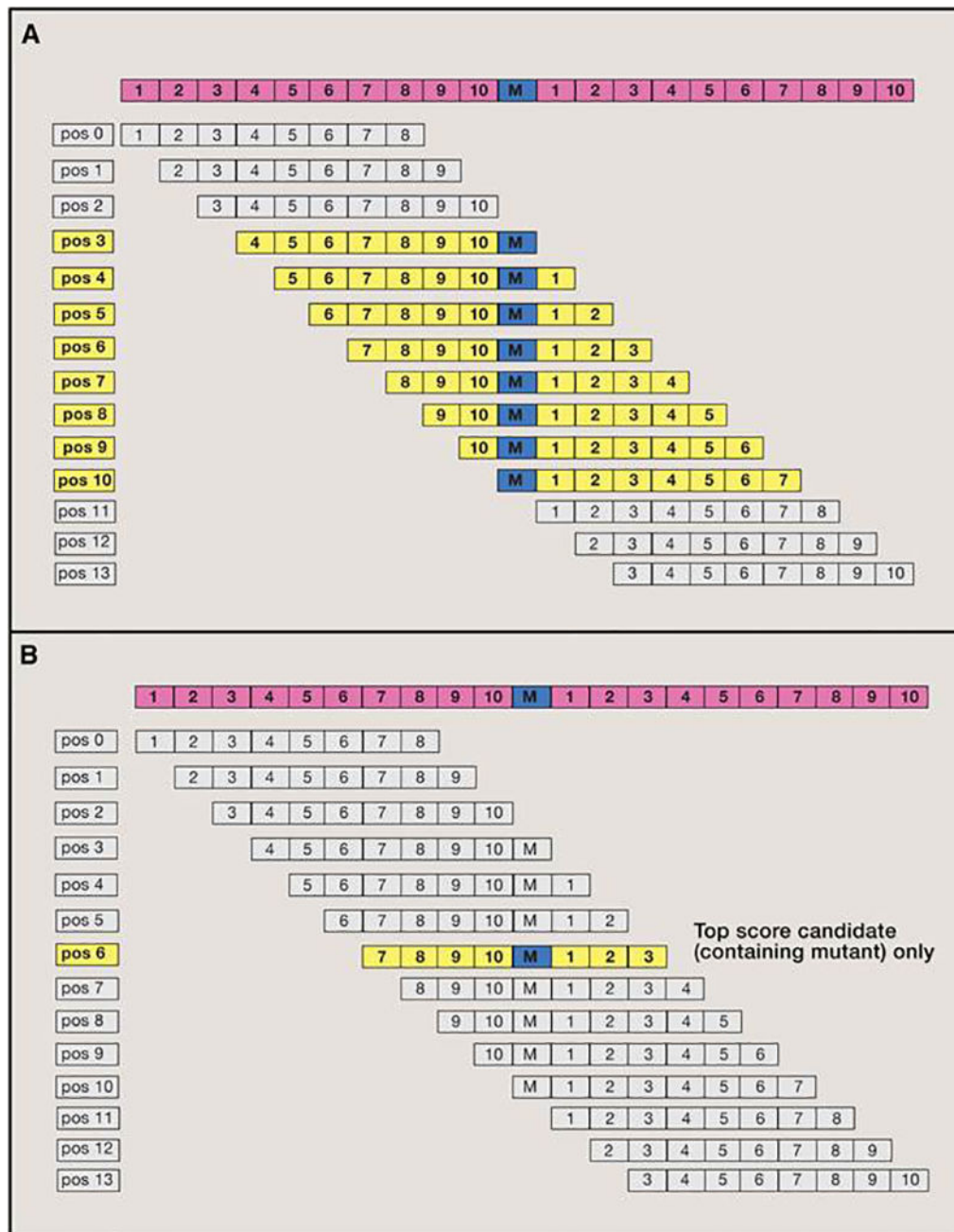
favorable prognosis in hepatocellular carcinoma. *Clin. Cancer Res.* 2013; 19:5994–6005. [PubMed: 24056784]

