frontiers
in Oncology

Cancer Neoantigens: Challenges and Future Directions for Prediction, Prioritization, and Validation

Elizabeth S. Borden[1,2], Kenneth H. Buetow[3,4], Melissa A. Wilson[3,4] and Karen Taraszka Hastings[1,2]*

[1] Department of Basic Medical Sciences, College of Medicine-Phoenix, University of Arizona, Phoenix, AZ, United States, [2] Department of Research and Internal Medicine (Dermatology), Phoenix Veterans Affairs Health Care System, Phoenix, AZ, United States, [3] School of Life Sciences, Arizona State University, Tempe, AZ, United States, [4] Center for Evolution and Medicine, Arizona State University, Tempe, AZ, United States

Prioritization of immunogenic neoantigens is key to enhancing cancer immunotherapy through the development of personalized vaccines, adoptive T cell therapy, and the prediction of response to immune checkpoint inhibition. Neoantigens are tumor-specific proteins that allow the immune system to recognize and destroy a tumor. Cancer immunotherapies, such as personalized cancer vaccines, adoptive T cell therapy, and immune checkpoint inhibition, rely on an understanding of the patient-specific neoantigen profile in order to guide personalized therapeutic strategies. Genomic approaches to predicting and prioritizing immunogenic neoantigens are rapidly expanding, raising new opportunities to advance these tools and enhance their clinical relevance. Predicting neoantigens requires acquisition of high-quality samples and sequencing data, followed by variant calling and variant annotation. Subsequently, prioritizing which of these neoantigens may elicit a tumor-specific immune response requires application and integration of tools to predict the expression, processing, binding, and recognition potentials of the neoantigen. Finally, improvement of the computational tools is held in constant tension with the availability of datasets with validated immunogenic neoantigens. The goal of this review article is to summarize the current knowledge and limitations in neoantigen prediction, prioritization, and validation and propose future directions that will improve personalized cancer treatment.

Keywords: neoantigens (neoAgs), MHC class I, MHC class II, neoantigen prioritization, neoantigen prediction

1 INTRODUCTION

Neoantigens are tumor-specific mutated peptides that are key targets of the anti-cancer immune response, because neoantigens are not subject to immune tolerance (non-reactivity to self) (1–4). Three classes of cancer therapies reliant on the neoantigen expression and presentation by MHC are personalized neoantigen vaccines, adoptive T cell therapy, and immune checkpoint inhibitors. Personalized neoantigen vaccines have gained momentum in recent years because of their early success (5–9). Several approaches to vaccination have been attempted to date, including direct exposure to neoantigens (6), neoantigen-encoding RNA vaccines (7), and neoantigen-loaded

dendritic cell vaccines (5). Regardless of the vaccination strategy, all personalized neoantigen vaccines rely on accurate prediction of immunogenic neoantigens, neoantigens that are presented by MHC and elicit a T cell-mediated immune response.

Adoptive T cell therapy has also demonstrated promise as a targeted immunotherapy. Adoptive T cell therapy includes transfer of tumor-infiltrating lymphocytes and T cells genetically modified to express a T cell receptor (TCR) or chimeric antigen receptor. Early attempts at adoptive T cell therapy focused on introducing T cells specific for tumor associated antigens including MAGE-A3 in melanoma and carcinoembryonic antigen (CEA) in colorectal cancer (10, 11). However, the lack of tumor-specificity of these antigens led to significant off target effects and severe toxicity. There is, therefore, growing interest in the application of neoantigen-specific adoptive T cell therapy to enhance T cell mediated tumor-destruction while reducing off target effects (1, 12–15). As for personalized neoantigen vaccines, adoptive T cell therapy specific to neoantigens relies on accurate prediction of immunogenic neoantigens.

Tumor-specific neoantigens are also the target of cancer immunotherapy with immune checkpoint inhibitors (16, 17). Immune checkpoint inhibitors, including monoclonal antibodies against PD-1 and CTLA-4, block inhibitory signals to the T cells to increase T cell-mediated tumor destruction (18). Unfortunately, immune checkpoint inhibitors are only effective in a subset of patients and are associated with immune-related adverse events. Thus, there is interest in predicting which patients will respond to treatment with a single immune checkpoint inhibitor and which would benefit from combination therapy. Several recent studies have demonstrated that the predicted immunogenic neoantigens are more strongly associated with response to immune checkpoint inhibition than mutational burden (19–23). Accurately determining the association of neoantigen immunogenicity with response to

immune checkpoint inhibition relies on accurate prioritization of immunogenic neoantigens.

Successful identification of immunogenic neoantigens using traditional genomic approaches requires a combination of neoantigen prediction and neoantigen prioritization (**Figure 1**). Neoantigen prediction requires sample acquisition, high quality sequencing data, prediction of the somatic mutations present in the tumor cell (variant calling), and accurate prediction of the neoantigens resulting from these somatic mutations (variant annotation). A few considerations for neoantigen identification include the tissue types to be sequenced, the best collection/preservation method for the tissues, and the type of sequencing data to be obtained. Additionally, one should decide on the types of mutations to be considered, appropriate methods by which to identify these mutations, and the most accurate annotation methods. Prioritization of immunogenic neoantigens relies on a thorough understanding of the characteristics of a neoantigen and the optimal ways of combining these characteristics to predict the potential of the neoantigen to elicit an immune response. For both MHC class I- and II-restricted neoantigens, characteristics that have been considered include expression of the neoantigen of interest, processing of the peptide including proteasomal cleavage and transport into the endoplasmic reticulum, binding of the neoantigen to MHC class I or II, and TCR recognition. Several tools are available for predicting each of these characteristics, and a variety of models have synthesized the characteristics into an overall immunogenicity score (21–26). We will review the available literature to guide decisions for each step in neoantigen prediction and prioritization and highlight areas for future research.

Datasets containing neoantigens that are validated to bind MHC class I or II and elicit a CD8+ or CD4+ T cell response are critical for assessing the overall performance of genomic pipelines and driving improved computational neoantigen identification. As the available datasets increase, models for

FIGURE 1 | Overview of neoantigen prediction, prioritization, and validation. Neoantigen prediction relies on sample acquisition, high quality sequencing data, variant calling, and variant annotation. Neoantigen prioritization requires predicting some combination of the potential for the neoantigen to be expressed, processed, bound by MHC, and recognized by the T cell receptor (TCR). The development of neoantigen prioritization models relies on the availability of validated datasets of neoantigen immunogenicity. Figure created with BioRender.com.

neoantigen prioritization will be refined. Currently, many datasets are available for MHC class I-restricted neoantigens derived from single nucleotide variants (SNVs) and small insertions and deletions (indels). There are limited datasets available for MHC class II-restricted neoantigens and a lack of datasets for neoantigens derived from large indels, frameshifts, and gene fusions. We will summarize the available datasets and highlight ways to enhance future validation sets for continued improvement of neoantigen prediction and prioritization.

2 NEOANTIGEN PREDICTION

2.1 Sample Acquisition
Sample collection and sequencing are the first steps in performing neoantigen prediction and prioritization from DNA- or RNA-level mutations. While proteomic-based methods, which do not universally require the sequencing data presented here, have been created with direct profiling of peptides bound to MHC class I or II molecules, these are beyond the scope of this review article and have been discussed (27). Decisions related to sequencing can broadly be classified into the types of tissue needed, tissue collection methods, and types of sequencing. Here, we provide an up-to-date review of the literature to help guide each of these decisions (summarized in **Figure 2**).

The first consideration for sample acquisition is the tissues that need to be used to generate accurate somatic variant calls; specifically, whether a germline reference sample is required for variant calling. Typically, tumor and germline samples are compared to identify tumor-specific, somatic mutations. However, germline samples are not always available, especially for archived samples, though they can be collected in clinical settings. Therefore, there is continued interest in whether neoantigens can be identified in the absence of a germline

sample. A patient-specific germline sample is currently the best available method to ensure that the variants being detected are due to true somatic mutations, rather than germline mutations (28). A few novel approaches have been suggested to reduce the number of germline variants identified without a germline reference sample (29–31). One such method uses tumor tissue only and assumptions about differences in the allelic frequency of the germline variants, compared to somatic variants to filter results. A germline heterozygous mutation should be closer to a 50% allelic frequency, whereas a somatic heterozygous mutation is likely to have less than a 50% allelic frequency because they won't be present in germline tissue, only tumor samples. However, the assumption that a somatic heterozygous mutation will have less than a 50% allele frequency is complicated by copy number variations and stromal contamination, which are also accounted for in the model. Across seven test tumors, the sensitivity of the method ranged from 44-87%, which the authors acknowledge is too low to currently be applicable for clinical use (30). Another method performed variant calling for a tumor compared to 20 unmatched normal samples and kept variants that were identified in 90% of the comparisons. The variants were also filtered by 1) elimination of variants with the same allelic frequency as known germline mutations, 2) removal of variants from hard to map regions of the genome and 3) elimination of C→T and G→A mutations with low allelic frequencies. This method reported a 94% sensitivity, 99% specificity, and 76% positive predictive value (31). These numbers exceed those of other available tools and indicate that tumor-only variant calling may be an option for clinical applications in the future. However, these results were only validated for a set of stringently selected somatic mutations suggesting that further analysis would be needed to ensure that the results are stable for a more comprehensive set of tumor mutations. While the field of somatic variant calling is constantly

FIGURE 2 | Sample collection and sequencing considerations. Here we describe considerations for obtaining sequencing data for neoantigen prediction including tissues needed, tissue collection method, and sequencing types. Figure created with BioRender.com.
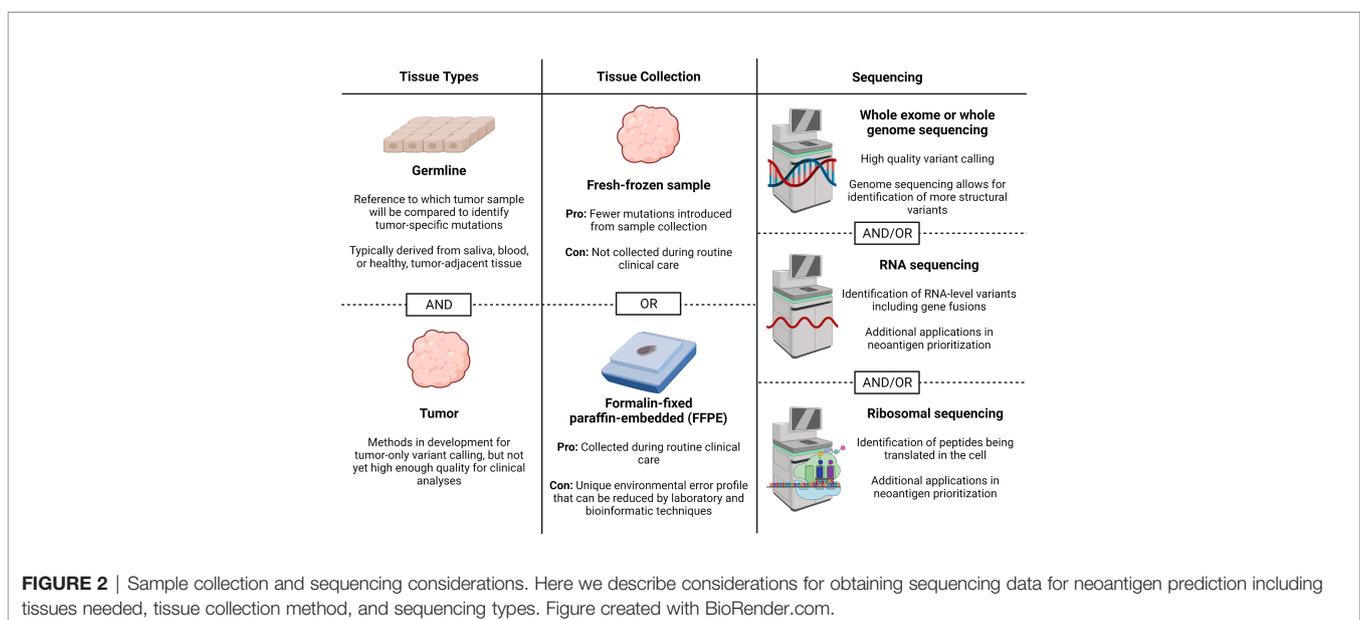
improving, until the sensitivity and specificity of available methods improve, a germline sample is recommended.

When using a germline sample, a second question is which tissue source is the most accurate to use as the germline reference. Options that have been frequently employed in the literature include saliva, blood, or tumor-adjacent tissue, but the source of germline tissue can affect which variants are called as tumor-specific. Each tissue has its advantages and disadvantages. Saliva has the advantage of being a readily available, non-invasive method for obtaining a germline DNA reference. However, two recent studies using whole genome sequencing (WGS) on saliva samples demonstrated a risk for contamination from bacteria and food DNA that can influence the read mapping and variant calling (32, 33). When aggregated across four patients, saliva resulted in the identification of 776 unique coding variants compared to 157 from blood. Manual inspection of a sampling of the saliva-only variants demonstrated that most were attributable to bacterial contamination (32). The risk of bacterial contamination may be lessened in whole exome sequencing (WES) where hybridization methods are used to capture exons; however, an older study demonstrated bacterial contamination in WES data (34). To our knowledge, there are no studies that assess the impact of bacterial contamination on variant calling from WES data.

While slightly more invasive than saliva collection, blood still has the advantage of being minimally invasive. A study on optimizing cancer genomics experiments suggests that blood may be the best germline reference for solid tumors. Blood is a different tissue origin from most solid tumors and may have a lower risk for tumor-in-normal contamination than tumor-adjacent tissue (35). While the advantage of no tumor-in-normal contamination could be undermined by circulating tumor cell contamination, examination of ten cancer types from the Cancer Genome Atlas (TCGA) demonstrated no detection of tumor-in-normal contamination across the 304 blood samples tested (36). The tested blood samples were from patients with untreated primary tumors, so the risk of circulating tumor cells may be greater in advanced and metastatic disease. A recently developed tool, DeTiN has also been suggested as a means of removing the tumor-in-normal contamination (36). DeTiN demonstrated increased true positive variant detection with no significant change in the false positive rate (36).

With regards to tumor-adjacent tissue, a factor to consider is the potential for shared mutations between the tumor and tumor-adjacent tissue (37). One cause of these shared mutations could be exposure to a shared carcinogen. For example, recent work in skin cancer has demonstrated that there are early mutations in non-tumor sun-exposed skin due to exposure to ultraviolet radiation (38). Recent evidence has demonstrated the presence of somatic mutant clones within normal tissue (39–41). These somatic mutant clones can have numerous somatic mutations, a portion of which overlap with tumor mutations (41). It is unclear whether the presence of shared mutations within the tumor-adjacent sample will benefit or hinder the therapeutic utility of identified neoantigens. Shared mutations between tumor and tumor-adjacent tissue introduce

the risk of eliminating neoantigens that occur in the cancer field and pre-cancerous lesions. However, the elimination of shared mutations may better facilitate tumor-specific targeting. Overall, if the goal is to maximize the number of tumor mutations identified, blood is the best available germline comparison for solid tumors, since it minimizes the risk of bacterial contamination and is the least likely to have a shared mutational profile. Additional research will be needed to assess the relative therapeutic benefits of neoantigens shared with the cancer field compared to unique tumor mutations.

