



OPEN ACCESS

EDITED BY

Andreas Mayer,
University College London,
United Kingdom

REVIEWED BY

Zlatko Trajanoski,
Innsbruck Medical University, Austria
Ningbo Zheng,
University of Houston, United States
Paul F. Robbins,
Clinical Center (NIH), United States

*CORRESPONDENCE

Anthony E. Zamora

✉ azamora@mcw.edu

RECEIVED 24 September 2023

ACCEPTED 29 November 2023

PUBLISHED 12 December 2023

CITATION

Shah RK, Cygan E, Kozlik T, Colina A and Zamora AE (2023) Utilizing immunogenomic approaches to prioritize targetable neoantigens for personalized cancer immunotherapy. *Front. Immunol.* 14:1301100. doi: 10.3389/fimmu.2023.1301100

COPYRIGHT

© 2023 Shah, Cygan, Kozlik, Colina and Zamora. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Utilizing immunogenomic approaches to prioritize targetable neoantigens for personalized cancer immunotherapy

Ravi K. Shah¹, Erin Cygan², Tanya Kozlik¹, Alfredo Colina² and Anthony E. Zamora^{1,2*}

¹Department of Medicine, Medical College of Wisconsin, Milwaukee, WI, United States, ²Department of Microbiology and Immunology, Medical College of Wisconsin, Milwaukee, WI, United States

Advancements in sequencing technologies and bioinformatics algorithms have expanded our ability to identify tumor-specific somatic mutation-derived antigens (neoantigens). While recent studies have shown neoantigens to be compelling targets for cancer immunotherapy due to their foreign nature and high immunogenicity, the need for increasingly accurate and cost-effective approaches to rapidly identify neoantigens remains a challenging task, but essential for successful cancer immunotherapy. Currently, gene expression analysis and algorithms for variant calling can be used to generate lists of mutational profiles across patients, but more care is needed to curate these lists and prioritize the candidate neoantigens most capable of inducing an immune response. A growing amount of evidence suggests that only a handful of somatic mutations predicted by mutational profiling approaches act as immunogenic neoantigens. Hence, unbiased screening of all candidate neoantigens predicted by Whole Genome Sequencing/Whole Exome Sequencing may be necessary to more comprehensively access the full spectrum of immunogenic neopeptides. Once putative cancer neoantigens are identified, one of the largest bottlenecks in translating these neoantigens into actionable targets for cell-based therapies is identifying the cognate T cell receptors (TCRs) capable of recognizing these neoantigens. While many TCR-directed screening and validation assays have utilized bulk samples in the past, there has been a recent surge in the number of single-cell assays that provide a more granular understanding of the factors governing TCR-pMHC interactions. The goal of this review is to provide an overview of existing strategies to identify candidate neoantigens using genomics-based approaches and methods for assessing neoantigen immunogenicity. Additionally, applications, prospects, and limitations of some of the current single-cell technologies will be discussed. Finally, we will briefly summarize some of the recent models that have been used to predict TCR antigen specificity and analyze the TCR receptor repertoire.

KEYWORDS

cancer immunotherapy, immunogenomics, neoantigens, genomics-based approaches, single-cell technologies, immunogenic neopeptides, predicting TCR antigen specificity, personalized medicine

1 Introduction

Cancer significantly impacts human health and quality of life, and is one of the leading causes of death worldwide, with approximately 10 million deaths in 2020, according to WHO (1). Gene mutations caused by genomic instability during carcinogenesis (2), viral infections (3), alternative splicing (4), and gene rearrangements (5) can alter the protein coding regions of the genome resulting in aberrant proteins that are not found in normal cells. On occasion, these protein variants may be processed into small peptides by a tumor cell's proteasome and bind the major histocompatibility complex (MHC) molecules with sufficient affinity to serve as novel tumor antigens (i.e., tumor neoantigens) that can then be recognized by CD4+ or CD8+ T cells and elicit an antitumor response (6, 7). Collectively, the repertoire of peptides that are displayed on the surface of tumor cells are referred to as the immunopeptidome (8) and each neoepitope (neoantigen bound to a specific MHC molecule) can be recognized by a collection of TCRs resulting in neoantigen-specific TCR repertoires of varying diversity. Neoantigens resulting from missense or fusion mutations aren't expressed by healthy cells making these neoantigens safe targets for T-cell based immunotherapies due to the ability to generate robust T cell responses and decreases the likelihood of off target toxicity. If neoantigens play a significant role in T cell-mediated tumor resolution, one would posit that tumors with higher mutational burdens would have correspondingly greater frequencies of neoantigen-specific T cells. Along these lines, several clinical trials have explored whether tumor mutational burden and T cell infiltration correlate with efficacy of immune checkpoint blockade (ICB) and/or adoptive cell therapy (ACT) (9–12). These clinical investigations have shed light on the crucial link between tumor neoantigens and the efficacy of immunotherapeutic strategies. Tumors with higher mutational burdens have been associated with greater responsiveness to ICB, such as anti-PD-1 or anti-CTLA-4 antibodies, as the abundance of mutations has been associated with an increased likelihood of there being immunogenic neoantigens that can be recognized by infiltrating T cells, resulting in improved antitumor immune responses. A positive correlation exists between the tumor mutational burden (TMB) and the abundance of neoantigen-specific T cells within the tumor microenvironment, resulting in an elevated rate of response to immunotherapeutic interventions for some cancers (13). Despite these cases, it is noteworthy that low TMB can still give rise to neoantigen-reactive lymphocytes, particularly in hematological malignancies and specific epithelial cancers such as gastrointestinal cancers (14–16). More recently, personalized vaccinations in the form of mRNA, peptides, or peptide-loaded antigen presenting cells (APCs) have been shown to be safe, immunogenic and capable of generating durable clinical response (17–19). Although the tumor mutational burden is highly variable across different types of cancer, immunogenic neoantigens (i.e., neoantigens that induce T-cell activation and proliferation) have been identified in several cancer types, implying that although a patient might have a lower tumor mutational burden (and consequently fewer presented neoantigens) they may still derive

benefit from ICB and personalized immunotherapies (20, 21). Also, the majority of tumor neoantigens are private (i.e., unique to an individual rather than being present across the population), suggesting that future decisions regarding the best cancer therapy will need to be made on a case-by-case basis. To date, neoantigens present the most promising target for cell-based immunotherapies; however, the challenges associated with their identification requires adoption of reliable high-throughput neoantigen discovery pipelines to expedite the time it takes to manufacture a personalized cancer immunotherapy. This review will provide a comprehensive overview of currently used methods for identification and validation of candidate immunogenic neoantigens. Initially, we will cover the workflows for identifying candidate neoantigens using genomics-based methods. Next, we will review screening and validation approaches for the candidate neoantigens using antigen-directed and TCR-directed approaches. Furthermore, we will provide overview of reporter methods used to identify neoantigen-TCR pairs and introduce some of the novel single cell platforms that provide a more granular picture of how immune cells respond to tumors. Additionally, to bridge the experimental and computational realms, this review will also showcase the advancements in computational tools for predicting TCR antigen specificity. We will describe some of the innovative bioinformatics approaches and machine learning algorithms that leverage genomics data and experimental insights to predict TCR-PMHC interactions and identify potential target antigens.

2 Genomics-based neoantigen identification

Advances in high-throughput sequencing technologies, including greater sequencing speeds and higher accuracy, combined with significant reductions in sequencing cost have enabled researchers to identify an increasing number of somatic mutations and convert the mutated variants to putative neoantigens. The price for genetic sequencing has declined at an astonishing rate, with per genome costs dropping from \$50,000 in late 2000 to roughly \$600 in 2022 – the price of genome sequencing has recently been predicted to drop as low as \$100 per genome with the development of new high-throughput, low-cost sequencing platforms (22). The most common method for identifying putative neoantigens comes from sequencing DNA mutations and/or corresponding RNA. Additional approaches, such as eluting peptides bound to MHC (class I or II) molecules, do not require sequencing data, but these approaches have been covered at length in recent reviews (23–26). Whole genome sequencing (WGS) and whole exome sequencing (WES) combined with RNA-seq are the most widely used approaches for neoantigen discovery, each having their own advantages and limitations. WGS and WES have traditionally been the methods of choice for variant identification. WGS has the advantage of allowing for identification of variants in non-coding regions such as untranslated regions (UTRs) which are missed by WES approaches, but WGS is considerably more expensive than WES due to the extra sequencing coverage

required. WES, on the other hand, provides a fine balance between cost and benefits (27) – although this sequencing approach only targets approximately 2–3% of the whole genome, it provides data on the protein-coding regions believed to be the major source of somatic mutations. Combining RNA-seq with WGS or WES provides an additional layer of resolution by incorporating gene expression levels into the process of selecting genes that are more likely to produce translated proteins and filter out candidates that do not meet a predefined threshold. Moreover, other types of neoantigens such as gene fusions, alternative splicing isoforms, and RNA editing events can be revealed from RNA-seq data (28, 29). Although PCR methods have been considered the gold standard for human leukocyte antigen (HLA) typing, RNA-seq data can be used as an alternative given the recent improvements in prediction power now reaching levels comparable to PCR-based approaches (30, 31). Recently, an alternative approach to traditional mRNA sequencing, Ribo-Seq (32), has allowed for rapid identification of neoantigens by providing a snapshot of only the mRNA bound by ribosomes, which depicts all proteins being translated at the time of cell lysis and allows for identification of an expanded set of open reading frames (33, 34). Ribo-seq provides an RNA-sequencing based readout of mRNA translation by isolating ribosome bound RNA fragments, thereby offering a genome-wide footprint of ribosome-RNA interactions. As such, this approach circumvents the experimental difficulties of working with protein molecules and readily identifies translations which might have been missed by other methods. The advantage of Ribo-seq technology in predicting targetable neoantigens is that it identifies only those variants likely to generate proteins and it also provides a more reliable estimation of protein expression. However, it should also be noted that not all the translations reported by Ribo-seq will actually result in expressed proteins and a wide range of functional scenarios may be possible for Ribo-Seq ORF's (35) including making stable proteins, mediators of gene regulation and having medical implications.

The growing number of peptide:MHC (pMHC) neoepitopes that have been validated using traditional wet lab assays, as outlined below, has resulted in tremendous advancements in the field of neoantigen discovery, and have allowed new computational algorithms to be developed that more accurately predict which peptides bind to specific HLA molecules. While many of these bioinformatic pipelines rely on the same series of steps to prioritize targetable neoepitopes, namely filtering sequence quality, reference mapping, somatic mutation calling, mutated peptide sequence identification, and peptide ranking based on cellular processing and presentation pathways, the accuracy in neoepitope prediction is significantly impacted by the quality and quantity of information contained in the training sets used to develop these algorithms.

