








Characterization of Tumor Antigens from Multi-omics Data: Computational Approaches and Resources

Yunzhe Wang (王韞哲) ¹, James Wengler ², Yuzhu Fang (房钰竹) ¹, Joseph Zhou ²,
Hang Ruan ³, Zhao Zhang ^{1,*}, Leng Han ^{2,4,5,*}

¹MOE Key Laboratory of Metabolism and Molecular Medicine, Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Fudan University, Shanghai 200032, China

²Center for Epigenetics and Disease Prevention, Institute of Biosciences and Technology, Texas A&M University, Houston, TX 77030, USA

³Jiangsu Key Laboratory of Infection and Immunity, Institutes of Biology and Medical Sciences, Soochow University, Suzhou 215123, China

⁴Brown Center for Immunotherapy, School of Medicine, Indiana University, Indianapolis, IN 46202, USA

⁵Department of Biostatistics and Health Data Science, School of Medicine, Indiana University, Indianapolis, IN 46202, USA

*Corresponding authors: zhaozhang@fudan.edu.cn (Zhang Z), lenghan@iu.edu (Han L).

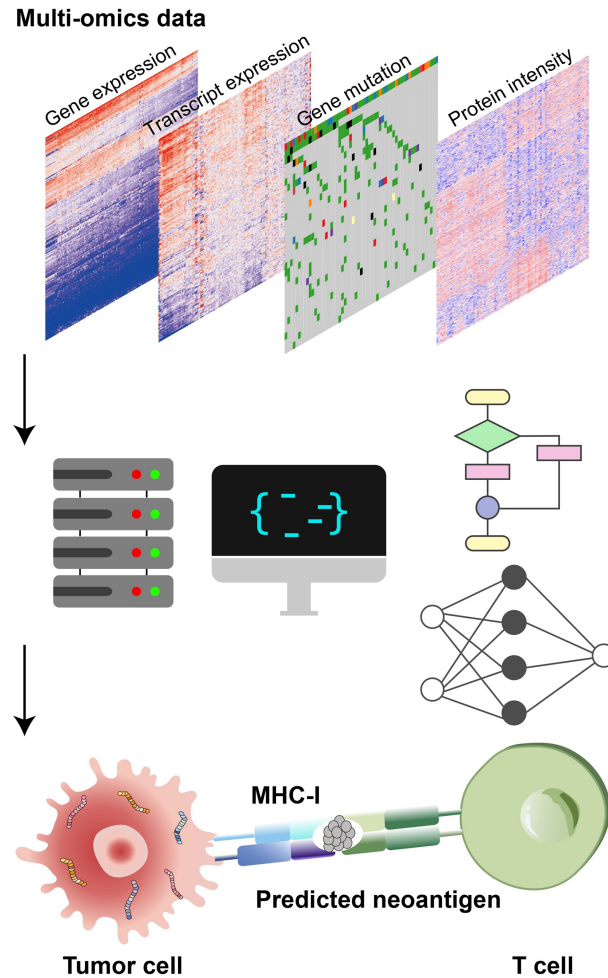
Handling Editor: Song Liu

Abstract

Tumor-specific antigens, also known as neoantigens, have potential utility in anti-cancer immunotherapy, including immune checkpoint blockade (ICB), neoantigen-specific T cell receptor-engineered T (TCR-T), chimeric antigen receptor T (CAR-T), and therapeutic cancer vaccines (TCVs). After recognizing presented neoantigens, the immune system becomes activated and triggers the death of tumor cells. Neoantigens may be derived from multiple origins, including somatic mutations (single nucleotide variants, insertions/deletions, and gene fusions), circular RNAs, alternative splicing, RNA editing, and polymorphic microbiomes. An increasing amount of bioinformatics tools and algorithms are being developed to predict tumor neoantigens derived from different sources, which may require inputs from different multi-omics data. In addition, calculating the peptide–major histocompatibility complex (MHC) affinity can aid in selecting putative neoantigens, as high binding affinities facilitate antigen presentation. Based on these approaches and previous experiments, many resources have been developed to reveal the landscape of tumor neoantigens across multiple cancer types. Herein, we summarize these tools, algorithms, and resources to provide an overview of computational analysis for neoantigen discovery and prioritization, as well as the future development of potential clinical utilities in this field.

Key words: Antigen; Tumor; Multi-omics; Computational approach; Resource.

Graphical abstract



Introduction

The adaptive immune system, with custom-tailored receptors, evolved to provide specific and flexible responses, immunologic memory, and rapid reactions upon re-exposure [1]. Cell-mediated immunity and humoral immunity work together to form the adaptive immune system, in which antigen presentation (AP) and recognition are fundamental processes. Antigens — molecules with immunogenic properties — originate from two main sources: foreign pathogens and autoantigens originating within the body [2]. Lurquin et al. demonstrated that cancerous cells present altered antigens that are not regularly found on non-cancerous cells, defined as “neoantigens” [3,4]. Neoantigens were widely detected across multiple cancer types and derived from aberrant alterations in somatic mutations, circular RNAs (circRNAs), alternative splicing (AS), RNA editing, and polymorphic microbiomes. The uniqueness of neoantigens would make cancer cells more targetable in anti-cancer therapy.

The presenting efficiency of tumor neoantigens is greatly determined by the high heterogeneity of major histocompatibility complexes (MHCs), also known as human leukocyte antigens (HLAs) in humans [5,6]. Each MHC is a heterodimer of an α and a β polypeptide chain: α chains give MHCs great potential to bind to different peptides, and β chains

serve as the stabilizers of MHC structures [7–9]. Generally, MHC molecules can be divided into two classes, MHC-I and MHC-II. MHC-I molecules, encompassing HLA-A, HLA-B, and HLA-C in humans, are expressed in all nucleated cells and platelets, and present peptides derived from cytoplasmic proteins [10,11]. Within HLA-I groups, polymorphic alleles of HLAs introduce complexities into the AP process and subsequent immune responses by performing allele-specific preferences [11–13]. For example, allele HLA-A2 shares peptide specificity with aliphatic hydrophobic residues at the C-terminus [12]. Class II molecules are categorized by different chain alleles, including HLA-DP, HLA-DQ, HLA-DR, HLA-DX, HLA-DM, HLA-DOA, and HLA-DOB in humans. The contributors of polymorphism are different across these alleles. For example, the DR α chain is largely conserved, but the DQ locus contains an extensively polymorphic α chain gene [14]. HLA-II molecules present peptides derived from extracellular sources that have undergone absorption into endosomes or lysosomes [10,15].

After the AP process, MHC molecules can be recognized by T cell receptors (TCRs) on the surface of T cells, resulting in cell fate determination: the cell either remains stable or undergoes apoptosis [16]. CD8⁺ T cells recognize autologous peptides from 8 to 11 amino acids presented by MHC-I molecules. If the peptides are identified as “self” signals, T cells

stay tolerant [10,17]. However, mutant sequences or microbe-derived peptides, once presented to the cell surface, can be recognized by CD8⁺ T cells as abnormal (or “non-self”) signals. These CD8⁺ T cells can then become activated and exert cytolytic functions [10]. Differently, MHC-II molecules on antigen presentation cells (APCs) present antigens to naïve CD4⁺ T cells. The recognition activates CD4⁺ T cells and transforms them into effector T cells. Subsequently, MHC-II molecules mediate interactions between B cells and macrophages with these newly activated antigen-specific CD4⁺ effector T cells [18]. Accumulated studies have focused on the impacts of utilizing neoantigen-induced immune responses against cancer, which may aid in cancer treatment.

Clinical utility of neoantigens in cancer treatment

Recent clinical endeavors in utilizing neoantigens include immune checkpoint blockade (ICB) responses, neoantigen-specific T cell receptor-engineered T (TCR-T) and chimeric antigen receptor T (CAR-T) therapies, and therapeutic cancer vaccines (TCVs) (Figure 1). In personalized cancer immunotherapies, neoantigens can also serve as predictors or adjuvants for ICB responses [19]. Immune checkpoint inhibitors (ICIs), such as anti-programmed cell death protein 1 (anti-PD-1) and anti-cytotoxic T lymphocyte-associated antigen 4 (anti-CTLA-4), are drugs designed to block these checkpoints, allowing the immune system to recognize and attack cancer cells more effectively [20,21].

Due to immunogenic properties, neoantigens may improve responses toward ICBs. For example, frameshift mutations generated by insertions/deletions (InDels) could produce neoantigens in non-small cell lung cancer (NSCLC), resulting in increased infiltration of activated T cells and predicting prolonged progression-free survival [22] (Figure 1A). Recent studies also found that patients with a sustained response to anti-PD-1 showed diverse neoantigen-specific T cell responses in melanoma and lung cancers [23–25]. Furthermore, neoantigens can serve as ideal targets for TCR-based immunotherapy [26]. For example, a patient with progressive metastatic pancreatic cancer was treated with autologous T cells expressing HLA-C*08:02-restricted TCRs targeting the neoantigen derived from mutant KRAS G12D, resulting in the recovery of other immune cells and regression of metastatic lung lesions [27] (Figure 1B).

CAR-T therapies genetically engineer autologous T cells by fusing an antigen-specific single-chain variable fragment (scFv) with TCR to form the chimeric antigen receptor (CAR) [28–30]. Traditional CAR-T therapies recognize antigens through HLA-independent and antibody-like-mediated processes [31]. For example, anti-GD2 CAR T cells can target glioma cells expressing K27M mutation in the gene of histone H3 (H3K27M-mutated glioma cells), thus performing antigen-specific cytotoxicity [32]. Recent studies also designed scFvs to target neoantigen–HLA complexes to enhance immunological responses [33,34]. For example, CAR-T cells target a nucleophosmin-derived neoantigen, NPM1c, which could be presented by HLA-A2 in patients with acute myeloid leukemia, thus mediating anti-tumor immunity [34] (Figure 1C).

In addition, neoantigens can potentially be engineered as TCVs in therapy [35–37]. A peptide vaccine derived from the reverse transcriptase subunit of telomerase (hTERT) has been

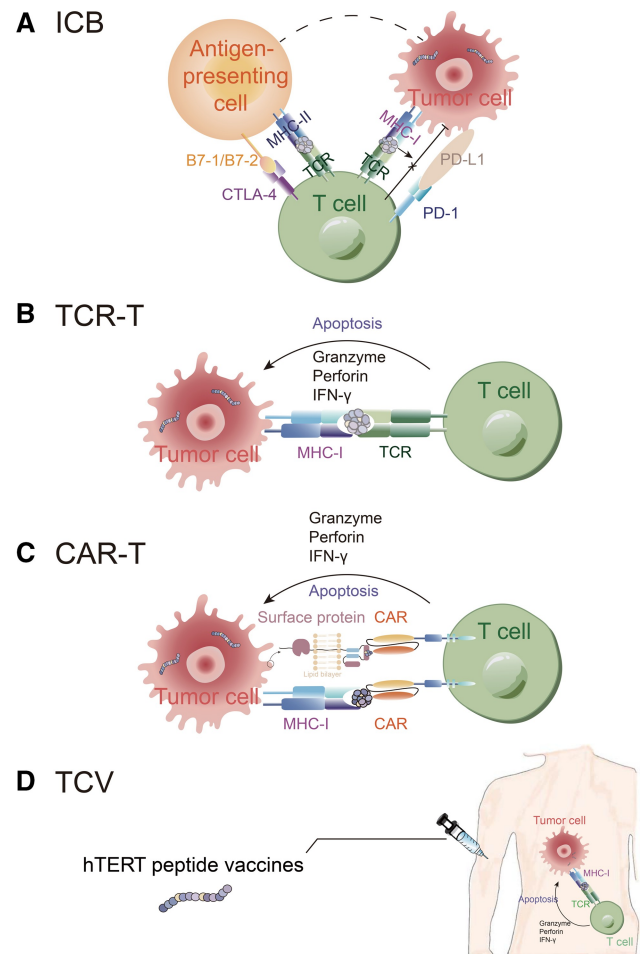


Figure 1 Tumor neoantigens in anti-cancer therapy

A. Neoantigens in ICI treatment. **B.** Neoantigens in TCR-T. **C.** Neoantigens in CAR-T. **D.** Neoantigens in TCVs. ICB, immune checkpoint blockade; ICI, immune checkpoint inhibitor; TCR-T, T cell receptor-engineered T; CAR-T, chimeric antigen receptor T; TCV, therapeutic cancer vaccine; TCR, T cell receptor; CAR, chimeric antigen receptor; MHC, major histocompatibility complex; PD, programmed cell death protein; CTLA, cytotoxic T lymphocyte-associated antigen; IFN, interferon; hTERT, human telomerase reverse transcriptase.

shown to induce immune responses in metastatic hormone-naïve prostate cancer [38] (Figure 1D). Some neoantigen-based TCV platforms have also been put into clinical trials (NCT02348320 and NCT03122106) to enhance immune responses against cancer [39,40].

Mining potential neoantigens from multi-omics data could provide effective information for clinical utility. The origins and immunogenicity of potential neoantigens, the types of HLA alleles, and the binding affinity between them all affect the *in silico* prediction of neoantigens. Therefore, these characteristics should be carefully considered when developing tools and algorithms. Using these tools and algorithms, neoantigens could be identified across various cancer types.