Two common options for storing tissues used for neoantigen identification are fresh-frozen and formalin-fixed, paraffin-embedded (FFPE) samples, and each sample type has strengths and weaknesses. Fresh-frozen samples are attractive, because the samples have minimal processing that can affect DNA integrity; and typically, fresh-frozen samples can be used for both DNA and RNA isolation. However, fresh-frozen samples require a biobank setup to collect and are not part of routine clinical care. FFPE samples have the distinct advantage of being routinely collected in clinical settings, but have a characteristic set of mutations due to the preservation method and lack reliable RNA. A recent side-by-side comparison of variant calling in FFPE compared to fresh-frozen samples found that FFPE samples have ~5% more variants called than paired, fresh-frozen samples (42). The false discovery rate (FDR) was highly concentrated in the variants with low allelic frequency and is also predominated by C→T and G→A transitions due to deamination of methylated cytosine position (42), introduced by the FFPE process. Going forward, new approaches have been developed with DNA extraction kits that include enzymatic removal of cytosine deamination artifacts. Extraction with enzymatic removal of artifacts was shown to decrease the estimated FDR in low allelic frequency variants from 94.8% to 69.8% (42). Thus, fresh-frozen samples are preferred when possible; and when FFPE samples are used, DNA extraction protocols specific for FFPE samples are recommended.

Three possible bioinformatic methods can be applied to reduce false positives from FFPE damage: 1) taking the overlap from multiple variant callers, 2) eliminating low variant allele frequency variants, and 3) eliminating characteristic FFPE mutations. de Schaetzen et al. demonstrated increased agreement of variants called from FFPE with a set of high-confidence variants from fresh-frozen tumors when analyzing the overlap of at least two of the four variant callers employed: Strelka2, GATK Mutect2, Shimmer, and VarScan2 (43). A limitation to the consensus approach is that it may emphasize specificity over sensitivity and eliminate variants with potential clinical significance. Another approach (to be taken independently or in combination) is to filter out low allelic frequency calls, as one study suggested that the bulk of the false positives occur at lower frequency (42). However, a separate study demonstrated that highly reliable variants from fresh-frozen samples were generally represented at a lower allelic frequency in FFPE samples than in fresh-frozen samples (43). Additionally, the discrepancies between FFPE and fresh-frozen samples that remained after using the overlap of two out of four

variant callers were demonstrated to be due to differences in the subclonal population (43). If low frequency variants are not errors, but rather represent subclonal populations, then eliminating the low frequency variants will result in a reduced ability to predict neoantigens. A newer method called Ideafix uses machine learning to consider a range of characteristics of the mutation to determine the likelihood of that mutation being an artifact of FFPE preservation. These characteristics include the variant allele frequency, the C→T mutational signature, the genomic context of the variant (based on flanking nucleotides that may increase the risk of deamination), and strand bias (whether the mutation is only identified on forward or reverse strand reads). Combining these features with a machine learning algorithm demonstrated an area under the receiver operator characteristics curve (AUC) of over 0.96 in two independent test datasets (44). Other recently developed models have taken a similar approach (45, 46). One potential challenge for these approaches is the inability to distinguish between true C→T mutations, including those enriched in ultraviolet light-induced tumors, and those that are artifacts due to FFPE processing. Overall, application of a newer model such as Ideafix may be helpful in eliminating FFPE artifacts while sacrificing minimal clinically relevant variants.

2.2 Sequencing

Each type of sequencing, including RNA sequencing (RNAseq), WGS, WES and combined approaches have potential advantages and limitations with regards to neoantigen prediction and prioritization. Traditionally, WGS or WES have been the preferred sequencing types for variant calling. WGS has the advantage of allowing for the identification of certain structural variants that are excluded by WES data (discussed below) but has the disadvantage of being more expensive than WES data (47). RNAseq data is a potential alternative to WGS or WES sequencing as it would allow for variant calling, as well as differential expression analysis and incorporation of mRNA expression data into neoantigen prioritization. Using RNAseq data for both variant calling and expression is a potentially attractive measure to reduce sequencing costs. However, methods for variant calling from RNAseq data have not been traditionally considered high enough quality to be used in isolation (48). Recent benchmarking demonstrated a low level of agreement between WES and RNAseq variants (49). One of the likely causes of the discrepancy between WES and RNAseq variants is that WES does not include all areas of the genome that may be transcribed, as another study demonstrated that ~71% of RNAseq-only variants occurred in regions not covered by the WES capture (48). Other possible causes include RNA-level modifications or differences in the read depth (49). To assess the performance of each method, variants called were compared to the COSMIC and dbSNP databases. The COSMIC database is a set of known cancer-specific mutations, whereas the dbSNP database is a database of variants known to exist in a healthy population. Therefore, enrichment of COSMIC-only variants reflects an increase in the likelihood of the mutation being a somatic mutation (49). Taking the intersection of RNAseq and WES variants led to enrichment of COSMIC-only variants, with

87.7% being COSMIC-only in the intersection approach compared to 39.5% in the WES and 3.0% in the RNAseq approach (49). A limitation acknowledged by the authors is that the COSMIC database is limited primarily to variants previously identified by WES analysis, so many RNAseq mutations may not be included in the COSMIC database. Overall, although WES-only approaches have been the most popular to date, there are possible advantages to RNAseq-based variant calling approaches. Further work is warranted to document the rates of true positive and false positive variant calls with RNAseq and WES approaches.

One final sequencing type to consider is a newer method for ribosomal profiling known as Ribo-seq, which allows for the specific transcription of all proteins being actively translated at the time of cell lysis (50). Ribo-seq has two potential advantages in the space of neoantigen prediction and prioritization. First, it has been proposed as a novel approach for detecting neoantigens derived from open reading frames by providing a snapshot of the reading frames of all proteins being translated in the cell (51). Secondly, Ribo-seq has the potential to give a more accurate expression profile for the purposes of neoantigen prioritization (discussed below). Given the novelty of the Ribo-seq approach, it does have the downside of being expensive and less readily available (51). Overall, further research is needed to fully explore the potential applications and advantages of Ribo-seq technology in neoantigen prioritization.

2.3 Variant Calling

While SNVs and small indels have been the only sources of neoantigens considered in most studies to date, a growing body of evidence demonstrates the need to consider a much broader set of neoantigens (52–54) (summarized in **Figure 3**). T cells specific to a single gene fusion mutation led to complete clinical response for a patient treated with immune checkpoint inhibition, even in the absence of any other immunogenic neoantigens (54). Large indels, particularly those that induce a frameshift, have a significantly enriched percentage of neoantigens predicted to bind with high affinity to MHC class I (52). Additionally, the number of indels with a frameshift was significantly associated with response to immune checkpoint inhibitors across three independent melanoma cohorts (52). Other targets of the immune response to cancer have been suggested, including peptides selectively expressed in tumors, foreign peptides in the case of viral-mediated cancers, and peptides derived from antigen presentation in the absence of transporter associated with antigen processing (TAP) (53, 55–57). However, this review focuses specifically on neoantigens derived from mutated peptides. This section will assess the literature on identifying SNVs, small indels, DNA-level structural variants (including gene fusions, large indels, and frameshifts) and RNA-level gene fusions.

Software for identifying SNVs continue to offer very disparate reports of the mutational profile, despite being the most common variant to be identified (22, 23, 58). For example, recent work across five patients compared the SNV and indel results from Strelka2, VarScan2, and GATK Mutect2 and demonstrated that an average of 84.41% (range 77.48-92.23%) of mutations were
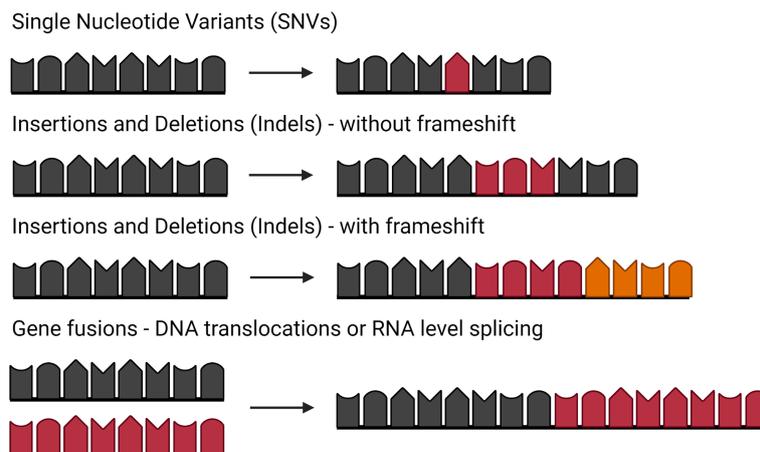
FIGURE 3 | Types of mutations that can lead to neoantigens. Single nucleotide variants (SNVs) caused by a point mutation in a single nucleic acid. Insertions and deletions (indels) caused by addition of nucleic acids or loss of nucleic acids. Indels with a frameshift occur when the number of nucleic acids is not a multiple of three, changing the reading frame. Gene fusions can be caused by either translocations at the DNA level or RNA splicing of independent transcripts. Figure created with BioRender.com.

identified by only one of the three callers, while 13.75% (range 7.21-22.17%) of mutations were identified by two of the three, and 1.83% (0.35-5.30%) by all three (23). Because of these disparities, selection of an SNV caller is a critical component of neoantigen prediction.

Despite a large number of confounding variables, a few software for identifying SNVs and small indels stand out across multiple benchmarking studies (**Tables 1**, **2**). Several confounding factors were shown to influence the performance of variant callers, including the type of validation sets employed (63), tumor purity (61), read depth of the sequencing data (59), and upstream features of the bioinformatic pipeline, such as read mapping software (62). Additionally, across the six studies summarized in **Table 1** (59–64), only nine of the 21 SNV software were tested in more than one study. Even with these confounding factors, GATK Mutect2 and Deep Variant were routinely rated as the top or second to top programs in terms of their sensitivity and specificity for detecting SNVs and small indels. When tested at different tumor purities, all variant callers demonstrated decreased performance with decreased tumor purity (61). However, TNscope and GATK Mutect2 maintained high performance for significantly lower tumor purities than the other variant callers. Deep Variant was not evaluated in this study. Overall, GATK Mutect2 and Deep Variant showed consistently high performance across multiple benchmarking studies, with GATK Mutect2 showing high performance even at lower tumor purities.

Consensus approaches have also been suggested, but highlight the need to balance sensitivity and specificity, especially for potential clinical applications. Wang et al. demonstrated that a majority voting approach with LoFreq, Mutect2, Strelka, and VarDict demonstrated an improved balance of precision (false discovery rate) and recall (true positive discovery rate) compared to any of the individual methods (63). Wang et al. further enhanced these results by giving increased voting power to Strelka and MuTect2 for

variants with low variant allele frequency, as these variant callers demonstrated stronger performance for low frequency variants (63). For indels, results were also improved with the majority voting approach, but showed even greater improvement if a greater number of software identified the indel, suggesting higher rates of false positives among indels (63). Bian et al. demonstrated improved results, as measured by the average of sensitivity and specificity, for SNVs using a majority voting approach between FreeBayes, VarDict, and Mutect compared to individual programs (60). FreeBayes, VarDict, and Mutect were selected because they could be run with an integrated Python package, but these programs were the three callers with the worst balance of sensitivity and specificity when run individually (60). Consensus approaches present a trade-off, as they often improve specificity while decreasing sensitivity. While increasing the specificity is important to avoid testing a large number of false positive variants, lowered sensitivity increases the risk of missing a clinically important variant. Therefore, an important area for future research is to compare different combination approaches and their influence on downstream neoantigen prioritization.

Structural variants, defined as genomic alterations encompassing at least 50 base pairs, can be identified well by a single, high-quality software, and do not demonstrate a benefit from a consensus approach (65). Structural variant types include large indels (with or without a frameshift) and gene fusions (66). Several software packages have been created for the identification of structural variants. GRIDSS and Manta perform consistently well across samples, as shown by Cameron et al. in a benchmarking study evaluating precision and recall (65). An advantage of Manta is that it works well with WES or WGS, whereas GRIDSS is only applicable to WGS (67). Cameron et al. also points out the risks of a combination approach with respect to structural variants: a simple union approach can drive up the false positives significantly, whereas conservative combinations, such as

TABLE 1 | Comparison of the ranking of single nucleotide variant (SNV) callers across six benchmarking studies that have been released since 2017.

Software	Supernat et al., 2018 (59)[1]	Bian et al., 2018 (60)	Pei et al., 2020 (61)	Kumaran et al., 2019 (62)	Wang et al., 2020 (63)[4]	Wang et al., 2020 (63)[5]	Hofmann et al., 2017 (64)
Deep Variant	#1			#1			
GATK MuTect2	#1	#1	#2[2]		#2	#3[7]	
SpeedSeq	#1						
TNscopeS[4]			#1[2]				
MuSE		#2			#1	#2	
Strelka					#1	#1	
LoFreq					#1	#2	
JointSNVMix2							#1
SAMtools				#2			#4
MuTect		#3			#2	#2	
DeepSNV							#2
NeuSomatic			#3[2]				
SomaticSniper					#3		#3
GATK							#3
UnifiedGenotyper							
GATK Halotype				#3			#4
Caller							
VarDict	#4				#3	#4[8]	
VarScan					#3		
FreeBayes	#4						
VarScan2			#4[3]				
Strelka2			#4[3]				
TNseq[6]			#4				

Benchmarking papers from before 2017 were excluded as they typically compared outdated software versions or compared software that are no longer maintained. Numbers and colors indicate the relative ranking based on the individual paper, with one (green) being the highest two (yellow), three (orange), and four (red) being the lowest.
[1]*These rankings are based on 30x data. In 15x data, the improved performance of DeepVariant was enhanced.*
[2]*At 20% purity.*
[3]*Good performance at high purity, but poor performance for low purity samples.*
[4]*Results based on DREAM WGS datasets as ground truth.*
[5]*Results based on WES and deep sequencing spike in studies.*
[6]*Software not free.*
[7]*High performance at low VAF, low performance at high VAF.*
[8]*High sensitivity, but with very high false positive rate.*

intersections of two software, can lead to extremely low sensitivity. No combination approach was able to consistently outperform the results from Manta or GRIDSS independently (65). Therefore, for structural variants, the current recommendation is to employ a single, highly rated caller such as Manta or GRIDSS.

A complementary RNAseq approach to detecting gene fusions allows for both confirmation of DNA-level structural variants and the identification of RNA-level splicing events. A 2019 benchmarking study recommended the use of STAR_Fusion, Arriba, or STAR-SEQR for the identification of

TABLE 2 | Comparison of the ranking of insertion and deletion (indel) callers across four benchmarking studies that have been released since 2017.