2.1 Identification of somatic mutations (variants)

Most sequencing efforts have focused on the identification of in-frame insertions and deletions (indels), out of frame insertions and deletions (frameshift mutations), and single nucleotide variants

(SNVs), with much less attention given to gene fusions, gene inversions, and gene duplications despite these mutations accounting for a significant number of the mutational landscape and potentially serving as neoantigens with greater immunogenicity. Non-SNV mutations have recently been shown to account for 15% of all neoantigens (36). Somatic variant identification is one of the most critical steps in any pipeline focused on neoantigen discovery as samples from tumor and matched non-tumor DNA sequencing data provides a list of all targetable neoantigens uniquely present in tumor samples. Table 1 lists some of the widely used mutation callers. It should be noted that NGS sequencing depth directly influences the reproducibility of variant detection; specifically, the higher the number of aligned sequenced reads, the higher the confidence to base call at a specific position, regardless of whether the base call is the same as the reference or mutated nucleotide. Also, higher sequencing depth achieves more sensitive detection of variants at low allele frequency. Numerous studies have directly compared the performance of somatic variant callers (SVCs) (51–53). Somatic variant callers are challenged by the need to balance between accurately identifying true low-allelic somatic mutations and the stringency of the calling procedure to reduce the number of false positive calls. Because of disparities in the number of mutations predicted by different calling algorithms, selection of an SVC is an important component of neoantigen prediction. Running multiple SVC algorithms simultaneously and using the consensus results can result in improved specificity for the variant detection (54). Consensus approaches present a trade-off as they often improve the specificity at the cost of sensitivity, as the increased specificity will decrease the number of false positives to be tested, but decreased sensitivity could result in missing clinically relevant variants. Regardless of which variant caller or approach is used to predict mutations, it is always recommended to validate the putative mutations by manually reviewing the matched tumor-normal samples in Integrated Genomics Viewer (IGV) (55).

In addition to SNVs, there has been increasing demand for the development of tools that can provide proper identification of other neoantigen sources including large INDELS (46, 47) and gene fusions (48–50). Structural variants, typically large indels (with or without frameshift) and gene fusions can be identified using a single tool and do not benefit from the consensus approach (56).

2.2 Variant annotation

Variant annotation is the process of labeling the variants with genomic or genetic characteristics which can be categorized or prioritized for further investigation. Variant annotation is critical for neoantigen prediction as the mutations can impact the corresponding amino acid sequence and may result in silent variants, missense mutations, frameshifts, mutations in non-coding regions and gain or loss of stop codons, each of which can result in neoantigens with varying immunogenicities. Variant Effects Predictor (VEP) (57) and ANNOVAR (58) are two common variant annotation software programs and have been compared directly using the same set of transcripts (59).

TABLE 1 List of Bioinformatics tools used for variant calling and gene fusion detection.

| Tools | Approach/Method | URL | Type of call | Reference |
|----------------------------|--|---|--|-----------|
| Small Variants | | | | |
| EBCall | Allele Frequency Analysis | https://github.com/friend1ws/EBCall | SNV, Indel | (37) |
| Mutect | Allele Frequency Analysis | https://github.com/broadinstitute/mutect | SNV | (38) |
| Strelka | Allele Frequency Analysis | https://github.com/Illumina/strelka | SNV, Indel | (39) |
| Varscan2 | Heuristic Threshold | http://varscan.sourceforge.net/ | SNV, Indel | (40) |
| SomaticSniper | Joint Genotype analysis | https://github.com/genome/somatic-sniper | SNV | (41) |
| Virmid | Joint Genotype analysis | https://sourceforge.net/projects/virmid/ | SNV | (42) |
| VarDict | Heuristic Threshold | https://github.com/AstraZeneca-NGS/VarDict | SNV, Indel, Structural Variants | (43) |
| SnooPer | Machine Learning | https://sourceforge.net/projects/snooper/ | SNV, Indel | (44) |
| SomaticSeq | Machine Learning | https://github.com/bioinform/somaticseq | SNV | (45) |
| Structural Variants | | | | |
| DELLY | Paired end, read depth and Split-read analysis | https://github.com/dellytools/delly | Indels, duplications, inversions and translocation | (46) |
| Pindel | Paired end and split-read analysis | https://github.com/genome/pindel | Indels, Large deletions | (47) |
| Gene Fusions | | | | |
| STAR-Fusion | STAR Alignment, artifact filtering | https://github.com/STAR-Fusion/STAR-Fusion | Gene fusion | (48) |
| Arriba | STAR alignment, read and event level filtering | https://github.com/suhrig/arriba | Gene fusion, Structural variant | (49) |
| defuse | Paired end, split read alignment, filtering | https://github.com/amcpherson/defuse | Gene fusion | (50) |

Although it is difficult to accurately benchmark the success of different programs, in the comparison study VEP aligned more consistently with manually curated variants. However, ANNOVAR and VEP are command line driven and Perl based tools which can be inherently complicated for researchers without programming backgrounds to use. To interactively annotate the data without programming knowledge, a new R shiny based interactive application called ShAn (60) was developed and has demonstrated greater speed and online accessibility compared to VEP and ANNOVAR with comparable predictive capabilities.

2.3 Antigen processing algorithms

While neoepitope prediction algorithms have resulted in tremendous advancements in our ability to predict which neoantigens will likely bind to specific HLA molecules, greater accuracy can be achieved when upstream events involved in antigen presentation are incorporated. It is widely known that peptide-pulsing studies that rely solely on MHC-binding algorithms tend to overestimate the number of immunogenic neoantigens that would be naturally processed and presented due to the fact that not all full-length proteins will be cleaved by the proteasome or immunoproteasome and result in peptides with mutated amino acids being loaded onto specific MHCs (61, 62). Before MHC class I

presentation, peptides are transported by the transporter associated with antigen processing (TAP) protein to the endoplasmic reticulum (ER) and then trimmed by ER-related aminopeptidases (ERAP) present there (63). There are several tools that predict both proteosomal cleavage and account for TAP's peptide transportation efficiency. For MHC class I processing and presentation, NetChop20S (64) and ProteaSMM (65) have been shown to reliably predict *in vitro* cleavage patterns owing to the large proteasome digestion training sets used to train these models (66). For MHC class II (MHCII) processing and presentation PepCleaveCD4 (67) and MHCII-NP (68) are two tools that predict antigen excision positions resulting in epitopes that can be recognized by CD4+ T cells. While these tools have generally improved the accuracy of neoepitope prediction pipelines, a major limitation is that genes coding for proteins of various components of the antigen presentation machinery, such as TAP1, TAP2, B2M, are known sites of mutation in cancer. Therefore, the utility of these algorithms in identifying cancer neoepitopes needs to be evaluated on a case-by-case basis.

The affinity of peptide to a given MHC molecule is an important contributor to neoantigen immunogenicity and is a major factor that should be weighed by major epitope prediction algorithms. It is also crucial to know the HLA type of the patient before ranking the peptides as it is widely established that different MHC allotypes differ in specificity with respect to peptide binding. State-of-the-art

methods currently available to predict peptide presentation are based on artificial neural networks trained on large datasets and are known to perform best on frequent and well-characterized alleles (e.g. HLA-A*02:01) and their accuracy is decreased for rare or not well-studied HLA alleles. Also, HLA gene expression (69) and somatic mutation patterns (70) in this locus need to be evaluated as HLA downregulation or loss of heterozygosity are known mechanisms to disrupt neoantigen presentation and can result in immune escape. Currently, HLA class I typing algorithms relying on NGS data can accurately predict HLA class I alleles with up to 99% (71) accuracy when using WES/RNA-seq data, however HLA class II algorithms are less reliable and require further development to improve their prediction accuracy. Although many class I typing algorithms exist, Optitype (72), Polysolver (73) and PHLAT (74) are the most frequently used and have the highest reported accuracy. More recent algorithms, such as xHLA (75) and HLA-HD (76), have expedited the HLA typing process and show comparable accuracies to those methods mentioned above.

By combining peptide:MHC binding affinity and HLA typing algorithms or peptide elution and mass spectrometry-based assays, researchers can quickly generate a list of putative neoepitopes (Figure 1). Several recent reviews (77–82) and articles (83, 84) have provided comprehensive coverage and new considerations when developing computational algorithms for identifying cancer neoantigens. In Table 2, we provide an overview of the different genomic-based methods used for neoantigen prediction and describe their advantages and limitations. Incorporation of the antigen processing steps may provide a more accurate depiction of putative MHC-binding peptides; however, most of these algorithms cannot predict whether the bound peptides will be immunogenic (i.e., induce an active response by T cells) (85–87). Ultimately, for a neoepitope to serve as a therapeutic target, it must first be recognized by a cognate TCR and result in a productive TCR signaling cascade. While some MHC class I neoepitope prediction algorithms are beginning to factor in immunogenicity as a parameter for neoantigen prioritization, in most cases robust data sets are still lacking and traditional assays to perturb T cell

functionality (e.g. cytokine secretion and/or cytotoxicity assays using peptide-pulsed APCs) are necessary. These approaches are even more essential for MHC class II neoepitopes which are more difficult to predict using in silico algorithms. In Table 3, we provide an overview of frequently used neoantigen prediction tools and highlight key features for each.

3 Screening and validation steps to identify neoantigen-specific T cells

3.1 Antigen-directed approaches

Antigen-directed approaches have rapidly gained prominence as a method for identifying antigen-specific T cells. Early approaches to determine T cell reactivity against neoantigens included culturing T cell clones with target cells containing the appropriate HLA molecules and transfected with tumor complementary DNA (cDNA) library pools (97, 98). This approach has resulted in the identification of several well-known antigens including MAGE (99) and MART-1 (100). Serological analysis of recombinant cDNA expression helped identify NYESO-1 (101), another widely known tumor antigen. However, the major limitation of these approaches is that they require large numbers of neoantigen-specific T cells to perform the screens and to validate immunogenic neoantigens, which may be a limiting factor if cells obtained from biopsies or clinical specimens are scarce.