- Shugay M, Britanova OV, Merzlyak EM, Turchaninova MA, Mamedov IZ, Tuganbaev TR, Bolotin DA, Staroverov DB, Putintseva EV, Plevova K, et al. Towards error-free profiling of immune repertoires. *Nat. Methods.* 2014; 11:653–655. [PubMed: 24793455]
- Shugay M, Bagaev DV, Turchaninova MA, Bolotin DA, Britanova OV, Putintseva EV, Pogorelyy MV, Nazarov VI, Zvyagin IV, Kirgizova VI, et al. VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput. Biol.* 2015; 11:e1004503. [PubMed: 26606115]
- Shukla SA, Rooney MS, Rajasagi M, Tiao G, Dixon PM, Lawrence MS, Stevens J, Lane WJ, Dellagatta JL, Steelman S, et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* 2015; 33:1152–1158. [PubMed: 26372948]
- Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, Walsh LA, Postow MA, Wong P, Ho TS, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* 2014; 371:2189–2199. [PubMed: 25409260]
- Srivastava A, Ghosh S, Anantharaman N, Jayaraman VK. Hybrid biogeography based simultaneous feature selection and MHC class I peptide binding prediction using support vector machines and random forests. *J. Immunol. Methods.* 2013; 387:284–292. [PubMed: 23058675]
- Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics.* 2014; 30:3310–3316. [PubMed: 25143287]
- Taube JM, Anders RA, Young GD, Xu H, Sharma R, McMiller TL, Chen S, Klein AP, Pardoll DM, Topalian SL, et al. Colocalization of inflammatory response with B7-h1 expression in human melanocytic lesions supports an adaptive resistance mechanism of immune escape. *Sci. Transl. Med.* 2012; 4:127ra37.
- Tirosh I, Izar B, Prakadan SM, Wadsworth MH 2nd, Treacy D, Trombetta JJ, Rotem A, Rodman C, Lian C, Murphy G, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science.* 2016; 352:189–196. [PubMed: 27124452]
- Tran E, Turcotte S, Gros A, Robbins PF, Lu Y-C, Dudley ME, Wunderlich JR, Somerville RP, Hogan K, Hinrichs CS, et al. Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science.* 2014; 344:641–645. [PubMed: 24812403]
- Tran E, Robbins PF, Lu Y-C, Prickett TD, Gartner JJ, Jia L, Pasetto A, Zheng Z, Ray S, Groh EM, et al. T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. *N. Engl. J. Med.* 2016; 375:2255–2262. [PubMed: 27959684]
- Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJM, Robert L, Chmielowski B, Spasic M, Henry G, Ciobanu V, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature.* 2014; 515:568–571. [PubMed: 25428505]
- Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, Sucker A, Hillen U, Geukes Foppen MH, Goldinger SM, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science.* 2015; 350:207–211. [PubMed: 26359337]
- Van den Eynde B, Hainaut P, Hérin M, Knuth A, Lemoine C, Weynants P, van der Bruggen P, Fauchet R, Boon T. Presence on a human melanoma of multiple antigens recognized by autologous CTL. *Int. J. Cancer.* 1989; 44:634–640. [PubMed: 2529220]
- Vormehr M, Schrörs B, Boegel S, Löwer M, Türeci Ö, Sahin U. Mutanome Engineered RNA Immunotherapy: Towards Patient-Centered Tumor Vaccination. *J Immunol Res.* 2015; 2015:595363. [PubMed: 26844233]
- Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics.* 2010; 11:568. [PubMed: 21092157]
- Warren RL, Choe G, Freeman DJ, Castellarin M, Munro S, Moore R, Holt RA. Derivation of HLA types from shotgun sequence datasets. *Genome Med.* 2012; 4:95. [PubMed: 23228053]
- Yaari G, Kleinstein SH. Practical guidelines for B-cell receptor repertoire sequencing analysis. *Genome Med.* 2015; 7:121. [PubMed: 26589402]
- Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 2013; 4:2612. [PubMed: 24113773]

- Zaretsky JM, Garcia-Diaz A, Shin DS, Escuin-Ordinas H, Hugo W, Hu-Lieskovan S, Torrejon DY, Abril-Rodriguez G, Sandoval S, Barthly L, et al. Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. *N. Engl. J. Med.* 2016; 375:819–829. [PubMed: 27433843]
- Zhang H, Lund O, Nielsen M. The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. *Bioinformatics.* 2009; 25:1293–1299. [PubMed: 19297351]
- Zhang J, White NM, Schmidt HK, Fulton RS, Tomlinson C, Warren WC, Wilson RK, Maher CA. INTEGRATE: gene fusion discovery using whole genome and transcriptome data. *Genome Res.* 2016a; 26:108–118. [PubMed: 26556708]
- Zhang J, Mardis ER, Maher CA. INTEGRATE-neo: a pipeline for personalized gene fusion neoantigen discovery. *Bioinformatics.* 2016b
- Zhang N, Wu H-J, Zhang W, Wang J, Wu H, Zheng X. Predicting tumor purity from methylation microarray data. *Bioinformatics.* 2015; 31:3401–3405. [PubMed: 26112293]
- Zhang X, Sharma PK, Peter Goedegebuure S, Gillanders WE. Personalized cancer vaccines: Targeting the cancer mutanome. *Vaccine.* 2016c
- Zheng X, Zhao Q, Wu H-J, Li W, Wang H, Meyer CA, Qin QA, Xu H, Zang C, Jiang P, et al. MethylPurify: tumor purity deconvolution and differential methylation detection from single tumor DNA methylomes. *Genome Biol.* 2014; 15:419. [PubMed: 25103624]



**Fig 1. An Overall Workflow for Neoantigen Discovery and Personalized Cancer Vaccine Design** Starting from next-generation sequencing of DNA exomes to compare tumor to normal DNA, and of tumor RNA to evaluate gene expression, this figure illustrates the steps outlined in the primer to identify tumor-specific mutant antigens (neoantigens) from NGS data, to evaluate the neoantigens, and to design a personalized neoantigen vaccine.