Software	Supernat et al., 2018 (59)	Pei et al., 2020[1] (61)	Kumaran et al., 2019 (62)	Wang et al., 2020 (63)
Deep Variant	#1		#1	
GATK MuTect2	#2	#1		#1
LoFreq				#1
Strelka				#1
TNscope		#2		
GATK Haplotype Caller			#2	
VarDict				#2
VarScan				#2
SpeedSeq	#3			
TNseq		#3		
VarScan2		#3		
Strelka2		#3		
NeuSomatic		#3		
SAMtools			#3	

Benchmarking papers from before 2017 were excluded as they typically compared outdated software versions or compared software that are no longer maintained. Numbers and colors indicate the relative ranking based on the individual paper, with one (green) being the highest, two (yellow), and three (orange) being the lowest.
[1]*40% purity.*

gene fusions from RNAseq, due to their combination of fast speed and high accuracy, as measured by the AUC (68). At this time, there have not been reported studies of benefits from combining gene fusion prediction results.

While frameshift mutations are traditionally accounted for using structural variant software, an alternative approach that may allow for identification of an expanded set of open reading frames is the use of Ribo-seq data. Ribo-seq identifies the triplet shifts of actively translating ribosomes, which allows the reading frame to be identified for all proteins being translated at the time of cell lysis (50). Ribo-seq has been proposed as a novel approach for detecting neoantigens derived from open reading frames by providing a snapshot of all active translation (51). An advantage to Ribo-seq data is that it may be able to identify novel open reading frames caused by translational dysregulation rather than by frameshifts. Further evaluation of the neoantigens identified by Ribo-seq compared to other sequencing technologies may clarify the implications of Ribo-seq technology to the clinical setting.

2.4 Variant Annotation

Annotating the effects of a variant on the resulting peptide sequence has high accuracy for SNVs and small indels, but accuracy drops for more complex variants such as splicing variants. Nucleotide mutations can have many potential impacts on the amino acid sequences including silent variants, variants in a non-coding region, missense mutations, frameshifts, and stop codon gain or loss. Each of these results in a significantly different set of neoantigen predictions, and therefore, variant annotation is essential to determine the neoantigen profile. Between the two most common variant annotation software, the Variant Effects Predictor (VEP) (69) and ANNOVAR (70), there was an 86.5% exact match rate overall, dropping to a low of 57.27% for splicing variants (71). Because of the difficulty in determining a "correct answer" for each variant, it is very difficult to benchmark the success of different programs. Nonetheless, based on a 2014 benchmarking study, VEP more consistently aligned with the best available, manually curated results (71). Since this benchmarking was performed before the most recent versions of either software, the results may be different with a repeated benchmarking analysis. The most recently released software, ShAn and Nirvana, have demonstrated an increase in speed and online accessibility compared to VEP with the same level of predictive abilities (72, 73). Therefore, the best available software by current recommendations is VEP for command-line applications and ShAn or Nirvana for online applications.

3 NEOANTIGEN PRIORITIZATION

3.1 MHC Class I-Restricted Neoantigen Characteristics

Once the neoantigens are predicted, each neoantigen can be prioritized for therapeutic use by predicting their potential to elicit a CD8+ T cell response. The experimentally validated

potential for MHC class I and II-restricted neoantigens to elicit a CD8+ or CD4+ T cell response, respectively, will be referred to as the "immunogenicity" of the neoantigen. One driving hypothesis in prioritization of immunogenic neoantigens is that the ability to predict the potential of the neoantigen to undergo each requisite step in the antigen presentation pathway will lead to improved prediction of the neoantigen immunogenicity. Tools have therefore been created to predict the expression of the neoantigen, the percentage of the tumor that contains the neoantigen of interest, the proteasomal cleavage potential, the potential for transport in the endoplasmic reticulum *via* TAP, the potential to bind the MHC class I molecule, the stability of the neoantigen:MHC class I interaction, and the potential to be recognized by a TCR (Summarized in **Figure 4**). Another body of work has focused on how to best summarize these individual tools into overall predictive models for CD8+ T cell response. Here, we will summarize the tools available for predicting each characteristic individually and then the different models available for integrating the characteristics into an overall score of the neoantigen immunogenicity.

3.1.1 Expression

A neoantigen needs to be expressed within the cell in order to elicit an immune response, but the best technology to assess the expression of the neoantigen is an ongoing question. Options for assessing expression can be broken down broadly into mRNA expression, protein level expression, or active translation. mRNA expression can be assessed through RNAseq, targeted sequencing, or microarray data. RNAseq data has the advantages that it is a readily available sequencing technique and can serve as a multi-purpose dataset, contributing to variant calling and neoantigen prioritization. One limitation in the use of mRNA expression data has been isolating only the expression of the specific allele in which the variant occurs. Identification of the specific variant allele is important as there are demonstrated cases where transcription of either the mutant allele or wildtype allele is favored (74). Novel methods of selecting the allele-specific expression have been published but have not yet been applied to the field of neoantigen prioritization (74). A second limitation to the RNAseq approach is that translational regulation may lead to discrepancies between the mRNA expression in the cell and the availability of the resulting peptide to be presented by MHC. Protein level expression can be assessed through various array-based methods, as well as mass spectrometry. One method growing in popularity is complete proteomic analysis, wherein a cell is lysed, and the full protein profile of the cell is analyzed with mass spectrometry (75). The advantage of a proteomics-based approach is that it validates the presence of the mutation on the protein level, eliminating variants that may never be translated within the cell. Some limitations to the use of mass spectrometry include low sensitivity for mutated peptides and high false positives in peptide identification algorithms (75). The rapid improvement in these methods may soon ameliorate these concerns (76), but a remaining limitation is that important neoantigens may come from translational products that are rapidly degraded and would

FIGURE 4 | Steps of MHC class I-restricted neoantigen prioritization and summary of characteristics considered for each step. Mutations in the DNA of a tumor cell are transcribed into RNA and translated into a protein. At the end of the life cycle of the protein, the protein is broken down into peptides by the proteasome and transported into the endoplasmic reticulum by the transporter associated with antigen presentation (TAP). Once inside the endoplasmic reticulum, the peptide has the opportunity to be loaded on MHC class I. If the peptide is successfully bound to MHC class I, the peptide:MHC complex is transported to the cell surface where the peptide:MHC complex has the opportunity to be recognized by the T cell receptor (TCR). Characteristics of the neoantigen encompassing expression, processing, MHC class I binding, and TCR recognition potential have been assessed to enhance prioritization of MHC class I-restricted neoantigens and are summarized in each of the boxes in the figure.

not be detected by proteomic techniques (57). Therefore, another option is the analysis of active translation occurring within a cell through the newer Ribo-seq technology. Ribo-seq data allows for quantification of all transcripts being actively translated at the time of cell lysis. Advantages of the use of Ribo-seq data are that it eliminates consideration of variants that are not translated by the cell, but also will detect translational products that are too rapidly degraded to be detected by traditional proteomic approaches (57). An important direction for future research is the comparison of RNA level, protein level, and translational level data on quantifying expression and their impact on neoantigen prioritization.

Following overall expression level, the next characteristic to consider in prioritizing immunogenicity is the percentage of the tumor that contains the variant of interest - also termed the clonality of the variant. Clonality is thought to be of particular importance for cancer therapeutics, since a variant expressed by a small, sub-clonal population of the tumor is a less attractive candidate for tumor therapy. There are a few possible ways by which to approach estimating the clonality of the variant. The ideal approach would be to use a clonal deconvolution software and then assign each neoantigen a value based on the percentage of the tumor that contains that neoantigen. Until recently, PyClone was the software most widely used (77). Recently, a newer model called FastClone was released, which demonstrated enhanced performance compared to PyClone (78). While clonal deconvolution is ideal, the programs do not always converge on a solution, especially depending on the purity and read depth of the samples. An alternative approach is to use the variant allele frequency (VAF) as a proxy for the clonality of a neoantigen, although the VAF does not account for the copy number

variation, germline tissue contamination, or sample purity. Overall, as estimating the clonality of a variant is a rapidly evolving field, it is likely that enhanced deconvolution methods will continue to develop and improve.

A unique system for applying the clonality has been put forward called the CSiN score, which is applied across both MHC class I- and II-restricted neoantigens (79). The CSiN score is calculated by first calculating the product of the variant allele frequency (VAF) of each somatic mutation and the number of neoantigens that can be generated from that mutation. The overall score for the tumor is then calculated by taking the average across all mutations, weighted by the binding affinity of the neoantigens. The CSiN score is associated with survival in response to immune checkpoint inhibitors, suggesting that clonality may play a significant role in determining the potential of MHC class I and II-restricted neoantigens to elicit immune-mediated tumor destruction (79).

3.1.2 Processing

One of the first steps in MHC class I-restricted antigen processing is proteasomal cleavage of proteins in the cytoplasm; incorporation of the enzyme specificity of the proteasome may lead to enhanced neoantigen prioritization. The first available model for predicting proteasomal C-terminal cleavage was NetChop (80), the method incorporated in the popular NetCTLpan model for predicting the processing and MHC binding of neoantigens. NetChop enhances the specificity of binding predictions (81). A newer model, the Proteasome Cleavage Prediction Server (PCPS), demonstrated enhanced sensitivity (0.89 vs. 0.79), but diminished specificity (0.55 vs. 0.60), compared to NetChop for discriminating known

CD8+ T cell epitopes from random peptides (82). While these results are not sufficient to recommend proteasomal cleavage as an independent metric for immunogenicity, they indicate that proteasomal C-terminal cleavage may play a role in determining the neoantigen profile.

Once small peptides are generated through proteasomal cleavage, TAP transports peptides into the endoplasmic reticulum for loading onto MHC class I; predicting the specificity of TAP for certain peptide motifs may enhance neoantigen prioritization. Prediction tools for TAP transport potential are less established. Currently, the only available program for predicting TAP specificity is that integrated into the NetCTL program (81). TAP transport potential was demonstrated by the NetCTL paper to enhance specificity for MHC class I binding predictions, but decreased sensitivity at lower specificity thresholds (81). Assessment of the association between TAP transport potential and neoantigen immunogenicity has not been directly assessed. Overall, TAP transport may prove to be a useful addition to other tools, but has not shown evidence of individual predictive value for neoantigen immunogenicity.

3.1.3 MHC Class I Binding

MHC class I binding affinity is one of the central neoantigen characteristics considered for prediction of neoantigen immunogenicity. Many studies have shown that MHC class I binding affinity alone has strong predictive ability for neoantigen immunogenicity (83–86). There is an abundance of models to predict MHC class I binding affinity that are summarized in **Table 3**. Binding affinity is defined as the inverse of the dissociation constant and models created to predict the binding affinity have been trained on either binding affinity alone, or binding affinity in combination with peptides eluted from MHC class I molecules and assessed by mass spectrometry. Since peptide elution does not give quantitative information regarding the binding affinity of the peptide, mass spectrometry data is included in these models as a categorical

TABLE 3 | Comparison of available neoantigen: MHC class I binding prediction tools.

Software	Model Type	Data Type	Published Comparisons	Performance metrics
NetMHCpan4.1 (86**)**	Artificial neural network	Mass spectrometry eluted peptides and binding affinity measurements	Outperformed MHCflurry1.2 and MixMHCpred, outperformed NetMHC4.0 for HLA-B and -C	Immunogenicity predictions
MHCflurry2.0 (87**)**	Artificial neural network	Mass spectrometry eluted peptides and binding affinity measurements	Outperformed NetMHCpan4.0 and MixMHCpred	Binding vs. non-binding predictions
MHCnuggets (88**)**	Artificial neural network	Binding affinity measurements	Comparable performance to MHCflurry1.2 and NetMHCpan3.0	Binding vs. non-binding predictions
NNAlign (85**)**	Artificial neural network	Mass spectrometry eluted peptides and binding affinity measurements	Comparable performance to NetMHCpan4.0, outperformed MHCflurry1.2 and MixMHCpred	Binding vs. non-binding predictions
ForestMHC (89**)**	Artificial neural network	Mass spectrometry eluted peptides	Outperformed original NetMHC, original NetMHCpan, and MixMHCpred	Binding vs. non-binding predictions
ACME (90**)**	Artificial neural network	Binding affinity measurements	Outperformed NetMHCpan4.0	Correlation with validated binding affinity
NetMHC4.0 (84**)**	Artificial neural network	Binding affinity measurements	None provided	None provided
MixMHCpred (91**)**	Matrix approach	Mass spectrometry eluted peptides	Outperformed NetMHC3.0 and NetMHCpan3.0	Binding vs. non-binding predictions
MSIntrinsic (92**)**	Artificial neural network	Mass spectrometry eluted peptides	Outperformed NetMHC4.0 and NetMHCpan2.8	Binding vs. non-binding predictions
ConvMHC (93**)**	Artificial neural network	Binding affinity measurements	Outperformed Pickpocket, IEDB SMM, and original NetMHCpan model	Binding vs. non-binding predictions
PAComplex (83**)**	Binding models	Database of known binding peptides	None provided	None provided
IEDB SMMPMBEC (94**)**	Matrix approach	Binding affinity measurements	Outperformed IEDB SMM, underperformed original NetMHC model	Binding vs. non-binding predictions
PickPocket1.1 (95**)**	Matrix approach	Binding affinity measurements	Underperformed original NetMHCpan model	Binding vs. non-binding predictions and correlation with validated binding affinity
IEDB recommended (96**)**	Positional scanning peptide libraries	Binding affinity measurements	Only compared to older models not included in this summary, performed better than 10/16 available methods	Binding vs. non-binding predictions
ARB (97**)**	Matrix approach	Binding affinity measurements	None provided	None provided
IEDB SMM (98**)**	Matrix approach	Binding affinity measurements	None provided	None provided
SYFPEITHI (99**)**	Binding motifs	Binding affinity measurements	None provided	None provided

Models included that met the following criteria 1) released since 2012 or included in a benchmarking study since 2012, 2) published in a peer reviewed journal, 3) available for web-based or command-line application, and 4) the most recent versions of a given software. Published comparisons are based on the comparisons reported in the publication of the new model. Performance metrics summarize whether the published comparisons were based on the ability of the model to predict immunogenicity, categorize each neoantigen as a binder vs. non-binder, or on the correlation between the predicted and experimentally validated binding affinity.

value that is integrated with the continuous binding affinity data. Many of the top performing models assess their performance with metrics such as the AUC, which measure the success of their ability to classify neoantigens as binders compared to non-binders. However, the top performing models based on AUC underperform when assessed with correlation coefficients between true and predicted binding affinities (100). As noted in **Table 3**, many models self-reported relative performance compared to other available models. In addition to the self-reported performance, three benchmarking studies have been published since 2012 which report the relative performance of the available tools. The first study found that no tool emerged as the best across all HLA alleles and all peptide lengths, but generally, artificial neural network tools outperformed those trained with other models (101). A second benchmarking study found that MHCflurry, NN_align, and NetMHCpan4.0 performed best for binding/non-binding classification. When tested specifically on mass spectrometry data, NetMHCpan4.0 and MixMHCpred show enhanced predictive power (100). Consistent with the first benchmarking study, all of these except MixMHCpred are artificial neural networks. The third benchmarking study assessed a large number of tools in terms of their ability to distinguish peptides that elicited a CD8+ T cell response. They found, similarly to the first two benchmarking studies, that NetMHCpan4.0 and MHCflurry outperformed other available models (102). Overall, neural network approaches including NetMHCpan4.0, MHCflurry, and NN_align consistently emerge as the top performing binding affinity models currently available.