Another common neoantigen-specific T cell screening approach is to use pMHC tetramers (102) or multimers, which are oligomers formed from four (tetramer) or more (multimer) MHC subunits containing the neoantigen of interest. While tetramers are often sufficient to stain T cells with high-affinity TCR:pMHC interactions, multimers are often necessary for detecting TCRs that bind pMHC with lower affinities (KD < 10 μ M) and avidities (103, 104). Although laborious to produce, as unique multimers must be generated for every MHC:peptide combination, they have been used to identify thousands of

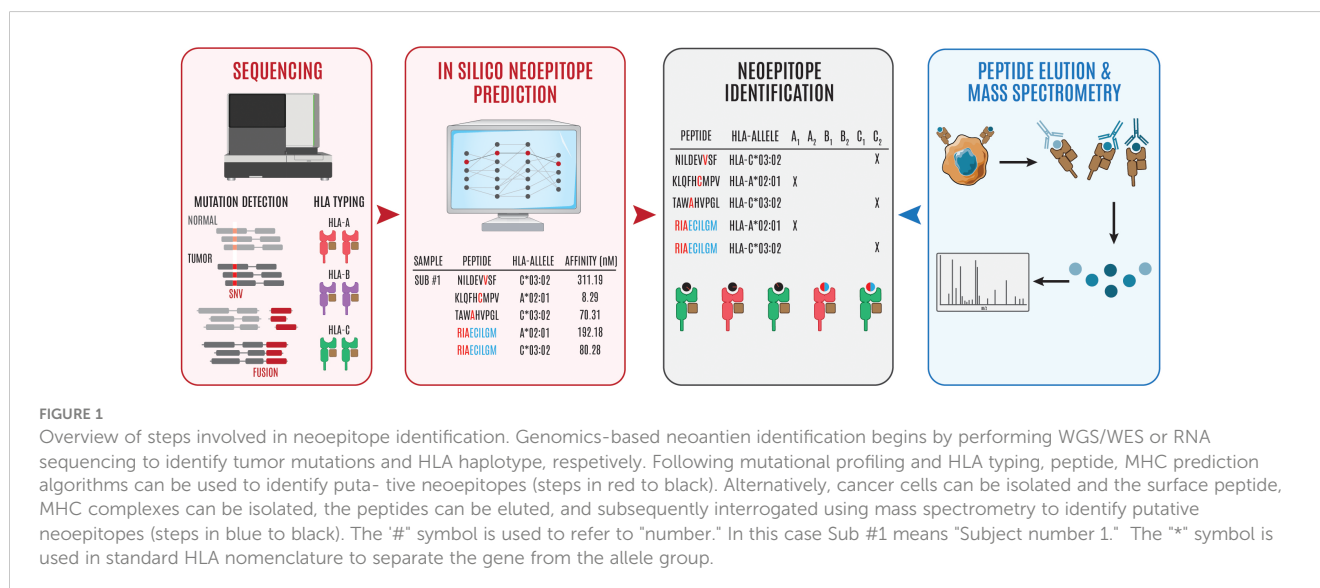


TABLE 2 Genomic and transcriptomic-based methods for identifying candidate neoantigens.

| Methods | Advantages | Limitations |
|------------------------------------|--|--|
| Whole Genome Sequencing (WGS) | Identifies variants in non-coding regions, providing comprehensive genomic information | Costlier due to the requirement for extra sequencing coverage, potentially leading to higher expenses. |
| Whole Exome Sequencing (WES) | Balances cost and benefits by targeting protein-coding regions, capturing essential genomic information at a more manageable expense | Targets only 2-3% of the genome, potentially missing variations in non-coding regions that could be biologically significant. |
| RNA-seq and Alternative Approaches | <ul style="list-style-type: none"> - RNA-seq reveals gene fusions, alternative splicing, and RNA editing events, offering insights into post-transcriptional regulation. - Ribo-seq provides a snapshot of mRNA bound by ribosomes, offering a genome-wide footprint of ribosome-RNA interactions. | <ul style="list-style-type: none"> - Not all translations reported by Ribo-seq may result in expressed proteins, necessitating careful interpretation. - Different functional scenarios are possible for Ribo-Seq open reading frames (ORFs), requiring detailed analysis. |
| Somatic Variant Identification | <ul style="list-style-type: none"> - Critical for neoantigen discovery by providing a list of targetable neoantigens. - Running multiple somatic variant callers improves specificity. | <ul style="list-style-type: none"> - Balancing between accurate identification and reducing false positives is challenging. - Consensus approaches trade off sensitivity for specificity, impacting the comprehensiveness of variant calling. |
| Variant Annotation | - Critical for categorizing variants for further investigation, aiding in prioritization. | - Difficult to benchmark the success of different programs, making it challenging to determine the optimal annotation strategy for specific analyses. |
| Antigen Processing Algorithms | <ul style="list-style-type: none"> - Incorporates upstream events in antigen presentation for greater accuracy. - Predicts proteasomal cleavage and TAP's peptide transportation efficiency. | - Limitations in predicting mutations in genes involved in the antigen presentation machinery, potentially leading to incomplete assessments. |
| Peptide : MHC Binding Affinity | <ul style="list-style-type: none"> - Affinity to MHC molecules is crucial for neoantigen immunogenicity. - Artificial neural networks offer accurate predictions for frequent alleles. | - Accuracy decreased for rare or not well-studied HLA alleles, emphasizing the need for further advancements in prediction models to encompass a broader range of human leukocyte antigen (HLA) diversity. |

antigen-specific T cells across several studies (104, 105). For cases where greater avidity between the pMHC and TCR are necessary, pentamers (106, 107) or dodecamers (108) can be used in common immunological workflows such as flow cytometry. To improve on the throughput of tetramer-binding assays, T cell responses against large peptide libraries can be screened using UV-sensitive peptides

TABLE 3 List of commonly used bioinformatics tools for neoepitope prediction.

| Tool | Key features | References |
|------------------|---|------------|
| NetMHC/NetMHCpan | <ul style="list-style-type: none"> -Predicts peptide-MHC binding affinity - Widely used tool for MHC binding prediction -NetMHCpan covers a broad range of MHC alleles | (88–90) |
| MuPeXi | <ul style="list-style-type: none"> -Extracts mutated peptides and predicts MHC binding affinity -Integrates multiple tools for a comprehensive analysis | (91) |
| pVAC-Seq | <ul style="list-style-type: none"> -Identifies and prioritizes neoantigens based on mutation data - Considers MHC binding, transcript expression and clonality | (92) |
| INTEGRATE-Neo | -Integrates RNA-seq and DNA sequencing data for immunogenic neoantigen prediction | (93) |
| NeoPredPipe | <ul style="list-style-type: none"> -Predicts neoantigens considering MHC binding, T-cell receptor recognition and gene expression -Multiparameter prediction approach | (94) |
| NetCTLpan | <ul style="list-style-type: none"> Predicts CTL epitopes based on MHC-I binding, proteasomal cleavage and Tap transport efficiency -comprehensive approach for CTL epitope prediction | (95) |
| MHCflurry | Deep learning-based tool for predicting peptide-MHC binding affinity | (96) |

(109) or thrombin cleavage (110) of peptide linkers which allows for the production of thousands of different monomeric molecules from the same pMHC molecule. Nevertheless, to overcome the limitations imposed by cellular input from clinical biopsies, additional multiplexed and combinatorial tetramer staining approaches are needed. With combinatorial staining, a specific T cell population is identified using tetramers containing the same peptide but conjugated to multiple fluorochromes rather than just a single fluorochrome. Using this approach with just two different fluorochromes per peptide drastically increases the number of specificities that can be examined simultaneously from a single sample using flow cytometry (e.g., if each peptide:MHC specificity is identified using a dual fluorochrome approach and a total of 8 fluorochromes are used across antigen specificities, this will result in a two-dimensional matrix capable of identifying up to 28 unique specificities) (111). While this approach significantly increases the number of antigen specificities that can be screened from a single sample, other conditions may be limiting, such as the configuration of flow analyzers and the number of fluorochromes incorporated in phenotypic panels to robustly characterize antigen-specific T cells with sufficient resolution (i.e., with low enough spillover). To overcome this limitation, mass spectrometry can be used as an alternative, in which tetramers are labelled with isotopically purified metal conjugates providing little spillover between labels, less variation in signal intensity between parameters, and many more surface markers to include phenotypic characterization of

neoantigen-specific T cells (112, 113). Although mass-spec based approaches such as CyTOF address the issue of multiplexing, cellular material cannot be recovered from these workflows as cells are incinerated and impossible to retrieve for downstream applications. More recently, the use of DNA barcoded pMHC multimers has extended this toolkit (114). While the increased capability of labelling multimers with DNA barcodes offers flexibility to screen entire cancer mutanomes in one single reaction, it only provides information about the frequency of corresponding T cells but lacks the inclusion of functional readouts as the cells are lysed for sequencing. An upgraded version of DNA barcoded multimers called tetramer-associated TCR sequencing (tetTCRSeq) (115) can further link the antigen specificity to the TCR sequences in single cells at high throughput. This technology allows for the simultaneous generation of peptide libraries and DNA barcodes using *in vitro* transcription and translation (IVTT), thereby reducing labor and cost but still allowing the recall of antigen specificity and TCR sequences.

Although pMHC multimers have revolutionized the field of neoantigen discovery, screening approaches using multimers require prior knowledge of the antigens being targeted (i.e., these approaches aren't completely unbiased and not truly *de novo* screening approaches). Additionally, pMHC multimers aren't commercially available for every HLA molecule, which has been a major impedance in the field. Moreover, the low accuracy of *in silico* prediction of HLA class II-restricted epitopes and technical issues with the production of pMHCI multimers possess challenges for screening and identifying neoantigens presented by MHCI molecules. Furthermore, TCRs that bind pMHC tetramers with low affinity can be difficult to detect by flow cytometry and may not accurately reflect the affinity needed for T cell activation, which can underestimate the frequency of neoantigen specific T cells. Other potential bottlenecks attributed to the use of pMHC multimers include limited throughput of peptide synthesis, maximum library size, instability of pMHC multimers, and PCR amplification bias.

With recent advances in microfluidics, additional tools to study TCR:pMHC interactions with higher throughput have started to emerge. Microfluidic approaches can be divided broadly into three categories (1): trap-based devices (2), valve-based devices, and (3) droplet microfluidics. In this review we will mainly focus on droplet microfluidics for single cell analysis, but we would highly encourage readers to read a recent review that provides additional coverage of additional approaches (116). Droplet microfluidics separates individual cells into low volume oil droplets, making precise characterization of complex immune responses possible at single cell resolution. Since this technique has minimal sample loss and requires very low cell input, having a limited number of cells from tumor biopsies or clinical specimens doesn't exclude this approach from being used for screening applications. Building on droplet microfluidics, the nuclear factor of T cells (NFAT)-eGFP reporter system was used for functional screening and real-time monitoring of T cell activation kinetics upon recognition of tumor cells, enabling quick identification of the responding T cells and verification of corresponding TCR sequences that could then be used as therapies for cancer (117). In a similar study, nanoparticle barcoded nucleic acid cell sorting (NACS) (118) was used to count

and isolate neoantigen specific T cells. MATE-Seq (119), another technique that builds on the approaches used in NACS (118), allowed for high-throughput isolation and single cell TCR sequencing of neoantigen specific T cells using magnetic nanoparticle-barcoded pMHC tetramers linked to photocleavable TCR-specific primers. These antigen-directed approaches can simultaneously increase the number of putative neoantigens that can be screened and identify TCRs capable of binding, but they often underestimate immunogenicity due to the lack of a functional readout.

