Published in final edited form as:

Curr Opin Biotechnol. 2016 December ; 42: 92–97. doi:10.1016/j.copbio.2016.03.001.

Unmasking targets of T cell-mediated antitumor immunity through high-throughput antigen profiling

Sebastiano Battaglia¹ and Jason B Muhitch²

¹Center for Immunotherapy, Roswell Park Cancer Institute, Buffalo, NY, United States

²Department of Urology, Roswell Park Cancer Institute, Buffalo, NY, United States

Abstract

More than three decades of evidence has established that antitumor immune responses, initially shown with IL-2 treatment, can result in complete, durable eradication of malignant disease in metastatic patients. Recent studies have demonstrated that immune checkpoint blockade as well as cellular therapies, including dendritic cell activation of T cells and adoptive T cell transfer, can induce long-lasting responses. To elicit cytotoxicity of tumor cells, effector T cells rely on tumor expression of target antigens. However, the antigens targeted during antitumor responses are largely unknown. Technological advancements and availability of sequencing data have paved the way for more efficient screening and validation of tumor-associated antigens and neoantigens derived from non-synonymous mutations targeted by T cells under baseline conditions and in the context of immunotherapy.

Elimination of antigen expressing tumors: A common link in immunotherapeutic regimens

Recent innovations for immune-based treatment of cancer have led to a resurgence of enthusiasm in immunotherapies due to a broadening of responsive tumor types and more well-tolerated treatment programs. In this regard, 8 out of 19 cancer therapies that received FDA breakthrough status in 2015 were designed to improve antitumor immune responses. The identification of novel molecular immune checkpoints [1], along with innovative adoptive T cell transfer [2,3] and dendritic cell vaccine protocols in clinical trials [4], as well as combinatorial approaches that utilize conventional immunotherapy [5], define immunotherapy as an evolving treatment option with the documented potential to eradicate metastatic malignancies.

The majority of immunotherapy protocols are designed to activate effector T cell responses in order to mediate tumor regression, yet the targets or tumor antigens that are recognized by T cells are often ill-defined. Tumor antigens are divided into two classes: First, tumor-associated antigens that are overexpressed in tumors and have restricted expression in normal tissues, and second, neoantigens originating from non-synonymous mutations in the

tumor microenvironment. Pioneering studies performed more than twenty years ago utilized autologous T cell clones and patient-derived melanoma cell lines to identify the first tumor-associated antigen (melanoma antigen-1 or MAGE-1) and the first neoantigen [6–8]. In these studies, autologous T cell clones were painstakingly cultured with melanoma cell lines transfected with potential antigens and the candidate antigen identified on the basis of T cell reactivity. Other methods of antigen identification include mass spectrometry [4], transcript comparison between tumor and normal tissues [9,10], and screening of cDNA library products for reactivity with antibodies present in cancer patient serum [11]. Tumor-associated antigens identified through these conventional methods have been utilized in immunotherapies that in some cases resulted in durable complete responses [3,12]. Importantly, complete and durable responses following therapy designed to target a single tumor-associated antigen can be observed in tumors with heterogeneous expression of that particular antigen [12]. This is speculated to occur due to antigen spreading whereby distinct antigens not targeted by the initial therapy are released during tumor cell cytolysis and processed by antigen presenting cells (APC) which then activate endogenous T cell responses.

Current studies have leveraged unprecedented access to next generation sequencing data and tumor specimens with *in silico* analysis algorithms to profile the neoantigens targeted by intratumoral T cells. Whole exome sequencing opens the possibility of efficiently characterizing neoantigens arising from somatic mutations in coding regions, which are likely to cause changes in the amino acid sequence. Neoantigens represent potentially superior immune targets since neoantigen-specific T cell clones have not been deleted during negative selection of self-reactive cells that occurs during T cell development. However, in order for an antigen to be targeted by T cells, it must be presented in the context major histocompatibility complex (MHC, known as human leukocyte antigen (HLA) in humans) on tumor cells. Furthermore, clinical tumor samples contain stromal cells, macrophages, tumor infiltrating lymphocytes and other immune cells, which can present technical challenges to efficiently identify truly actionable neoantigens. Therefore, a coordinated workflow, such as the one described below (Figure 1), is required.