Tools and algorithms for predicting neoantigens from different origins

The current landscape of bioinformatics tools and algorithms designed for neoantigen prediction exhibits diverse

techniques. They generally share a common workflow to streamline the identification process [41]. This typical workflow involves several key steps: (1) identification of neoantigen candidates: extraction of putative neoantigen candidates from the input genomic, transcriptomic, or proteomic data, considering different origins of neoantigens; (2) detection of HLA alleles: determination of the patient's HLA alleles, providing essential information for subsequent analyses; (3) prediction of peptide–HLA binding affinity and stability: assessment of the binding affinity and stability of the identified neoantigen candidates with the patients' HLA molecules, ensuring compatibility for effective presentation to immune cells; (4) prioritization of neoantigen candidates: selection and prioritization of the most promising neoantigen candidates based on various criteria, such as predicted immunogenicity, likelihood of eliciting immune responses, and expression validations [39]. These computational approaches hold the promise of unlocking the potential of neoantigens in personalized cancer treatment.

Neoantigens may arise from multiple origins, which must be considered in the development of tools and algorithms. Somatic mutations, referring to alterations at the cellular level in somatic tissues occurring after fertilization, have long been regarded as the most frequent origin of neoantigens. Among mutations, single nucleotide variants (SNVs) and InDels could generate aberrant open reading frames (ORFs), which may be transcribed and translated into neoantigens. Another type of somatic mutation, fusion genes, can also generate neoantigens due to the erroneous fusions of different sequences. Other abnormal transcriptional and/or post-transcriptional events, including circRNAs, AS, and RNA editing, can also produce neoantigens, especially when the RNAs contain and alter ORFs within their sequences. Additionally, the polymorphic microbiome, recently listed as one of the hallmarks of cancer, is likely to be lysed into peptides serving as neoantigens. These highly specific neoantigens from different origins necessitate individualized identification for each patient [42]. Experimental breakthroughs incorporating identified neoantigens have shown anti-tumoral effects across cancer types. In parallel, the field of computational-level identification of neoantigens has also experienced rapid advancements in recent years. The computational approach is crucial in providing a systematic and efficient means of predicting neoantigens that can selectively stimulate T cell responses to target and eliminate cancer cells.

Predicting neoantigens from somatic mutations

Somatic mutations are spontaneously occurring mutations that accumulate in somatic cells, including SNVs, InDels, copy number variations (CNVs), and gene fusions [43,44]. In this section, we review the tools focusing on SNVs, InDels, and fusion genes, as these types of mutations are more likely to produce aberrant peptide sequences with AP potentials. As tumors with a mutation load above ten somatic mutations per megabase (Mb) tend to share common formations of neoantigens potentially recognized by T cells, somatic mutations are thought to be the primary origin of neoantigens [35]. A diverse array of tools has been developed to automate these processes.

Predicting neoantigens from SNVs and InDels

SNVs and InDels may result in nonsynonymous mutations and/or novel ORFs that alter the amino acid sequences and generate abnormal peptides [45–48]. For example, explicitly

targeting the neoantigen derived from the most common *TP53* mutation, R175H, in complex with HLA-A has been proved to effectively activate T cells, leading to the lysis of cancer cells by presenting the neoantigen *in vitro* and in mice [49] (Figure 2A). More importantly, neoantigens derived from mutant *AIM2*, *HT001*, and *TAF1B* have already undergone clinical trials, providing promising prospects for utilizing somatic mutation-derived neoantigens [50].

The workflow for targeting SNVs and InDels usually takes two types of inputs: (1) genome sequencing data of paired tumor and normal samples and (2) pre-analyzed mutation files (also derived from genome sequencing) [47,51]. In Table 1, we summarize the tools capable of predicting neoantigens originating from SNVs and InDels. Some tools specifically predict neoantigens from SNVs and InDels, such as personalized variant antigens by cancer sequencing (pVAC-Seq), tumor-specific neoantigen detector (TSNAD), prioritizing tumor neoantigens (pTuneos), Neopepsee, PGNneo, neoantigen prediction pipeline (NeoPredPipe), mutant peptide extractor and informer (MuPeXI), CloudNeo, TIminer, and proteogenomics prediction of neoantigen (ProGeo-neo) [52–61]. Most tools only take sequencing data as input; however, PGNneo requires mass spectrometry (MS) data as input to further confirm the existence of candidate neoantigens at the proteomic level [56,62]. On the other hand, some tools can predict neoantigens derived from different origins (e.g., fusion genes, AS, and/or microbiome, see below), such as landscape of effective neoantigens software (LENS), ScanNeo2, ProGeo-neo v2.0, nextflow neoantigen prediction pipeline (nextNEOpI), TruNeo, TSNAD v2.0, and Seq2Neo [63–69]. The multiple origins provided by these tools also require different sorts of omics data.

Though the tools exhibit shared workflows, they may have tool-specific processes. Taking pVAC-Seq as an example, it requires properly formatted lists of annotated variants explaining the amino acid changes and transcript sequences to predict binding affinity. Specifically, differences between tumor and normal peptides are compared to enhance the prediction performance. Isoform-level expression information from RNA sequencing (RNA-seq) is also incorporated to accurately exclude variants that are not actively expressed in tumor cells. Altogether, multi-omics data and information from different aspects are analyzed in the pipeline of pVAC-Seq [52]. Importantly, neoantigens identified by pVACtools were designed as TCVs using a newly developed DNA vaccine platform. On average, each vaccine contained 11 neoantigens per patient and was then put into clinical trial (NCT02348320) [40]. Furthermore, another clinical trial (NCT03122106) based on pVACtools-predicted neoantigens was also conducted in pancreatic cancer patients ($n = 15$) [39]. Some tools incorporated multiple machine learning algorithms to improve the prediction of neoantigens from SNVs/InDels. For example, Neopepsee takes a list of mutations and the raw RNA-seq data as inputs and then automatically predicts neoantigens based on four machine learning models, including Gaussian naïve Bayes, locally weighted naïve Bayes, random forest, and support vector machine. After the calculation, it outputs rich annotations for candidate peptides, such as peptide half-life, peptide sequences, neoantigen probabilities in three levels, and expression levels of neoantigens with immune regulatory genes [55].

Together, these integrated pipelines and newly developed algorithms provide *in silico* prediction of neoantigens originating from SNVs and InDels according to genome data and

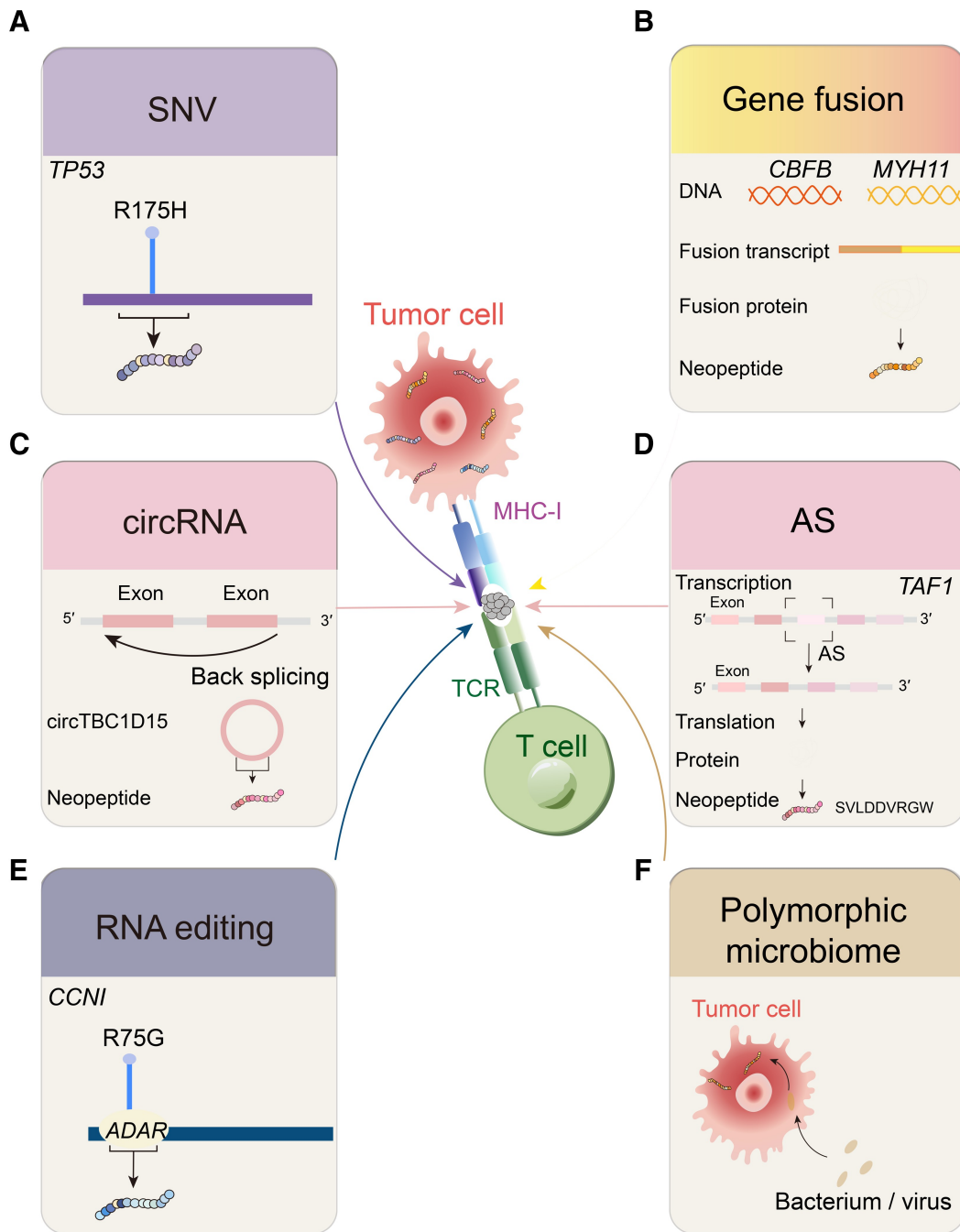


Figure 2 Diverse origins of tumor neoantigens

A. Neoantigens derived from SNVs, taking the R175H mutation in *TP53* as an example. **B.** Neoantigens derived from gene fusions, taking *CBFB-MYH11* fusion as an example. **C.** Neoantigens derived from circRNAs, taking circTBC1D15 as an example. **D.** Neoantigens derived from AS events, taking SVLDDVRGW from the *TAF1* gene in leukemia cell line HL-60 as an example. **E.** Neoantigens derived from RNA editing events, taking the edited peptide from *CCNI* as an example. **F.** Neoantigens derived from polymorphic microbiomes, including bacteria and viruses. circRNA, circular RNA; SNV, single nucleotide variant; AS, alternative splicing.

may validate results on other omics data. Differently, the computational-level identification of fusion genes is mainly based on RNA-seq data, which makes neoantigen prediction from SNVs/InDels and fusion genes differ in inputs and processing. Therefore, we separate the prediction of fusion-derived neoantigens from SNVs/InDels and will introduce the related tools in the next section.

Predicting neoantigens from fusion genes

Gene fusions occur when two genes become physical neighbors in the DNA sequence due to mutations and are expressed as a singular RNA or when two adjacent genes are erroneously translated together as a single gene due to transcription read-through [70,71]. Peptides spanning the breakpoint regions of different genes compose fusion proteins that