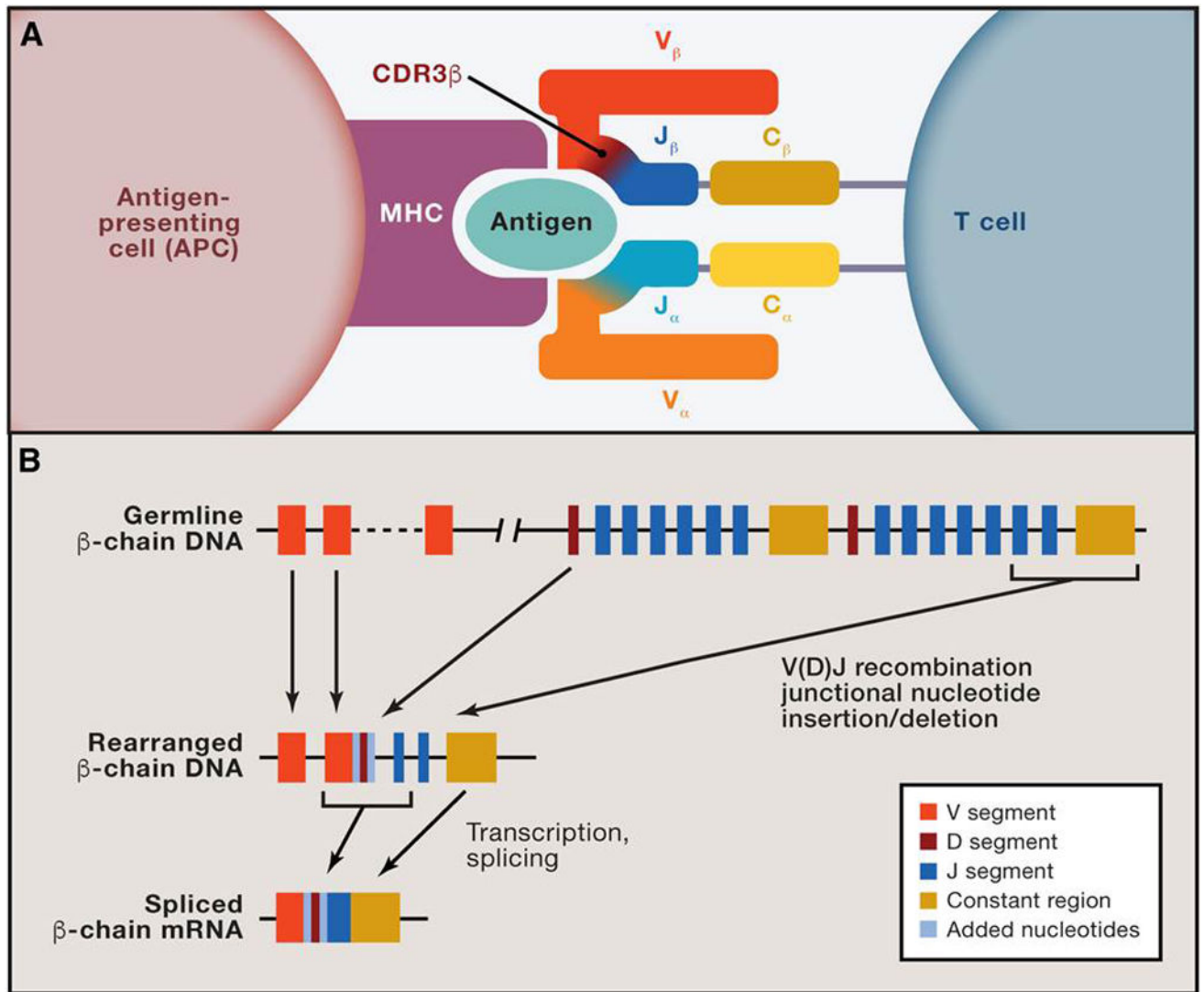


**Figure 2. Idealized Selection of Mutant-Containing Peptides for Neoantigen Prediction**

(A) The localized peptides that tile across and contain the mutated amino acid substitution are identified and parsed into the neoantigen prediction pipeline. Each peptide is considered for HLA binding strength relative to its non-mutant (wild-type) counterpart.

(B) Shown is the top scoring candidate peptide that was selected across all specified k-mers and between all HLA types that were input to the neoantigen prediction pipeline.





**Fig 3. Structure and Diversity in the T Cell Receptor**

(A) The mature T cell heterodimer, consisting of  $\alpha$ - and  $\beta$ -subunit chains. The  $\alpha$  subunit chains consist of variable (V), joining (J), and constant (C) regions, whereas the  $\beta$  subunit includes an additional diversity (D) region.

(B) V-D-J recombination and post-transcriptional processing of a TCR- $\beta$  subunit chain.