3.2 TCR-directed approaches

To overcome challenges associated with peptide-pulsing using individual peptides per condition, which requires high cellular input, and the limitations of peptide-MHC binding predictions for specific HLAs, a new screening assay was designed which allowed T cell responses to any non-synonymous mutation across HLAs to be evaluated (120). For this approach, each mutation is expressed by a single minigene designed to encode the mutated amino acid flanked by 12 amino acids of the wildtype sequence on both sides. Tandem minigenes (TMG) were generated by stringing together 6 to 24 minigenes in a single open reading frame, *in-vitro* transcribing to generate mRNA, and then transfecting the mRNA into autologous APCs. As an alternative to the TMG screening approach, peptides (25 amino acid residues in length) can be synthesized and pooled together to generate peptide pools (PP). Pools for all putative neoantigens can be screened using either the TMG approach, which utilizes the natural antigen processing and presentation machinery (APPM) of a cell, or by peptide pulsing, which bypasses the APPM and instead directly binds surface MHC molecules. Once T cell reactivity is detected, either through cytokine secretion or upregulation of a specific activation marker (e.g., 4-1BB), against a specific pool of TMGs or PP, further deconvolution of each immunogenic neoantigen found within the pool of reactive peptides can be screened individually. This unbiased approach has been used to identify multiple mutated antigens and neoantigen specific TILs that have been shown to induce antitumor responses in patients (85, 120–122). The biggest advantage of using TMGs and peptide pools is that it mimics processing and presentation of neopeptides on both class I and II HLA molecules without bias, overcoming the need for *in silico* prediction and enabling the identification of neoantigen-reactive CD4 T cells since the prediction algorithms for HLA class II molecules aren't fully optimized. Although in theory one could screen every mutation identified by NGS using TMGs and/or using peptide pools, in practice, this isn't feasible due to the cost associated with peptide synthesis and the time required to carry out *in-vitro* transcription, especially in cases for tumors with high mutational burdens. As with the tetramer-binding assays, the deconvolution steps needed to identify the specific immunogenic neoantigens requires large numbers of target and effector cells which can be limiting for clinical specimens. To compensate for the lack of autologous APCs, HLA matched cells or monoallelic antigen presenting cell lines can be generated. Alternatively, cell-free antigen presenting

beads, generated by coupling the peptide-HLA complex of interest with costimulatory antibodies on the surface of microbeads can be used to determine T cell reactivity (123, 124). Even with the limitations surrounding the number of cells needed to adequately perform these screens, these approaches allow for unbiased identification of candidate neoantigens without prior knowledge of whether the corresponding peptides are capable of binding patient specific HLAs.

3.3 Reporter systems to identify neoantigen-TCR pairs

Novel reporter systems have recently been described using mammalian cell surface display, which results in a detectable signal as soon as a cognate TCR binds to a cancer neoepitope presented by an APC. Signaling and antigen-presenting bifunctional receptor (SABR) (125) is a cell-based platform for T-cell antigen discovery that relies on screening of large numbers of antigens through expression of chimeric receptors consisting of an extracellular pMHC complex fused to an intracellular CD28 costimulatory and CD3 ζ signaling domain in NFAT-GFP Jurkat cells. Upon recognition by a TCR, the CD28-CD3 ζ signaling triggers the expression of GFP in APCs for the identification of the presented peptide by downstream sequencing. This technology has been demonstrated to successfully identify the cognate antigen for TCRs from a large library of epitopes and for the discovery of personalized neoantigens.

Another system, T-Scan (126), is a high-throughput platform for systematic identification of antigens recognized by T cells that incorporates an engineered reporter for granzyme B activity by tagging APCs expressing epitope coding minigenes. Upon recognition of antigen by cognate TCR, granzyme B secreted from the T cells leads to the reconstitution of a fluorescent reporter in the APCs. Utility of T-scan was validated by showing that it can identify known peptide epitopes of both viral and human-genome libraries (126). Another similar system leveraging the specificity of the granzyme-perforin pathway (127), used a reporter-fusion protein consisting of cyan fluorescent protein (CFP) and Yellow Fluorescent protein (YFP) separated by a peptide linker harboring a granzyme B recognition site. Upon recognition of target cells by T cells, cleavage of the fusion protein by granzyme B causes a loss in the fluorescence resonance energy transfer (FRET) signal generated from YFP and results in a concomitant gain of CFP signal that is easily identifiable and allows for the isolation of recognized target cells by fluorescence-activated cell sorting (FACS).

Another similar system based on the NFAT reporter was developed to identify tumor specific CD4 T cells by fusing murine MHCII with the signaling domains of TCR resulting in generation of pMHC-TCR (MCR) hybrid molecules (128). MCR libraries were generated by cloning tumor cDNA into MCR sequences and transducing reporter cells which can then be used as APCs in multiple rounds of coculture with T cells. Upon interaction of specific TCR with the MCR, NFAT activation results in the expression of a reporter gene, which ultimately

allows the activated reporter cells to be identified, single cell sorted by FACS, and the recognized peptides can be identified by sequencing the corresponding DNA.

Lastly, trogocytosis, which is the transfer of membrane protein from one cell to another was used to identify the sequence of peptides (129). Upon successful interaction of TCR expressing Jurkat cells and pMHC presenting K562 cells, transfer of TCR from the Jurkat cells to K562 cells can be detected by identification of K562 cells with TCR using FACS. Following cell sorting of K562 cells containing transferred TCRs, PCR amplification and deep sequencing is used to identify the corresponding epitope of interest.

Table 4 succinctly summarizes some of the most common methods, assays, and techniques that can be used to screen and validate neoantigen specific T cells.

4 Single-cell platforms

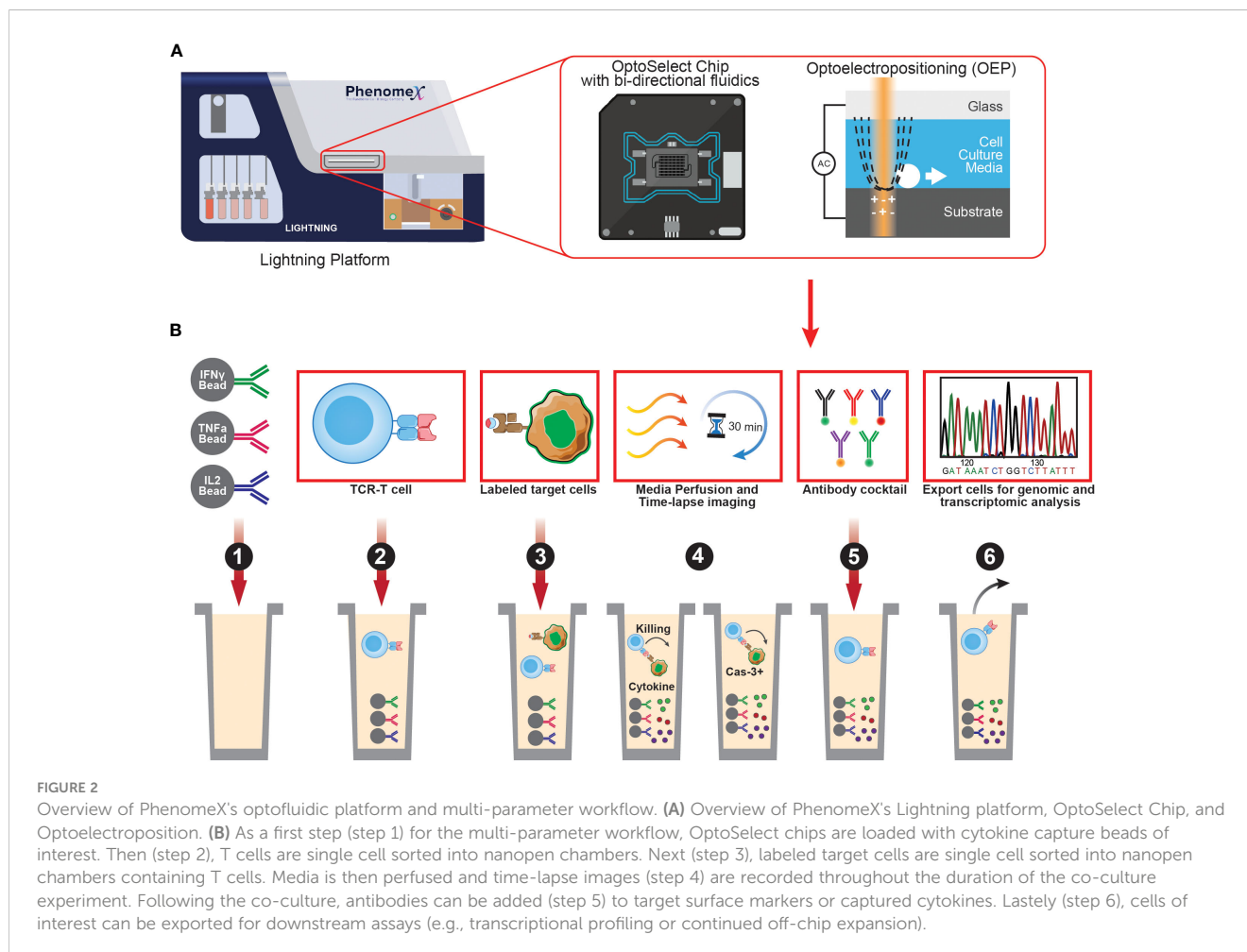
Single-cell sequencing is a powerful tool that enables high-throughput analysis of genetic or transcript material for single cells. One of the most widely used platforms for this purpose is the Chromium system by 10X Genomics, which can sequence both the transcriptome and TCRs of single cells by incorporating cell barcodes. While scDNA-seq can reveal mutations and structural changes in cell genomes, it is not as frequently utilized due to limited DNA copies available in single cells, making scRNA-seq a more popular option for measuring gene expression across multiple

TABLE 4 Common methods, assays, and techniques that can be used to screen and validate neoantigen specific T cells.

| Approach Type | Methods/Techniques |
|------------------------------------|---|
| Antigen-Directed Approaches | <ul style="list-style-type: none"> - Utilization of pMHC tetramers or multimers for precise detection of T cells with high-affinity TCR-pMHC interactions. - Screening large peptide libraries using UV-sensitive peptides and thrombin cleavage. - Combinatorial staining employing multiple fluorochromes to enhance specificity. - Adoption of mass spectrometry for high-throughput analysis with minimal spillover. - Integration of DNA barcoded pMHC multimers for flexibility and high throughput. - Leveraging droplet microfluidics for single-cell analysis with minimal sample loss |
| TCR-Directed Approaches | <ul style="list-style-type: none"> - Implementation of Tandem minigenes (TMG) and peptide pools for unbiased screening of T cell reactivity. - Functional assays measuring T cell reactivity against TMGs or peptide pools and application of deconvolution techniques to identify specific immunogenic neoantigens. - Use of autologous APCs, HLA-matched cells, or cell-free antigen-presenting beads for enhanced flexibility in screening. |
| Reporter Systems | <ul style="list-style-type: none"> - Utilization of SABR and T-Scan systems, incorporating chimeric receptors and measuring granzyme B activity for high-throughput T-cell antigen discovery. - Integration of NFAT reporter systems for the identification of CD4 T cells. - Application of Trogocytosis for TCR transfer, facilitating the identification of specific peptides involved in immune responses. |

transcripts. It is important to note that many single-cell sequencing technologies currently only allow assessment of up to 10,000 cells, which is several orders of magnitude lower than conventional bulk sequencing methods. Nevertheless, scRNA-seq can be leveraged to identify multiple layers of information such as cell phenotype using oligo-tagged antibodies in combination with transcriptome. In comparison to bulk RNA-Seq approaches, scRNA-seq provides more detailed and comprehensive characterization of individual cells. Recent advances in single-cell platforms have enabled additional multiplexing capabilities to interrogate neoantigen-specific T cells. The PhenomeX Lightning and Beacon platforms combine optics and fluidics to link phenotypic, functional, and transcriptomic profiles to single neoantigen-reactive T cells (Figure 2) (130). These technologies can characterize and isolate single cells from a larger population of cells that display desired phenotypes and/or functional characteristics. There are three main features that make the PhenomeX's optofluidic platforms uniquely suited for antigen discovery. First, at the core of the technology is the OptoSelect Chip which enables isolation of single cells using Opto-electropositioning (OEP) and nanopen chambers. OEP functions to guide cells in and out of selected nanopens using a non-destructive localized light-induced electric field that can move cells using a weak repulsive force against charged membranes or particles. This feature can be used to sort or "pen" single antigen-