Table 1 Bioinformatics tools for predicting neoantigens

Tool	Input omics	Origin type	Origin prediction	HLA typing	Binding prediction	Neoantigen prioritization	System	Language	Link	Ref.
LENS	Genome, transcriptome	SNVs/InDels, gene fusions, AS, viruses	GATK Mutect2 [165], Strelka2 [166], ABRAX2 [167], NeoSplice [103], STAR-Fusion [168], VirDetect [131]	NA	NetMHCpan-4.0 [94], NetMHCstabpan [143], MHCFI [140], DeepHLApan [141], antigen.garnish [169]	Harmonization metric [63]	Linux	Python	https://gitlab.com/landscape-of-effective-neoantigens-software	[63]
ScanNeo2	Genome, transcriptome	SNVs/InDels, gene fusions, AS	GATK Mutect2 [165], transIndel [170], SplAdder [171], ScanExtron [162]	OptiType [133], HLA-HD [135]	NetMHC [138], NetMHCpan-4.0 [116], NetMHCII [139], NetMHCIIpan-3.0 [172]	IEDB immunogenicity [144]	Linux	Python	https://github.com/yilab-hi/ScanNeo2/	[64]
ProGeo-neo v2.0	Genome, transcriptome, proteome	SNVs/InDels, gene fusions	GATK Mutect2 [165], STAR-Fusion [168]	OptiType [133], HL-Aminer [136]	NetMHCpan-4.1 [94], NetMHCIIpan-4.0 [116]	Expression validation [65], BLASTP [147]	Linux	Python	https://github.com/kbystmd/ProGeo-neo2.0	[65]
nextNEOpi	Genome, transcriptome	SNVs/InDels, gene fusions	pVACseq [52], NeoFuse [77]	OptiType [133], HLA-HD [135]	NetMHCpan-4.0 [116], MHCFI [140]	MiXCR [145]	Linux	Nextflow, R, Python	https://github.com/icblab/nextNEOpi	[66]
TruNeo	Genome, transcriptome	SNVs/InDels, gene fusions	VarScan 2 [173], GATK somatic InDel detector [165], STAR-Fusion [168]	Polysolver [134], BWA-HLA	NetMHCpan-3.0 [137], MHCFI [140]	Score combining VAF [67], BA [67], TPM [67]	Linux	Perl, Python	https://github.com/yucebio/TruNeo	[67]
TSNAD v2.0	Genome, transcriptome	SNVs/InDels, gene fusions	GATK Mutect2 [165], Arriba [174]	OptiType [133]	DeepHLApan [141]	Binding score [68], immunogenic score [68], expression validation [68]	Linux	Perl, Python	https://github.com/jujujazz/tsnad	[68]
Seq2Neo	Genome, transcriptome	SNVs/InDels, gene fusions	GATK Mutect2 [165], STAR-Fusion [168]	HLA-HD [135]	NetMHCpan-4.0 [116]	Seq2Neo-CNN [69]	Linux	Python	https://github.com/XSLiuLab/Seq2Neo	[69]
pVAC-Seq	Genome, transcriptome	SNVs/InDels	VarScan 2 [173]	HL-Aminer [136], ATHLETES [175]	NetMHC 3.4 [138]	MT binding score [52], bam-read-count [52], expression validation [52]	Linux, MacOS, Windows	Python	https://github.com/griffithlab/pVAC-Seq	[52]
TSNAD	Genome, transcriptome	SNVs/InDels	GATK Mutect2 [165]	SOAP-HLA [176]	NetMHCpan-2.8 [177]	NA	Linux	Python, Perl	https://github.com/jujujazz/tsnad	[178]
pTuneos	Genome, transcriptome	SNVs/InDels	GATK Mutect2 [165], VarScan 2 [173], Strelka2 [166]	OptiType [133]	NetMHCpan-4.0 [116]	MHC multimer analysis [179]	Linux	Python	https://github.com/bm2-lab/pTuneos	[54]
Neopepsee	Genome, transcriptome	SNVs/InDels	NA	NA	NetCTLpan [180]	Immunogenicity classifier [55]	Windows	Java	https://sourceforge.net/projects/neopepsee/	[55]
PGNeo	Transcriptome, proteome	SNVs/InDels	GATK Mutect2 [165]	OptiType [133]	NetMHCpan-4.1 [94]	BLASTP [150]	Linux	Python, R	https://github.com/tanxiaoxiu/PGNeo	[56]

(continued)

Table 1 (continued)

Tool	Input omics	Origin type	Origin prediction	HLA typing	Binding prediction	Neoantigen prioritization	System	Language	Link	Ref.
NeoPredPipe	Genome, transcriptome	SNVs/InDels	NA	NA	NetMHCpan-4.0 [116]	Peptide Match [148], immune fitness model [146] Priority score [59]	Linux	Python, R	https://github.com/MathOnco/NeoPredPipe	[57]
MuPeXI	Genome, transcriptome	SNVs/InDels	GATK Mutect2 [165]	NA	NetMHCpan-4.0 [116]	NA	Linux, MacOS	Python	https://github.com/ambj/MuPeXI	[58]
CloudNeo	Genome, transcriptome	SNVs/InDels	NA	Polysolver [134], HL-Aminer [136]	NetMHCpan-3.0 [137]	NA	Linux, MacOS, Windows	Common workflow language	https://github.com/TheJacksonLaboratory/CloudNeo	[59]
TIminer	Genome, transcriptome	SNVs	NA	OptiType [133]	NetMHCpan-3.0 [137]	Expression validation [60]	Linux, MacOS	Python	https://icbi.i-med.ac.at/software/timiner/	[60]
ProGeo-neo	Genome, transcriptome, proteome	SNVs	BCFrools [181]	OptiType [133]	NetMHCpan-4.0 [116]	Expression validation [61], BLASTP [147], OmicsBean-Cancer workflow [61], SAAV mapping [61]	Linux	Python	https://github.com/timiner.shtml https://github.com/kbystmd/ ProGeo-neo	[61]
INTEGRATE-Neo	Genome, transcriptome	Gene fusions	INTEGRATE [182]	HL-Aminer [136]	NetMHC 4.0 [138]	NA	Linux	C++, Python	https://github.com/ChrisMaHerLab/	[76]
NeoFuse	Transcriptome	Gene fusions	Arriba [174]	OptiType [133]	MHCflurry [140]	IC50 [77], confidence score [77]	Linux, MacOS	Shell, Python	https://github.com/icbi-lab/NeoFuse	[77]
neoFusion	Proteome	Gene fusions	neoFusion search	NA	NetMHCpan-4.0 [116]	Target-decoy based FDR [78]	Linux	Python	https://github.com/bm2-lab/neoFusion	[78]
pVACfuse	Transcriptome	Gene fusions	AGFusion [183], Arriba [174], STAR-Fusion [168]	IEDB RESTful [150]	NetMHCpan-3.0 [137], NetMHC [138], NetMHCcons [184], PickPocket [185], SMM-align [186], SMMPMBEC [187], MHCflurry [140], NetMHCIIpan-3.0 [172], NN-align [188]	NetChop [189], BLASTP [147]	Linux, MacOS, Windows	Python, Perl	https://pvactools.readthedocs.io/en/latest/pvacfuse.html	[79]
CIRC_neo	Transcriptome	circRNA	find_circ, CIRCexplorer2 [190], IREFinder [191], CPAT [92], CPC2 [192]	OptiType [133], HLA-HD [135]	NetMHCpan-4.0 [116], MixxMHC2pred [193]	NetChop [189]	NA	Python, R	https://github.com/summerjiaqi/CIRC_neo	[82]
CICADA	Transcriptome, proteome	circRNA	CICADA	NA	NetMHCpan-4.1 [94]	NA	Linux	Python, R	https://github.com/SunLab-bio/tool/CICADA	[88]

(continued)

Table 1 (continued)

Tool	Input omics	Origin type	Origin prediction	HLA typing	Binding prediction	Neoantigen prioritization	System	Language	Link	Ref.
SNAF	Transcriptome	AS	AltAnalyze [194]	OptiType [133]	NetMHCpan-4.1 [94], MHCflurry [140]	DeepImmuno [195], BayesTS [196]	Linux	Python, Shell	https://github.com/franklity/SNAF	[197]
Retained-intron-neoantigen-pipeline	Genome, transcriptome	AS	Retained-intron-neoantigen-pipeline	Polysover [134]	NetMHCpan-3.0 [137]	NA	Linux	Python, Shell	https://github.com/vanallenlab/retained-intron-neoantigen-pipeline	[101]
ASNEO	Genome, transcriptome, proteome	AS	STAR [198]	OptiType [133]	NetMHCpan-4.0 [116]	Immune score [102]	Linux	Python	https://github.com/bm2-lab/ASNEO	[102]
NeoSplice	Transcriptome	AS	BWT-based algorithm, depth-first search, graph traversal	OptiType [133]	NetMHCpan-4.0 [116]	NA	Linux	Emacs Lisp	https://github.com/Benjamin-Vincent-Lab/NeoSplice	[103]
IRIS	Transcriptome, proteome	AS	STAR [198], rMATS [199]	seq2HLA [200]	IEDB predictor [150]	PSI-based screening [104], SJC-based screening [104]	Linux	Python	https://github.com/Xinglab/IRIS	[104]
SpliceMutr	Transcriptome	AS	STAR [198], LeafCutter [201]	arcasHLA [202]	MHCnuggets [203]	Splicing antigenicity metric [105]	Linux	R, Python	https://github.com/FertigLab/splicemutr	[105]

Note: SNV, single nucleotide variant; InDel, insertion/deletion; circRNA, circular RNA; AS, alternative splicing; NA, not available.

differ from self-antigens and can serve as neoantigens pending T cell recognition [72,73]. Although the formation of fusion genes is a relatively rare event compared to other somatic mutations, the neoantigens derived from fusion genes exhibit higher immunogenicity than those derived from pure mutations. This heightened immunogenicity makes fusion genes a compelling area of interest for immunotherapy development [70]. For instance, the *CBFB-MYH11* fusion neoantigen was found to be immunogenic and capable of enabling T cells to kill acute myeloid leukemia cells [74] (Figure 2B). In fibrolamellar carcinoma, *DNAJB1-PRKACA* fusion neoantigens, exhibiting similar immunogenic characteristics, have also been utilized with paired fusion-TCRs to better enhance the adoptive T cell therapies [75].

In tools and algorithms predicting fusion-derived neoantigens, RNA-seq data are requisite. Some tools, such as LENS, ScanNeo2, ProGeo-neo v2.0, nextNEOpi, TruNeo, TSNAD v2.0, and Seq2Neo, are originally designed for predicting SNV/InDel-derived neoantigens, as mentioned above. However, they also have the capability to predict neoantigens originating from fusion genes [63–69] (Table 1). After the prediction is performed, some tools also establish scoring systems to aid in selecting neoantigen candidates. For example, TruNeo incorporates deep learning models, the transport efficiency of transporter associated with antigen processing (TAP), and the peptide-MHC affinity to form a comprehensive scoring system. Candidate peptides passing the scoring system would be identified as neoantigens [67]. Other tools, such as INTEGRATE-Neo, NeoFuse, neoFusion, and pVACfuse, treat the prediction of fusion gene-derived neoantigens as an independent process [76–79]. For example, INTEGRATE-Neo first predicts gene fusion peptides and then performs neoantigen discovery within these candidate peptides based on HLA-binding affinity [76]. The ability to discern and predict neoantigens arising from fusion genes underscores the importance of these computational tools in helping researchers understand fusion events and their potential influence on cancer immunology.

Predicting neoantigens from circRNAs

The circRNAs, single-stranded continuous loops of RNAs spliced from linear RNAs, are usually derived from back-splicing events [80]. The peptide-encoding potentials of circRNAs, if they contain position-altered ORFs within the sequences, suggest their ability to produce neoantigens [81–85]. For example, the circTBC1D15-derived neoantigen has been proved to be presented by HLA-A, HLA-B, and HLA-C, drastically reducing the survival rate of the tumor organoid [86] (Figure 2C). Advancements in RNA-seq techniques, particularly those utilizing enriched non-polyadenylated transcriptomes, make the detection of circRNAs much easier [87].

While various methods have been developed to discover circRNAs, bioinformatics tools capable of predicting immunogenic peptides derived from circRNAs remain relatively scarce. One comprehensive pipeline for predicting such neoantigens is circRNA-derived neopeptide prediction pipeline (CIRC_neo). It integrates circRNA finders, coding potential identifiers, HLA-typing methods, and binding affinity predictors to form the whole pipeline, which is designed to sift through total RNA-seq or circRNA sequencing data to identify putative neopeptides from circRNAs [82]. Another tool, named circRNA coding capability and product detection

algorithm (CICADA), employs a machine learning model to discern circRNAs with high coding capability and high HLA-binding affinity. This approach enables the prediction of neoantigen potentials arising from circRNAs [88] (Table 1). As the understanding of circRNAs and their potential immunological significance continues to deepen, these computational tools play an increasingly important role in facilitating circRNA detection and exploring their role in the intricate network of neoantigen formation. Accumulated studies demonstrated that neoantigens could also be translated from some long non-coding RNAs (lncRNAs) [89,90]. Studies have integrated current tools like ORFfinder, CPAT, OrthoMCL, and NetMHCpan with algorithms like multi-state thermodynamic models to form the pipelines for detecting neoantigens from lncRNAs [91–96] (see below).

Predicting neoantigens from AS

AS refers to a regulated process in which a strand of pre-mRNA can be variably spliced into different mature messenger RNAs (mRNAs) through adjoining different exons, which gives rise to the phenomenon wherein a singular gene encodes multiple proteins [97]. Due to the highly regulated properties of AS in different tissues, some isoforms generated from the events can serve as tissue-specific or state-specific markers. AS events mainly include five modes: exon skipping, alternative 5' splice-site selection, alternative 3' splice-site selection, intron retention, and mutually exclusive exons [97]. Specifically, among these modes, intron retention may arise either through a dysregulated process impacting specific junctions or incomplete processing across the entire gene [98]. The alterations and dysregulation of AS events may contribute to increased complexities of transcriptomes as well as their translated products [99,100]. For example, Smart et al. identified peptides translated from retained introns as a source of cancer-specific neoantigens, such as peptide sequence SVLDDVRGW from the *TAF1* gene in leukemia cell line HL-60 [101]. A comprehensive study focused on neoantigens derived from AS neo-junctions has also highlighted the expanded target field for immunotherapy based on pan-cancer analysis [11]. Identifying such events and utilizing abnormal proteins as neoantigens might aid in finding new immunogenic targets.

Tools predicting neoantigens from AS include LENS, retained-intron-neoantigen-pipeline, alternative splicing neoantigens (ASNEO), NeoSplice, isoform peptides from RNA splicing for immunotherapy target screening (IRIS), and SpliceMutr [63,101–105] (Table 1). IRIS provides AS-derived neoantigen predictions together with CAR-T annotations, making the results more applicable to a clinical setting [104]. Among the tools, retained-intron-neoantigen-pipeline was the first to specifically focus on intron retentions. It generates peptide sequences after identifying retained intron loci, taking both the ORF orientation and normal retention events into account [101]. A newly developed pipeline, pVACsplice, will soon be incorporated into pVACtools to further expand the comprehensive neoantigen-predicting toolkit [79]. In these tools, researchers could use tumor-normal paired samples to distinguish those AS events specifically happening in tumor tissues [106]. For example, NeoSplice requires matched tumor and normal RNA-seq data as inputs to identify isoforms specifically appearing in tumor [103]. Tools like ASNEO and splicing neo antigen finder (SNAF) use the expression of normal tissues based on Genotype-Tissue

Expression (GTEx) portal data as references in case there is a lack of paired normal data [102,107].