A few studies have also suggested consensus approaches to the prediction of MHC class I binding, though none are currently optimized for application. For example, MHCcombine is a web application which runs 13 prediction algorithms and provides the outcome from each (100). Given that no model consistently outperformed across all peptide lengths and HLA alleles, MHCcombine may allow the user to apply the best result for the particular peptide length and HLA allele being tested. Additional research is needed on how to scale this approach for application to large lists of peptides and how these results would impact the overall performance. Another study averaged the results from early versions of NetMHCpan and NetMHC and showed a small performance enhancement (103). However, as these results predate many of the high-performing software summarized in **Table 3**, further research is needed to see how combined methods may impact performance.

Another characteristic of MHC class I binding that has been less studied is the binding stability, which is not directly assessed in any of the tools summarized above. While the binding affinity and binding stability are mathematically related, they may provide complementary information. Whereas the binding affinity, which is assessed by most available tools, is the inverse of the concentration at which 50% of the MHC class I molecules will be bound to the neoantigen, the binding stability is the half-life of the binding interaction. The binding affinity is the best metric for reactions in which the interaction of the two molecules is instantaneous. But, for the prediction of neoantigens, which

must stay bound until a circulating T cell is able to recognize them, the stability of the interaction may also be important. Therefore, tools predicting the binding affinity and binding stability have been proposed to be synergistic in predicting the potential for a neoantigen to be meaningfully presented on an MHC class I molecule. There is only one program that predicts MHC class I:peptide binding stability, NetMHCstabpan (104). As noted by the creators of NetMHCstabpan, the creation of a binding stability model was limited by the relative lack of training data for stability compared to binding affinity. Despite the limited training data, recent work has demonstrated enhanced neoantigen prioritization by combining both binding affinity and binding stability predictions (22, 23). Prediction of binding stability is an area where future work may lead to substantial improvements.

The hydrophobicity of a neoantigen is an additional characteristic with the potential to impact MHC binding and TCR recognition, but has demonstrated inconsistent predictive value for neoantigen immunogenicity. Since the binding cleft of the MHC class I molecule and the CD8+ TCR contact residues are both hydrophobic, one hypothesis is that a more hydrophobic neoantigen would be more likely to bind the MHC binding cleft and TCR (105). Two independent neural network approaches demonstrated a significant association of increased hydrophobicity with increased neoantigen immunogenicity (21, 105). In contrast, the TESLA consortium calculated a hydrophobicity fraction as the number of hydrophobic neoantigens divided by the length of the neoantigen and found a significantly higher hydrophobicity fraction among non-immunogenic neoantigens (22). When the hydrophobicity fraction was applied across four independent datasets, no consistent association of hydrophobicity with immunogenicity was observed (23). The differences in the observed associations of hydrophobicity with immunogenicity may be due to differences in the hydrophobicity of different HLA alleles. Published binding motifs for peptides known to bind different HLA alleles have demonstrated dramatic differences in the conserved amino acids. For example, HLA-A02:01 has several conserved hydrophobic amino acids, whereas HLA-A01:01 has predominantly polar and charged conserved amino acids (91). The neural network models from Chowell et al. and Zhou et al. were trained on known T cell epitopes from the immune epitope database (IEDB) (106), which has an HLA-A2 allelic bias since HLA-A2 is the most common class I allele, particularly in Caucasian populations (107). HLA-A2 also has more available experimental tools, which has expanded the bias towards this allele. A similar HLA-A2 allelic bias was observed in the TESLA dataset, with HLA-A2 alleles comprising 39.3% of the data, but there was also a high percentage of several alleles known to have conserved amino acid residues that are polar or charged, including HLA-A01:01 (22). Additional research is needed to fully understand the association of hydrophobicity with immunogenicity in the context of a diverse set of HLA alleles.

For all considerations of MHC class I binding, an understanding of the HLA alleles present in the tumor is critical. Predictions of dissociation constants and stability rely

on the specific HLA allele to which the neoantigen is binding. Additionally, as discussed above, there is evidence that the impact of hydrophobicity may be allele specific. Beyond the facilitation of binding and hydrophobicity predictions, changes in the HLA alleles such as mutations or loss of heterozygosity are a known mechanism of immune evasion in cancers (108). In addition, intact antigen processing machinery is required for presentation of the neoantigen and subsequent destruction by CD8+ T cells (109). Loss of functional components of the MHC class I antigen processing pathway including beta-2-microglobulin (110), TAP (111, 112), and tapasin (113, 114) have been implicated in immune-evasion or resistance to immunotherapy. Therefore, the HLA allelic profile of the tumor and the status of the antigen presentation pathways are critical to understanding which neoantigens can be presented to facilitate immune-mediated tumor destruction.

3.1.4 T Cell Receptor Recognition

Another characteristic of neoantigens that has been considered for impact on neoantigen immunogenicity is the TCR recognition potential. As T cells develop in the thymus, they are exposed to self peptides. T cells that recognize self peptides with high avidity undergo apoptosis. Therefore, T cell recognition has been broadly evaluated as either the similarity of the neoantigen to a normal human peptide or the similarity of neoantigens to known T cell epitopes.

The first method, similarity of the neoantigen to a normal human peptide has been shown to decrease the likelihood of the neoantigen eliciting an immune response. Increased sequence similarity was demonstrated to be highly associated with a decreased chance of eliciting an immune response across a large set of peptides known to elicit a T cell response from the IEDB (106). Sequence similarity alone was able to predict immunogenicity with an AUC of 0.85 (115). Importantly, these peptides derive from a variety of diseases including viruses, bacteria, and cancer neoantigens. In subsequent studies restricted to tumor neoantigens, the sequence similarity has not shown a significant association with neoantigen immunogenicity (23, 26). The observed differences may be explained by the much smaller range of sequence similarity available in the tumor neoantigens tested for immunogenicity. Since most tumor-derived neoantigens that have been tested for immunogenicity derive from SNVs (discussed below), they differ by a single amino acid. By contrast, peptides from viruses could be 100% distinct from normal human peptides. Further research is needed to determine if sequence similarity is more important in predicting neoantigen immunogenicity of tumor neoantigens when a broader set of neoantigens is considered.

Another method for accounting for TCR recognition is a model developed by Łuksza et al., which integrates three neoantigen characteristics into an aggregate fitness score for the tumor and demonstrated significant association of a lower fitness score with improved response to immune checkpoint inhibition. The overall fitness score is defined by the product of the T cell recognition probability, anchor residue hydrophobicity, amplitude, and a factor of negative one. A higher value for T cell recognition, amplitude, or

hydrophobicity all contribute to a lower fitness score (more negative value) and a neoantigen that is more likely to be visible to the immune system. The first characteristic, the T cell recognition potential, applies a probabilistic model for the binding of the neoantigen to the TCR by using the sequence similarity between the neoantigen and the closest matched T cell epitope from the IEDB (19, 106). The second characteristic accounts for the hydrophobicity of the neoantigen by giving the neoantigen a hydrophobicity of zero if an anchor residue is mutated from a hydrophobic residue to a hydrophilic residue, and all other changes are given a score of one. The third characteristic is called the "amplitude" and is intended to adjust for self-recognition. The amplitude is calculated as the ratio of the dissociation constant for the wildtype peptide and the neoantigen (19). The amplitude is higher for neoantigens that have a lower dissociation constant (higher binding affinity) and are derived from a wildtype peptide with a high dissociation constant. Neoantigens derived from a wildtype peptide with a high dissociation constant are predicted to be less likely to be subject to immune tolerance, since the wildtype peptide is less likely to be presented to developing T cells in the thymus. The integrated Łuksza model demonstrated a significant association of lower tumor fitness score with improved survival in patients treated with immunotherapy but was not assessed as a predictive measure for the immunogenicity of individual neoantigens (19).

Capietto et al. independently assessed the amplitude characteristic and suggested that the amplitude may be of greatest importance in predicting neoantigen immunogenicity for mutations in anchor residues (116). Capietto et al. found that the amplitude was a better predictor of immunogenicity for neoantigens with a mutation in the anchor residue than was the dissociation constant alone (116). These results suggest that the difference based on mutation position may be due to a greater change in T cell recognition when the mutation is in a non-anchor residue. However, the unadjusted binding affinity was significantly associated with immunogenicity in neoantigens with mutations in either anchor or non-anchor residues in this study and an independent study (23, 116). Further research will be needed to isolate the role of mutation position on immunogenicity predictions.

3.1.5 Integrated Models

Given the large number of neoantigen characteristics and tools to consider in prioritizing immunogenicity, several papers have focused on integrating neoantigen characteristics into an overall immunogenicity score. **Table 4** summarizes six recent models based on the characteristics they include and their reported performance as an AUC when provided. Of interest, the one commonality among all models is the inclusion of the binding affinity calculated by NetMHCpan (21–26). The consistent inclusion of MHC binding affinity across all available studies highlights the importance of MHC class I binding in determining the immunogenicity of at least a subset of neoantigens. Three models (TESLA, NeoScore and Neopepsee) focused specifically on reducing the characteristics included to only those most necessary for prioritizing neoantigens (22, 23, 25). TESLA and NeoScore were trained on the same training dataset and selected the same three

TABLE 4 | Summary of MHC class I-restricted neoantigen prioritization models.

	MuPeXI (24)	Neoepitope novelty (26)	Neopepsee (25)	pTuneos (21)	TESLA (22)	NeoScore (23)
Expression	RNA	–	–	RNA	RNA	RNA
Clonality	VAF[1]	–	–	PyClone	–	–
Cleavage	–	–	–	NetCTLpan	–	–
TAP[2]	–	–	–	NetCTLpan	–	–
Kd[3]	NetMHCpan	NetMHCpan	NetMHCpan	NetMHCpan	NetMHCpan	NetMHCpan
Stability	–	–	–	–	NetMHCstabpan	NetMHCstabpan
Hydrophobicity	–	–	Chowell et al.	Trained neural network	–	–
Polarity	–	–	Chowell et al.	–	–	–
T cell Recognition	–	–	Sequence similarity to known epitopes	Łuksza et al.	–	–
Sequence Similarity	Number of mismatches	BLOSUM62 matrix	–	BLOSUM62 matrix	–	–
Amplitude[4]	X	–	X	X	–	–
Viral Similarity	–	BLOSUM62 matrix	–	–	–	–
Amino Acid Contact Potentials	–	–	Saethang et al.	–	–	–
AUC[5] (if reported)	AUC = 0.635 in test set	AUC = 0.66 in training set	Not reported	AUC = 0.833 with 10-fold cross-validation	Cannot be calculated	AUC = 0.845 in test set

Where applicable, the tool used for each characteristic is specified. Dashes indicate that the characteristic was not included in the model and "X" indicates that the characteristic was included in a model, but that the characteristic is a fixed quantity with no specific tools to report.

[1]*VAF, variant allele frequency.*

[2]*TAP, transporter associated with antigen processing.*

[3]*Kd, dissociation constant.*

[4]*Amplitude, ratio of the dissociation constant of the wild type peptide and neoantigen.*

[5]*AUC, area under the receiver operator characteristics curve.*

characteristics, with the difference being that TESLA provides a series of thresholds across the three characteristics, while NeoScore provides a continuous score. The three selected characteristics were MHC class I binding affinity, MHC class I binding stability, and mRNA expression level (22, 23). In contrast, Neopepsee selected hydrophobicity, polarity, T cell recognition potential, amplitude, and the amino acid contact potentials (25). The striking difference in the selected characteristics may reflect a difference in the underlying training datasets. Neopepsee was trained on a set of known T cell epitopes from across diseases compared to common human variants presumed to not be immunogenic, whereas the other models were trained on tumor-specific neoantigens (22, 23, 25). The final Neopepsee score was demonstrated to be associated with immunogenicity in a test set derived exclusively from tumor mutations (25). Continued research is needed to select and validate the best set of characteristics to prioritize immunogenic neoantigens

While the models summarized above focus on predicting the immunogenicity of individual neoantigens, a few models trained specifically on the response to immune checkpoint inhibition. These models include the model from Łuksza et al. and the CSiN model which are both summarized in prior sections. The model from Łuksza et al. and the CSiN model demonstrate significant association with the response to immune checkpoint inhibition but were not tested for their potential to discriminate between individual neoantigens and their potential to elicit an immune response (19, 79). NeoScore and pTuneos have also demonstrated a significant association with response to immune checkpoint inhibition, despite being trained on the immune response to individual neoantigens (21, 23). Additional work is needed to understand the relative predictive value and clinical utility of each integrated model.

3.2 MHC Class II-Restricted Neoantigen Prioritization

The therapeutic applications of a neoantigen are also directly impacted by the potential of the neoantigen to bind to MHC class II and elicit a CD4+ T cell response, as CD4+ T cells have been demonstrated to play a critical role in initiating and maintaining a successful immune-mediated tumor destruction (6, 117). Prioritization of MHC class II-restricted neoantigens can incorporate many of the same characteristics as MHC class I: expression, processing, binding, and TCR recognition. Though the body of literature is smaller for prioritizing MHC class II-restricted neoantigens, tools are available to predict the expression of the neoantigen, the percentage of the tumor that contains the neoantigen of interest, the N/C-terminal cleavage potential, the potential to bind the MHC class II molecule, and the potential to be recognized by a CD4+ TCR (summarized in **Figure 5**). As for MHC class I-restricted neoantigens, there is also a body of work focused on integrating these tools into overall neoantigen immunogenicity scores. We will summarize the individual tools for each characteristic and the models available for integrating these characteristics into an overall score of the neoantigen immunogenicity.

3.2.1 Expression
Expression and clonality of MHC class II-restricted neoantigens can be calculated with the same tools as for MHC class I-restricted neoantigens.