presenting cells encoding a library of neoepitopes and a single T cell in each nanopen, avoiding the need to set up bulk coculture experiments as required for earlier assays. Second the platform contains fluidics capable of bi-directional flow through the chip which can function to help transfer single cells into nanopens, diffuse media throughout the chip during culturing assays, and also recover specific cells of interest from the chip. Third is the optics system which acts as a microscope and fluorescent analyzer and allows for brightfield and fluorescent images to be acquired, cell counts to be made and phenotypic and functional assays to be performed using fluorescent readouts. By modifying these assays, multiple parameters can be assessed in single nanopens to assess T cell activation upon recognition of cognate pMHC complexes (e.g., similar to T-scan, capture beads for Granzyme B can be added in each nanopen). Upon successful interaction of T cell and pMHC expressing APCs, Granzyme B secreted by neoantigen-specific T cells can be captured by capture beads, and subsequently stained by a fluorescently labeled anti-granzyme B antibody binding a different epitope. Alternatively, additional functional readouts such as IFN-gamma or IL-2 secretion can also be assessed. Future studies using PhenomeX's platforms could also provide screening workflows to test libraries of T-cell clones and pMHC simultaneously as the user has greater control over the loading of specific cells into distinct nanopens on the chip, expanding the number of effectors and target



cell pools that can be screened at once. Depending on the required throughput, either the Lightning (for more validation-based workflows requiring 1,500 nanopens or less) or the Beacon (for either validation- or discovery-based workflows requiring greater than 1,500 nanopens) platform could be used. However, further advancements in the development of fluidics-based methods would be required to dissect the T-cell specificity at larger scales and to scale up to those routinely performed in bulk coculture assays. Finally, desired cells recovered from these platforms can be sequenced using single-cell sequencing, providing a comprehensive understanding of the genetic and transcriptomic features of individual cells of interest and their roles in complex biological processes.

In the field of cancer immunotherapy, single-cell technologies have begun to take center stage and have helped lead to a greater appreciation of the dynamic interactions between cancer and immune cells. A notable advancement involves the concurrent evaluation of T cell receptor (TCR) sequences and gene expression profiles from individual T cells, highlighting the intricate connections between T cell functionality and gene expression within the tumor microenvironment. Recent studies have shed light on how the TCR repertoire significantly influences the functional attributes of individual T cells (131–135), which can help guide newer approaches for developing cell-based therapies for cancer. Previous challenges in properly identifying rare and elusive cell types within the TME, such as cancer stem cells or immunosuppressive cell subsets, have also been addressed with single-cell technologies by empowering researchers to isolate and analyze these rare cell types with unprecedented precision and resolution. Techniques like single-cell genomics and proteomics provide key insights into how these cells drive tumor progression, metastasis, and therapy resistance. By performing scRNA-seq on tumor biopsies, the corresponding transcriptional profiles of individual cells can be used to identify global and local signatures (136–140), and determine whether specific subsets of cells, such as effector T cells, regulatory T cells, dendritic cells, and macrophages, have infiltrated specific regions of a tumor. This high-resolution analysis provides crucial insights into the functional states, activation statuses, and potential immunosuppressive features of distinct immune cell populations, which if used on longitudinally collected samples, can also help determine whether a patient is responding to a specific therapeutic intervention. The seamless integration of single-cell technologies propels cancer immunotherapy into a realm of precision and depth previously unattainable, promising a future where therapeutic strategies are finely tuned to each patient's immune landscape.

5 Advances in computational tools for prediction of TCR antigen specificity

As the number of high throughput workflows for sequencing and TCR functional characterization have increased, several databases such as McPAS-TCR (141), VDJdb (142), and the

TBAdB subset of PIRD (143) have emerged as repositories for TCR and epitope information. Building upon the unprecedented amount of experimental data generated by single-cell platforms, recent advances in digital biology and machine learning have expanded our ability to predict T-cell receptor (TCR) antigen specificity and analyzing TCR repertoires. While traditional approaches like X-ray crystallography, Nuclear Magnetic Resonance (NMR), Surface Plasmon Resonance (SPR) and Mass Spectrometry (MS) are still used for confirming conformational and structural interactions (144–148), they are limited by the time-consuming process of characterizing pMHC-TCR interactions one by one. However, cutting-edge approaches that leverage computational tools and machine learning algorithms are bridging the gap between experimental data and predictive modeling, offering new possibilities for understanding immune responses and developing personalized immunotherapies.

These innovative tools analyze the sequence and structural features of TCRs and their corresponding antigen epitopes, establishing patterns and predicting potential interactions. They can be broadly classified into two categories: supervised predictive models (SPMs) and unsupervised clustering models (UCMs) (149) based on their use of supervised and unsupervised learning, respectively. Arsenal of tools have emerged recently that uses the training data set generated by the experimental approaches to general a model. Some of the widely used tools have been summarized in Table 5.

One widely used tool is GLIPH (150), which applies clustering techniques to identify functionally related TCRs by identifying shared binding motifs in their CDR3 sequences. By uncovering these motifs, GLIPH and its successor GLIPH2 (151) enable the inference of potential epitopes and TCR-pMHC interactions. Another commonly employed tool, TCRdist (152, 153), calculates distances between TCR CDR3 sequences, providing a measure of similarity and aiding in the inference of antigen specificity. IGoR (154) is an additional tool frequently utilized in the field, estimating the probability of generating a specific CDR3 sequence through V(D)J recombination. By accounting for the recombination process, IgoR offers a more accurate estimation of TCR repertoire diversity and antigen specificity, enhancing the understanding of the T cell repertoire and the likelihood of eliciting a specific TCR response. TCRex (155), another prominent tool, is a computational framework trained on large-scale TCR-pMHC binding datasets that employs deep learning algorithms to predict TCR-pMHC interaction specificity, facilitating the identification of potential target antigens. Furthermore, DeepTCR (156), a deep learning-based tool, analyzes TCR repertoires to predict antigen specificity of TCRs and identify potential epitope targets. By integrating experimental data, computational techniques, and machine learning algorithms, these tools have revolutionized TCR sequencing, enabling researchers to rapidly predict potential targets.

These innovative tools and approaches have not only provided a solid foundation for understanding basic biological principles surrounding TCR repertoire diversity and the influence of pMHC on this diversity but also hold great potential for transforming

TABLE 5 List of bioinformatics tools for TCR repertoire analysis and TCR-pMHC specificity prediction.

| Tool | Description | Key Features | Availability | URL | References |
|---------|---|---|---|--|------------|
| GLIPH | Cluster TCR sequences based on shared motifs | Motif based clustering, identification of hotspots | Open-source software, Web based service | https://github.com/immunoengineer/gliph | (150, 151) |
| TCRdist | Measures similarity between TCR sequences based on CDR motifs | Repertoire wide TCR distance calculation | Python Package | https://github.com/phbradley/tcr-dist https://tcrdist3.readthedocs.io/en/latest/ | (152, 153) |
| IGoR | Estimation of TCR repertoire diversity | Probability of CDR3 generation, V(D)J recombination | C++ package | https://github.com/qmarcou/IGoR | (154) |
| TCRex | Deep learning based framework for TCR specificity prediction | Integration of diverse TCR features, deep learning | Web based tool | https://tcrex.biodatamining.be/ | (155) |
| DeepTCR | Deep learning based tool for antigen-specific TCR prediction | Integration of TCR sequence and binding data | Python Package | https://github.com/sidhomj/DeepTCR | (156) |

personalized therapies and clinical care. Their utilities extend to fields such as transplantation (GVHD), autoimmunity (cross-reactivity), and immunotherapy (cell-based therapies and cancer vaccines), offering valuable insights and paving the way for personalized immunotherapies.

6 Concluding remarks

Neoantigens have emerged as promising targets in cancer immunotherapy. In silico identification of candidate neoantigens is a critical step in personalized immunotherapy that relies on the development and application of bioinformatics and computational approaches. Neoantigen prediction algorithms have continued to evolve and improve in recent years, which has been aided by the ever-growing training data sets. However, more robust measures are needed to improve the identification of immunogenic neoantigens, as not all predicted neoantigens elicit an immune response. Screening and validation of neoantigen specific T cells using antigen- or TCR-directed approaches provides greater efficiency and accuracy in identifying targetable neoantigens, but they are expensive and time consuming. The recent advances in single cell technologies have offered new ways to integrate multiple information layers at the level of individual cell. While these technologies offer exciting prospects, they also present challenges in data analysis and interpretation. Additionally, the prediction of TCR-pMHC interactions plays a vital role in translating neoantigen into actionable targets for cell-based therapies. Cutting edge approaches, that leverage the computational tools and machine learning algorithms, have been developed recently offering new possibilities for understanding immune responses and developing personalized immunotherapies. Overall, these innovative tools, in combination with experimental data, pave the way for more effective and tailored immunotherapeutic approaches, offering valuable insights for personalized immunotherapies.

Author contributions

RS: Conceptualization, Writing – original draft, Writing – review & editing. EC: Writing – review & editing. TK: Writing – review & editing. AC: Writing – review & editing. AZ: Writing – review & editing, Conceptualization, Funding acquisition, Visualization, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Paula and Rodger Riney Foundation. AC is supported in part by the Medical College of Wisconsin Cancer Center's Graduate Fellowship.