Predicting neoantigens from RNA editing

RNA editing, whose dysregulation is frequent across different cancer types, involves chemically modifying RNA nucleotides [108–110] (Figure 2D). One common type of RNA editing converts A to I or C to U / U to C, which could change the amino acid sequences and produce “edited” peptides with potential immunogenic properties. The conversion of A to I is the most common editing pattern, producing accordingly altered peptides [110]. Recent studies have proved that the edited peptide from *CCNI* showed elevated abundances in several tumors and was capable of stimulating T cell responses in melanoma tumors when presented by HLA-A*02:01 [111] (Figure 2E).

The conversion can be predicted by a variety of algorithms, such as RNA/DNA difference prediction (RDDpred), RNA-DNA differences with support vector machines (RDDSVM), and plant RNA editing prediction & analysis computer tool (PREPACT; v2.0) [112–114]. Some tools, like A-to-I editing predictor (ATTIC), also provide “edited” peptide sequences after predicting RNA editing events [115]. Researchers have incorporated peptide-HLA affinity predictors, like NetMHCpan, used in aforementioned pipelines to design an RNA editing neoantigen immunogenicity score schema [116,117]. However, tools combining RNA editing prediction and subsequent neoantigen prediction are still lacking, calling for further development.

Predicting neoantigens from polymorphic microbiomes

In the latest dimensions of cancer hallmarks, polymorphic microbiomes have been regarded as new hallmarks [118]. Due to the dysfunctional status of tumor cells, various types of microbes may enter tumor cells to produce “non-self” proteins [119–122]. An example of such proteins is viral oncoproteins, defined as viral proteins with oncogenic properties and common in viral infection-induced cancers, such as cervical cancers derived from human papillomaviruses (HPVs) [123–125]. As typical pathogens markedly differ from human peptides, viral antigens tend to have a high affinity with TCRs [126]. Similar to viral oncoproteins, intratumoral bacteria and the gut microbiome can also produce specific peptides (Figure 2F). For example, neoantigen pACP1780 and pACP2283, discovered from gut microbiome-derived peptides, have been proved to inhibit tumor growth both *in vivo* and in mice [127]. In addition, some engineered bacteria, such as *Escherichia coli* genetically engineered to carry bacteria derived vesicle-neoantigen (BDVs-Neo), have also shown their potency in producing specific neoantigens and eliciting systemic anti-tumor immunity after entering host cells [128,129]. The discovery of the intratumor microbiome, spanning a diverse range of cancer types, has unveiled its dual functions in influencing cancer progression [122,130]. Some microbiome-derived proteins may be lysed to oligopeptides and presented by HLAs on tumor cell surface. Based on their exogenous characteristics, tools and algorithms predicting microbiome-derived neoantigens could skip the processes of identifying changed ORFs and immunogenic properties. Accordingly, assigning sequences to different microbial taxonomies becomes a main upstream process in predicting such neoantigens.

Among the aforementioned algorithms, LENS incorporates VirDetect to detect viral sequences from RNA-seq data [63,131]. Additionally, it uses BCFTools to detect homozygous germline variants and generate viral peptides. After the peptide has been identified, the binding affinity and stability can be tested to finally select neoantigens from candidate peptides [63]. Another tool, pVAC-Seq, originally developed to predict neoantigens from somatic mutations, has also demonstrated its potential utility in predicting bacterial and/or viral neoantigens [52,132]. Tools like ProGeo-neo have taken a comprehensive approach by including bacterial and viral peptides in their consideration of immunogenic HLA-binding peptides during prediction processes [61] (Table 1). Despite these advancements, it is noteworthy that with our current knowledge, there is a lack of specific tools or algorithms predicting intratumor bacteria-derived neoantigens, indicating a necessity for future research and tool development.

Incorporated tools and algorithms aiding in neoantigen prediction

HLA typing is usually the primary step for sample-specific neoantigen–HLA binding predictions (Table 1). OptiType and Polysolver are widely used for predicting HLA-I alleles; HLA-HD and HLaminer are similarly popular for predicting HLA-II alleles [133–136]. For neoantigen prediction tools that don't include HLA identification, they either require users to provide clinical files of sample-specific HLA alleles or use the most frequent HLA alleles among the human populations [57,58].

Predicting the binding affinity of peptide–MHC is a critical step in the selection and determination of candidate neoantigens, which decides whether these predicted peptides could be presented as antigens by HLAs. Almost all the tools for neoantigen discovery incorporated at least one algorithm to predict the binding affinity. Indeed, in the workflow of neoantigen predictors, 19 algorithms were included and shared by them, such as NetMHCpan, NetMHC, NetMHCII, and MHCflurry [116,94,137–140] (Figure 3; Table 1). Machine learning methods or neural networks are essential in these algorithms, in which MS data and/or sequencing data are included as weights for training processes to enhance the prediction performance of binding affinity scores. NetMHCpan, one of the most popular algorithms, was incorporated in 22 listed tools (Figure 3; Table 1). NetMHCpan performs artificial neural networks (ANNs) on the binding affinity values and MS eluted ligand data to build its model. With any given HLA typing or sequence with candidate peptides, it can predict allele-specific affinity scores based on this model, which can be used as an important filter excluding those peptides exhibiting low affinities toward HLA alleles [116]. Another commonly used algorithm is MHCflurry, incorporated in 7 listed tools (Figure 3; Table 1). To improve prediction performance, it separates the predictors into two categories: one for MHC allele-dependent effects (binding affinity prediction) and the other for allele-independent effects (AP prediction), and then combines the two predictors in a logistic regression model. Any MHC-I alleles with known sequences could be predicted by MHCflurry based on the combined predictor [140]. DeepHLApan, incorporated in TSNAD v2.0 and tumor-specific neoantigen database (TSNadb; v2.0), combines predicting peptide–MHC binding affinity and immunogenicity together, thus providing more options for

neoantigen selection [68,141,142]. Additionally, some tools mainly target the affinity between peptide sequences and MHC-II alleles. Taking NetMHCII as an example, it could perform predictions on 25 HLA-DR, 20 HLA-DQ, 9 HLA-DP, and 7 mouse H2 class II alleles once the peptide sequences are provided [139]. Besides the binding affinity, binding stability is also important for a peptide to be translocated from MHCs to cytotoxic T lymphocytes (CTLs), thus inducing an immune response. This type of algorithm, such as NetMHCstabpan, is also incorporated into some tools [143]. NetMHCstabpan has been trained on quantitative stability data and uses ANNs to build its predictor. Similar to NetMHCpan, given any HLA typing or sequence with the candidate peptides, it can provide the stability score of putative neoantigens [116,143]. With the binding affinity and stability predicted, the possibility of the AP process can be determined *in silico*. Peptides with high binding affinities and stabilities to HLA molecules are prioritized.

There are various criteria to prioritize neoantigen candidates. Immunogenicity, potential immune responses, and expression validation are widely considered (Table 1). Different methods may define their own criteria based on these aspects. Immunogenicity can be predicted by tools like immune epitope database and analysis resources (IEDB) immunogenicity and DeepHLApan [141,144]. Immune responses can be predicted by tools (*e.g.*, MiXCR) and algorithms (*e.g.*, immune fitness model) [145,146]. Some expression validation is represented by transcript per million (TPM) in transcriptome if the former prediction of neoantigen candidates is performed in genome data. Validation can also be represented by peptide search in proteome database or cohort-specific proteome datasets provided by users, using basic local alignment search tool-protein (BLASTP) or Peptide Match [147,148]. Tools like pVAC-Seq, ProGeo-neo, TruNeo, and TSNAD v2.0, could quantify and assign values to some aforementioned criteria, integrating HLA-binding affinity and stability with assigned weights to provide comprehensive scoring systems [52,61,67,142]. These algorithms, together with peptide performance predictors, are classified in Table 1 and featured in constructed data portals (Figure 3).

Data resources of tumor neoantigens

To date, plenty of neoantigens from different origins and across cancer types have been discovered and collected in databases based on the development of both computational and experimental strategies. Herein, we summarize these databases, including NeoTCR, IEDB, TSNadb v2.0, dbPepNeo2.0, NEPdb, Neodb, TSNadb, dbPepNeo, GNIFdb, cancer antigens database (CAD) v1.0, small peptides encoded by ncRNA from cancer patients (SPENCER), and TransLnc to provide references and resources for neoantigen search [53,95,96,142,149–155] (Table 2). These databases share some similar methods and provide general function modules, such as searching for HLAs and genes, browsing neoantigens, and downloading databases. Some databases also developed specific methods or modules. For example, dbPepNeo2.0 incorporates DeepCNN-Ineo — a deep learning model for predicting the immunogenicity of neoantigens — to aid users in identifying their own candidate neoantigens [151]. GNIFdb and CAD provide the visualization of the genome across cancer subtypes and MHC–peptide structures, respectively, to help users get a better

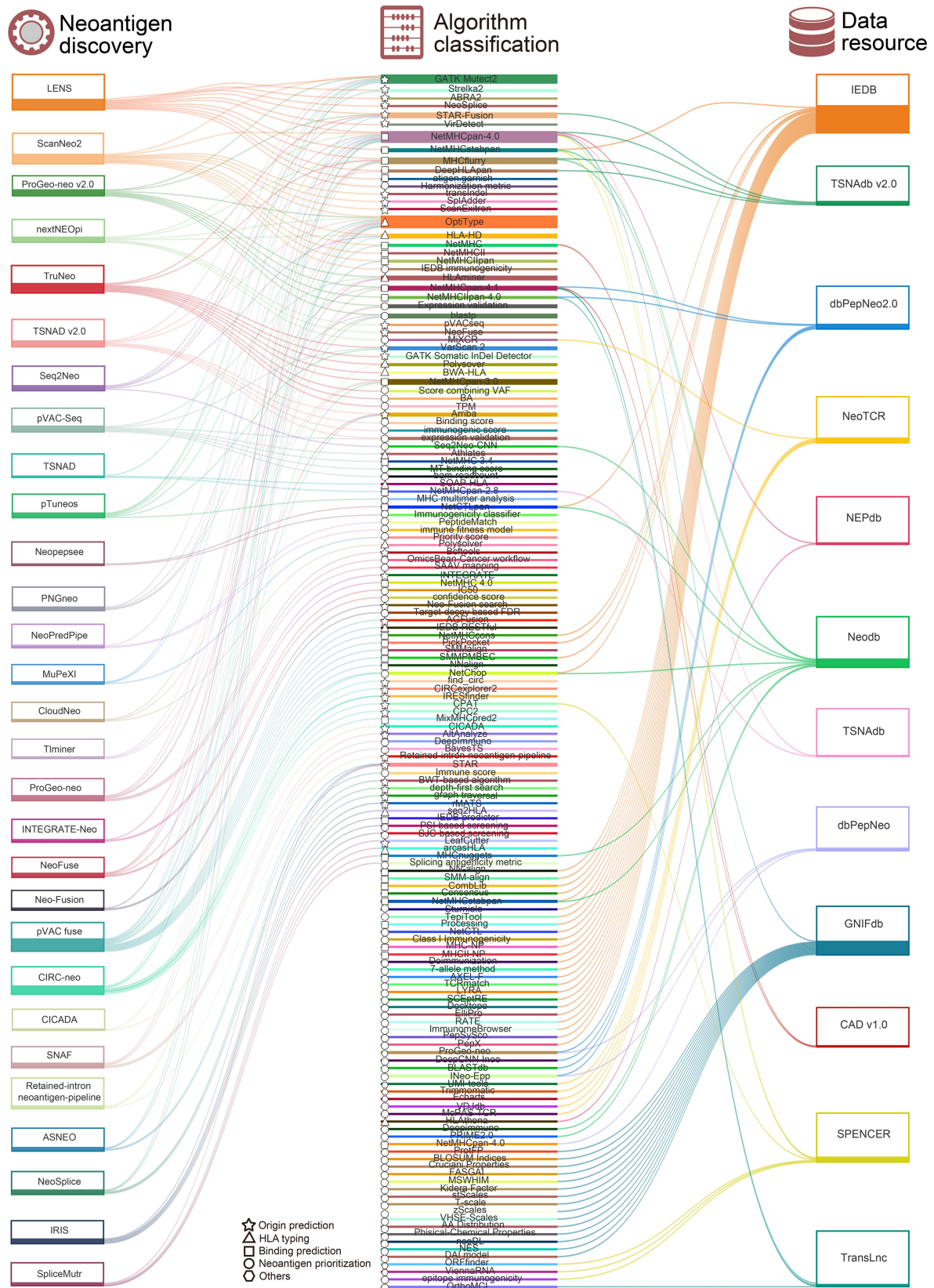


Figure 3 Summary of neoantigen discovery tools, affinity prediction methods, and neoantigen data portals

The figure illustrates the methods (middle bars) shared by neoantigen discovery tools (left bars) and databases (right bars). Symbols denote origin prediction (star), HLA typing (triangle), binding prediction (rectangle), neoantigen prioritization (circle), and others (hexagon). HLA, human leukocyte antigen; LENS, landscape of effective neoantigens software; nextNEOpI, nextflow neoantigen prediction pipeline; pVAC-Seq, personalized variant antigens by cancer sequencing; TSNAD, tumor-specific neoantigen detector; pTuneos, prioritizing tumor neoantigens; NeoPredPipe, neoantigen prediction pipeline; MuPeXI, mutant peptide extractor and informer; ProGeo-neo, proteogenomics prediction of neoantigen; CIRC_neo, circRNA-derived neopeptide prediction pipeline; CICADA, circRNA coding capability and product detection algorithm; ASNEO, alternative splicing neoantigens; SNAF, splicing neo antigen finder; GTeX, Genotype-Tissue Expression; TSNADB, tumor-specific neoantigen database; IEDB, immune epitope database and analysis resources; CAD, cancer antigens database; SPENCER, small peptides encoded by ncRNA from cancer patients.