Non-synonymous mutation discovery through sequencing and data processing

Restricting sequencing efforts to exons (~1% of the genome) using whole exome sequencing is a cost-effective method to evaluate non-synonymous mutations in patient samples. An important consideration for using this technology is that while the vast majority of neoantigens are likely contained within exons [13], neoantigens resulting from mutations in adjacent intronic sequences have also been identified [7]. High cellular heterogeneity and the rarity of neoantigens contained within tumor samples make sequencing coverage (or depth) a key factor for reliable single nucleotide variant (SNV) detection. An average coverage of 100–150 × is desirable and paired end is preferred over single end sequencing since the DNA fragments are sequenced from both sides. Each sequencing experiment produces short strings of nucleotides called reads and tens of millions of reads are produced by a single sequencing run per sample. To correctly place reads on the genomic space, a known

reference genome is needed. The choice of the reference genome must be consistent across the whole experiment, from reads mapping to variant calls and annotation, since coordinates vary across different builds. For the human genome, the latest build available at the time of writing is hg38/GRCh38, downloadable from NCBI and UCSC [14]. STAR [15], BWA [16] and Bowtie [17,18] are some of the most commonly used and freely available tools to map raw reads to a reference genome. The output from these mapping tools will often be sequence alignment files (SAM) [19] or their binary version, BAM files. These files indicate the region of the genome where the read was mapped, the detected sequence in the reference genome, and a series of quality scores for the alignment [16]. At this point numerous algorithms used to identify SNVs termed variant callers, including Samtools [16,20], VarScan [21], SomaticSniper [22], Strelka [23], GATK [24,25,26] and MuTect [27], can be used to identify SNVs. When making somatic SNV predictions it is paramount to use a normal control tissue, often blood, to discriminate somatic from germline SNVs. Germline mutations can be described as variations from the reference genome that are present in both normal and tumor samples. In contrast, somatic mutations are SNVs present in the tumor sample but not in the normal tissue or reference genome. The most common output from this analysis is a variant call format file (VCF), containing information on the position of the variant, reference and alternative sequence as well as quality data on the variant call.

When multiple callers are used the risk of over-filtering the data has to be taken into account. Each algorithm has pros and cons (summarized in previous literature [28,29]), so it is often required to choose educated thresholds to filter potential false positives. An example of such an approach can be taken from Rizvi *et al.* who utilized four different mutation callers and sequential filtering steps to identify actionable neoantigens in primary lung cancer samples [30*]. The first filtering step was based upon coverage (>7X), frequency of the variant nucleotide in the tumor (>10%), and frequency of the reference nucleotide in normal tissue (>97%). Variants identified by two or more callers that passed this initial filtering step were accepted. Excluded variants and variants called by a single caller that passed the initial screening, were further filtered by removing known shared single-nucleotide polymorphisms (SNP) and manually curating the SNVs with the integrative genomics viewer or IGV [30*]. The advantage of this filtering approach lies on the use of different thresholds to obtain a comprehensive profile of somatic SNVs that considers sample heterogeneity and different callers' sensitivity and specificity. Once variants have been identified, the resulting SNVs have to be filtered based on whether they are non-synonymous and their potential immunogenicity. Annovar [31] is a commonly used tool for variant annotation, and others, like SnpEff [32] and Oncotator [33], can predict deleterious SNVs (or SNVs likely to change gene function) using scores created by SIFT [34] and Polyphen2 [35] algorithms. Deleterious SNVs are more likely than neutral SNVs to cause changes in protein conformation, function, and downstream signaling pathways that are often deregulated in cancer and thus make for attractive immunotherapy targets.