Conflict of interest

AE Zamora reports research support from PhenomeX.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J Clin* (2021) 71:209–49. doi: 10.3322/caac.21660
- Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability — an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* (2010) 11:220–8. doi: 10.1038/nrm2858
- Wang G, Kang X, Chen KS, Jehng T, Jones L, Chen J, et al. An engineered oncolytic virus expressing PD-L1 inhibitors activates tumor neoantigen-specific T cell responses. *Nat Commun* (2020) 11:1395. doi: 10.1038/s41467-020-15229-5
- Zhang Y, Qian J, Gu C, Yang Y. Alternative splicing and cancer: a systematic review. *Sig Transduct Target Ther* (2021) 6:78. doi: 10.1038/s41392-021-00486-7
- Rathe SK, Popescu FE, Johnson BE, Watson AL, Marko TA, Moriarity BS, et al. Identification of candidate neoantigens produced by fusion transcripts in human osteosarcomas. *Sci Rep* (2019) 9:358. doi: 10.1038/s41598-018-36840-z
- Jhunjhunwala S, Hammer C, Delamarre L. Antigen presentation in cancer: insights into tumour immunogenicity and immune evasion. *Nat Rev Cancer* (2021) 21:298–312. doi: 10.1038/s41568-021-00339-z
- Zamora AE, Crawford JC, Thomas PG. Hitting the target: how T cells detect and eliminate tumors. *Jl* (2018) 200:392–9. doi: 10.4049/jimmunol.1701413
- Istrail S, Florea L, Halldórsson B, Kohlbacher O, Schwartz RS, Yap VB, et al. Comparative immunopeptidomics of humans and their pathogens. *Proc Natl Acad Sci USA* (2004) 101:13268–72. doi: 10.1073/pnas.0404740101
- Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med* (2014) 371:2189–99. doi: 10.1056/NEJMoa1406498
- Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* (2015) 350:207–11. doi: 10.1126/science.aad0095
- Lauss M, Donia M, Harbst K, Andersen R, Mitra S, Rosengren F, et al. Mutational and putative neoantigen load predict clinical benefit of adoptive T cell therapy in melanoma. *Nat Commun* (2017) 8:1738. doi: 10.1038/s41467-017-01460-0
- Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* (2015) 348:124–8. doi: 10.1126/science.aaa1348
- Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang T-H, et al. The immune landscape of cancer. *Immunity* (2018) 48:812–830.e14. doi: 10.1016/j.immuni.2018.03.023
- Yang W, Lee K-W, Srivastava RM, Kuo F, Krishna C, Chowell D, et al. Immunogenic neoantigens derived from gene fusions stimulate T cell responses. *Nat Med* (2019) 25:767–75. doi: 10.1038/s41591-019-0434-2
- Parkhurst MR, Robbins PF, Tran E, Prickett TD, Gartner JJ, Jia L, et al. Unique neoantigens arise from somatic mutations in patients with gastrointestinal cancers. *Cancer Discovery* (2019) 9:1022–35. doi: 10.1158/2159-8290.CD-18-1494
- Maleki Vareki S. High and low mutational burden tumors versus immunologically hot and cold tumors and response to immune checkpoint inhibitors. *J Immunotherapy Cancer* (2018) 6:157. doi: 10.1186/s40425-018-0479-7
- Carreno BM, Magrini V, Becker-Hapak M, Kaabinejadian S, Hundal J, Petti AA, et al. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* (2015) 348:803–8. doi: 10.1126/science.aaa3828
- Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* (2017) 547:217–21. doi: 10.1038/nature22991
- Sahin U, Derhovanessian E, Miller M, Kloke B-P, Simon P, Löwer M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* (2017) 547:222–6. doi: 10.1038/nature23003
- Yarchoan M, Hopkins A, Jaffee EM. Tumor mutational burden and response rate to PD-1 inhibition. *N Engl J Med* (2017) 377:2500–1. doi: 10.1056/NEJMc1713444
- Motzer RJ, Tannir NM, McDermott DF, Arén Frontera O, Melichar B, Choueiri TK, et al. Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma. *N Engl J Med* (2018) 378:1277–90. doi: 10.1056/NEJMoa1712126
- Almogly G, Pratt M, Oberstrass F, Lee L, Mazur D, Beckett N, et al. Cost-efficient whole genome-sequencing using novel mostly natural sequencing-by-synthesis chemistry and open fluidics platform. [preprint]. *Genomics* (2022). doi: 10.1101/2022.05.29.493900
- Illing PT, Ramarathinam SH, Purcell AW. New insights and approaches for analyses of immunopeptidomes. *Curr Opin Immunol* (2022) 77:102216. doi: 10.1016/j.coi.2022.102216
- Purcell AW, Ramarathinam SH, Ternette N. Mass spectrometry-based identification of MHC-bound peptides for immunopeptidomics. *Nat Protoc* (2019) 14:1687–707. doi: 10.1038/s41596-019-0133-y
- Bassani-Sternberg M, Coukos G. Mass spectrometry-based antigen discovery for cancer immunotherapy. *Curr Opin Immunol* (2016) 41:9–17. doi: 10.1016/j.coi.2016.04.005
- Kote S, Pirog A, Bedran G, Alfaro J, Dapic I. Mass spectrometry-based identification of MHC-associated peptides. *Cancers (Basel)* (2020) 12:E535. doi: 10.3390/cancers12030535
- Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet* (2012) 20:490–7. doi: 10.1038/ejhg.2011.258
- Zhang J, White NM, Schmidt HK, Fulton RS, Tomlinson C, Warren WC, et al. INTEGRATE: gene fusion discovery using whole genome and transcriptome data. *Genome Res* (2016) 26:108–18. doi: 10.1101/gr.186114.114
- Park J, Chung Y-J. Identification of neoantigens derived from alternative splicing and RNA modification. *Genomics Inform* (2019) 17:e23. doi: 10.5808/GI.2019.17.3.e23
- Orenbuch R, Filip I, Comito D, Shaman J, Pe'er I, Rabadan R. arcasHLA: high-resolution HLA typing from RNAseq. *Bioinformatics* (2020) 36:33–40. doi: 10.1093/bioinformatics/btz474
- Buchkovich ML, Brown CC, Robasky K, Chai S, Westfall S, Vincent BG, et al. HLAProfiler utilizes k-mer profiles to improve HLA calling accuracy for rare and common alleles in RNA-seq data. *Genome Med* (2017) 9:86. doi: 10.1186/s13073-017-0473-6
- Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* (2009) 324:218–23. doi: 10.1126/science.1168978
- Ruiz Cuevas MV, Hardy M-P, Holly J, Bonneil É, Durette C, Courcelles M, et al. Most non-canonical proteins uniquely populate the proteome or immunopeptidome. *Cell Rep* (2021) 34:108815. doi: 10.1016/j.celrep.2021.108815
- Ouspenskaia T, Law T, Clauser KR, Klaeger S, Sarkizova S, Aguet F, et al. Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nat Biotechnol* (2022) 40:209–17. doi: 10.1038/s41587-021-01021-3
- Mudge JM, Ruiz-Orera J, Prensner JR, Brunet MA, Calvet F, Jungreis I, et al. Standardized annotation of translated open reading frames. *Nat Biotechnol* (2022) 40:994–9. doi: 10.1038/s41587-022-01369-0
- Wood MA, Nguyen A, Struck AJ, Ellrott K, Nellore A, Thompson RF. neopepisco improves neopeptide prediction with multivariant phasing. *Bioinformatics* (2020) 36:713–20. doi: 10.1093/bioinformatics/btz653
- Shiraishi Y, Sato Y, Chiba K, Okuno Y, Nagata Y, Yoshida K, et al. An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data. *Nucleic Acids Res* (2013) 41:e89–9. doi: 10.1093/nar/gkt126
- Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* (2013) 31:213–9. doi: 10.1038/nbt.2514
- Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods* (2018) 15:591–4. doi: 10.1038/s41592-018-0051-x
- Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* (2012) 22:568–76. doi: 10.1101/gr.129684.111
- Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics* (2012) 28:311–7. doi: 10.1093/bioinformatics/btr665
- Kim S, Jeong K, Bhutani K, Lee J, Patel A, Scott E, et al. Virmid: accurate detection of somatic mutations with sample impurity inference. *Genome Biol* (2013) 14:R90. doi: 10.1186/gb-2013-14-8-r90
- Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res* (2016) 44:e108–8. doi: 10.1093/nar/gkw227
- Spinella J-F, Mehanna P, Vidal R, Saillour V, Cassart P, Richer C, et al. SNooPer: a machine learning-based method for somatic variant identification from low-pass next-generation sequencing. *BMC Genomics* (2016) 17:912. doi: 10.1186/s12864-016-3281-2
- Fang LT, Afshar PT, Chhibber A, Mohiyuddin M, Fan Y, Mu JC, et al. An ensemble approach to accurately detect somatic mutations using SomaticSeq. *Genome Biol* (2015) 16:197. doi: 10.1186/s13059-015-0758-2
- Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* (2012) 28:i333–9. doi: 10.1093/bioinformatics/bts378
- Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* (2009) 25:2865–71. doi: 10.1093/bioinformatics/btp394
- Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. Accuracy assessment of fusion transcript detection via read-mapping and *de novo* fusion transcript assembly-based methods. *Genome Biol* (2019) 20:213. doi: 10.1186/s13059-019-1842-9
- Uhrig S, Ellermann J, Walther T, Burkhardt P, Fröhlich M, Hutter B, et al. Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res* (2021) 31:448–60. doi: 10.1101/gr.257246.119

50. McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MGF, et al. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput Biol* (2011) 7:e1001138. doi: 10.1371/journal.pcbi.1001138
51. Bian X, Zhu B, Wang M, Hu Y, Chen Q, Nguyen C, et al. Comparing the performance of selected variant callers using synthetic data and genome segmentation. *BMC Bioinf* (2018) 19:429. doi: 10.1186/s12859-018-2440-7
52. 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. doi: 10.1371/journal.pone.0151664
53. Chen Z, Yuan Y, Chen X, Chen J, Lin S, Li X, et al. Systematic comparison of somatic variant calling performance among different sequencing depth and mutation frequency. *Sci Rep* (2020) 10:3501. doi: 10.1038/s41598-020-60559-5
54. Callari M, Sammut S-J, De Mattos-Arruda L, Bruna A, Rueda OM, Chin S-F, et al. Intersect-then-combine approach: improving the performance of somatic variant calling in whole exome sequencing data using multiple aligners and callers. *Genome Med* (2017) 9:35. doi: 10.1186/s13073-017-0425-1
55. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant review with the integrative genomics viewer. *Cancer Res* (2017) 77:e31–4. doi: 10.1158/0008-5472.CAN-17-0337
56. Cameron DL, Di Stefano L, Papenfuss AT. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nat Commun* (2019) 10:3240. doi: 10.1038/s41467-019-11146-4
57. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The ensemble variant effect predictor. *Genome Biol* (2016) 17:122. doi: 10.1186/s13059-016-0974-4
58. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* (2010) 38:e164–4. doi: 10.1093/nar/gkq603
59. McCarthy DJ, Humburg P, Kanapin A, Rivas MA, Gaulton K, Zier J-B, et al. Choice of transcripts and software has a large effect on variant annotation. *Genome Med* (2014) 6:26. doi: 10.1186/gm543
60. Rathinakannan VS, Schukov H-P, Heron S, Schleutker J, Sipeky C, ShAn: An easy-to-use tool for interactive and integrated variant annotation. *PLoS One* (2020) 15:e0235669. doi: 10.1371/journal.pone.0235669
61. Sijts EJM, Kloetzel P-M. The role of the proteasome in the generation of MHC class I ligands and immune responses. *Cell Mol Life Sci* (2011) 68:1491–502. doi: 10.1007/s00018-011-0657-y
62. Friedman J, Gunti S, Lee M, Bai K, Hinrichs C, Allen CT. Determining if T cell antigens are naturally processed and presented on HLA class I molecules. *BMC Immunol* (2022) 23:5. doi: 10.1186/s12865-022-00478-4
63. Rock KL, Reits E, Neefjes J. Present yourself! By MHC class I and MHC class II molecules. *Trends Immunol* (2016) 37:724–37. doi: 10.1016/j.it.2016.08.010
64. Nielsen M, Lundegaard C, Lund O, Keşmir 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. doi: 10.1007/s00251-005-0781-7
65. Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM, et al. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *CMLS Cell Mol Life Sci* (2005) 62:1025–37. doi: 10.1007/s00018-005-4528-2
66. Calis JJA, Reinink P, Keller C, Kloetzel PM, Keşmir C. Role of peptide processing predictions in T cell epitope identification: contribution of different prediction programs. *Immunogenetics* (2015) 67:85–93. doi: 10.1007/s00251-014-0815-0
67. Hoze E, Tsaban L, Maman Y, Louzoun Y. Predictor for the effect of amino acid composition on CD4+ T cell epitopes preprocessing. *J Immunol Methods* (2013) 391:163–73. doi: 10.1016/j.jim.2013.02.006
68. Paul S, Karosiene E, Dhanda SK, Jurtz V, Edwards L, Nielsen M, et al. Determination of a predictive cleavage motif for eluted major histocompatibility complex class II ligands. *Front Immunol* (2018) 9:1795. doi: 10.3389/fimmu.2018.01795
69. Paulson KG, Tegeder A, Willmes C, Iyer JG, Afanasiev OK, Schrama D, et al. Downregulation of MHC-I expression is prevalent but reversible in Merkel cell carcinoma. *Cancer Immunol Res* (2014) 2:1071–9. doi: 10.1158/2326-6066.CIR-14-0005
70. Paulson KG, Voillet V, McAfee MS, Hunter DS, Wagener FD, Perdicchio M, et al. Acquired cancer resistance to combination immunotherapy from transcriptional loss of class I HLA. *Nat Commun* (2018) 9:3868. doi: 10.1038/s41467-018-06300-3
71. Bauer DC, Zadoorian A, Wilson LOWMelbourne Genomics Health Alliance, Thorne NP. Evaluation of computational programs to predict HLA genotypes from genomic sequencing data. *Brief Bioinform* (2016) 19(2):bbw097. doi: 10.1093/bib/bbw097
72. Zsolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* (2014) 30:3310–6. doi: 10.1093/bioinformatics/btu548
73. Shukla SA, Rooney MS, Rajasagi M, Tiao G, Dixon PM, Lawrence MS, et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat Biotechnol* (2015) 33:1152–8. doi: 10.1038/nbt.3344
74. Bai Y, Wang D, Fury W. PHLAT: inference of high-resolution HLA types from RNA and whole exome sequencing. In: Boegel S, editor. *HLA typing. Methods in molecular biology*. New York, NY: Springer New York (2018). p. 193–201. doi: 10.1007/978-1-4939-8546-3_13
75. Xie C, Yeo ZX, Wong M, Piper J, Long T, Kirkness EF, et al. Fast and accurate HLA typing from short-read next-generation sequencing data with xHLA. *Proc Natl Acad Sci USA* (2017) 114:8059–64. doi: 10.1073/pnas.1707945114
76. Kawaguchi S, Higasa K, Shimizu M, Yamada R, Matsuda F. HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data. *Hum Mutat* (2017) 38:788–97. doi: 10.1002/humu.23230
77. Fotakis G, Trajanoski Z, Rieder D. Computational cancer neoantigen prediction: current status and recent advances. *Immuno-Oncology Technol* (2021) 12:100052. doi: 10.1016/j.iotech.2021.100052
78. Schaap-Johansen A-L, Vujović M, Borch A, Hadrup SR, Marcatili P. T cell epitope prediction and its application to immunotherapy. *Front Immunol* (2021) 12:712488. doi: 10.3389/fimmu.2021.712488
79. De Mattos-Arruda L, Vazquez M, Finotello F, Lepore R, Porta E, Hundal J, et al. Neoantigen prediction and computational perspectives towards clinical benefit: recommendations from the ESMO Precision Medicine Working Group. *Ann Oncol* (2020) 31:978–90. doi: 10.1016/j.annonc.2020.05.008
80. Castro A, Zanetti M, Carter H. Neoantigen controversies. *Annu Rev BioMed Data Sci* (2021) 4:227–53. doi: 10.1146/annurev-biodatasci-092820-112713
81. Gopanenko AV, Kosobokova EN, Kosorukov VS. Main strategies for the identification of neoantigens. *Cancers* (2020) 12:2879. doi: 10.3390/cancers12102879
82. Xie N, Shen G, Gao W, Huang Z, Huang C, Fu L. Neoantigens: promising targets for cancer therapy. *Sig Transduct Target Ther* (2023) 8:9. doi: 10.1038/s41392-022-01270-x
83. Xia H, McMichael J, Becker-Hapak M, Onyeador OC, Buchli R, McClain E, et al. Computational prediction of MHC anchor locations guides neoantigen identification and prioritization. *Sci Immunol* (2023) 8:eabg2200. doi: 10.1126/sciimmunol.abg2200
84. Wells DK, Van Buuren MM, Dang KK, Hubbard-Lucey VM, Sheehan KCF, Campbell KM, et al. Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. *Cell* (2020) 183:818–834.e13. doi: 10.1016/j.cell.2020.09.015
85. Gros A, Parkhurst MR, Tran E, Pasetto A, Robbins PF, Ilyas S, et al. Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nat Med* (2016) 22:433–8. doi: 10.1038/nm.4051
86. Tran E, Ahmadzadeh M, Lu Y-C, Gros A, Turcotte S, Robbins PF, et al. Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science* (2015) 350:1387–90. doi: 10.1126/science.aad1253
87. Yossef R, Tran E, Deniger DC, Gros A, Pasetto A, Parkhurst MR, et al. Enhanced detection of neoantigen-reactive T cells targeting unique and shared oncogenes for personalized cancer immunotherapy. *JCI Insight* (2018) 3:e122467. doi: 10.1172/jci.insight.122467
88. Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* (2003) 12:1007–17. doi: 10.1110/ps.0239403
89. 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* (2008) 36:W509–12. doi: 10.1093/nar/gkn202
90. Nielsen M, Andreatta M. NetMHCpan-3.0: improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med* (2016) 8:33. doi: 10.1186/s13073-016-0288-x
91. Bjerregaard A-M, Nielsen M, Hadrup SR, Szallasi Z, Eklund AC. MuPeXI: prediction of neo-epitopes from tumor sequencing data. *Cancer Immunol Immunother* (2017) 66:1123–30. doi: 10.1007/s00262-017-2001-3
92. Hundal J, Carreno BM, Petti AA, Linette GP, Griffith OL, Mardis ER, et al. pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. *Genome Med* (2016) 8:11. doi: 10.1186/s13073-016-0264-5
93. Zhang J, Mardis ER, Maher CA. INTEGRATE-neo: a pipeline for personalized gene fusion neoantigen discovery. *Bioinformatics* (2017) 33:555–7. doi: 10.1093/bioinformatics/btw674
94. Schenck RO, Lakatos E, Gatenbee C, Graham TA, Anderson ARA. NeoPredPipe: high-throughput neoantigen prediction and recognition potential pipeline. *BMC Bioinf* (2019) 20:264. doi: 10.1186/s12859-019-2876-4
95. Stranzl T, Larsen MV, Lundegaard C, Nielsen M. NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics* (2010) 62:357–68. doi: 10.1007/s00251-010-0441-4
96. O'Donnell TJ, Rubinsteyn A, Bonsack M, Riemer AB, Laserson U, Hammerbacher J. MHCflurry: open-source class I MHC binding affinity prediction. *Cell Syst* (2018) 7:129–132.e4. doi: 10.1016/j.cels.2018.05.014
97. Coullie PG, Lehmann F, Lethé B, Herman J, Lurquin C, Andrawiss M, et al. A mutated intron sequence codes for an antigenic peptide recognized by cytolytic T lymphocytes on a human melanoma. *Proc Natl Acad Sci USA* (1995) 92:7976–80. doi: 10.1073/pnas.92.17.7976
98. Lu Y-C, Yao X, Li YF, El-Gamil M, Dudley ME, Yang JC, et al. Mutated PPP1R3B is recognized by T cells used to treat a melanoma patient who experienced a durable complete tumor regression. *JCI* (2013) 190:6034–42. doi: 10.4049/jimmunol.1202830