Table 2 Resources of tumor neoantigens

Database	Algorithm and tool	Neoantigen origin	Link	Ref.
IEDB	NN-align [188], SMM-align [186], TepiTool [204], NetCTL [205], NetCTLpan [189], MHC-NP [206], MHCII-NP [207], DockTope [208], PepX [209]	Integrating multiple resources and origins	https://www.immuneepitope.org/	[144]
TSNAdb v2.0	NetMHCpan-4.0 [116], DeepHLApan [210], MHCflurry [140]	SNVs/InDels, gene fusions	https://pgx.zju.edu.cn/tsnadb/	[142]
dbPepNeo2.0	NetMHCpan-4.1 [94], NetMHCIIpan-4.0 [94], ProGeo-neo [61], DeepCNN-Ineo, BLAST [147], INeo-Epp [211]	SNVs/InDels, gene fusions	http://118.31.70.55/dbPepNeo2.0/	[151]
NeoTCR	UMI-tools [212], Trimmomatic [213], MiXCR [145], Echarts [214], VDjdb [215], McPAS-TCR [216]	SNVs/InDels, gene fusions, AS	http://neotcrdb.bioxai.cn/home	[149]
NEPdb	NetMHCpan-4.0 [116], HLAthena [217]	SNVs/InDels	http://nep.whu.edu.cn/	[152]
Neodb	IEDB [144], NetMHCpan-4.0 [116], NetMHCstabpan [218], MHCflurry [140], MHCnuggets [203], NetChop [219], NetCTLpan [180], Seq2Neo-CNN [69], DeepImmuno [195], PRIME2.0 [220]	SNVs/InDels	https://github.com/XSLiuLab/Neodb	[153]
TSNAdb	NetMHCpan-4.0 [116], NetMHCpan-2.8 [177]	SNVs/InDels	http://biopharm.zju.edu.cn/tsnadb/	[142]
dbPepNeo	NetMHCpan-4.0 [116], ProGeo-neo [61], INeo-Epp [211]	SNVs/InDels	http://www.biostatistics.online/dbPepNeo/index.php	[154]
NeoPeptide	NA	SNVs/InDels	https://github.com/lyotvincent/NeoPeptide	[221]
GNIFdb	protFP [157], BLOSUM indices [158], Cruciani Properties [159], FASGAI [222], MS-WHIM [157], Kidera Factor [223], stScales [224], T-scale [222], zScales [225], VHSE-scales [226], AA distribution [157], physical and chemical Properties [160], neoDL [227], NES [228]	SNVs/InDels	http://www.oncoimmunobank.cn/index.php	[155]
CAD v1.0	NetMHC [138], NetMHCpan-4.1 [94]	SNVs/InDels	http://cad.bio-it.cn/	[156]
SPENCER	CPAT [92], ViennaRNA [229], NetMHCpan-4.0 [116], NetMHCstabpan [143]	lncRNAs	http://spencer.renlab.org	[96]
TransLnc	NetMHCpan-4.1 [94], NetMHCIIpan-4.0 [94], OrthoMCL [93]	lncRNAs	http://bio-bigdata.hrbmu.edu.cn/TransLnc/	[95]

Note: lncRNA, long non-coding RNA.

understanding of sequences and products of different neoantigens [155,156]. GNIFdb also provides a list of non-antigens, including 482,109 peptides derived from 100 non-antigens by 9-mer sliding window, which would save users' time by excluding these non-antigen peptides [155].

Some database websites provide tools predicting affinities between neoantigens and HLAs (Figure 2; Table 2), such as NetMHCpan-4.0 in NEPdb, TSNAdb, Neodb, dbPepNeo, and CAD [53,116,142,152–154,156]. Additionally, some databases incorporate aforementioned algorithms and/or tools capable of predicting neoantigens, such as Neodb incorporating Seq2Neo and dbPepNeo incorporating ProGeo-neo [61,69,153,154]. Crossing and merging of algorithms, tools, and databases provide abundant resources for neoantigen discovery and exploration.

Among the databases, IEDB gathered antigen data from multiple resources and has been widely used in neoantigen predicting tools, such as IRIS, which identifies AS-derived neoantigens [104,150]. Multiple analysis tools targeting different aspects of neoantigen immunogenicity are presented by IEDB, in which users could choose the preferred tools for analyses. Another pan-cancer database, TSNAdb, incorporated predicated neoantigens derived from somatic mutations across 16 tumor types with 7748 tumor samples and provided shared neoantigens based on recurrent mutations among cancer types and prevalent HLA alleles [53]. The updated version, TSNAdb v2.0, also provided information on mutation types and experimental validations [142].

Unlike the broad information incorporation, GNIFdb specifically gathered neoantigens in glioma while providing a more detailed analysis for neoantigen selection, such as protFP descriptor, blocks substitution matrix (BLOSUM) indices, Cruciani properties, and physical and chemical property predictions [155,157–160]. SPENCER and TransLnc provide predicted neoantigen sequences derived from translative long non-coding regions [95,96]. Incorporating predicted and/or validated neoantigens into these data portals can aid researchers in discerning cancer-specific antigens. The comparative analysis between newly identified peptides and existing epitopes within these resources may also provide a more elucidating approach for exploring novel treatment.

Conclusion and perspectives

In conclusion, researchers have made remarkable advancements in neoantigen discovery and prediction in recent years, propelled by techniques like high-throughput sequencing, multi-omics profiling, and innovative computational tools and algorithms leveraging machine learning and deep learning [161]. The expansion of these tools and algorithms enriches our understanding of the immunogenic landscape associated with diverse origins of neoantigens, including somatic mutations (SNVs/InDels and gene fusions), circRNAs, AS, RNA editing, and polymorphic microbiomes. Most current tools focus on peptides derived from one or two origins, such as TIminer (SNVs/InDels) and NeoFuse (gene fusions)

[60,77]. However, only a few tools have integrated functions to predict peptides from multiple origins, such as LENS (SNVs/InDels, gene fusions, AS, and viruses) and ScanNeo2 (SNVs/InDels, gene fusions, and AS) [63,162]. Additionally, there is a lack of objective benchmarking in this field, which may make the selection of tools, algorithms, and data resources difficult for users. In this review, we conclude some of the characteristics, while a more detailed benchmark work is needed.

As for more specific predictions, existing tools have showcased adaptability in predicting neoantigens derived from RNA editing and polymorphic microbiomes. However, intact pipelines or novel algorithms predicting such neoantigens require further development. With regard to exogenous microbiome, the integration of these neoantigens with host immune profiling remains unexplored. Additionally, considering peptide homology with human proteins, it remains unclear whether such homology will affect later immune responses [163]. Advanced algorithms like deep learning have not yet been widely integrated into neoantigen prediction despite pioneering tools like TruNeo [67]. Perhaps integrating such algorithms could help improve the predictive accuracy and sensitivity of current algorithms. In summary, optimizing tools and algorithms for predicting these specific tumor neoantigens contributes to expanding the application of this field. In addition, the translation potentials of some regions formerly known as non-coding regions have recently been explored. Databases and resources are gathering predicted neoantigens from such regions [95,96], while the predicting processes were not integrated into user-friendly tools to make neoantigen discovery in other cancer datasets more efficient.

The consideration of incorporating different sequencing data is usually based on the origins of neoantigens. For example, the detection of somatic mutations requires genome data like whole-genome sequencing (WGS) or whole-exome sequencing (WES), while the detection of AS requires RNA-seq data. Expression validation for neoantigen prioritization is a great example of combining multi-omics data. Furthermore, proteome data would be the major contributor to improving reliability and decreasing false discovery rates for confirming the existence of translated peptides. Integrative multi-omics data, including genome, transcriptome, and proteome, could improve the accuracy and reliability of neoantigen predictions. For example, nextNEOpI utilizes WES/WGS data to predict mutations and HLA types. It also utilizes RNA-seq data to predict gene fusions and to reconfirm the HLA types predicted by WES/WGS, which further improves the accuracy of later affinity prediction [66]. The combination of multi-omics data not only serves data-specific prediction needs but also refines the accuracy of predictions that could obtain different sources of data. Additionally, the enhancement of resolution based on single-cell RNA sequencing (scRNA-seq) may provide a better understanding of neoantigens. Recent studies have shown the utility of scRNA-seq, *i.e.*, Smart-seq2, in neoantigen prediction [107,164]. For example, researchers tried to use SNAF to identify neo-junctions that could become neoantigens in terms of tumor cells in multiple cell types from both bulk RNA-seq and scRNA-seq data [107]. Specific tools or algorithms in mining neoantigens from scRNA-seq data still need further exploration.

Furthermore, higher-throughput exploration based on experiments is required to fully identify the functions of

neoantigens and provide reliable background on *in silico* predictions afterward. Database resources incorporating experimental details like TSNAdb v2.0 also call for more combinations of computational predictions and experimental validations [142]. With multi-omics data support, integrated predictors, improved algorithms, and experimental validations, more tumor neoantigens with potential clinical functions will be identified to advance drug discovery and develop novel therapeutical strategies.

CRediT author statement

Yunzhe Wang: Writing – original draft, Writing – review & editing, Visualization, Data curation. **James Wengler:** Writing – original draft. **Yuzhu Fang:** Visualization, Data curation. **Joseph Zhou:** Writing – review & editing. **Hang Ruan:** Writing – original draft. **Zhao Zhang:** Funding acquisition, Supervision, Writing – review & editing. **Leng Han:** Conceptualization, Supervision, Project administration, Writing – review & editing. All authors have read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 32270708 and 82203195). This work was supported by the Medical Science Data Center at Shanghai Medical College of Fudan University, China.

ORCID

0009-0000-8726-3409 (Yunzhe Wang)
 0000-0002-9641-4560 (James Wengler)
 0009-0004-2086-8170 (Yuzhe Fang)
 0009-0006-6994-2639 (Joseph Zhou)
 0000-0001-7506-3805 (Hang Ruan)
 0000-0002-3757-8469 (Zhao Zhang)
 0000-0002-7380-2640 (Leng Han)

References

- [1] Bonilla FA, Oettgen HC. Adaptive immunity. *J Allergy Clin Immunol* 2010;125:S33–40.
- [2] Moser M, Leo O. Key concepts in immunology. *Vaccine* 2010; 28:C2–13.
- [3] Lurquin C, Van Pel A, Mariamé B, De Plaen E, Szikora JP, Janssens C, et al. Structure of the gene of tum^r transplantation antigen P91A: the mutated exon encodes a peptide recognized with L^d by cytolytic T cells. *Cell* 1989;58:293–303.
- [4] Lu YC, Robbins PF. Cancer immunotherapy targeting neoantigens. *Semin Immunol* 2016;28:22–7.
- [5] Jhunjhunwala S, Hammer C, Delamarre L. Antigen presentation in cancer: insights into tumour immunogenicity and immune evasion. *Nat Rev Cancer* 2021;21:298–312.
- [6] Dhatchinamoorthy K, Colbert JD, Rock KL. Cancer immune evasion through loss of MHC class I antigen presentation. *Front Immunol* 2021;12:636568.