Immunogenic prediction and validation of non-synonymous mutations

The majority of mutations identified in protein coding sequences are unlikely to represent antigens that can be detected by autologous T cells. A recent study has shown that autologous intratumoral T cells respond to approximately 0.5% of mutated peptides

expressed by the tumor [36*]. A critical limiting step to T cell activation is the capacity of a particular peptide to bind MHC with enough affinity to be stabilized and subsequently presented to T cells. Approximately 1 in 200 peptides will bind to a particular class I MHC with the requisite affinity to elicit T cell activation [37,38]. Interestingly, mutations of class I or class II MHC, that may limit antigen presentation and T cell-mediated antitumor responses have been noted [39,40], and can be evaluated during this step in the neoantigen discovery process. To correctly predict HLA binding and potential immunogenicity, it is required to know the patient's HLA type, which can be profiled using commercial services or next generation sequencing data. On the computational side, HLAMiner is a tool developed by Warren *et al.* commonly used to predict the HLA type by parsing RNA-Seq [41], WES or WGS data and was successfully used to identify the HLA type from TCGA data [42]. Previous studies described the relationship between peptide length and affinity to HLA [43,44], suggesting that length of the potential antigen should be taken into account. NetChop [45] and NetMHC [46,47] are neuronal network-based algorithms that predict peptide cleavage and affinity to the HLA, respectively. SYFPEITHI [48] is a database for MHC ligands and predicts peptide immunogenicity by providing a numeric score based on the binding strength of the peptide at 2 key positions: the anchor and auxiliary anchor. NetMHC and SYFPEITHI are based on a sliding window approach. The algorithms utilize a predetermined peptide of length n and slide the 'window' of size n one residue at a time from the first amino acid to the end of the peptide, thus analyzing each possible n -mer. Since NetMHC and SYFPEITHI analyses do not consider protein digestion, integrating a predicted protein cleavage step could increase the likelihood of identifying peptides likely to be recognized by a particular HLA complex. As for identifying somatic SNVs, accuracy of the prediction can be greatly improved by comparing the predicted immunogenicity of normal versus mutated peptides, with desirable higher affinity of the mutated antigen for HLA.

Even after using HLA prediction algorithms, the percentage of remaining candidate antigens that are actually recognized by tumor infiltrating T cells can be less than 5% [49*]. Validation of potential peptides is therefore a critical step towards identifying true neoantigens. This lengthy process requires coculture of autologous T cells with APC that have been loaded with either synthesized peptide or transfected to present the potential antigen. Although the percentage of tumor infiltrating CD4 and/or CD8 T cells that are specific for neoantigens is highly variable (<1–65% of T cells at a tumor site) [13,50], these cells are utilized due to the increased likelihood that neoantigen-specific clones will be present in this population compared to circulating lymphocytes [36*,49*,51]. There are fewer restrictions on the source of APC which often are B cells [36*] or dendritic cells derived from circulating pools of monocytes [2*].

Putative tumor antigens can be functionally screened using a number of different methods. Work from Steven Rosenberg's group has utilized a minigene approach where the mutation is flanked on each side by 12 amino acids from the reference sequence-derived protein and multiple minigenes are loaded into tandem minigene constructs (TMGs) [2*,49*] which are *in vitro* transcribed into RNA, which is then electroporated into APC. TMG-presenting APC are co-cultured with T cells and T cell responses are measured by IFN- γ secretion. To elucidate the specific mutation that the T cells respond to, new TMGs are constructed each

with a single reference sequence reversion for a particular mutation. Other groups have utilized synthesized mutated peptides and loaded onto APC to evaluate T cell reactivity [36*].

Utilizing antigen discovery for prognosis and therapeutic intervention

Immunogenic antigens expressed by patient tumors have already been characterized under baseline conditions and following immunologic therapies such as checkpoint inhibitors [52], adoptive T cell transfer [49*], dendritic cell vaccination [4], and hematopoietic stem cell transplantation [51,53]. Importantly, overall survival [42] and clinical response to therapy is associated with neoantigen load within tumors [52,54]. However, the greater promise of these antigen discovery strategies is to reach beyond informing clinicians of potential responders to immunotherapies that target known antigens by offering additional therapeutic targets and truly personalized medicine. In this respect, an exciting recent report of a single patient with metastatic disease experiencing tumor regression following adoptive transfer of neoantigens-specific CD4 T cells provides evidence that these approaches are feasible [2*]. As bioinformatics and sequencing approaches have streamlined the path towards identification of candidate antigens, challenges remain in optimizing validation protocols that still involve isolation, expansion, and monitoring of tumor infiltrating lymphocytes upon stimulation with candidate antigens.