3.2.2 Processing
Cleavage of peptides for MHC class II occurs in the endocytic pathway and is performed by cathepsins. The current

FIGURE 5 | Steps of MHC class II-restricted neoantigen prioritization and summary of characteristics considered for each step. Mutations in the DNA of a tumor cell are transcribed into RNA and translated into a protein. The protein can either be taken up into the endocytic compartment of an antigen presenting cell or processed and presented by the tumor cell if the tumor cell expresses MHC class II (not pictured). In the late endosomes, protein cleavage and MHC class II loading occurs. The protein is cleaved by cathepsins at the N- and C-termini before and after binding to the MHC class II molecule. If the peptide is successfully bound to MHC class II, the peptide:MHC complex is transported to the cell surface where the peptide: MHC complex has the opportunity to be recognized by the T cell receptor (TCR). Characteristics of the neoantigen encompassing expression, processing, MHC class II binding, and TCR recognition potential that may enhance prioritization of MHC class II-restricted neoantigens are summarized in each of the boxes in the figure. * indicates characteristics that, to our knowledge, have not been assessed for the prioritization of MHC class II-restricted neoantigens.

understanding of cleavage of peptides for MHC class II is that cleavage occurs both before and after binding of the peptide to the MHC class II molecule (118). Cleavage before and after binding is supported by binding of large proteins with exposed binding motifs in the absence of any protease activity (119) and dominant binding of accessible regions of proteins over high-affinity binders that are not solvent accessible (120). Abelin et al. applied the current understanding of MHC class II-restricted neoantigen processing to create models for predicting MHC class II-restricted neoantigens. Abelin et al. assessed the solvent accessibility of different regions of the protein at the pH of the late endosome to account for binding before processing and the N- and C-terminal motifs to account for enzyme specificity of the cathepsins (121). Abelin et al. demonstrated enhanced prioritization of neoantigens that bind MHC class II based on specific N- and C-terminal motifs, but did not find an impact of solvent accessibility (121). These results are in concordance with several other models which have demonstrated an ability to improve neoantigen prioritization by identifying specific motifs (122, 123). These studies combine to suggest the importance of considering N- and C-terminal motifs in the prioritization of MHC class II-restricted neoantigens.

3.2.3 MHC Class II Binding
There are many tools to predict the MHC class II binding affinity and a few that stand out as top candidates. A potentially helpful resource is the IEDB MHC II automated server benchmarks (124). The IEDB automated benchmarking system releases weekly scoring reports, ranking available MHC class II binding predictions based on the performance of the model in the most recently updated IEDB test datasets. While the IEDB automated

benchmarking system has the potential to be useful for research purposes, it is currently limited by only having six software registered, all of which were published by or before 2015. A benchmarking study of a set of older tools compared to two newer tools, NetMHCIIpan3.2 and DeepSeqPanII, demonstrated a distinct jump in performance between the older and newer tools (125). Because of the large gap in performance, only a set of the five newest models is included in **Table 5**. Based on the published data, NetMHCIIpan4.0 and DeepSeqPanII are likely the best performing models currently available (125). One newer method, NeonMHC, demonstrated enhanced positive predictive value compared to NetMHCIIpan3.1 (121). However, no direct comparison has been done of NeonMHC to updated versions of the other software, suggesting that further comparison studies between these techniques may be beneficial. While MHC class II binding affinity models have improved dramatically in the last few years, side-by-side comparisons of MHC class I and II binding affinity prediction models demonstrate that MHC class II binding affinity predictions still have lower performance than binding affinity predictions for MHC class I (86). This highlights the importance of continued research and model development for MHC class II binding affinity prediction.

There is evidence that a hydrophobicity-type approach to MHC class II binding may be worth exploring. To our knowledge a hydrophobicity model has not yet been attempted for MHC class II-restricted neoantigens. For HLA-DR, crystal structures have demonstrated that the binding cleft is hydrophobic (127). Additional, complementary evidence has demonstrated that there are two cooperative, hydrophobic binding pockets on HLA-DR which are thought to be primarily responsible for binding of MHC class II-restricted neoantigens (128). Similar

TABLE 5 | Comparison of available neoantigen: MHC class II binding prediction tools.

Software	Model Type	Data Type	Published Comparisons	Performance Metrics
DeepSeqPanII (125)	Artificial neural network	Binding affinity measurements	Comparable performance to NetMHCIIpan3.2, outperformed models from before 2012	Binding vs. non-binding predictions and correlation with validated binding affinity
NetMHCIIpan4.0 (86)	Artificial neural network	Mass spectrometry eluted peptides and binding affinity measurements	Outperformed NetMHCIIpan3.2, MixMHC2Pred, MHCnuggets, and DeepSeqPanII	Immunogenicity predictions
MHCnuggets (88)	Artificial neural network	Binding affinity measurements	Comparable performance to NetMHCIIpan3.2	Binding vs. non-binding predictions
MixMHCIIpred (126)	Matrix approach	Mass spectrometry eluted peptides	Outperformed NetMHCIIpan3.2	Binding vs. non-binding predictions
NeonMHC (121)	Artificial neural network	Mass spectrometry eluted peptides	Outperformed NetMHCpan3.1	Binding vs. non-binding predictions

Five of the newest methods summarized here due to recent benchmarking demonstrating that these methods highly outperformed earlier models. Only the most recent version of each software is included. Published comparisons are based on the comparisons reported in the publication of the new model. Performance metrics summarize whether the published comparisons were based on the ability of the model to predict immunogenicity, categorize each neoantigen as a binder vs. non-binder, or on the correlation between the predicted and experimentally validated binding affinity.

to MHC class I-restricted neoantigen prediction, the hypothesis is that, given the hydrophobicity of key binding pockets in the MHC class II binding groove and the TCR contact residues, increased neoantigen hydrophobicity may lead to increased immunogenicity. Overall, hydrophobicity is a characteristic of MHC class II-restricted neoantigens that will require additional research.

3.2.4 T Cell Receptor Recognition

Studies predicting the T cell recognition of MHC class II-restricted neoantigens have also been limited to date. Dhanda et al. trained neural networks using known T cell epitopes and demonstrated an AUC of 0.725 (129). This study suggests that there is potential for using known T cell epitopes to determine the probability of eliciting a T cell response, although more work is needed to enhance these predictions. An integrated model from Alspach et al. (discussed below) considered the amplitude characteristic and demonstrated that a neoantigen with a high amplitude was validated to be immunogenic (117). Additional work is needed to assess the impact of T cell recognition and immune tolerance on MHC class II-restricted neoantigen immunogenicity, whether measured by sequence similarity, amplitude, or a novel method.

3.2.5 Integrated Models

A few models have integrated multiple MHC class II-restricted neoantigen characteristics into an overall predictive model for MHC class II-restricted neoantigen immunogenicity. Three of the most recent models are summarized in **Table 6**. All three models have demonstrated particularly strong performance in predicting the presentation of neoantigens on MHC class II, with Abelin et al. demonstrating the strongest predictive value (AUC = 0.98) (121). Because of the limited data available for experimentally validated CD4+ T cell responses, all three of these models were built on predicting MHC class II presentation rather than neoantigen immunogenicity. The MARIA model was subsequently tested on two available datasets for MHC class II-restricted neoantigen immunogenicity and demonstrated significant association with T cell responses when split into a high, medium, and low immunogenicity score (130). The model by Abelin et al. was used to predict top candidates for immunogenicity and 8/12 tested neoantigens elicited a CD4+ T cell response, suggesting good predictive ability for immunogenic neoantigens (121). Finally, the Alspach et al. model (trained in mouse data) was integrated with amplitude and expression data, and a CD4+ T cell response was observed for the top predicted neoantigen candidate (117). Testing of these models on expanded sets of neoantigens validated to elicit

TABLE 6 | Summary of MHC class II-restricted neoantigen prioritization models.

	MARIA (130)	Abelin (121)	Alspach (117)
Expression	RNA	RNA	–
Processing	Neural network for N-/C-terminal motifs	Neural network for N-/C-terminal motifs	–
MHC class II Binding	Neural network for MHC class II binding scores	NeonMHC	Hidden Markov model
		Overlap with known HLA-DQ peptides	
Gene Bias	–	Weighted for genes over-represented on MHC class II	–
AUC[1] for Predicted MHC class II Presentation	AUC = 0.89	AUC = 0.98	AUC = 0.90

Where applicable, the tool used for each characteristic is specified. Dashes indicate that the characteristic was not included in the model.
[1]*AUC stands for area under the receiver operator characteristics curve.*

a CD4+ T cell response would be useful to further understand their performance capabilities and areas for improvement.

4 NEOANTIGEN VALIDATION

The development of prioritization models for MHC class I- and II-restricted neoantigens is reliant on the availability of datasets with validated CD8+ and CD4+ T cell responses, respectively. Generating a neoantigen validation dataset requires identification of mutations, prioritization of neoantigens to test, and testing of the neoantigens. A number of validation sets are available for MHC class I-restricted neoantigens (**Table 7**), but a far more limited selection of validation sets are available for MHC class II-restricted neoantigens (**Table 8**). The creation of a neoantigen validation set requires a number of choices regarding the mutations to be validated, the methods by which to select which neoantigens to test, and the experimental validation methods employed. This section will summarize the standard methods used for generating validation datasets to date and highlight potential areas for further research.

Neoantigen validation sets have traditionally focused on SNV and small indel-derived neoantigens, though expansion to a larger set of mutations may be an important future direction in the field. As demonstrated in **Tables 7**, **8**, all available datasets have validated SNVs and small indels (5–7, 22, 131–138). The abundance of data has allowed for the creation and testing of many models for neoantigen prioritization. However, expanding the mutations tested has the potential to illuminate if there are different characteristics that are important for neoantigens

derived from a broader set of mutations. One characteristic that may be particularly impacted by expanded sets of mutations is the sequence similarity. SNVs change only a single amino acid in a protein, leaving most of the neoantigen unaltered. While indels may have slightly greater changes, these represent a minority of neoantigens validated to date. By contrast, neoantigens from novel open reading frames, gene fusions, or large indels may have over 50% of the neoantigen changed compared to the corresponding wildtype peptide. Given recent evidence of the increased immunogenicity of large indels compared to SNVs (52), the inclusion of these neoantigens may enhance the importance of the sequence similarity, which has a small range when considering only SNVs and small indels. Additionally, inclusion of an expanded set of mutations may enhance the clinical applications of available neoantigen prioritization models. Currently available MHC class I and II models are trained on mutations derived from SNVs and indels and are not trained on other mutations such as frameshifts and gene fusions (21–26, 117, 121, 130). Therefore, expanding validation sets would pave the way to allow these models to expand the neoantigens considered. A recent report demonstrated that a single gene fusion neoantigen was able to drive complete disease response in a patient (54), which further underscores the importance of considering additional sources of mutations beyond SNVs and indels as candidates for personalized cancer vaccines.

Expanding the subsets of neoantigens tested may also contribute to enhanced models for MHC class I- and II-restricted neoantigen prioritization. In order to select a reasonable number of neoantigens for validation, candidates are typically prioritized by one or more neoantigen characteristics

TABLE 7 | Available sets of MHC class I-restricted neoantigens validated to elicit a CD8+ T cell response.

Author and Year	Tumor Type	Tested Neoantigens: Immunogenic Neoantigens	Available Sequencing Data	Mutations Tested	Prioritization Method	Validation Method
Robbins et al., 2013 (131)	Melanoma	227:10	WES	SNVs and small indels	NetMHCpan2.4	ELISPOT
Wick et al., 2014 (132)	Ovarian cancer	114:1	WES	SNVs	NetMHCpan2.4	ELISPOT
Rajasagi et al., 2014 (133)	Chronic lymphocytic leukemia	48:3	WES	SNVs	NetMHCpan2.4	ELISPOT
Cohen et al., 2015 (134)	Melanoma	357:9	WES, RNAseq	SNVs	Expression >1 FPKM and MHC binding by IEDB	ELISA
Carreno et al., 2015 (5)	Melanoma	21:11	WES, RNAseq	SNVs	NetMHC2.4	ELISA
McGranahan et al., 2016 (135)	Lung cancer	355:2	WES	SNVs	NetMHCpan2.8	Multimers
Strønen et al., 2016 (136)	Melanoma	57:11	WES, RNAseq	SNVs	Expression >0 FPKM and NetMHC3.2, NetMHCpan2.0	Multimers
Bentzen et al., 2016 (137)	Lung cancer	702:9	WES	SNVs	NetMHCpan2.8	Multimers
Gros et al., 2016 (138)	Melanoma	27:6	WES, RNAseq	SNVs	VAF>10%, mutation in DNA and RNA	ELISPOT
Ott et al., 2017 (6)	Melanoma	165:18	WES, RNAseq	SNVs and small indels	NetMHCpan2.4 and oncogene mutations	ELISPOT
Wells et al., 2020 (22)	Melanoma and Lung cancer	347:27 (available)	WES, RNAseq	SNVs and small indels	Consensus from 25 groups	Multimers

Datasets were only included if they included a minimum of ten neoantigens.

TABLE 8 | Available sets of MHC class II-restricted neoantigens validated to elicit a CD4+ T cell response.

Author and Year	Tumor Type	Neoantigens Tested : Immuno-genic Neoantigens	Available Sequencing Data	Mutations Tested	Prioritization Method	Validation Method
Ott et al., 2017 (6)	Melanoma	165:80	WES, RNAseq	SNVs and small indels	NetMHCpan2.4 and oncogene mutations	ELISPOT
Sahin et al., 2017 (7)	Melanoma	125:60	WES, RNAseq	SNVs	RNA expression > 10 RPKM and IEDB binding predictions	ELISPOT

Datasets were only included if they included a minimum of ten neoantigens.

before validation. As demonstrated in **Tables 7**, **8**, neoantigens are nearly universally prioritized by MHC binding, and in the majority of cases, by NetMHCpan predicted binding. Given the low prevalence of immunogenic neoantigens, pre-filtering is important to ensure that some immunogenic neoantigens are identified. However, the pre-filtering of neoantigens does represent a bias in the selection of optimal neoantigen characteristics. The use of the binding as a criterion is limited by the predictive power of the MHC class I binding prediction tool employed. Furthermore, work in mice has demonstrated that neoantigens with a dissociation constant experimentally validated to be orders of magnitude above 500 nM (the typical cutoff used) successfully elicited a CD8+ T cell response (139). The observation that neoantigens with low binding affinity (high dissociation constants) can elicit a CD8+ T cell response suggests that there may be additional characteristics at play in determining the neoantigen immunogenicity. However, building neoantigen prioritization models on existing datasets cannot assess these other characteristics as effectively since none of the tested neoantigens have low predicted binding affinity. While model building with the same validation datasets may enhance our ability to prioritize the neoantigens that are known to be candidates, it also has the potential to bias the field away from classes of neoantigens that have not been explored in as great of depth.