- [7] Falk K, Rötzschke O, Stevanović S, Jung G, Rammensee HG. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* 1991;351:290–6.
- [8] Thorsby E. MHC structure and function. *Transplant Proc* 1999; 31:713–6.
- [9] Arase H. Rheumatoid rescue of misfolded cellular proteins by MHC class II molecules: a new hypothesis for autoimmune diseases. *Adv Immunol* 2016;129:1–23.
- [10] Rock KL, Reits E, Neefjes J. Present yourself! By MHC class I and MHC class II molecules. *Trends Immunol* 2016;37:724–37.
- [11] Shukla SA, Rooney MS, Rajasagi M, Tiao G, Dixon PM, Lawrence MS, et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat Biotechnol* 2015;33:1152–8.
- [12] Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. *BMC Immunol* 2008;9:1.
- [13] Marty R, Kaabinejadian S, Rossell D, Slifker MJ, van de Haar J, Engin HB, et al. MHC-I genotype restricts the oncogenic mutational landscape. *Cell* 2017;171:1272–83.e15.
- [14] Le Gall I, Marcadet A, Font MP, Auffray C, Strominger J, Lalouel JM, et al. Exuberant restriction fragment length polymorphism associated with the DQ alpha-chain gene and the DX alpha-chain gene. *Proc Natl Acad Sci U S A* 1985;82:5433–6.
- [15] Falk K, Rötzschke O, Stevanović S, Jung G, Rammensee HG. Pool sequencing of natural HLA-DR, DQ, and DP ligands reveals detailed peptide motifs, constraints of processing, and general rules. *Immunogenetics* 1994;39:230–42.
- [16] Marrack P, Kappler J. The T cell receptor. *Science* 1987; 238:1073–9.
- [17] Jojic N, Reyes-Gomez M, Heckerman D, Kadie C, Schueler-Furman O. Learning MHC I-peptide binding. *Bioinformatics* 2006;22:e227–35.
- [18] Unanue ER, Turk V, Neefjes J. Variations in MHC class II antigen processing and presentation in health and disease. *Annu Rev Immunol* 2016;34:265–97.
- [19] Jing Y, Yang J, Johnson DB, Moslehi JJ, Han L. Harnessing big data to characterize immune-related adverse events. *Nat Rev Clin Oncol* 2022;19:269–80.
- [20] Topalian SL, Drake CG, Pardoll DM. Immune checkpoint blockade: a common denominator approach to cancer therapy. *Cancer Cell* 2015;27:450–61.
- [21] Hu Q, Ye Y, Chan LC, Li Y, Liang K, Lin A, et al. Oncogenic lncRNA downregulates cancer cell antigen presentation and intrinsic tumor suppression. *Nat Immunol* 2019;20:835–51.
- [22] Chae YK, Viveiros P, Lopes G, Sukhadia B, Sheikh MM, Saravia D, et al. Clinical and immunological implications of frameshift mutations in lung cancer. *J Thorac Oncol* 2019;14:1807–17.
- [23] Puig-Saus C, Sennino B, Peng S, Wang CL, Pan Z, Yuen B, et al. Neoantigen-targeted CD8⁺ T cell responses with PD-1 blockade therapy. *Nature* 2023;615:697–704.
- [24] Caushi JX, Zhang J, Ji Z, Vaghasia A, Zhang B, Hsiue EHC, et al. Transcriptional programs of neoantigen-specific TIL in anti-PD-1-treated lung cancers. *Nature* 2021;596:126–32.
- [25] Zhang Z, Luo M, Li Q, Liu Y, Lussier C, Zhang J, et al. Genetic, pharmacogenomic, and immune landscapes of enhancer RNAs across human cancers. *Cancer Res* 2022;82:785–90.
- [26] Blankenstein T, Leisegang M, Uckert W, Schreiber H. Targeting cancer-specific mutations by T cell receptor gene therapy. *Curr Opin Immunol* 2015;33:112–9.
- [27] Leidner R, Sanjuan Silva N, Huang H, Sprott D, Zheng C, Shih YP, et al. Neoantigen T-cell receptor gene therapy in pancreatic cancer. *N Engl J Med* 2022;386:2112–9.
- [28] Yang J, Chen Y, Jing Y, Green MR, Han L. Advancing CAR T cell therapy through the use of multidimensional omics data. *Nat Rev Clin Oncol* 2023;20:211–28.
- [29] Mehrabadi AZ, Ranjbar R, Farzanehpour M, Shahriary A, Dorostkar R, Hamidinejad MA, et al. Therapeutic potential of CAR T cell in malignancies: a scoping review. *Biomed Pharmacother* 2022;146:112512.
- [30] Jing Y, Liu Y, Li Q, Ye Y, Diao L, Huang Y, et al. Expression of chimeric antigen receptor therapy targets detected by single-cell sequencing of normal cells may contribute to off-tumor toxicity. *Cancer Cell* 2021;39:1558–9.
- [31] Ott PA, Dotti G, Yee C, Goff SL. An update on adoptive T-cell therapy and neoantigen vaccines. *Am Soc Clin Oncol Educ Book* 2019;39:e70–8.
- [32] Xie N, Shen G, Gao W, Huang Z, Huang C, Fu L. Neoantigens: promising targets for cancer therapy. *Signal Transduct Target Ther* 2023;8:9.
- [33] Mount CW, Majzner RG, Sundaresh S, Arnold EP, Kadapakkam M, Haile S, et al. Potent antitumor efficacy of anti-GD2 CAR T cells in H3-K27M⁺ diffuse midline gliomas. *Nat Med* 2018;24:572–9.
- [34] Xie G, Ivica NA, Jia B, Li Y, Dong H, Liang Y, et al. CAR-T cells targeting a nucleophosmin neoepitope exhibit potent specific activity in mouse models of acute myeloid leukaemia. *Nat Biomed Eng* 2021;5:399–413.
- [35] Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science* 2015;348:69–74.
- [36] Shemesh CS, Hsu JC, Hosseini I, Shen BQ, Rotte A, Twomey P, et al. Personalized cancer vaccines: clinical landscape, challenges, and opportunities. *Mol Ther* 2021;29:555–70.
- [37] Wirth TC, Kühnel F. Neoantigen targeting — dawn of a new era in cancer immunotherapy? *Front Immunol* 2017;8:1848.
- [38] Lilleby W, Gaudernack G, Brunsvig PF, Vlatkovic L, Schulz M, Mills K, et al. Phase I/IIa clinical trial of a novel hTERT peptide vaccine in men with metastatic hormone-naïve prostate cancer. *Cancer Immunol Immunother* 2017;66:891–901.
- [39] De Mattos-Arruda L, Vazquez M, Finotello F, Lepore R, Porta E, Hundal J, et al. Neoantigen prediction and computational perspectives towards clinical benefit: recommendations from the ESMO Precision Medicine Working Group. *Ann Oncol* 2020; 31:978–90.
- [40] Zhang X, Goedegebuure SP, Chen MY, Mishra R, Zhang F, Yu YY, et al. Neoantigen DNA vaccines are safe, feasible, and induce neoantigen-specific immune responses in triple-negative breast cancer patients. *Genome Med* 2024;16:131.
- [41] Boegel S, Castle JC, Kodysh J, O'Donnell T, Rubinsteyn A. Bioinformatic methods for cancer neoantigen prediction. *Prog Mol Biol Transl Sci* 2019;164:25–60.
- [42] Peng M, Mo Y, Wang Y, Wu P, Zhang Y, Xiong F, et al. Neoantigen vaccine: an emerging tumor immunotherapy. *Mol Cancer* 2019;18:128.
- [43] Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res* 2017;45:D777–83.
- [44] Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. *Science* 2015;349:1483–9.
- [45] Parkhurst MR, Robbins PF, Tran E, Prickett TD, Gartner JJ, Jia L, et al. Unique neoantigens arise from somatic mutations in patients with gastrointestinal cancers. *Cancer Discov* 2019; 9:1022–35.
- [46] Ishino T, Kawashima S, Tanji E, Ueno T, Ueda Y, Ogasawara S, et al. Somatic mutations can induce a noninflamed tumour microenvironment via their original gene functions, despite deriving neoantigens. *Br J Cancer* 2023;128:1166–75.
- [47] Chu D, Wei L. Nonsynonymous, synonymous and nonsense mutations in human cancer-related genes undergo stronger purifying selections than expectation. *BMC Cancer* 2019;19:359.
- [48] Greenman C, Stephens P, Smith R, Dalglish GL, Hunter C, Bignell G, et al. Patterns of somatic mutation in human cancer genomes. *Nature* 2007;446:153–8.
- [49] Hsiue EHC, Wright KM, Douglass J, Hwang MS, Mog BJ, Pearlman AH, et al. Targeting a neoantigen derived from a common TP53 mutation. *Science* 2021;371:eabc8697.

- [50] Kloor M, Reuschenbach M, Pauligk C, Karbach J, Rafiyan MR, Al-Batran SE, et al. A frameshift peptide neoantigen-based vaccine for mismatch repair-deficient cancers: a phase I/IIa clinical trial. *Clin Cancer Res* 2020;26:4503–10.
- [51] Alban TJ, Chan TA. Immunotherapy biomarkers: the long and winding road. *Nat Rev Clin Oncol* 2021;18:323–4.
- [52] Hundal J, Carreno BM, Petti AA, Linette GP, Griffith OL, Mardis ER, et al. pVAC-Seq: a genome-guided *in silico* approach to identifying tumor neoantigens. *Genome Med* 2016; 8:11.
- [53] Wu J, Zhao W, Zhou B, Su Z, Gu X, Zhou Z, et al. TSNADB: a database for tumor-specific neoantigens from immunogenomics data analysis. *Genomics Proteomics Bioinformatics* 2018; 16:276–82.
- [54] 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;11:67.
- [55] Kim S, Kim HS, Kim E, Lee MG, Shin EC, Paik S, et al. NeoPeepsee: accurate genome-level prediction of neoantigens by harnessing sequence and amino acid immunogenicity information. *Ann Oncol* 2018;29:1030–6.
- [56] Tan X, Xu L, Jian X, Ouyang J, Hu B, Yang X, et al. PGNNeo: a proteogenomics-based neoantigen prediction pipeline in non-coding regions. *Cells* 2023;12:782.
- [57] Schenck RO, Lakatos E, Gatenbee C, Graham TA, Anderson ARA. NeoPredPipe: high-throughput neoantigen prediction and recognition potential pipeline. *BMC Bioinformatics* 2019; 20:264.
- [58] Bjerregaard AM, Nielsen M, Hadrup SR, Szallasi Z, Eklund AC. MuPeXI: prediction of neo-epitopes from tumor sequencing data. *Cancer Immunol Immunother* 2017;66:1123–30.
- [59] Bais P, Namburi S, Gatti DM, Zhang X, Chuang JH. CloudNeo: a cloud pipeline for identifying patient-specific tumor neoantigens. *Bioinformatics* 2017;33:3110–2.
- [60] Tappeiner E, Finotello F, Charoentong P, Mayer C, Rieder D, Trajanoski Z. TIminer: NGS data mining pipeline for cancer immunology and immunotherapy. *Bioinformatics* 2017;33:3140–1.
- [61] Li Y, Wang G, Tan X, Ouyang J, Zhang M, Song X, et al. ProGeo-neo: a customized proteogenomic workflow for neoantigen prediction and selection. *BMC Med Genomics* 2020; 13:52.
- [62] Chen C, Liu Y, Li Q, Zhang Z, Luo M, Liu Y, et al. The genetic, pharmacogenomic, and immune landscapes associated with protein expression across human cancers. *Cancer Res* 2023; 83:3673–80.
- [63] Vensko SP, Olsen K, Bortone D, Smith CC, Chai S, Beckabir W, et al. LENS: Landscape of Effective Neoantigens Software. *Bioinformatics* 2023;39:btad322.
- [64] Schäfer RA, Guo Q, Yang R. ScanNeo2: a comprehensive workflow for neoantigen detection and immunogenicity prediction from diverse genomic and transcriptomic alterations. *Bioinformatics* 2023;39:btad659.
- [65] Liu C, Zhang Y, Jian X, Tan X, Lu M, Ouyang J, et al. ProGeo-neo v2.0: a one-stop software for neoantigen prediction and filtering based on the proteogenomics strategy. *Genes (Basel)* 2022;13:783.
- [66] Rieder D, Fotakis G, Ausserhofer M, René G, Paster W, Trajanoski Z, et al. nextNEOpI: a comprehensive pipeline for computational neoantigen prediction. *Bioinformatics* 2022; 38:1131–2.
- [67] Tang Y, Wang Y, Wang J, Li M, Peng L, Wei G, et al. TruNeo: an integrated pipeline improves personalized true tumor neoantigen identification. *BMC Bioinformatics* 2020;21:532.
- [68] Zhou Z, Wu J, Ren J, Chen W, Zhao W, Gu X, et al. TSNAD v2.0: a one-stop software solution for tumor-specific neoantigen detection. *Comput Struct Biotechnol J* 2021;19:4510–6.
- [69] Diao K, Chen J, Wu T, Wang X, Wang G, Sun X, et al. Seq2Neo: a comprehensive pipeline for cancer neoantigen immunogenicity prediction. *Int J Mol Sci* 2022;23:11624.
- [70] Latysheva NS, Babu MM. Discovering and understanding oncogenic gene fusions through data intensive computational approaches. *Nucleic Acids Res* 2016;44:4487–503.
- [71] Luo M, Liu Y, Hermida LC, Gertz EM, Zhang Z, Li Q, et al. Race is a key determinant of the human intratumor microbiome. *Cancer Cell* 2022;40:901–2.
- [72] Wang Y, Shi T, Song X, Liu B, Wei J. Gene fusion neoantigens: emerging targets for cancer immunotherapy. *Cancer Lett* 2021; 506:45–54.
- [73] Yotnda P, Firat H, Garcia-Pons F, Garcia Z, Gourru G, Vernant JP, et al. Cytotoxic T cell response against the chimeric p210 BCR-ABL protein in patients with chronic myelogenous leukemia. *J Clin Invest* 1998;101:2290–6.
- [74] Biernacki MA, Foster KA, Woodward KB, Coon ME, Cummings C, Cunningham TM, et al. CFBF-MYH11 fusion neoantigen enables T cell recognition and killing of acute myeloid leukemia. *J Clin Invest* 2020;130:5127–41.
- [75] Kirk AM, Chou CH, Crawford JC, Awad W, Allen EK, Zhang X, et al. Characterization of CD8 T cell responses to DNAJB1-PRKACA fusion neoantigens in fibrolamellar carcinoma. *Cancer Res* 2022;82:1383.
- [76] Zhang J, Mardis ER, Maher CA. INTEGRATE-neo: a pipeline for personalized gene fusion neoantigen discovery. *Bioinformatics* 2017;33:555–7.
- [77] Fotakis G, Rieder D, Haider M, Trajanoski Z, Finotello F. NeoFuse: predicting fusion neoantigens from RNA sequencing data. *Bioinformatics* 2020;36:2260–1.
- [78] Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009;25:1754–60.
- [79] Hundal J, Kiwala S, McMichael J, Miller CA, Xia H, Wollam AT, et al. pVACtools: a computational toolkit to identify and visualize cancer neoantigens. *Cancer Immunol Res* 2020;8:409–20.
- [80] Chen LL. The expanding regulatory mechanisms and cellular functions of circular RNAs. *Nat Rev Mol Cell Biol* 2020; 21:475–90.
- [81] Tan S, Gou Q, Pu W, Guo C, Yang Y, Wu K, et al. Circular RNA F-circEA produced from *EML4-ALK* fusion gene as a novel liquid biopsy biomarker for non-small cell lung cancer. *Cell Res* 2018;28:693–5.
- [82] Xia J, Li S, Ren B, Zhang P. Circular RNAs as a potential source of neoepitopes in cancer. *Front Oncol* 2023;13:1098523.
- [83] Li M, Wang Y, Wu P, Zhang S, Gong Z, Liao Q, et al. Application prospect of circular RNA-based neoantigen vaccine in tumor immunotherapy. *Cancer Lett* 2023;563:216190.
- [84] Su M, Xiao Y, Ma J, Tang Y, Tian B, Zhang Y, et al. Circular RNAs in cancer: emerging functions in hallmarks, stemness, resistance and roles as potential biomarkers. *Mol Cancer* 2019; 18:90.
- [85] Dong Y, Gao Q, Chen Y, Zhang Z, Du Y, Liu Y, et al. Identification of circRNA signature associated with tumor immune infiltration to predict therapeutic efficacy of immunotherapy. *Nat Commun* 2023;14:2540.
- [86] Wang W, Ma L, Xing Z, Yuan T, Bao J, Zhu Y, et al. Tumor-specific circRNA-derived antigen peptide identification for hepatobiliary tumors. *Engineering* 2023;22:159–70.
- [87] Liu CX, Chen LL. Circular RNAs: characterization, cellular roles, and applications. *Cell* 2022;185:2016–34.
- [88] Fan L, Zhou X, Li M, Gao A, Yu H, Tian H, et al. CICADA: a circRNA effort to ghost proteome. *Nucleic Acids Res* 2025; 53:gkae1179.
- [89] Li G, Kryczek I, Nam J, Li X, Li S, Li J, et al. LIMIT is an immunogenic lncRNA in cancer immunity and immunotherapy. *Nat Cell Biol* 2021;23:526–37.
- [90] Barczak W, Carr SM, Liu G, Munro S, Nicastrì A, Lee LN, et al. Long non-coding RNA-derived peptides are immunogenic and drive a potent anti-tumour response. *Nat Commun* 2023; 14:1078.
- [91] Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, et al. Database resources of the National Center for