Acknowledgments

This work was supported by an award from the Roswell Park Alliance Foundation, NCI grant P30 CA016056, and the Fraternal Order of Eagles. The authors thank M. Appenheimer and S. Hess for critical reading of the manuscript.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Sarter K, Leimgruber E, Gobet F, Agrawal V, Dunand-Sauthier I, Barras E, Mastelic-Gavillet B, Kamath A, Fontannaz P, Guery L, et al. B2m2, a T cell immunomodulatory molecule coregulated with MHC class II genes. *J Exp Med*. 2016; 213:177–187. [PubMed: 26809444]
- 2*. Tran E, Turcotte S, Gros A, Robbins PF, Lu YC, Dudley ME, Wunderlich JR, Somerville RP, Hogan K. Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science*. 2014; 344:641–645. A first of its kind clinical report demonstrating the identification of endogenous neoantigen T cell reactivity in a patient with a metastatic epithelial tumor followed by adoptive transfer therapy utilizing a mostly mutation-specific T cells. Tumor regression was noted in this patient following adoptive transfer and maintained at followup six months following treatment. This manuscript documents the first known treatment and tumor regression in human disease with personalized immunotherapy based on a mutated antigen. [PubMed: 24812403]
3. Rapoport AP, Stadtmauer EA, Binder-Scholl GK, Golubeva O, Vogl DT, Lacey SF, Badros AZ, Garfall A, Weiss B, Finklestein J, et al. NY-ESO-1-specific TCR-engineered T cells mediate sustained antigen-specific antitumor effects in myeloma. *Nat Med*. 2015; 21:914–921. [PubMed: 26193344]

4. Carreno BM, Magrini V, Becker-Hapak M, Kaabinejadian S, Hundal J, Petti AA, Ly A, Lie WR, 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]
5. Seung SK, Curti BD, Crittenden M, Walker E, Coffey T, Siebert JC, Miller W, Payne R, Glenn L, Bageac A, et al. Phase 1 study of stereotactic body radiotherapy and interleukin-2—tumor and immunological responses. *Sci Transl Med*. 2012; 4:137ra174.
6. 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]
7. Coulie PG, Lehmann F, Lethe B, Herman J, Lurquin C, Andrawiss M, Boon T. A mutated intron sequence codes for an antigenic peptide recognized by cytolytic T lymphocytes on a human melanoma. *Proc Natl Acad Sci U S A*. 1995; 92:7976–7980. [PubMed: 7644523]
8. Wolfel T, Hauer M, Schneider J, Serrano M, Wolfel C, Klehmann-Hieb E, De Plaen E, Hankeln T, Meyer zum Buschenfelde KH, Beach D. A p16INK4a-insensitive CDK4 mutant targeted by cytolytic T lymphocytes in a human melanoma. *Science*. 1995; 269:1281–1284. [PubMed: 7652577]
9. Yadav M, Jhunjhunwala S, Phung QT, Lupardus P, Tanguay J, Bumbaca S, Franci C, Cheung TK, Fritsche J, Weinschenk T, et al. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature*. 2014; 515:572–576. [PubMed: 25428506]
10. Weinschenk T, Gouttefangeas C, Schirle M, Obermayr F, Walter S, Schoor O, Kurek R, Loeser W, Bichler KH, Wernet D, et al. Integrated functional genomics approach for the design of patient-individual antitumor vaccines. *Cancer Res*. 2002; 62:5818–5827. [PubMed: 12384544]
11. Chen YT, Gure AO, Tsang S, Stockert E, Jager E, Knuth A, Old LJ. Identification of multiple cancer/testis antigens by allogeneic antibody screening of a melanoma cell line library. *Proc Natl Acad Sci U S A*. 1998; 95:6919–6923. [PubMed: 9618514]
12. Hunder NN, Wallen H, Cao J, Hendricks DW, Reilly JZ, Rodmyre R, Jungbluth A, Gnjatich S, Thompson JA, Yee C. Treatment of metastatic melanoma with autologous CD4+ T cells against NY-ESO-1. *N Engl J Med*. 2008; 358:2698–2703. [PubMed: 18565862]
13. van Buuren MM, Calis JJ, Schumacher TN. High sensitivity of cancer exome-based CD8 T cell neo-antigen identification. *Oncoimmunology*. 2014; 3:e28836. [PubMed: 25083320]
14. Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, Nejad P, Lee BT, Learned K, Karolchik D, Hinrichs AS, et al. The UCSC Genome Browser database: 2016 update. *Nucleic Acids Res*. 2016; 44:D717–D725. [PubMed: 26590259]
15. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013; 29:15–21. [PubMed: 23104886]
16. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–1760. [PubMed: 19451168]
17. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9:357–359. [PubMed: 22388286]
18. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009; 10:R25. [PubMed: 19261174]
19. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome Project Data Processing S. The sequence alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25:2078–2079. [PubMed: 19505943]
20. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011; 27:2987–2993. [PubMed: 21903627]
21. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012; 22:568–576. [PubMed: 22300766]

22. Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, Ley TJ, Mardis ER, Wilson RK, Ding L. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics*. 2012; 28:311–317. [PubMed: 22155872]
23. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012; 28:1811–1817. [PubMed: 22581179]
24. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010; 20:1297–1303. [PubMed: 20644199]
25. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011; 43:491–498. [PubMed: 21478889]
26. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;11.10.11–11.10.33.
27. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013; 31:213–219. [PubMed: 23396013]
28. Wang Q, Jia P, Li F, Chen H, Ji H, Hucks D, Dahlman KB, Pao W, Zhao Z. Detecting somatic point mutations in cancer genome sequencing data: a comparison of mutation callers. *Genome Med*. 2013; 5:91. [PubMed: 24112718]
29. Kim SY, Speed TP. Comparing somatic mutation-callers: beyond Venn diagrams. *BMC Bioinform*. 2013; 14:189.
30. 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. Rizvi et al. demonstrated that patients with advanced non-small cells lung cancer (NSCLC) with a smoking signature have higher non-synonymous mutational load and correlated with improved progression-free survival after pembrolizumab treatment. The authors further demonstrated that ~91% of non-small cell lung cancer patients with low PD-1 expression and high non-synonymous mutational load had a beneficial long-term clinical response, compared to the ~10% of patients with high PD-1 expression and low non-synonymous mutational load. These results suggest that neoantigen load is the main driver in clinical response to pembrolizumab and cancer vaccines in NSCLC patients. [PubMed: 25765070]
31. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010; 38:e164. [PubMed: 20601685]
32. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012; 6:80–92. [PubMed: 22728672]
33. Ramos AH, Lichtenstein L, Gupta M, Lawrence MS, Pugh TJ, Saksena G, Meyerson M, Getz G. Oncotator: cancer variant annotation tool. *Hum Mutat*. 2015; 36:E2423–E2429. [PubMed: 25703262]
34. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009; 4:1073–1081. [PubMed: 19561590]
35. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010; 7:248–249. [PubMed: 20354512]
36. Linnemann C, van Buuren MM, Bies L, Verdegaal EM, Schotte R, Calis JJ, Behjati S, Velds A, Hilkman H, Atmioui DE, et al. High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4+ T cells in human melanoma. *Nat Med*. 2015; 21:81–85. Schumacher and colleagues utilized a mostly bioinformatic approach to identify approximately 100 candidate neoantigens per tumor in melanoma patients. They further demonstrated very few of these candidates are in fact recognized by tumor infiltrating CD4 T cells in patients following adoptive

T cell transfer. However, it is important to note that reactivity against neoantigens was demonstrated in 80% of patients and could be observed more than 6 months following adoptive T cell transfer. [PubMed: 25531942]