Validation of immunogenic neoantigens can be done in multiple ways, which all provide slightly different and complementary information. In this review, we focus on methods that involve the direct challenge of a T cell with a neoantigen. Other methods such as TCR profiling are available and have been recently reviewed (140). The standard validation techniques employed are mass spectrometry, tetramer/multimer staining, and ELISpot, ELISA, or intracellular cytokine staining, which are illustrated in **Figure 6**. These methods measure three different features of the neoantigen, and therefore, provide information about different aspects of neoantigen binding and immunogenicity. Mass spectrometry has been employed to directly profile the neoantigens presented on MHC class I and II by eluting bound peptides and identifying them using tumor-specific variant libraries (121). Mass spectrometry of eluted peptides validates MHC class I or II presentation, but must be combined with one of the other techniques to provide T cell recognition data. MHC multimers (sets of multiple MHC molecules complexed to a neoantigen of interest) bind the TCR and can be fluorescently labeled and used to stain T cells that recognize the neoantigen, a process called "multimer staining." Multimer staining directly measures the presence of neoantigen-

specific T cells that have expanded populations after activation. One feature of multimer staining that is important to keep in mind is that smaller multimers, such as tetramers, have a tendency not to stain low affinity T cells (141, 142). Given that the affinity of T cells responsive to cancer has been shown to be much lower than anti-viral neoantigens (143), using advanced methodologies for increasing sensitivity might be particularly useful in the study of tumor-specific neoantigens. The final group of techniques, ELISpot, ELISA, and intracellular cytokine staining, all test for cytokine production after stimulation of the TCR, a sign of T cell activation. One potential limitation of techniques that measure cytokine production is that these methods can give false negatives if a neoantigen-specific T cell becomes exhausted. Overall, each of the techniques provides valuable information regarding the binding or immunogenicity of the neoantigen. Where possible, combining two or more techniques may provide the best confirmation of immunogenicity.
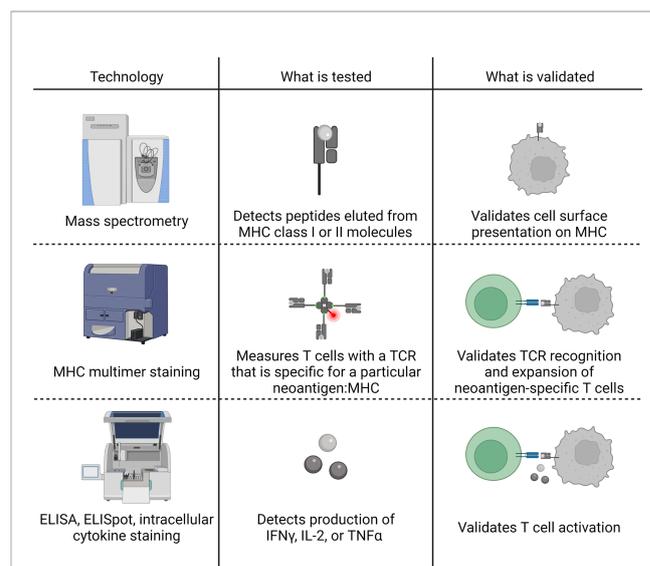
FIGURE 6 | Summary of three commonly applied validation techniques for the immunogenicity of MHC class I or II-restricted neoantigens. Mass spectrometry is performed by eluting peptides directly from tumor cells and validates the *in vivo* presentation of the neoantigen on the cell surface. MHC multimers (most commonly a tetramer) bind T cell receptors (TCR) specific for the particular neoantigen: MHC, validating TCR recognition of the neoantigen and expansion of neoantigen-specific T cells. ELISA, ELISpot, and intracellular cytokine staining detect the production of cytokines, typically interferon-gamma (IFNγ), interleukin-2 (IL-2), or tumor necrosis factor alpha (TNFα), to validate T cell activation. Figure created with BioRender.com.

Another consideration, both in the generation and application of neoantigen validation sets, is the differences in the neoantigen characteristics that can be validated in vaccine datasets. A common source of validated neoantigens is from vaccination studies (5–7, 136). While testing after vaccination has the advantage of demonstrating whether a given neoantigen has the potential to elicit a T cell response, it limits the validation of key, tumor-level characteristics such as expression. As recently demonstrated, models that incorporate expression underperform on these datasets (23). The underperformance of models incorporating expression is likely because the presence of a T cell response in a vaccinated patient is not necessarily due to the T cells encountering that neoantigen within the tumor. Rather, the T cells could have been activated by the vaccine, even if the tumor did not express the neoantigen of interest. Therefore, vaccination with neoantigens prior to testing presents an important consideration, both in the creation of neoantigen validation sets and in their application to validating the impact of various neoantigen characteristics on neoantigen immunogenicity.

Overall, a significant body of work has been done particularly for MHC class I-restricted neoantigen validation. Moving forward, there are several key areas that may enhance the development of clinically useful prioritization models. Specifically, these areas include 1) expansion of validation sets for MHC class II-restricted neoantigens, 2) expansion of the types of mutations considered in neoantigen validation, and 3) careful selection of which neoantigens to test for immunogenicity. Further research in these areas has the potential to build on the work already done to advance the utility of neoantigen prioritization models.

5 CONCLUSION

The field of neoantigen prediction and prioritization for cancer therapeutics has made tremendous strides and is still rapidly expanding. Prioritization of immunogenic neoantigens can be largely broken down into data acquisition and variant calling, neoantigen prioritization, and neoantigen validation. High quality sequencing data is becoming ever more accessible, and techniques for artifact removal in FFPE data and tumor-only variant calling are rapidly expanding, increasing what is feasible in each of these areas. One of the central questions in variant calling is how to find the appropriate balance between sensitivity

and specificity for the clinical applications of neoantigens. While using consensus approaches between several variant calling software has the potential to enhance specificity, it may do so at the expense of missing clinically important variants. Within neoantigen prioritization, a wide range of high-performance tools are available for prioritizing MHC class I- and II-restricted neoantigens. However, MHC class II tools generally have not been assessed to the same degree as those for MHC class I, representing a key area for future research. Other key areas for enhancing neoantigen prioritization models include 1) training models directly on predicting the potential of an MHC class II-restricted neoantigen to elicit a CD4+ T cell response and 2) expanding models to include neoantigens derived from other sources of mutations. Advances in these areas will rely on the expansion of available neoantigen validation datasets with a specific focus on MHC class II-restricted neoantigens and neoantigens derived from large indels, gene fusions, or frameshifts. Overall, a combination of expanding datasets and continued improvement of computational modelling will build on past successes to create more clinically relevant models moving forward.

AUTHOR CONTRIBUTIONS

Writing – original draft preparation, EB. Writing – review and editing, KH, MW, KB, and EB. Visualization, EB. Supervision, KH, MW, and KB. Project administration, KH. Funding acquisition, KH and EB. All authors have read and agreed to the published version of the manuscript.

FUNDING

REFERENCES

1. Tran E, Turcotte S, Gros A, Robbins PF, Lu YC, Dudley ME, et al. Cancer Immunotherapy Based on Mutation-Specific CD4+ T Cells in a Patient With Epithelial Cancer. *Science* (2014) 344(6184):641–5. doi: 10.1126/science.1251102

2. Schumacher TN, Schreiber RD. Neoantigens in Cancer Immunotherapy. *Science* (2015) 348(6230):69–74. doi: 10.1126/science.aaa4971

3. Ward JP, Gubin MM, Schreiber RD. The Role of Neoantigens in Naturally Occurring and Therapeutically Induced Immune Responses to Cancer. *Adv Immunol* (2016) 130:25–74. doi: 10.1016/bs.ai.2016.01.001

4. Yarchoan M, Hopkins A, Jaffee EM. Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *N Engl J Med* (2017) 377(25):2500–1. doi: 10.1056/nejmc1713444

5. Carreno BM, Magrini V, Becker-Hapak M, Kaabinejadian S, Hundal J, Petti AA, et al. Cancer Immunotherapy. A Dendritic Cell Vaccine Increases the Breadth and Diversity of Melanoma Neoantigen-Specific T Cells. *Science* (2015) 348(6236):803–8. doi: 10.1126/science.aaa3828

6. 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(7662):217–21. doi: 10.1038/nature22991

7. Sahin U, Derhovanessian E, Miller M, Kloke BP, Simon P, Lower M, et al. Personalized RNA Mutanome Vaccines Mobilize Poly-Specific Therapeutic Immunity Against Cancer. *Nature* (2017) 547(7662):222–6. doi: 10.1038/nature23003

8. Hilf N, Kuttruff-Coqui S, Frenzel K, Bukur V, Stevanovic S, Gouttefangeas C, et al. Actively Personalized Vaccination Trial for Newly Diagnosed Glioblastoma. *Nature* (2019) 565(7738):240–5. doi: 10.1038/s41586-018-0810-y

9. Keskin DB, Anandappa AJ, Sun J, Tirosh I, Mathewson ND, Li S, et al. Neoantigen Vaccine Generates Intratumoral T Cell Responses in Phase Ib Glioblastoma Trial. *Nature* (2019) 565(7738):234–9. doi: 10.1038/s41586-018-0792-9

10. Parkhurst MR, Yang JC, Langan RC, Dudley ME, Nathan DA, Feldman SA, et al. T Cells Targeting Carcinoembryonic Antigen Can Mediate Regression of Metastatic Colorectal Cancer But Induce Severe Transient Colitis. *Mol Ther* (2011) 19(3):620–6. doi: 10.1038/mt.2010.272

11. Morgan RA, Chinnasamy N, Abate-Daga D, Gros A, Robbins PF, Zheng Z, et al. Cancer Regression and Neurological Toxicity Following Anti-MAGE-A3 TCR Gene Therapy. *J Immunother* (2013) 36(2):133–51. doi: 10.1097/CJI.0b013e3182829903

12. Stevanovic S, Draper LM, Langhan MM, Campbell TE, Kwong ML, Wunderlich JR, et al. Complete Regression of Metastatic Cervical Cancer After Treatment With Human Papillomavirus-Targeted Tumor-Infiltrating T Cells. *J Clin Oncol* (2015) 33(14):1543–50. doi: 10.1200/JCO.2014.58.9093

13. Zacharakis N, Chinnasamy H, Black M, Xu H, Lu YC, Zheng Z, et al. Immune Recognition of Somatic Mutations Leading to Complete Durable Regression in Metastatic Breast Cancer. *Nat Med* (2018) 24(6):724–30. doi: 10.1038/s41591-018-0040-8

14. Bianchi V, Harari A, Coukos G. Neoantigen-Specific Adoptive Cell Therapies for Cancer: Making T-Cell Products More Personal. *Front Immunol* (2020) 11:1215. doi: 10.3389/fimmu.2020.01215

15. Wang Z, Cao YJ. Adoptive Cell Therapy Targeting Neoantigens: A Frontier for Cancer Research. *Front Immunol* (2020) 11:176. doi: 10.3389/fimmu.2020.00176

16. Gubin MM, Zhang X, Schuster H, Caron E, Ward JP, Noguchi T, et al. Checkpoint Blockade Cancer Immunotherapy Targets Tumour-Specific Mutant Antigens. *Nature* (2014) 515(7528):577–81. doi: 10.1038/nature13988

17. Lommatzsch M, Bratke K, Stoll P. Neoadjuvant PD-1 Blockade in Resectable Lung Cancer. *N Engl J Med* (2018) 379(9):e14. doi: 10.1056/NEJMc1808251

18. Rausch MP, Hastings KT. "Immune Checkpoint Inhibitors in the Treatment of Melanoma: From Basic Science to Clinical Application". In: WH Ward and JM Farma, editors. *Cutaneous Melanoma: Etiology and Therapy*. Brisbane (AU): Codon Publications (2017).

19. Łuksza M, Riaz N, Makarov V, Balachandran VP, Hellmann MD, Solovyov A, et al. A Neoantigen Fitness Model Predicts Tumour Response to Checkpoint Blockade Immunotherapy. *Nature* (2017) 551(7681):517–20. doi: 10.1038/nature24473

20. Liu D, Schilling B, Liu D, Sucker A, Livingstone E, Jerby-Arnon L, et al. Integrative Molecular and Clinical Modeling of Clinical Outcomes to PD1 Blockade in Patients With Metastatic Melanoma. *Nat Med* (2019) 25 (12):1916–27. doi: 10.1038/s41591-019-0654-5

21. Zhou C, Wei Z, Zhang Z, Zhang B, Zhu C, Chen K, et al. Ptuneos: Prioritizing Tumor Neoantigens From Next-Generation Sequencing Data. *Genome Med* (2019) 11(1):67. doi: 10.1186/s13073-019-0679-x

22. 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(3):818–34. doi: 10.1016/j.cell.2020.09.015

23. Borden ES, Ghafoor S, Buetow KH, LaFleur BJ, Wilson MA, Hastings KT. NeoScore Integrates Characteristics of the Neoantigen:MHC Class I Interaction and Expression to Accurately Prioritize Immunogenic Neoantigens. *J Immunol Accepted* (2022). doi: 10.4049/jimmunol.2100700

24. Bjerregaard AM, Nielsen M, Hadrup SR, Szallasi Z, Eklund AC. MuPeXI: Prediction of Neo-Epitopes From Tumor Sequencing Data. *Cancer Immunol Immunother* (2017) 66(9):1123–30. doi: 10.1007/s00262-017-2001-3

25. Kim S, Kim HS, Kim E, Lee MG, Shin EC, Paik S, et al. Neopepsee: Accurate Genome-Level Prediction of Neoantigens by Harnessing Sequence and Amino Acid Immunogenicity Information. *Ann Oncol* (2018) 29(4):1030–6. doi: 10.1093/annonc/mdy022

26. Wood MA, Paralkar M, Paralkar MP, Nguyen A, Struck AJ, Ellrott K, et al. Population-Level Distribution and Putative Immunogenicity of Cancer Neoepitopes. *BMC Cancer* (2018) 18(1):414. doi: 10.1186/s12885-018-4325-6

27. Hayes SA, Clarke S, Pavlakis N, Howell VM. The Role of Proteomics in the Age of Immunotherapies. *Mamm Genome* (2018) 29(11-12):757–69. doi: 10.1007/s00335-018-9763-6

28. Koboldt DC. Best Practices for Variant Calling in Clinical Sequencing. *Genome Med* (2020) 12(1):91. doi: 10.1186/s13073-020-00791-w

29. Halperin RF, Carpten JD, Manojlovic Z, Aldrich J, Keats J, Byron S, et al. A Method to Reduce Ancestry Related Germline False Positives in Tumor Only Somatic Variant Calling. *BMC Med Genomics* (2017) 10(1):61. doi: 10.1186/s12920-017-0296-8

30. Halperin RF, Liang WS, Kulkarni S, Tassone EE, Adkins J, Enriquez D, et al. Leveraging Spatial Variation in Tumor Purity for Improved Somatic Variant Calling of Archival Tumor Only Samples. *Front Oncol* (2019) 9:119. doi: 10.3389/fonc.2019.00119

31. Little P, Jo H, Hoyle A, Mazul A, Zhao X, Salazar AH, et al. UNMASC: Tumor-Only Variant Calling With Unmatched Normal Controls. *NAR Cancer* (2021) 3(4):zcab040. doi: 10.1093/narcan/zcab040

32. Trost B, Walker S, Haider SA, Sung WWL, Pereira S, Phillips CL, et al. Impact of DNA Source on Genetic Variant Detection From Human Whole-Genome Sequencing Data. *J Med Genet* (2019) 56(12):809–17. doi: 10.1136/jmedgenet-2019-106281