- Biotechnology Information. *Nucleic Acids Res* 2011; 39:D38–51.
- [92] Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res* 2013;41:e74.
- [93] Li L, Stoekert CJ Jr, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 2003; 13:2178–89.
- [94] 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:W449–54.
- [95] Lv D, Chang Z, Cai Y, Li J, Wang L, Jiang Q, et al. TransLnc: a comprehensive resource for translatable lncRNAs extends immunopeptidome. *Nucleic Acids Res* 2022;50:D413–20.
- [96] Luo X, Huang Y, Li H, Luo Y, Zuo Z, Ren J, et al. SPENCER: a comprehensive database for small peptides encoded by noncoding RNAs in cancer patients. *Nucleic Acids Res* 2022; 50:D1373–81.
- [97] Frankiw L, Baltimore D, Li G. Alternative mRNA splicing in cancer immunotherapy. *Nat Rev Immunol* 2019;19:675–87.
- [98] Wong JJJ, Ritchie W, Ebner OA, Selbach M, Wong JW, Huang Y, et al. Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* 2013;154:583–95.
- [99] Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, et al. Function of alternative splicing. *Gene* 2013;514:1–30.
- [100] Wang L, Shamardani K, Babikir H, Catalan F, Nejo T, Chang S, et al. The evolution of alternative splicing in glioblastoma under therapy. *Genome Biol* 2021;22:48.
- [101] Smart AC, Margolis CA, Pimentel H, He MX, Miao D, Adeegbe D, et al. Intron retention is a source of neoepitopes in cancer. *Nat Biotechnol* 2018;36:1056–8.
- [102] Zhang Z, Zhou C, Tang L, Gong Y, Wei Z, Zhang G, et al. ASNEO: identification of personalized alternative splicing based neoantigens with RNA-seq. *Aging* 2020;12:14633–48.
- [103] Chai S, Smith CC, Kochar TK, Hunsucker SA, Beck W, Olsen KS, et al. NeoSplice: a bioinformatics method for prediction of splice variant neoantigens. *Bioinform Adv* 2022;2:vbac032.
- [104] Pan Y, Phillips JW, Zhang BD, Noguchi M, Kutschera E, McLaughlin J, et al. IRIS: discovery of cancer immunotherapy targets arising from pre-mRNA alternative splicing. *Proc Natl Acad Sci U S A* 2023;120:e2221116120.
- [105] Palmer T, Kessler MD, Shao XM, Balan A, Yarchoan M, Zaidi N, et al. SpliceMutr enables pan-cancer analysis of splicing-derived neoantigen burden in tumors. *Cancer Res Commun* 2024;4:3137–50.
- [106] Richters MM, Cotto KC, Kiwala S, Xia H, Carreno BM, Dunn GP, et al. pVACsplice: predicting neoantigens from tumor-specific alternative splicing events derived from *cis*-acting regulatory mutations using whole exome and RNA sequencing data. *Cancer Res* 2022;82:1215.
- [107] Li G, Mahajan S, Ma S, Jeffery ED, Zhang X, Bhattacharjee A, et al. Splicing neoantigen discovery with SNAF reveals shared targets for cancer immunotherapy. *Sci Transl Med* 2024; 16:eade2886.
- [108] Han L, Diao L, Yu S, Xu X, Li J, Zhang R, et al. The genomic landscape and clinical relevance of A-to-I RNA editing in human cancers. *Cancer Cell* 2015;28:515–28.
- [109] Fumagalli D, Gacquer D, Rothé F, Lefort A, Libert F, Brown D, et al. Principles governing A-to-I RNA editing in the breast cancer transcriptome. *Cell Rep* 2015;13:277–89.
- [110] Paz-Yaacov N, Bazak L, Buchumenski I, Porath HT, Danan-Gotthold M, Knisbacher BA, et al. Elevated RNA editing activity is a major contributor to transcriptomic diversity in tumors. *Cell Rep* 2015;13:267–76.
- [111] Zhang M, Fritsche J, Roszik J, Williams LJ, Peng X, Chiu Y, et al. RNA editing derived epitopes function as cancer antigens to elicit immune responses. *Nat Commun* 2018;9:3919.
- [112] Kim M, Hur B, Kim S. RDDpred: a condition-specific RNA-editing prediction model from RNA-seq data. *BMC Genomics* 2016;17:5.
- [113] Tac HA, Koroglu M, Sezerman U. RDDSV: accurate prediction of A-to-I RNA editing sites from sequence using support vector machines. *Funct Integr Genomics* 2021;21:633–43.
- [114] Lenz H, Knoop V. PREPACT 2.0: predicting C-to-U and U-to-C RNA editing in organelle genome sequences with multiple references and curated RNA editing annotation. *Bioinform Biol Insights* 2013;7:1–19.
- [115] Chen R, Li F, Guo X, Bi Y, Li C, Pan S, et al. ATTIC is an integrated approach for predicting A-to-I RNA editing sites in three species. *Brief Bioinform* 2023;24:bbad170.
- [116] 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;199:3360–8.
- [117] Zhou C, Wei Z, Zhang L, Yang Z, Liu Q. Systematically characterizing A-to-I RNA editing neoantigens in cancer. *Front Oncol* 2020;10:593989.
- [118] Hanahan D. Hallmarks of cancer: new dimensions. *Cancer Discov* 2022;12:31–46.
- [119] D’Ippolito E, Schober K, Nauwerth M, Busch DH. T cell engineering for adoptive T cell therapy: safety and receptor avidity. *Cancer Immunol Immunother* 2019;68:1701–12.
- [120] Jung K, Son MJ, Lee SY, Kim JA, Ko DH, Yoo S, et al. Antibody-mediated delivery of a viral MHC-I epitope into the cytosol of targeted tumor cells repurposes virus-specific CD8⁺ T cells for cancer immunotherapy. *Mol Cancer* 2022;21:102.
- [121] Zhang Z, Gao Q, Ren X, Luo M, Liu Y, Liu P, et al. Characterization of intratumor microbiome in cancer immunotherapy. *Innovation (Camb)* 2023;4:100482.
- [122] Kalaora S, Nagler A, Nejman D, Alon M, Barbolin C, Barnea E, et al. Identification of bacteria-derived HLA-bound peptides in melanoma. *Nature* 2021;592:138–43.
- [123] Barrios K, Celis E. TriVax-HPV: an improved peptide-based therapeutic vaccination strategy against human papillomavirus-induced cancers. *Cancer Immunol Immunother* 2012;61:1307–17.
- [124] Moody CA, Laimins LA. Human papillomavirus oncoproteins: pathways to transformation. *Nat Rev Cancer* 2010;10:550–60.
- [125] Leko V, Rosenberg SA. Identifying and targeting human tumor antigens for T cell-based immunotherapy of solid tumors. *Cancer Cell* 2020;38:454–72.
- [126] 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;42:3174–9.
- [127] Ma Y, Liu X, Zhang X, Yu Y, Li Y, Song M, et al. Efficient mining of anticancer peptides from gut metagenome. *Adv Sci (Weinh)* 2023;10:e2300107.
- [128] Boesch M, Baty F, Rothschild SI, Tamm M, Joerger M, Früh M, et al. Tumour neoantigen mimicry by microbial species in cancer immunotherapy. *Br J Cancer* 2021;125:313–23.
- [129] Meng F, Li L, Zhang Z, Lin Z, Zhang J, Song X, et al. Biosynthetic neoantigen displayed on bacteria derived vesicles elicit systemic antitumour immunity. *J Extracell Vesicles* 2022; 11:e12289.
- [130] Sepich-Poore GD, Zitvogel L, Straussman R, Hasty J, Wargo JA, Knight R. The microbiome and human cancer. *Science* 2021; 371:eabc4552.
- [131] Selitsky SR, Marron D, Hollern D, Mose LE, Hoadley KA, Jones C, et al. Virus expression detection reveals RNA-sequencing contamination in TCGA. *BMC Genomics* 2020;21:79.
- [132] 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;18:414.
- [133] Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* 2014;30:3310–6.