37. Yewdell JW, Bennink JR. Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Annu Rev Immunol.* 1999; 17:51–88. [PubMed: 10358753]
38. Sette A, Vitiello A, Rehman B, Fowler P, Nayarsina R, Kast WM, Melief CJ, Oseroff C, Yuan L, Ruppert J, et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J Immunol.* 1994; 153:5586–5592. [PubMed: 7527444]
39. 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]
40. Kreiter S, Vormehr M, van de Roemer N, Diken M, Lower 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]
41. Warren RL, Choe G, Freeman DJ, Castellari M, Munro S, Moore R, Holt RA. Derivation of HLA types from shotgun sequence datasets. *Genome Med.* 2012; 4:95. [PubMed: 23228053]
42. Brown SD, Warren RL, Gibb EA, Martin SD, Spinelli JJ, Nelson BH, Holt RA. Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. *Genome Res.* 2014; 24:743–750. [PubMed: 24782321]
43. Rist MJ, Theodossis A, Croft NP, Neller MA, Welland A, Chen Z, Sullivan LC, Burrows JM, Miles JJ, Brennan RM, et al. HLA peptide length preferences control CD8+ T cell responses. *J Immunol.* 2013; 191:561–571. [PubMed: 23749632]
44. Bell MJ, Burrows JM, Brennan R, Miles JJ, Tellam J, McCluskey J, Rossjohn J, Khanna R, Burrows SR. The peptide length specificity of some HLA class I alleles is very broad and includes peptides of up to 25 amino acids in length. *Mol Immunol.* 2009; 46:1911–1917. [PubMed: 19157553]
45. Nielsen M, Lundegaard C, Lund O, Kesmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics.* 2005; 57:33–41. [PubMed: 15744535]
46. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics.* 2015
47. Nielsen M, Lundegaard C, Worning 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]
48. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics.* 1999; 50:213–219. [PubMed: 10602881]
49. Robbins PF, Lu YC, El-Gamil M, Li YF, Gross C, Gartner J, Lin JC, Teer JK, Clifton 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. This study addresses tumor infiltrating CD4 T cell reactivity against neoantigens through a novel discovery and validation platform. The group led by Steven Rosenberg utilized a combination of whole exome sequencing and in silico analysis to identify candidate antigens. Neoantigens were verified by monitoring cytokine secretion of expanded tumor infiltrating T cells that were cocultured with immortalized B cells loaded with candidate peptides. The group found that the CD4 T cells isolated from patients responding to immunotherapy were reactive against a low frequency of candidate antigens. [PubMed: 23644516]
50. Tran E, Ahmadzadeh M, Lu YC, Gros A, Turcotte S, Robbins PF, Gartner JJ, Zheng Z, Li YF, Ray S, et al. Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science.* 2015; 350:1387–1390. [PubMed: 26516200]
51. Rajasagi M, Shukla SA, Fritsch EF, Keskin DB, DeLuca D, Carmona E, Zhang W, Sougnez C, Cibulskis K, Sidney J, et al. Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood.* 2014; 124:453–462. [PubMed: 24891321]

52. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, Sucker A, Hillen U, Foppen MH, Goldinger SM, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science*. 2015; 350:207–211. [PubMed: 26359337]
53. Takahashi Y, Harashima N, Kajigaya S, Yokoyama H, Cherkasova E, McCoy JP, Hanada K, Mena O, Kurlander R, Tawab A, et al. Regression of human kidney cancer following allogeneic stem cell transplantation is associated with recognition of an HERV-E antigen by T cells. *J Clin Invest*. 2008; 118:1099–1109. [PubMed: 18292810]
54. McGranahan N, Furness AJS, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, Jamal-Hanjani M, Wilson GA, Birkbak NJ, Hiley CT, Watkins TBK, Shafi S, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science*. 2016

Figure 1.

Identification and validation of immunogenic neoantigens expressed in tumors. Tumor samples and peripheral blood provide tumor infiltrating lymphocytes (TIL) and antigen presenting cells (APC) respectively that are used for validation. Whole exome sequencing of tumor and comparison with autologous healthy tissue (i.e. circulating leukocytes) is utilized for discovery of somatic mutations. A number of mutation callers can be used in conjunction with data annotation tools. After data filtering, algorithms such as NetChop, NetMHC and SYFPEITHI are utilized to predict cleavage length and affinity for MHC since less than 1% of somatic mutations bind a particular class of MHC. Finally, candidate antigens (Ag) are loaded onto APC either by pulsing with peptide or electroporation of RNA. TIL are cultured with these APC and monitored for cytokine secretion (IFN- γ , GM-CSF, IL-2) to determine which neoantigens elicit productive T cell responses.