33. Samson CA, Whitford W, Snell RG, Jacobsen JC, Lehnert K. Contaminating DNA in Human Saliva Alters the Detection of Variants From Whole Genome Sequencing. *Sci Rep* (2020) 10(1):19255. doi: 10.1038/s41598-020-76022-4

34. Kidd JM, Sharpton TJ, Bobo D, Norman PJ, Martin AR, Carpenter ML, et al. Exome Capture From Saliva Produces High Quality Genomic and Metagenomic Data. *BMC Genomics* (2014) 15:262. doi: 10.1186/1471-2164-15-262

35. Griffith M, Miller CA, Griffith OL, Krysiak K, Skidmore ZL, Ramu A, et al. Optimizing Cancer Genome Sequencing and Analysis. *Cell Syst* (2015) 1 (3):210–23. doi: 10.1016/j.cels.2015.08.015

36. Taylor-Weiner A, Stewart C, Giordano T, Miller M, Rosenberg M, Macbeth A, et al. DeTiN: Overcoming Tumor-in-Normal Contamination. *Nat Methods* (2018) 15(7):531–4. doi: 10.1038/s41592-018-0036-9

37. Moore L, Cagan A, Coorens THH, Neville MDC, Sanghvi R, Sanders MA, et al. The Mutational Landscape of Human Somatic and Germline Cells. *Nature* (2021) 597(7876):381–6. doi: 10.1038/s41586-021-03822-7

38. Wei L, Christensen SR, Fitzgerald ME, Graham J, Hutson ND, Zhang C, et al. Ultradeep Sequencing Differentiates Patterns of Skin Clonal Mutations Associated With Sun-Exposure Status and Skin Cancer Burden. *Sci Adv* (2021) 7(1):eabd7703. doi: 10.1126/sciadv.abd7703

39. Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall MWJ, et al. Somatic Mutant Clones Colonize the Human Esophagus With Age. *Science* (2018) 362(6417):911–7. doi: 10.1126/science.aau3879

40. Brunner SF, Roberts ND, Wylie LA, Moore L, Aitken SJ, Davies SE, et al. Somatic Mutations and Clonal Dynamics in Healthy and Cirrhotic Human Liver. *Nature* (2019) 574(7779):538–42. doi: 10.1038/s41586-019-1670-9

41. Oh JH, Sung CO. Comprehensive Characteristics of Somatic Mutations in the Normal Tissues of Patients With Cancer and Existence of Somatic Mutant Clones Linked to Cancer Development. *J Med Genet* (2021) 58 (7):433–41. doi: 10.1136/jmedgenet-2020-106905

42. Bhagwate AV, Liu Y, Winham SJ, McDonough SJ, Stallings-Mann ML, Heinzen EP, et al. Bioinformatics and DNA-Extraction Strategies to Reliably Detect Genetic Variants From FFPE Breast Tissue Samples. *BMC Genomics* (2019) 20(1):689. doi: 10.1186/s12864-019-6056-8

43. de Schaetzen van Brienen L, Larmuseau M, van der Eecken K, De Ryck F, Robbe P, Schuh A, et al. Comparative Analysis of Somatic Variant Calling on Matched FF and FFPE WGS Samples. *BMC Med Genomics* (2020) 13(1):94. doi: 10.1186/s12920-020-00746-5

44. Tellaetxe-Abete M, Calvo B, Lawrie C. Ideafix: A Decision Tree-Based Method for the Refinement of Variants in FFPE DNA Sequencing Data. *NAR Genom Bioinform* (2021) 3(4):lqab092. doi: 10.1093/nargab/lqab092

45. Kim H, Lee AJ, Lee J, Chun H, Ju YS, Hong D. FIREVAT: Finding Reliable Variants Without Artifacts in Human Cancer Samples Using Etiologically Relevant Mutational Signatures. *Genome Med* (2019) 11(1):81. doi: 10.1186/s13073-019-0695-x

46. Diossy M, Sztupinszki Z, Krzystanek M, Borcsok J, Eklund AC, Csabai I, et al. Strand Orientation Bias Detector to Determine the Probability of FFPE Sequencing Artifacts. *Brief Bioinform* (2021) 22(6):bbab186. doi: 10.1093/bib/bbab186

47. Jegathisawaran J, Tsiplova K, Hayeems R, Ungar WJ. Determining Accurate Costs for Genomic Sequencing Technologies-A Necessary Prerequisite. *J Community Genet* (2020) 11(2):235–8. doi: 10.1007/s12687-019-00442-7

48. O'Brien TD, Jia P, Xia J, Saxena U, Jin H, Vuong H, et al. Inconsistency and Features of Single Nucleotide Variants Detected in Whole Exome Sequencing Versus Transcriptome Sequencing: A Case Study in Lung Cancer. *Methods* (2015) 83:118–27. doi: 10.1016/j.ymeth.2015.04.016

49. Coudray A, Battenhouse AM, Bucher P, Iyer VR. Detection and Benchmarking of Somatic Mutations in Cancer Genomes Using RNA-Seq Data. *PeerJ* (2018) 6:e5362. doi: 10.7717/peerj.5362

50. Erhard F, Halenius A, Zimmermann C, L'Hernault A, Kowalewski DJ, Weekes MP, et al. Improved Ribo-Seq Enables Identification of Cryptic Translation Events. *Nat Methods* (2018) 15(5):363–6. doi: 10.1038/nmeth.4631

51. Dersh D, Holly J, Yewdell JW. A Few Good Peptides: MHC Class I-Based Cancer Immunosurveillance and Immunoevasion. *Nat Rev Immunol* (2021) 21(2):116–28. doi: 10.1038/s41577-020-0390-6

52. Turajlic S, Litchfield K, Xu H, Rosenthal R, McGranahan N, Reading JL, et al. Insertion-And-Deletion-Derived Tumour-Specific Neoantigens and the Immunogenic Phenotype: A Pan-Cancer Analysis. *Lancet Oncol* (2017) 18 (8):1009–21. doi: 10.1016/S1470-2045(17)30516-8

53. Laumont CM, Vincent K, Hesnard L, Audemard E, Bonneil E, Laverdure JP, et al. Noncoding Regions Are the Main Source of Targetable Tumor-Specific Antigens. *Sci Transl Med* (2018) 10(470):eaau5516. doi: 10.1126/scitranslmed.aau5516

54. Yang W, Lee KW, Srivastava RM, Kuo F, Krishna C, Chowell D, et al. Immunogenic Neoantigens Derived From Gene Fusions Stimulate T Cell Responses. *Nat Med* (2019) 25(5):767–75. doi: 10.1038/s41591-019-0434-2

55. Marijt KA, van Hall T. To TAP or Not to TAP: Alternative Peptides for Immunotherapy of Cancer. *Curr Opin Immunol* (2020) 64:15–9. doi: 10.1016/j.coi.2019.12.004

56. Zhao Q, Laverdure JP, Lanoix J, Durette C, Cote C, Bonneil E, et al. Proteogenomics Uncovers a Vast Repertoire of Shared Tumor-Specific Antigens in Ovarian Cancer. *Cancer Immunol Res* (2020) 8(4):544–55. doi: 10.1158/2326-6066.CIR-19-0541

57. Ruiz Cuevas MV, Hardy MP, Holly J, Bonneil E, Durette C, Courcelles M, et al. Most Non-Canonical Proteins Uniquely Populate the Proteome or Immunopeptidome. *Cell Rep* (2021) 34(10):108815. doi: 10.1016/j.celrep.2021.108815

58. Kroigard AB, Thomassen M, Laenkholm AV, 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(3):e0151664. doi: 10.1371/journal.pone.0151664

59. Supernat A, Vidarsson OV, Steen VM, Stokowy T. Comparison of Three Variant Callers for Human Whole Genome Sequencing. *Sci Rep* (2018) 8 (1):17851. doi: 10.1038/s41598-018-36177-7

60. 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(1):429. doi: 10.1186/s12859-018-2440-7

61. Pei S, Liu T, Ren X, Li W, Chen C, Xie Z. Benchmarking Variant Callers in Next-Generation and Third-Generation Sequencing Analysis. *Brief Bioinform* (2021) 22(3):bbaa148. doi: 10.1093/bib/bbaa148

62. Kumaran M, Subramanian U, Devarajan B. Performance Assessment of Variant Calling Pipelines Using Human Whole Exome Sequencing and Simulated Data. *BMC Bioinf* (2019) 20(1):342. doi: 10.1186/s12859-019-2928-9

63. Wang M, Luo W, Jones K, Bian X, Williams R, Higson H, et al. SomaticCombiner: Improving the Performance of Somatic Variant Calling Based on Evaluation Tests and a Consensus Approach. *Sci Rep* (2020) 10 (1):12898. doi: 10.1038/s41598-020-69772-8

64. Hofmann AL, Behr J, Singer J, Kuipers J, Beisel C, Schraml P, et al. Detailed Simulation of Cancer Exome Sequencing Data Reveals Differences and Common Limitations of Variant Callers. *BMC Bioinf* (2017) 18(1):8. doi: 10.1186/s12859-016-1417-7

65. Cameron DL, Di Stefano L, Papenfuss AT. Comprehensive Evaluation and Characterisation of Short Read General-Purpose Structural Variant Calling Software. *Nat Commun* (2019) 10(1):3240. doi: 10.1038/s41467-019-11146-4

66. Mahmoud M, Gobet N, Cruz-Davalos DI, Mounier N, Dessimoz C, Sedlazeck FJ. Structural Variant Calling: The Long and the Short of It. *Genome Biol* (2019) 20(1):246. doi: 10.1186/s13059-019-1828-7

67. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Kallberg M, et al. Manta: Rapid Detection of Structural Variants and Indels for Germline and Cancer Sequencing Applications. *Bioinformatics* (2016) 32(8):1220–2. doi: 10.1093/bioinformatics/btv710

68. 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 (1):213. doi: 10.1186/s13059-019-1842-9

69. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol* (2016) 17(1):122. doi: 10.1186/s13059-016-0974-4

70. Yang H, Wang K. Genomic Variant Annotation and Prioritization With ANNOVAR and wANNOVAR. *Nat Protoc* (2015) 10(10):1556–66. doi: 10.1038/nprot.2015.105

71. McCarthy DJ, Humburg P, Kanapin A, Rivas MA, Gaulton K, Cazier JB, et al. Choice of Transcripts and Software Has a Large Effect on Variant Annotation. *Genome Med* (2014) 6(3):26. doi: 10.1186/gm543

72. Stromberg M, Roy R, Lajugie J, Jiang Y, Li H, Margulies E. "Nirvana: Clinical Grade Variant Annotator". In: *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (2017).

73. Rathinakannan VS, Schukov HP, Heron S, Schleutker J, Sipeky C. ShAn: An Easy-to-Use Tool for Interactive and Integrated Variant Annotation. *PLoS One* (2020) 15(7):e0235669. doi: 10.1371/journal.pone.0235669

74. Grant AD, Vail P, Padi M, Witkiewicz AK, Knudsen ES. Interrogating Mutant Allele Expression *via* Customized Reference Genomes to Define Influential Cancer Mutations. *Sci Rep* (2019) 9(1):12766. doi: 10.1038/s41598-019-48967-8

75. Wang D, Eraslan B, Wieland T, Hallstrom B, Hopf T, Zolg DP, et al. A Deep Proteome and Transcriptome Abundance Atlas of 29 Healthy Human Tissues. *Mol Syst Biol* (2019) 15(2):e8503. doi: 10.15252/msb.20188503

76. Wen B, Li K, Zhang Y, Zhang B. Cancer Neoantigen Prioritization Through Sensitive and Reliable Proteogenomics Analysis. *Nat Commun* (2020) 11 (1):1759. doi: 10.1038/s41467-020-15456-w

77. Gillis S, Roth A. PyClone-VI: Scalable Inference of Clonal Population Structures Using Whole Genome Data. *BMC Bioinf* (2020) 21(1):571. doi: 10.1186/s12859-020-03919-2

78. Xiao Y, Wang X, Zhang H, Ulintz PJ, Li H, Guan Y. FastClone Is a Probabilistic Tool for Deconvoluting Tumor Heterogeneity in Bulk-Sequencing Samples. *Nat Commun* (2020) 11(1):4469. doi: 10.1038/s41467-020-18169-2

79. Lu T, Wang S, Xu L, Zhou Q, Singla N, Gao J, et al. Tumor Neoantigenicity Assessment With CSiN Score Incorporates Clonality and Immunogenicity to Predict Immunotherapy Outcomes. *Sci Immunol* (2020) 5(44):eaaz3199. doi: 10.1126/sciimmunol.aaz3199

80. Saxova P, Buus S, Brunak S, Kesmir C. Predicting Proteasomal Cleavage Sites: A Comparison of Available Methods. *Int Immunol* (2003) 15(7):781–7. doi: 10.1093/intimm/dxg084

81. Stranzl T, Larsen MV, Lundegaard C, Nielsen M. NetCTLpan: Pan-Specific MHC Class I Pathway Epitope Predictions. *Immunogenetics* (2010) 62 (6):357–68. doi: 10.1007/s00251-010-0441-4

82. Gomez-Perosanz M, Ras-Carmona A, Lafuente EM, Reche PA. Identification of CD8(+) T Cell Epitopes Through Proteasome Cleavage Site Predictions. *BMC Bioinf* (2020) 21(Suppl 17):484. doi: 10.1186/s12859-020-03782-1

83. Liu IH, Lo YS, Yang JM. PAComplex: A Web Server to Infer Peptide Antigen Families and Binding Models From TCR-pMHC Complexes. *Nucleic Acids Res* (2011) 39(Web Server issue):W254–260. doi: 10.1093/nar/gkr434

84. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol* (2017) 199 (9):3360–8. doi: 10.4049/jimmunol.1700893

85. Alvarez B, Reynisson B, Barra C, Buus S, Ternette N, Connelley T, et al. NNAlign_MA; MHC Peptidome Deconvolution for Accurate MHC Binding Motif Characterization and Improved T-Cell Epitope Predictions. *Mol Cell Proteomics* (2019) 18(12):2459–77. doi: 10.1074/mcp.TIR119.001658

86. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved Predictions of MHC Antigen Presentation by

Concurrent Motif Deconvolution and Integration of MS MHC Eluted Ligand Data. *Nucleic Acids Res* (2020) 48(W1):W449–54. doi: 10.1093/nar/gkaa379

87. O'Donnell TJ, Rubinsteyn A, Laserson U. MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing. *Cell Syst* (2020) 11(4):418–9. doi: 10.1016/j.cels.2020.09.001

88. Shao XM, Bhattacharya R, Huang J, Sivakumar IKA, Tokheim C, Zheng L, et al. High-Throughput Prediction of MHC Class I and II Neoantigens With MHCnuggets. *Cancer Immunol Res* (2020) 8(3):396–408. doi: 10.1158/2326-6066.CIR-19-0464