- [134] Matey-Hernandez ML, Danish Pan Genome Consortium, Brunak S, Izarzugaza JMG. Benchmarking the HLA typing performance of Polysolver and OptiType in 50 Danish parental trios. *BMC Bioinformatics* 2018;19:239.
- [135] Kawaguchi S, Higasa K, Shimizu M, Yamada R, Matsuda F. HLA-HD: an accurate HLA typing algorithm for next-generation sequencing data. *Hum Mutat* 2017;38:788–97.
- [136] Warren RL, Choe G, Freeman DJ, Castellarin M, Munro S, Moore R, et al. Derivation of HLA types from shotgun sequence datasets. *Genome Med* 2012;4:95.
- [137] Nielsen M, Andreatta M. NetMHCpan-3.0: improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med* 2016;8:33.
- [138] Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acids Res* 2008;36:W509–12.
- [139] Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, et al. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* 2018;154:394–406.
- [140] O'Donnell TJ, Rubinsteyn A, Bonsack M, Riemer AB, Laserson U, Hammerbacher J. MHCflurry: open-source class I MHC binding affinity prediction. *Cell Syst* 2018;7:129–32.e4.
- [141] Wu J, Wang W, Zhang J, Zhou B, Zhao W, Su Z, et al. DeepHLApan: a deep learning approach for neoantigen prediction considering both HLA-peptide binding and immunogenicity. *Front Immunol* 2019;10:2559.
- [142] Wu J, Chen W, Zhou Y, Chi Y, Hua X, Wu J, et al. TSNAdb v2.0: the updated version of tumor-specific neoantigen database. *Genomics Proteomics Bioinformatics* 2023;21:259–66.
- [143] 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;197:1517–24.
- [144] 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:D339–43.
- [145] Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods* 2015;12:380–1.
- [146] Łuksza M, Riaz N, Makarov V, Balachandran VP, Hellmann MD, Solovoyov A, et al. A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature* 2017;551:517–20.
- [147] Mount DW. Using the basic local alignment search tool (BLAST). *CSH Protoc* 2007;2007:pdb.top17.
- [148] Chen C, Li Z, Huang H, Suzek BE, Wu CH, UniProt Consortium. A fast peptide match service for UniProt Knowledgebase. *Bioinformatics* 2013;29:2808–9.
- [149] Zhou W, Xiang W, Yu J, Ruan Z, Pan Y, Wang K, et al. NeoTCR: an immunoinformatic database of experimentally-supported functional neoantigen-specific TCR sequences. *Genomics Proteomics Bioinformatics* 2025;23:qzae010.
- [150] Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The Immune Epitope Database (IEDB) 3.0. *Nucleic Acids Res* 2015;43:D405–12.
- [151] Lu M, Xu L, Jian X, Tan X, Zhao J, Liu Z, et al. dbPepNeo2.0: a database for human tumor neoantigen peptides from mass spectrometry and TCR recognition. *Front Immunol* 2022;13:855976.
- [152] Xia J, Bai P, Fan W, Li Q, Li Y, Wang D, et al. NEPdb: a database of T-cell experimentally-validated neoantigens and pan-cancer predicted neoepitopes for cancer immunotherapy. *Front Immunol* 2021;12:644637.
- [153] Wu T, Chen J, Diao K, Wang G, Wang J, Yao H, et al. Neodb: a comprehensive neoantigen database and discovery platform for cancer immunotherapy. *Database* 2023;2023:baad041.
- [154] Tan X, Li D, Huang P, Jian X, Wan H, Wang G, et al. dbPepNeo: a manually curated database for human tumor neoantigen peptides. *Database* 2020;2020:baaa004.
- [155] Li W, Sun T, Li M, He Y, Li L, Wang L, et al. GNIFdb: a neoantigen intrinsic feature database for glioma. *Database* 2022;2022:baac004.
- [156] Yu J, Wang L, Kong X, Cao Y, Zhang M, Sun Z, et al. CAD v1.0: cancer antigens database platform for cancer antigen algorithm development and information exploration. *Front Bioeng Biotechnol* 2022;10:819583.
- [157] van Westen GJ, Swier RF, Wegner JK, IJzerman AP, van Vlijmen HW, Bender A. Benchmarking of protein descriptor sets in proteochemometric modeling (part 1): comparative study of 13 amino acid descriptor sets. *J Cheminform* 2013;5:41.
- [158] Georgiev AG. Interpretable numerical descriptors of amino acid space. *J Comput Biol* 2009;16:703–23.
- [159] Cruciani G, Baroni M, Carosati E, Clementi M, Valigi R, Clementi S. Peptide studies by means of principal properties of amino acids derived from MIF descriptors. *J Chemom* 2004;18:146–55.
- [160] Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 2000;16:276–7.
- [161] Liu Y, Chen Y, Han L. Bioinformatics: advancing biomedical discovery and innovation in the era of big data and artificial intelligence. *Innovation Medicine* 2023;1:100012.
- [162] Wang TY, Yang R. Integrated protocol for exitrone and exitrone-derived neoantigen identification using human RNA-seq data with ScanExitron and ScanNeo. *STAR Protoc* 2021;2:100788.
- [163] McGranahan N, Swanton C. Neoantigen quality, not quantity. *Sci Transl Med* 2019;11:eaax7918.
- [164] Tan CL, Lindner K, Boschert T, Meng Z, Rodriguez Ehrenfried A, De Roia A, et al. Prediction of tumor-reactive T cell receptors from scRNA-seq data for personalized T cell therapy. *Nat Biotechnol* 2025;43:134–42.
- [165] Benjamin D, Sato T, Cibulskis K, Getz G, Stewart C, Lichtenstein L. Calling somatic SNVs and indels with Mutect2. *bioRxiv* 2019;861054.
- [166] Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods* 2018;15:591–4.
- [167] Mose LE, Perou CM, Parker JS. Improved indel detection in DNA and RNA via realignment with ABRA2. *Bioinformatics* 2019;35:2966–73.
- [168] Haas BJ, Dobin A, Stransky N, Li B, Yang X, Tickle T, et al. STAR-Fusion: fast and accurate fusion transcript detection from RNA-seq. *bioRxiv* 2017;120295.
- [169] Richman LP, Vonderheide RH, Rech AJ. Neoantigen dissimilarity to the self-proteome predicts immunogenicity and response to immune checkpoint blockade. *Cell Syst* 2019;9:375–82.e4.
- [170] Yang R, Van Etten JL, Dehm SM. Indel detection from DNA and RNA sequencing data with transIndel. *BMC Genomics* 2018;19:270.
- [171] Kahles A, Ong CS, Zhong Y, Rättsch G. SplAdder: identification, quantification and testing of alternative splicing events from RNA-seq data. *Bioinformatics* 2016;32:1840–7.
- [172] Karosiene E, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. *Immunogenetics* 2013;65:711–24.
- [173] Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012;22:568–76.
- [174] Uhrig S, Ellermann J, Walther T, Burkhardt P, Fröhlich M, Hutter B, et al. Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res* 2021;31:448–60.

- [175] Liu C, Yang X, Duffy B, Mohanakumar T, Mitra RD, Zody MC, et al. ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. *Nucleic Acids Res* 2013; 41:e142.
- [176] Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, et al. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 2009;25:1966–7.
- [177] Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* 2009;61:1–13.
- [178] Zhou Z, Lyu X, Wu J, Yang X, Wu S, Zhou J, et al. TSNAD: an integrated software for cancer somatic mutation and tumour-specific neoantigen detection. *R Soc Open Sci* 2017;4:170050.
- [179] Bakker AH, Schumacher TN. MHC multimer technology: current status and future prospects. *Curr Opin Immunol* 2005; 17:428–33.
- [180] Stranzl T, Larsen MV, Lundegaard C, Nielsen M. NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics* 2010;62:357–68.
- [181] Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 2011; 27:2987–93.
- [182] Zhang J, White NM, Schmidt HK, Fulton RS, Tomlinson C, Warren WC, et al. INTEGRATE: gene fusion discovery using whole genome and transcriptome data. *Genome Res* 2016; 26:108–18.
- [183] Murphy C, Elemento O. AGFusion: annotate and visualize gene fusions. *bioRxiv* 2016;080903.
- [184] Karosiene E, Lundegaard C, Lund O, Nielsen M. NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics* 2012;64:177–86.
- [185] Zhang H, Lund O, Nielsen M. The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC–peptide binding. *Bioinformatics* 2009;25:1293–9.
- [186] Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics* 2007;8:238.
- [187] 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 Bioinformatics* 2009; 10:394.
- [188] Nielsen M, Lund O. *NN-align*. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics* 2009;10:296.
- [189] Kesmir C, Nussbaum AK, Schild H, Detours V, Brunak S. Prediction of proteasome cleavage motifs by neural networks. *Protein Eng* 2002;15:287–96.
- [190] Zhang XO, Dong R, Zhang Y, Zhang JL, Luo Z, Zhang J, et al. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res* 2016;26:1277–87.
- [191] Zhao J, Wu J, Xu T, Yang Q, He J, Song X. IRESfinder: identifying RNA internal ribosome entry site in eukaryotic cell using framed *k*-mer features. *J Genet Genomics* 2018;45:403–6.
- [192] Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res* 2017;45:W12–6.
- [193] 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;37:1283–6.
- [194] Fenix AM, Miyaoka Y, Bertero A, Blue SM, Spindler MJ, Tan KKB, et al. Gain-of-function cardiomyopathic mutations in RBM20 rewire splicing regulation and re-distribute ribonucleo-protein granules within processing bodies. *Nat Commun* 2021; 12:6324.
- [195] Li G, Iyer B, Prasath VBS, Ni Y, Salomonis N. DeepImmuno: deep learning-empowered prediction and generation of immunogenic peptides for T-cell immunity. *Brief Bioinform* 2021;22:bbab160.
- [196] Li G, Bhattacharjee A, Salomonis N. Quantifying tumor specificity using Bayesian probabilistic modeling for drug target discovery and prioritization. *bioRxiv* 2023;530994.
- [197] Li G, Mahajan S, Ma S, Jeffery ED, Zhang X, Bhattacharjee A, et al. Splicing neoantigen discovery with SNAF reveals shared targets for cancer immunotherapy. *Sci Transl Med* 2024; 16:eade2886.
- [198] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21.
- [199] Shen S, Park JW, Lu Z, Lin L, Henry MD, Wu YN, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-seq data. *Proc Natl Acad Sci U S A* 2014;111:E5593–601.
- [200] Boegel S, Löwer M, Schäfer M, Bukur T, De Graaf J, Boisguérin V, et al. HLA typing from RNA-seq sequence reads. *Genome Med* 2012;4:102.
- [201] Li YI, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, Im HK, et al. Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet* 2018;50:151–8.
- [202] Orenbuch R, Filip I, Comito D, Shaman J, Pe'er I, Rabadan R. arcasHLA: high-resolution HLA typing from RNA-seq. *Bioinformatics* 2020;36:33–40.
- [203] 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;8:396–408.
- [204] Paul S, Sidney J, Sette A, Peters B. TepiTool: a pipeline for computational prediction of T cell epitope candidates. *Curr Protoc Immunol* 2016;114:18.19.1–24.
- [205] Larsen MV, Lelic A, Parsons R, Nielsen M, Hoof I, Lamberth K, et al. Identification of CD8⁺ T cell epitopes in the West Nile virus polyprotein by reverse-immunology using NetCTL. *PLoS One* 2010;5:e12697.
- [206] Giguère S, Drouin A, Lacoste A, Marchand M, Corbeil J, Laviolette F. MHC-NP: predicting peptides naturally processed by the MHC. *J Immunol Methods* 2013;400–401:30–6.
- [207] Toulmin SA, Bhadiadra C, Paris AJ, Lin JH, Katzen J, Basil MC, et al. Type II alveolar cell MHCII improves respiratory viral disease outcomes while exhibiting limited antigen presentation. *Nat Commun* 2021;12:3993.
- [208] Rigo MM, Antunes DA, Vaz de Freitas M, Fabiano de Almeida Mendes M, Meira L, Sinigaglia M, et al. DockTope: a web-based tool for automated pMHC-I modelling. *Sci Rep* 2015;5:18413.
- [209] Vanhee P, Reumers J, Stricher F, Baeten L, Serrano L, Schymkowitz J, et al. PepX: a structural database of non-redundant protein–peptide complexes. *Nucleic Acids Res* 2010; 38:D545–51.
- [210] Wu J, Wang W, Zhang J, Zhou B, Zhao W, Su Z, et al. DeepHLApan: a deep learning approach for neoantigen prediction considering both HLA–peptide binding and immunogenicity. *Front Immunol* 2019;10:2559.
- [211] Wang G, Wan H, Jian X, Li Y, Ouyang J, Tan X, et al. INeoEpp: a novel T-cell HLA class-I immunogenicity or neoantigenic epitope prediction method based on sequence-related amino acid features. *Biomed Res Int* 2020;2020:5798356.
- [212] Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* 2017;27:491–9.
- [213] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; 30:2114–20.
- [214] Li D, Mei H, Shen Y, Su S, Zhang W, Wang J, et al. ECharts: a declarative framework for rapid construction of web-based visualization. *Vis Inform* 2018;2:136–46.
- [215] Bagaev DV, Vroomans RMA, Samir J, Stervbo U, Rius C, Dolton G, et al. VDJdb in 2019: database extension, new

- analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Res* 2020;48:D1057–62.
- [216] Tickotsky N, Sagiv T, Prilusky J, Shifrut E, Friedman N. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics* 2017; 33:2924–9.
- [217] Sarkizova S, Klaeger S, Le PM, Li LW, Oliveira G, Keshishian H, et al. A large peptidome dataset improves HLA class I epitope prediction across most of the human population. *Nat Biotechnol* 2020;38:199–209.
- [218] DeVette CI, Andreatta M, Welm A, Nielsen M, Hildebrand WH. A pan-allele class I MHC prediction tool in mice enables tumor immunology studies. *J Immunol* 2019;202: 195.12.
- [219] Nielsen M, Lundegaard C, Lund O, Keşmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* 2005;57:33–41.
- [220] Gfeller D, Schmidt J, Croce G, Guillaume P, Bobisse S, Genolet R, et al. Improved predictions of antigen presentation and TCR recognition with MixMHCpred2.2 and PRIME2.0 reveal potent SARS-CoV-2 CD8⁺ T-cell epitopes. *Cell Syst* 2023;14: 72–83.e5.
- [221] Zhou WJ, Qu Z, Song CY, Sun Y, Lai AL, Luo MY, et al. NeoPeptide: an immunoinformatic database of T-cell-defined neoantigens. *Database* 2019;2019:baz128.
- [222] Tian F, Zhou P, Li Z. T-scale as a novel vector of topological descriptors for amino acids and its application in QSARs of peptides. *J Mol Struct* 2007;830:106–15.
- [223] Kidera A, Konishi Y, Oka M, Ooi T, Scheraga HA. Statistical analysis of the physical properties of the 20 naturally occurring amino acids. *J Protein Chem* 1985;4:23–55.
- [224] Yang L, Shu M, Ma K, Mei H, Jiang Y, Li Z. ST-scale as a novel amino acid descriptor and its application in QSAM of peptides and analogues. *Amino Acids* 2010;38:805–16.
- [225] Sandberg M, Eriksson L, Jonsson J, Sjöström M, Wold S. New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *J Med Chem* 1998;41:2481–91.
- [226] Mei H, Liao ZH, Zhou Y, Li SZ. A new set of amino acid descriptors and its application in peptide QSARs. *Biopolymers* 2005;80:775–86.
- [227] Sun T, He Y, Li W, Liu G, Li L, Wang L, et al. neoDL: a novel neoantigen intrinsic feature-based deep learning model identifies IDH wild-type glioblastomas with the longest survival. *BMC Bioinformatics* 2021;22:382.
- [228] Zhang J, Caruso FP, Sa JK, Justesen S, Nam DH, Sims P, et al. The combination of neoantigen quality and T lymphocyte infiltrates identifies glioblastomas with the longest survival. *Commun Biol* 2019;2:135.
- [229] Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA websuite. *Nucleic Acids Res* 2008;36:W70–4.