99. van der Bruggen P, Traversari C, Chomez P, Lurquin C, De Plaen E, Van den Eynde B, et al. A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. *Science* (1991) 254:1643–7. doi: 10.1126/science.1840703
100. Kawakami Y, Elyahu S, Delgado CH, Robbins PF, Rivoltini L, Topalian SL, et al. Cloning of the gene coding for a shared human melanoma antigen recognized by autologous T cells infiltrating into tumor. *Proc Natl Acad Sci USA* (1994) 91:3515–9. doi: 10.1073/pnas.91.9.3515
101. Chen Y-T, Scanlan MJ, Sahin U, Türeci Ö, Gure AO, Tsang S, et al. A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening. *Proc Natl Acad Sci USA* (1997) 94:1914–8. doi: 10.1073/pnas.94.5.1914
102. Altman JD, Moss PAH, Goulder PJR, Barouch DH, McHeyzer-Williams MG, Bell JI, et al. Phenotypic analysis of antigen-specific T lymphocytes. *Science* (1996) 274:94–6. doi: 10.1126/science.274.5284.94
103. Hebeisen M, Oberle SG, Presotto D, Speiser DE, Zehn D, Rufer N. Molecular insights for optimizing T cell receptor specificity against cancer. *Front Immunol* (2013) 4:154. doi: 10.3389/fimmu.2013.00154
104. Wooldridge L, Lissina A, Cole DK, van den Berg HA, Price DA, Sewell AK. Tricks with tetramers: how to get the most from multimeric peptide-MHC. *Immunology* (2009) 126:147–64. doi: 10.1111/j.1365-2567.2008.02848.x
105. Dolton G, Tungatt K, Lloyd A, Bianchi V, Theaker SM, Trimby A, et al. More tricks with tetramers: a practical guide to staining T cells with peptide-MHC multimers. *Immunology* (2015) 146:11–22. doi: 10.1111/imm.12499
106. Svensson A, Nordström I, Sun J-B, Eriksson K. Protective immunity to genital herpes simplex virus type 2 infection is mediated by T-bet. *J Immunol* (2005) 174:6266–73. doi: 10.4049/jimmunol.174.10.6266
107. Binder RJ, Srivastava PK. Peptides chaperoned by heat-shock proteins are a necessary and sufficient source of antigens in the cross-priming of CD8+ T cells. *Nat Immunol* (2005) 6:593–9. doi: 10.1038/ni1201
108. Huang J, Zeng X, Sigal N, Lund PJ, Su LF, Huang H, et al. Detection, phenotyping, and quantification of antigen-specific T cells using a peptide-MHC dodecamer. *Proc Natl Acad Sci USA* (2016) 113:E1890–97. doi: 10.1073/pnas.1602488113
109. Toebes M, Coccoris M, Bins A, Rodenko B, Gomez R, Nieuwkoop NJ, et al. Design and use of conditional MHC class I ligands. *Nat Med* (2006) 12:246–51. doi: 10.1038/nm1360
110. Day CL, Seth NP, Lucas M, Appel H, Gauthier L, Lauer GM, et al. Ex vivo analysis of human memory CD4 T cells specific for hepatitis C virus using MHC class II tetramers. *J Clin Invest* (2003) 112:831–42. doi: 10.1172/JCI200318509
111. Hadrup SR, Bakker AH, Shu CJ, Andersen RS, van Veluw J, Hombrink P, et al. Parallel detection of antigen-specific T-cell responses by multidimensional encoding of MHC multimers. *Nat Methods* (2009) 6:520–6. doi: 10.1038/nmeth.1345
112. Ornatsky O, Baranov VI, Bandura DR, Tanner SD, Dick J. Multiple cellular antigen detection by ICP-MS. *J Immunol Methods* (2006) 308:68–76. doi: 10.1016/j.jim.2005.09.020
113. Bendall SC, Simonds EF, Qiu P, Amir ED, Krutzik PO, Finck R, et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* (2011) 332:687–96. doi: 10.1126/science.1198704
114. Bentzen AK, Marquard AM, Lyngaa R, Saini SK, Ramskov S, Donia M, et al. Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nat Biotechnol* (2016) 34:1037–45. doi: 10.1038/nbt.3662
115. Zhang S-Q, Ma K-Y, Schonnesen AA, Zhang M, He C, Sun E, et al. High-throughput determination of the antigen specificities of T cell receptors in single cells. *Nat Biotechnol* (2018) 36:1156–9. doi: 10.1038/nbt.4282
116. Zhou W, Yan Y, Guo Q, Ji H, Wang H, Xu T, et al. Microfluidics applications for high-throughput single cell sequencing. *J Nanobiotechnol* (2021) 19:312. doi: 10.1186/s12951-021-01045-6
117. Segaliny AI, Li G, Kong L, Ren C, Chen X, Wang JK, et al. Functional TCR T cell screening using single-cell droplet microfluidics. *Lab Chip* (2018) 18:3733–49. doi: 10.1039/C8LC00818C
118. Peng S, Zaretsky JM, Ng AHC, Chour W, Bethune MT, Choi J, et al. Sensitive detection and analysis of neoantigen-specific T cell populations from tumors and blood. *Cell Rep* (2019) 28:2728–2738.e7. doi: 10.1016/j.celrep.2019.07.106
119. Ng AHC, Peng S, Xu AM, Noh WJ, Guo K, Bethune MT, et al. MATE-Seq: microfluidic antigen-TCR engagement sequencing. *Lab Chip* (2019) 19:3011–21. doi: 10.1039/C9LC00538B
120. Lu Y-C, Yao X, Crystal JS, Li YF, El-Gamil M, Gross C, et al. Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions. *Clin Cancer Res* (2014) 20:3401–10. doi: 10.1158/1078-0432.CCR-14-0433
121. Tran E, Turcotte S, Gros A, Robbins PF, Lu Y-C, Dudley ME, et al. Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science* (2014) 344:641–5. doi: 10.1126/science.1251102
122. Zacharakis N, Chinnsamy H, Black M, Xu H, Lu Y-C, Zheng Z, et al. Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. *Nat Med* (2018) 24:724–30. doi: 10.1038/s41591-018-0040-8
123. Durai M, Krueger C, Ye Z, Cheng L, Mackensen A, Oelke M, et al. *In vivo* functional efficacy of tumor-specific T cells expanded using HLA-Ig based artificial antigen presenting cells (aAPC). *Cancer Immunol Immunother* (2009) 58:209–20. doi: 10.1007/s00262-008-0542-1
124. Lu X, Jiang X, Liu R, Zhang S. *In vivo* anti-melanoma efficacy of allo-restricted CTLs specific for melanoma expanded by artificial antigen-presenting cells. *Cancer Immunol Immunother* (2009) 58:629–38. doi: 10.1007/s00262-008-0573-7
125. Joglekar AV, Leonard MT, Jeppson JD, Swift M, Li G, Wong S, et al. T cell antigen discovery via signaling and antigen-presenting bifunctional receptors. *Nat Methods* (2019) 16:191–8. doi: 10.1038/s41592-018-0304-8
126. Kula T, Dezfulian MH, Wang CI, Abdelfattah NS, Hartman ZC, Wucherpfennig KW, et al. T-scan: A genome-wide method for the systematic discovery of T cell epitopes. *Cell* (2019) 178:1016–1028.e13. doi: 10.1016/j.cell.2019.07.009
127. Sharma G, Rive CM, Holt RA. Rapid selection and identification of functional CD8+ T cell epitopes from large peptide-coding libraries. *Nat Commun* (2019) 10:4553. doi: 10.1038/s41467-019-12444-7
128. Kisielow J, Obermair F-J, Kopf M. Deciphering CD4+ T cell specificity using novel MHC-TCR chimeric receptors. *Nat Immunol* (2019) 20:652–62. doi: 10.1038/s41590-019-0335-z
129. Li G, Bethune MT, Wong S, Joglekar AV, Leonard MT, Wang JK, et al. T cell antigen discovery via trogocytosis. *Nat Methods* (2019) 16:183–90. doi: 10.1038/s41592-018-0305-7
130. Rienzo M, Lin K-C, Mobilia KC, Sackmann EK, Kurz V, Navidi AH, et al. High-throughput optofluidic screening for improved microbial cell factories via real-time micron-scale productivity monitoring. *Lab Chip* (2021) 21:2901–12. doi: 10.1039/D1LC00389E
131. Han A, Glanville J, Hansmann L, Davis MM. Linking T-cell receptor sequence to functional phenotype at the single-cell level. *Nat Biotechnol* (2014) 32:684–92. doi: 10.1038/nbt.2938
132. Stubbington MJT, Lönnberg T, Proserpio V, Clare S, Speak AO, Dougan G, et al. T cell fate and clonality inference from single-cell transcriptomes. *Nat Methods* (2016) 13:329–32. doi: 10.1038/nmeth.3800
133. Zhang Z, Xiong D, Wang X, Liu H, Wang T. Mapping the functional landscape of T cell receptor repertoires by single-T cell transcriptomics. *Nat Methods* (2021) 18:92–9. doi: 10.1038/s41592-020-01020-3
134. Lagattuta KA, Kang JB, Nathan A, Pauken KE, Jonsson AH, Rao DA, et al. Repertoire analyses reveal T cell antigen receptor sequence features that influence T cell fate. *Nat Immunol* (2022) 23:446–57. doi: 10.1038/s41590-022-01129-x
135. Chen DG, Xie J, Su Y, Heath JR. T cell receptor sequences are the dominant factor contributing to the phenotype of CD8+ T cells with specificities against immunogenic viral antigens. *Cell Rep* (2023) 42(11):113279. doi: 10.1016/j.celrep.2023.113279
136. Oliveira G, Stromhaug K, Cieri N, Iorgulescu JB, Klaeger S, Wolff JO, et al. Landscape of helper and regulatory antitumor CD4+ T cells in melanoma. *Nature* (2022) 605:532–8. doi: 10.1038/s41586-022-04682-5
137. Oliveira G, Stromhaug K, Klaeger S, Kula T, Frederick DT, Le PM, et al. Phenotype, specificity and avidity of antitumor CD8+ T cells in melanoma. *Nature* (2021) 596:119–25. doi: 10.1038/s41586-021-03704-y
138. Krishna S, Lowery FJ, Copeland AR, Bahadiroglu E, Mukherjee R, Jia L, et al. Stem-like CD8 T cells mediate response of adoptive cell immunotherapy against human cancer. *Science* (2020) 370:1328–34. doi: 10.1126/science.abb9847
139. Lowery FJ, Krishna S, Yossef R, Parikh NB, Chatani PD, Zacharakis N, et al. Molecular signatures of antitumor neoantigen-reactive T cells from metastatic human cancers. *Science* (2022) 375:877–84. doi: 10.1126/science.abb5447
140. Hanada K, Zhao C, Gil-Hoyos R, Gartner JJ, Chow-Parmer C, Lowery FJ, et al. A phenotypic signature that identifies neoantigen-reactive T cells in fresh human lung cancers. *Cancer Cell* (2022) 40:479–493.e6. doi: 10.1016/j.ccell.2022.03.012
141. Tickotsky N, Sagiv T, Prilusky J, Shifrut E, Friedman N. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics* (2017) 33:2924–9. doi: 10.1093/bioinformatics/btx286
142. Shugay M, Bagaev DV, Zvyagin IV, Vroomans RM, Crawford JC, Dolton G, et al. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res* (2018) 46:D419–27. doi: 10.1093/nar/gkx760
143. Zhang W, Wang L, Liu K, Wei X, Yang K, Du W, et al. PIRD: pan immune repertoire database. *Bioinformatics* (2020) 36:897–903. doi: 10.1093/bioinformatics/btz614
144. Rudolph MG, Stanfield RL, Wilson IA. HOW TCRS BIND MHCS, PEPTIDES, AND CORECEPTORS. *Annu Rev Immunol* (2006) 24:419–66. doi: 10.1146/annurev.immunol.23.021704.115658
145. Mallis RJ, Brazin KN, Duke-Cohan JS, Hwang W, Wang J, Wagner G, et al. NMR: an essential structural tool for integrative studies of T cell development, pMHC ligand recognition and TCR mechanobiology. *J Biomol NMR* (2019) 73:319–32. doi: 10.1007/s10858-019-00234-8
146. Cole DK, Pumphrey NJ, Boulter JM, Sami M, Bell JI, Gostick E, et al. Human TCR-binding affinity is governed by MHC class restriction. *J Immunol* (2007) 178:5727–34. doi: 10.4049/jimmunol.178.9.5727
147. Piepenbrink KH, Gloor BE, Armstrong KM, Baker BM. Methods for quantifying T cell receptor binding affinities and thermodynamics. In: *Methods in*

enzymology. Cambridge, MA: Elsevier (2009). p. 359–81. doi: 10.1016/S0076-6879(09)66015-8

148. Newell EW, Davis MM. Beyond model antigens: high-dimensional methods for the analysis of antigen-specific T cells. *Nat Biotechnol* (2014) 32:149–57. doi: 10.1038/nbt.2783

149. Hudson D, Fernandes RA, Basham M, Ogg G, Koohy H. Can we predict T cell specificity with digital biology and machine learning? *Nat Rev Immunol* (2023) 23(8):511–21. doi: 10.1038/s41577-023-00835-3

150. Glanville J, Huang H, Nau A, Hatton O, Wagar LE, Rubelt F, et al. Identifying specificity groups in the T cell receptor repertoire. *Nature* (2017) 547:94–8. doi: 10.1038/nature22976

151. Huang H, Wang C, Rubelt F, Scriba TJ, Davis MM. Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat Biotechnol* (2020) 38:1194–202. doi: 10.1038/s41587-020-0505-4

152. Dash P, Fiore-Gartland AJ, Hertz T, Wang GC, Sharma S, Souquette A, et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* (2017) 547:89–93. doi: 10.1038/nature22383

153. Mayer-Blackwell K, Schattgen S, Cohen-Lavi L, Crawford JC, Souquette A, Gaevert JA, et al. TCR meta-clonotypes for biomarker discovery with tcrdist3 enabled identification of public, HLA-restricted clusters of SARS-CoV-2 TCRs. *eLife* (2021) 10:e68605. doi: 10.7554/eLife.68605

154. Marcou Q, Mora T, Walczak AM. High-throughput immune repertoire analysis with IGoR. *Nat Commun* (2018) 9:561. doi: 10.1038/s41467-018-02832-w

155. Gielis S, Moris P, Bittremieux W, De Neuter N, Ogunjimi B, Laukens K, et al. Detection of enriched T cell epitope specificity in full T cell receptor sequence repertoires. *Front Immunol* (2019) 10:2820. doi: 10.3389/fimmu.2019.02820

156. Sidhom J-W, Larman HB, Pardoll DM, Baras AS. DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires. *Nat Commun* (2021) 12:1605. doi: 10.1038/s41467-021-21879-w