89. Boehm KM, Bhinder B, Raja VJ, Dephoure N, Elemento O. Predicting Peptide Presentation by Major Histocompatibility Complex Class I: An Improved Machine Learning Approach to the Immunopeptidome. *BMC Bioinf* (2019) 20(1):7. doi: 10.1186/s12859-018-2561-z

90. Hu Y, Wang Z, Hu H, Wan F, Chen L, Xiong Y, et al. ACME: Pan-Specific Peptide-MHC Class I Binding Prediction Through Attention-Based Deep Neural Networks. *Bioinformatics* (2019) 35(23):4946–54. doi: 10.1093/bioinformatics/btz427

91. Bassani-Sternberg M, Chong C, Guillaume P, Solleder M, Pak H, Gannon PO, et al. Deciphering HLA-I Motifs Across HLA Peptidomes Improves Neo-Antigen Predictions and Identifies Allostery Regulating HLA Specificity. *PLoS Comput Biol* (2017) 13(8):e1005725. doi: 10.1371/journal.pcbi.1005725

92. Abelin JG, Keskin DB, Sarkizova S, Hartigan CR, Zhang W, Sidney J, et al. Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-Allelic Cells Enables More Accurate Epitope Prediction. *Immunity* (2017) 46(2):315–26. doi: 10.1016/j.immuni.2017.02.007

93. Han Y, Kim D. Deep Convolutional Neural Networks for Pan-Specific Peptide-MHC Class I Binding Prediction. *BMC Bioinf* (2017) 18(1):585. doi: 10.1186/s12859-017-1997-x

94. Kim Y, Sidney J, Pinilla C, Sette A, Peters B. Derivation of an Amino Acid Similarity Matrix for Peptide: MHC Binding and Its Application as a Bayesian Prior. *BMC Bioinf* (2009) 10:394. doi: 10.1186/1471-2105-10-394

95. Zhang H, Lund O, Nielsen M. The PickPocket Method for Predicting Binding Specificities for Receptors Based on Receptor Pocket Similarities: Application to MHC-Peptide Binding. *Bioinformatics* (2009) 25(10):1293–9. doi: 10.1093/bioinformatics/btp137

96. Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, et al. Quantitative Peptide Binding Motifs for 19 Human and Mouse MHC Class I Molecules Derived Using Positional Scanning Combinatorial Peptide Libraries. *Immunome Res* (2008) 4:2. doi: 10.1186/1745-7580-4-2

97. Bui HH, Sidney J, Peters B, Sathiamurthy M, Sinichi A, Purton KA, et al. Automated Generation and Evaluation of Specific MHC Binding Predictive Tools: ARB Matrix Applications. *Immunogenetics* (2005) 57(5):304–14. doi: 10.1007/s00251-005-0798-y

98. Peters B, Sette A. Generating Quantitative Models Describing the Sequence Specificity of Biological Processes With the Stabilized Matrix Method. *BMC Bioinf* (2005) 6:132. doi: 10.1186/1471-2105-6-132

99. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI: Database for MHC Ligands and Peptide Motifs. *Immunogenetics* (1999) 50(3-4):213–9. doi: 10.1007/s002510050595

100. Zhao W, Sher X. Systematically Benchmarking Peptide-MHC Binding Predictors: From Synthetic to Naturally Processed Epitopes. *PLoS Comput Biol* (2018) 14(11):e1006457. doi: 10.1371/journal.pcbi.1006457

101. Bonsack M, Hoppe S, Winter J, Tichy D, Zeller C, Kupper MD, et al. Performance Evaluation of MHC Class-I Binding Prediction Tools Based on an Experimentally Validated MHC-Peptide Binding Data Set. *Cancer Immunol Res* (2019) 7(5):719–36. doi: 10.1158/2326-6066.CIR-18-0584

102. Paul S, Croft NP, Purcell AW, Tscharke DC, Sette A, Nielsen M, et al. Benchmarking Predictions of MHC Class I Restricted T Cell Epitopes in a Comprehensively Studied Model System. *PLoS Comput Biol* (2020) 16(5):e1007757. doi: 10.1371/journal.pcbi.1007757

103. Karosiene E, Lundegaard C, Lund O, Nielsen M. NetMHCcons: A Consensus Method for the Major Histocompatibility Complex Class I Predictions. *Immunogenetics* (2012) 64(3):177–86. doi: 10.1007/s00251-011-0579-8

104. Rasmussen M, Fenoy E, Harndahl M, Kristensen AB, Nielsen IK, Nielsen M, et al. Pan-Specific Prediction of Peptide-MHC Class I Complex Stability, a

Correlate of T Cell Immunogenicity. *J Immunol* (2016) 197(4):1517–24. doi: 10.4049/jimmunol.1600582

105. Chowell D, Krishna S, Becker PD, Cocita C, Shu J, Tan X, et al. TCR Contact Residue Hydrophobicity is a Hallmark of Immunogenic CD8+ T Cell Epitopes. *Proc Natl Acad Sci USA* (2015) 112(14):E1754–1762. doi: 10.1073/pnas.1500973112

106. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The Immune Epitope Database (IEDB): 2018 Update. *Nucleic Acids Res* (2019) 47 (D1):D339–43. doi: 10.1093/nar/gky1006

107. Ellis JM, Henson V, Slack R, Ng J, Hartzman RJ, Katovich Hurley C. Frequencies of HLA-A2 Alleles in Five U.S. Population Groups. Predominance Of A*02011 and Identification of HLA-A*0231. *Hum Immunol* (2000) 61(3):334–40. doi: 10.1016/s0198-8859(99)00155-x

108. McGranahan N, Rosenthal R, Hiley CT, Rowan AJ, Watkins TBK, Wilson GA, et al. Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* (2017) 171(6):1259–71.e1211. doi: 10.1016/j.cell.2017.10.001

109. Dhatchinamoorthy K, Colbert JD, Rock KL. Cancer Immune Evasion Through Loss of MHC Class I Antigen Presentation. *Front Immunol* (2021) 12:636568:636568. doi: 10.3389/fimmu.2021.636568

110. Zaretsky JM, Garcia-Diaz A, Shin DS, Escuin-Ordinas H, Hugo W, Hu-Lieskovan S, et al. Mutations Associated With Acquired Resistance to PD-1 Blockade in Melanoma. *N Engl J Med* (2016) 375(9):819–29. doi: 10.1056/NEJMoa1604958

111. Seliger B, Ritz U, Abele R, Bock M, Tampe R, Sutter G, et al. Immune Escape of Melanoma: First Evidence of Structural Alterations in Two Distinct Components of the MHC Class I Antigen Processing Pathway. *Cancer Res* (2001) 61(24):8647–50.

112. Kloor M, Becker C, Benner A, Woerner SM, Gebert J, Ferrone S, et al. Immunoselective Pressure and Human Leukocyte Antigen Class I Antigen Machinery Defects in Microsatellite Unstable Colorectal Cancers. *Cancer Res* (2005) 65(14):6418–24. doi: 10.1158/0008-5472.CAN-05-0044

113. Belicha-Villanueva A, Golding M, McEvoy S, Sarvaiya N, Cresswell P, Gollnick SO, et al. Identification of an Alternate Splice Form of Tapasin in Human Melanoma. *Hum Immunol* (2010) 71(10):1018–26. doi: 10.1016/j.humimm.2010.05.019

114. Chang CC, Pirozzi G, Wen SH, Chung IH, Chiu BL, Errico S, et al. Multiple Structural and Epigenetic Defects in the Human Leukocyte Antigen Class I Antigen Presentation Pathway in a Recurrent Metastatic Melanoma Following Immunotherapy. *J Biol Chem* (2015) 290(44):26562–75. doi: 10.1074/jbc.M115.676130

115. Richman LP, Vonderheide RH, Rech AJ. Neoantigen Dissimilarity to the Self-Proteome Predicts Immunogenicity and Response to Immune Checkpoint Blockade. *Cell Syst* (2019) 9(4):375–82.e374. doi: 10.1016/j.cels.2019.08.009

116. Capietto A-H, Jhunjhunwala S, Pollock SB, Lupardus P, Wong J, Hänsch L, et al. Mutation Position Is an Important Determinant for Predicting Cancer Neoantigens. *J Exp Med* (2020) 217(4):e20190179. doi: 10.1084/jem.20190179

117. Alspach E, Lussier DM, Miceli AP, Kizhvatov I, DuPage M, Luoma AM, et al. MHC-II Neoantigens Shape Tumour Immunity and Response to Immunotherapy. *Nature* (2019) 574(7780):696–701. doi: 10.1038/s41586-019-1671-8

118. Sercarz EE, Maverakis E. Mhc-Guided Processing: Binding of Large Antigen Fragments. *Nat Rev Immunol* (2003) 3(8):621–9. doi: 10.1038/nri1149

119. Lee P, Matsueda GR, Allen PM. T Cell Recognition of Fibrinogen. A Determinant on the A Alpha-Chain Does Not Require Processing. *J Immunol* (1988) 140(4):1063–8.

120. Buus S, Sette A, Colon SM, Miles C, Grey HM. The Relation Between Major Histocompatibility Complex (MHC) Restriction and the Capacity of Ia to Bind Immunogenic Peptides. *Science* (1987) 235(4794):1353–8. doi: 10.1126/science.2435001

121. Abelin JG, Harjanto D, Malloy M, Suri P, Colson T, Goulding SP, et al. Defining HLA-II Ligand Processing and Binding Rules With Mass Spectrometry Enhances Cancer Epitope Prediction. *Immunity* (2021) 54 (2):388. doi: 10.1016/j.immuni.2020.12.005

122. Barra C, Alvarez B, Paul S, Sette A, Peters B, Andreatta M, et al. Footprints of Antigen Processing Boost MHC Class II Natural Ligand Predictions. *Genome Med* (2018) 10(1):84. doi: 10.1186/s13073-018-0594-6

123. 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) 9:1795:1795. doi: 10.3389/fimmu.2018.01795

124. Andreatta M, Trolle T, Yan Z, Greenbaum JA, Peters B, Nielsen M. An Automated Benchmarking Platform for MHC Class II Binding Prediction Methods. *Bioinformatics* (2018) 34(9):1522–8. doi: 10.1093/bioinformatics/btx820

125. Liu Z, Jin J, Cui Y, Xiong Z, Nasiri A, Zhao Y, et al. DeepSeqPanII: An Interpretable Recurrent Neural Network Model With Attention Mechanism for Peptide-HLA Class II Binding Prediction. *IEEE/ACM Trans Comput Biol Bioinform PP* (2021). doi: 10.1109/TCBB.2021.3074927

126. Racle J, Michaux J, Rockinger GA, Arnaud M, Bobisse S, Chong C, et al. Robust Prediction of HLA Class II Epitopes by Deep Motif Deconvolution of Immunopeptidomes. *Nat Biotechnol* (2019) 37(11):1283–6. doi: 10.1038/s41587-019-0289-6

127. Jones EY, Fugger L, Strominger JL, Siebold C. MHC Class II Proteins and Disease: A Structural Perspective. *Nat Rev Immunol* (2006) 6(4):271–82. doi: 10.1038/nri1805

128. Ferrante A, Gorski J. Cooperativity of Hydrophobic Anchor Interactions: Evidence for Epitope Selection by MHC Class II as a Folding Process. *J Immunol* (2007) 178(11):7181–9. doi: 10.4049/jimmunol.178.11.7181

129. Dhanda SK, Karosiene E, Edwards L, Grifoni A, Paul S, Andreatta M, et al. Predicting HLA CD4 Immunogenicity in Human Populations. *Front Immunol* (2018) 9:1369:1369. doi: 10.3389/fimmu.2018.01369

130. Chen B, Khodadoust MS, Olsson N, Wagar LE, Fast E, Liu CL, et al. Predicting HLA Class II Antigen Presentation Through Integrated Deep Learning. *Nat Biotechnol* (2019) 37(11):1332–43. doi: 10.1038/s41587-019-0280-2

131. Robbins PF, Lu YC, El-Gamil M, Li YF, Gross C, Gartner J, et al. Mining Exomic Sequencing Data to Identify Mutated Antigens Recognized by Adoptively Transferred Tumor-Reactive T Cells. *Nat Med* (2013) 19 (6):747–52. doi: 10.1038/nm.3161

132. Wick DA, Webb JR, Nielsen JS, Martin SD, Kroeger DR, Milne K, et al. Surveillance of the Tumor Mutanome by T Cells During Progression From Primary to Recurrent Ovarian Cancer. *Clin Cancer Res* (2014) 20(5):1125–34. doi: 10.1158/1078-0432.CCR-13-2147

133. Rajasagi M, Shukla SA, Fritsch EF, Keskin DB, DeLuca D, Carmona E, et al. Systematic Identification of Personal Tumor-Specific Neoantigens in Chronic Lymphocytic Leukemia. *Blood* (2014) 124(3):453–62. doi: 10.1182/blood-2014-04-567933

134. Cohen CJ, Gartner JJ, Horovitz-Fried M, Shamalov K, Trebska-McGowan K, Bliskovsky VV, et al. Isolation of Neoantigen-Specific T Cells From Tumor and Peripheral Lymphocytes. *J Clin Invest* (2015) 125(10):3981–91. doi: 10.1172/JCI82416

135. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal Neoantigens Elicit T Cell Immunoreactivity and Sensitivity to Immune Checkpoint Blockade. *Science* (2016) 351(6280):1463–9. doi: 10.1126/science.aaf1490

136. Strønen E, Toebes M, Kelderman S, van Buuren MM, Yang W, van Rooij N, et al. Targeting of Cancer Neoantigens With Donor-Derived T Cell Receptor Repertoires. *Science* (2016) 352(6291):1337–41. doi: 10.1126/science.aaf2288

137. 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 (10):1037–45. doi: 10.1038/nbt.3662

138. 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(4):433–8. doi: 10.1038/nm.4051

139. Duan F, Duitama J, Al Seesi S, Ayres CM, Corcelli SA, Pawashe AP, et al. Genomic and Bioinformatic Profiling of Mutational Neoepitopes Reveals New Rules to Predict Anticancer Immunogenicity. *J Exp Med* (2014) 211 (11):2231–48. doi: 10.1084/jem.20141308

140. Pai JA, Satpathy AT. High-Throughput and Single-Cell T Cell Receptor Sequencing Technologies. *Nat Methods* (2021) 18(8):881–92. doi: 10.1038/s41592-021-01201-8

141. 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(13):E1890–1897. doi: 10.1073/pnas.1602488113

142. Dolton G, Zervoudi E, Rius C, Wall A, Thomas HL, Fuller A, et al. Optimized Peptide-MHC Multimer Protocols for Detection and Isolation of Autoimmune T-Cells. *Front Immunol* (2018) 9:1378:1378. doi: 10.3389/fimmu.2018.01378

143. Aleksic M, Liddy N, Molloy PE, Pumphrey N, Vuidepot A, Chang KM, et al. Different Affinity Windows for Virus and Cancer-Specific T-Cell Receptors: Implications for Therapeutic Strategies. *Eur J Immunol* (2012) 42(12):3174–9. doi: 10.1002/eji.201242606

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.